```
* Stata Analysis Examples Replication for ASDA 2nd Edition
* Berglund April 2017
* Chapter 11

* use the wide format HRS 2006_2012 data set
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0
* prepare ln income
gen ln_inc06 = ln(H8ITOT + 1)
gen ln_inc08 = ln(H9ITOT + 1)
gen ln_inc10 = ln(H10ITOT + 1)
gen ln_inc12 = ln(H11ITOT + 1)

* histogram of each income variable
hist ln_inc06
hist ln_inc08
hist ln_inc10
hist ln_inc12

* 11.3.1 Example: Descriptive Estimation at a Single Wave
* Complete case analysis for 2008 log-income.
svyset secu [pweight = kwgtr], strata(stratum)
* Table 11.2
svy: mean ln_inc08
matrix list r(table)
matrix a = r(table)
di exp(a[1,1])
di exp(a[5,1])
di exp(a[6,1])

* Weight adjustment approach for 2008 log-income.
* Compute response indicator for 2008.
gen resp08 = 1 if ln_inc08 != .
replace resp08 = 0 if ln_inc08 == .
tab resp08, miss

* Modal imputation of missing covariate values.
replace selfrhealth_06 = 3 if selfrhealth_06 == .
replace marcat_06 = 2 if marcat_06 == .
replace diabetes_06 = 0 if diabetes_06 == .
replace arthritis_06 = 1 if arthritis_06 == .
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

* Response propensity model, with predictions.
svyset secu [pweight = kwgtr], strata(stratum)
svy: logit resp08 ln_inc06 i.selfrhealth_06 age_06 i.marcat_06 diabetes_06 arthritis_06 i.racecat
i.edcat
predict phat, p

* Form deciles of response propensity, and compute means
* within each decile for weight adjustment.
xtile dec = phat, nq(10)
egen mean_phat = mean(phat), by(dec)
mean phat, over(dec)

gen adj_kwgtr = kwgtr * (1 / mean_phat)

* Estimation using adjusted weights.
* Table 11.2
svyset secu [pweight = adj_kwgtr], strata(stratum)
svy: mean ln_inc08
matrix a = r(table)
di exp(a[1,1])
di exp(a[5,1])
di exp(a[6,1])
```

```
* Multiple Imputation method.
* Create deciles of the 2006 sampling weights.
xtile wgt_dec = kwgtr, nq(10)
sum wgt_dec
tab wgt_dec

* Set data structure and register both complete "regular"
* variables and variables with missing data to be imputed.
mi set flong
mi register imputed ln_inc08
mi register regular ln_inc06 selfrhealth_06 age_06 marcat_06 diabetes_06 arthritis_06 racecat edcat
stratum wgt_dec
mi describe

* Implement regression imputation after setting a seed.
set seed 41279

mi impute chained (regress) ln_inc08 = ln_inc06 i.selfrhealth_06 age_06 i.marcat_06 diabetes_06 ///
   arthritis_06 i.racecat i.edcat i.wgt_dec i.stratum, noisily augment add(5) burnin(5)

* Set complex sampling features in MI framework.
mi svyset secu [pweight = kwgtr], strata(stratum)

* Multiple Imputation estimation.
* Table 11.2
mi estimate, vartable: svy: mean ln_inc08
matrix list r(table)
matrix a = r(table)
di exp(a[1,1])
di exp(a[5,1])
di exp(a[6,1])

* Imputation Using a Selection Model. * use original non-imputed data for this model.
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0
* prepare ln income
gen ln_inc06 = ln(H8ITOT + 1)
gen ln_inc08 = ln(H9ITOT + 1)
gen ln_inc10 = ln(H10ITOT + 1)
gen ln_inc12 = ln(H11ITOT + 1)
gen resp08 = 1 if ln_inc08 != .
replace resp08 = 0 if ln_inc08 == .
* Modal imputation of missing covariate values
replace selfrhealth_06 = 3 if selfrhealth_06 == .
replace marcat_06 = 2 if marcat_06 == .
replace diabetes_06 = 0 if diabetes_06 == .
replace arthritis_06 = 1 if arthritis_06 == .
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .
svyset secu [pweight = kwgtr], strata(stratum)

* Output discussed in text for model results
svy: heckman ln_inc08 ln_inc06 selfrhealth_06 age_06 i.marcat_06 i.edcat, ///
select(resp08 = diabetes_06 arthritis_06 selfrhealth_06 age_06 i.marcat_06)
```

```
*****************************************************************************************************
* 11.3.2 Example: Change across Two Waves
* 1. Complete Case Analysis.
* Complete case analysis: mean change in income,
* 2006 to 2010.

* use the wide format HRS 2006_2012 data set
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

gen incdiff_06_10 = H10ITOT - H8ITOT
* Table 11.3 Complete Case
svyset secu [pweight = kwgtr], strata(stratum)
svy: mean incdiff_06_10

* 2. Weight Adjustment.
* prepare ln income
gen ln_inc06 = ln(H8ITOT + 1)
gen ln_inc08 = ln(H9ITOT + 1)
gen ln_inc10 = ln(H10ITOT + 1)
gen ln_inc12 = ln(H11ITOT + 1)
* Compute response indicator for 2010.
gen resp10 = 1 if ln_inc10 != .
replace resp10 = 0 if ln_inc10 == .
tab resp10, miss

* Modal imputation of missing covariate values.
replace selfrhealth_06 = 3 if selfrhealth_06 == .
replace marcat_06 = 2 if marcat_06 == .
replace diabetes_06 = 0 if diabetes_06 == .
replace arthritis_06 = 1 if arthritis_06 == .
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

* Response propensity model.
svyset secu [pweight = kwgtr], strata(stratum)
svy: logit resp10 ln_inc06 i.selfrhealth_06 age_06 i.marcat_06 diabetes_06 arthritis_06 i.racecat
i.edcat
predict phat, p

* Form deciles of response propensity, and compute means within each decile for weight adjustment.
xtile dec = phat, nq(10)
egen mean_phat = mean(phat), by(dec)
gen adj_kwgtr = kwgtr * (1 / mean_phat)

* Estimation of mean change using adjusted weights.
* Table 11.3 Response Propensity Adjustment
svyset secu [pweight = adj_kwgtr], strata(stratum)
svy: mean incdiff_06_10

* 3. Multiple Imputation.
* Multiple imputation of 2010 log-income.
* Create deciles of the 2006 sampling weights.
xtile wgt_dec = kwgtr, nq(10)
* Set data structure and register both complete "regular" variables and variables with missing data
to be imputed.
mi set flong
mi register imputed ln_inc10
mi register regular ln_inc06 selfrhealth_06 age_06 marcat_06 diabetes_06 arthritis_06 racecat edcat
stratum wgt_dec
mi describe

* Implement regression imputation
set seed 41279
mi impute chained (regress) ln_inc10 = ln_inc06 i.selfrhealth_06 age_06 i.marcat_06 diabetes_06
arthritis_06 i.racecat i.edcat ///
   i.wgt_dec i.stratum, noisily augment add(5) burnin(5)

* Compute bounded change scores in each imputed data set.
```

```
sum incdiff_06_10 if _mi_m == 0
sum ln_inc10 if _mi_m == 0
replace ln_inc10=14.92 if ln_inc10 > 14.92 & ln_inc10 != .
gen new_chg0610 = exp(ln_inc10) - exp(ln_inc06)
replace new_chg0610 = -12300000 if new_chg0610 < -12300000
replace new_chg0610 = 2062968 if new_chg0610 > 2062968 & new_chg0610 != .

* Set complex sampling features for MI analysis.
mi svyset secu [pweight = kwgtr], strata(stratum)

* Multiple Imputation estimation.
* Table 11.3 MI
mi estimate, vartable noisily: svy: mean new_chg0610

*4. Calibration.
* use the wide format HRS 2006_2012 data set
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

* Modal imputation of missing covariate values.
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

* Compute sums of 2006 weights in cross-classes
* defined by sex, race, and education.
egen cal_class = group(racecat edcat gender)
collapse (sum) kwgtr, by(cal_class)
rename kwgtr popsize

save "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\cal_pop_sizes.dta", replace

* Repeat process for cases with complete data.
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

* Modal imputation of missing covariate values.
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

gen ln_inc10 = ln(H10ITOT + 1)

* Compute response indicator for 2010.
gen resp10 = 1 if ln_inc10 != .
replace resp10 = 0 if ln_inc10 == .
tab resp10, miss

* Only keep cases with complete data.
keep if resp10 == 1

* Compute sums of 2006 weights in cross-classes
* defined by sex, race, and education.
egen cal_class = group(racecat edcat gender)
collapse (sum) kwgtr, by(cal_class)
rename kwgtr sumrespwgts

* save resp sizes data set.
save "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\cal_resp_sizes.dta", replace

* Merge the two data sets of estimated population sizes.
merge 1:1 cal_class using "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal
File\cal_pop_sizes.dta"

* Compute calibration adjustments for each cross-class.
gen cal_adj = popsize / sumrespwgts
drop _merge
save "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\cal_resp_sizes.dta", replace

* Open original data set and merge calibration adjustments.
```

```
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

* Modal imputation of missing covariate values.
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

egen cal_class = group(racecat edcat gender)
sort cal_class
gen ln_inc06 = ln(H8ITOT + 1)
gen ln_inc10 = ln(H10ITOT + 1)

merge m:1 cal_class using "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal
File\cal_resp_sizes.dta"

* Compute response indicator for 2010.
gen resp10 = 1 if ln_inc10 != .
replace resp10 = 0 if ln_inc10 == .
tab resp10, miss

* Compute calibrated weights for cases with complete data.
gen kwgtr_cal = kwgtr * cal_adj if resp10 == 1

* Verify that sums of calibrated weights for cases with complete data are equal to sums of base
weights for full sample.
total kwgtr, over(cal_class)
total kwgtr_cal if resp10 == 1, over(cal_class)

* Estimate mean change using complete cases.
* Table 11.3
gen incdiff_06_10 = H10ITOT - H8ITOT
svyset secu [pweight = kwgtr_cal], strata(stratum)
svy: mean incdiff_06_10

*******************************************************************************************
* 11.3.3 Example: Weighted Multilevel Modeling
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

* Reshape the original wide data.
gen ln_inc1 = ln(H8ITOT + 1)
gen ln_inc2 = ln(H9ITOT + 1)
gen ln_inc3 = ln(H10ITOT + 1)
gen ln_inc4 = ln(H11ITOT + 1)

* Initial wave-specific survey weights.
gen wgt1 = kwgtr
gen wgt2 = lwgtr
gen wgt3 = mwgtr
gen wgt4 = nwgtr

* Baseline sampling weight for respondents from 2006.
gen basewgt = kwgtr

* Keep variables of interest.
keep hhid pn gender ln_inc1-ln_inc4 wgt1-wgt4 secu stratum basewgt

* Reshape data set from wide to vertical format.
reshape long ln_inc wgt, i(hhid pn) j(year)

* Drop missing income values (wave nonresponse).
drop if ln_inc == .

* Form a unique ID variable.
egen newid = concat(hhid pn)
destring newid, gen(newid_num)

* Compute conditional Level 1 weights (Skinner and Holmes 2003), and rescale weights.
```

```
gen level1wgt = wgt / basewgt
gen level2wgt = basewgt

* "Method 2" rescaling (normalizing weights).
egen sumw = sum(level1wgt), by(newid)
egen nj = count(ln_inc), by(newid)
gen level1wgt_r = level1wgt * nj/sumw

* Years since 2006 variables.
gen yrssince06 = 0 if year == 1
replace yrssince06 = 2 if year == 2
replace yrssince06 = 4 if year == 3
replace yrssince06 = 6 if year == 4

save "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\chapter11_hrs_vert.dta", replace

* Plot data for small subsample of subjects.
* Figure 11.1
xtset newid_num year
xtline ln_inc if newid_num <= 10200000, overlay legend(off) ytitle(Log-Transformed Annual Income)
ttitle(HRS Wave (1 = 2006))

* Plot means across waves for males and females.
* Figure 11.2
collapse (mean) ln_inc, by(gender year)
xtset gender year
xtline ln_inc, overlay ytitle(Mean Log-Transformed Annual Income) ttitle(Wave (1 = 2006))

* Fit multilevel model of interest, Use long data set for modeling.

use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\chapter11_hrs_vert.dta", clear
gen yrs06sq = yrssince06*yrssince06
egen unique_secu = group(stratum secu)
* Model without stratum and secu as unique identifier.
* Output not shown in book.

mixed ln_inc c.yrssince06##i.gender c.yrs06sq##i.gender ///
   [pweight = level1wgt_r] ///
   || newid_num: yrssince06 yrs06sq, variance cov(unstruct) ///
   pweight(level2wgt) pwscale(size)

* Fully account for complex sampling features.
mixed ln_inc c.yrssince06##i.gender c.yrs06sq##i.gender i.stratum ///
   [pweight = level1wgt_r] ///
   || newid_num: yrssince06 yrs06sq, variance cov(unstruct) ///
   pweight(level2wgt) pwscale(size) vce(cluster unique_secu)

* Unweighted approach, note used "reml" option to avoid non definite matrix
mixed ln_inc c.yrssince06##i.gender c.yrs06sq##i.gender ///
   || newid_num: yrssince06 yrs06sq, reml variance cov(unstruct)

*******************************************************************************************
* Veiga et al. (2014) approach.
* Start with wide HRS data set
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

gen ln_inc1 = ln(H8ITOT + 1)
gen ln_inc2 = ln(H9ITOT + 1)
gen ln_inc3 = ln(H10ITOT + 1)
gen ln_inc4 = ln(H11ITOT + 1)

* Identify cases with complete data
gen comp = (ln_inc1 != . & ln_inc2 != . & ln_inc3 != . & ln_inc4 != .)

* Fit a response propensity model predicting probability of complete data

* Modal imputation of missing covariate values
replace selfrhealth_06 = 3 if selfrhealth_06 == .
```

```
replace marcat_06 = 2 if marcat_06 == .
replace diabetes_06 = 0 if diabetes_06 == .
replace arthritis_06 = 1 if arthritis_06 == .
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

svyset secu [pweight = kwgtr], strata(stratum)
svy: logit comp ln_inc1 i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat, p

* Adjust baseline sampling weights in 2006 by inverse of pred. prob of complete data
gen indwgt = kwgtr * (1 / phat)

* Set cluster-specific weight equal to 1 (weights not provided)
egen newclust = group(stratum secu)
gen clustwgt = 1

* Create unique person ID
egen newid = concat(hhid pn)
destring newid, gen(newid_num)

* Only keep cases with complete data
keep if comp == 1

* Only keep variables of interest
keep newid_num newclust gender ln_inc1-ln_inc4 indwgt clustwgt

* Reshape data to vertical format
reshape long ln_inc, i(newid_num) j(wave)

* Create unique indicators for each wave
gen wave1 = (wave == 1)
gen wave2 = (wave == 2)
gen wave3 = (wave == 3)
gen wave4 = (wave == 4)

* Compute constant variable for macro
gen cons = 1

* Recode gender
gen male = (gender == 1)

* modifications to the MATA do file to be made prior to running it, see do file below for
implementation
/*
s = 5
delta_matrix=J(s,16,0)
delta_matrix[2,]=
(1,0,0,0,
 0,1,0,0,
 0,0,1,0,
 0,0,0,1)

delta_matrix[3,]=
(0,1,0,0,
 1,0,1,0,
 0,1,0,1,
 0,0,1,0)

delta_matrix[4,]=
(0,0,1,0,
 0,0,0,1,
 1,0,0,0,
 0,1,0,0)

delta_matrix[5,]=
 (0,0,0,1,
 0,0,0,0,
 0,0,0,0,
```

```
 1,0,0,0)

rowshape(delta_matrix[1,],4)
rowshape(delta_matrix[2,],4)
rowshape(delta_matrix[3,],4)
rowshape(delta_matrix[4,],4)
rowshape(delta_matrix[5,],4)

name1 = tokens(varlist)

name3 = ("Sigma_u_2","Genlin(1)","Genlin(2)", ///
"Genlin(3)","Genlin(4)","Genlin(5)")

theta_genlin=theta[2,]*rowshape(delta_matrix[2,],4)+ theta[3,]*rowshape(delta_matrix[3,],4)+
theta[4,]*rowshape(delta_matrix[4,],4)+ theta[5,]*rowshape(delta_matrix[5,],4)

theta0_genlin=theta0[2,]*rowshape(delta_matrix[2,],4)+ theta0[3,]*rowshape(delta_matrix[3,],4)+
theta0[4,]*rowshape(delta_matrix[4,],4)+ theta0[5,]*rowshape(delta_matrix[5,],4)
*/

/*** Running do file with the written mata function ***/
* run modified file prepared by West
do "P:\ASDA 2\Draft Chapters\Chapter 11\Viega materials\pwigls_genlin_adcv_modAV1.do"

sort newclust newid_num wave
count

/*this is needed for the mata macro*/
gl wav wave1 wave2 wave3 wave4
gl x wave1 wave2 wave3 wave4 male

keep if newid_num < 20000000

mata: pwigls_genlin_adcvw4toep1("$x","ln_inc", "$wav", "cons", "newclust" ,"clustwgt", "indwgt" )
* Examine sensitivity to weights
gen indwgtone = 1

mata: pwigls_genlin_adcvw4toep1("$x","ln_inc", "$wav", "cons", "newclust" ,"clustwgt", "indwgtone"
)

************************************
*11.3.4 Example: Weighted GEE Analysis
* Use wide data set
use "P:\ASDA 2\Data sets\HRS 2012\HRS 2006_2012 Longitudinal File\hrs_2006_2012_15jul2016.dta",
clear
keep if kfinr == 1 & kwgtr != 0

* Modal imputation of missing covariate values.
replace selfrhealth_06 = 3 if selfrhealth_06 == .
replace marcat_06 = 2 if marcat_06 == .
replace diabetes_06 = 0 if diabetes_06 == .
replace arthritis_06 = 1 if arthritis_06 == .
replace racecat = 2 if racecat == .
replace edcat = 2 if edcat == .

gen ln_inc06 = ln(H8ITOT + 1)
gen ln_inc08 = ln(H9ITOT + 1)
gen ln_inc10 = ln(H10ITOT + 1)
gen ln_inc12 = ln(H11ITOT + 1)

* Compute response indicator for 2008.
gen resp08 = 1 if ln_inc08 != .
replace resp08 = 0 if ln_inc08 == .
tab resp08, miss

* Response propensity model (2008).
svyset secu [pweight = kwgtr], strata(stratum)
svy: logit resp08 ln_inc06 i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat08, p
```

```
sum phat08

* Compute response indicator for 2010.
gen resp10 = 1 if ln_inc10 != .
replace resp10 = 0 if ln_inc10 == .
tab resp10, miss

* Response propensity model (2010), respondents in 2008.
svyset secu [pweight = kwgtr], strata(stratum)
svy, subpop(if resp08 == 1): logit resp10 ln_inc06 ///
   i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat10_11, p
sum phat10_11


* Response propensity model (2010), nonrespondents in 2008.
svyset secu [pweight = kwgtr], strata(stratum)
svy, subpop(if resp08 == 0): logit resp10 ln_inc06 ///
   i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat10_10, p
mean phat10_10

* Compute response indicator for 2012.
gen resp12 = 1 if ln_inc12 != .
replace resp12 = 0 if ln_inc12 == .
tab resp12, miss

* Response propensity model (2012), 111 pattern.
svyset secu [pweight = kwgtr], strata(stratum)
svy, subpop(if resp08 == 1 & resp10 == 1): logit resp12 ///
   ln_inc06 i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat12_111, p
mean phat12_111

* Response propensity model (2012), 110 pattern.
svyset secu [pweight = kwgtr], strata(stratum)
svy, subpop(if resp08 == 1 & resp10 == 0): logit resp12 ///
   ln_inc06 i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat12_110, p
mean phat12_110

* Response propensity model (2012), 101 pattern.
svyset secu [pweight = kwgtr], strata(stratum)
svy, subpop(if resp08 == 0 & resp10 == 1): logit resp12 ///
   ln_inc06 i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat12_101, p
mean phat12_101

* Response propensity model (2012), 100 pattern.
svyset secu [pweight = kwgtr], strata(stratum)
svy, subpop(if resp08 == 0 & resp10 == 0): logit resp12 ///
   ln_inc06 i.selfrhealth_06 age_06 ///
   i.marcat_06 diabetes_06 arthritis_06 i.racecat i.edcat
predict phat12_100, p
mean phat12_100

* Compute cumulative response probabilities
gen cumprob1 = 1

gen cumprob2 = phat08 if resp08 == 1
replace cumprob2 = 1-phat08 if resp08 == 0

gen cumprob3 = phat08 * phat10_11 ///
   if resp08 == 1 & resp10 == 1
replace cumprob3 = phat08 * (1-phat10_11) ///
```

```
   if resp08 == 1 & resp10 == 0
replace cumprob3 = (1-phat08) * phat10_10 ///
   if resp08 == 0 & resp10 == 1
replace cumprob3 = (1-phat08) * (1-phat10_10) ///
   if resp08 == 0 & resp10 == 0

gen cumprob4 = phat08 * phat10_11 * phat12_111 ///
   if resp08 == 1 & resp10 == 1 & resp12 == 1
replace cumprob4 = phat08 * phat10_11 * (1-phat12_111) ///
   if resp08 == 1 & resp10 == 1 & resp12 == 0
replace cumprob4 = phat08 * (1-phat10_11) * phat12_110 ///
   if resp08 == 1 & resp10 == 0 & resp12 == 1
replace cumprob4 = phat08 *(1-phat10_11)*(1-phat12_110) ///
   if resp08 == 1 & resp10 == 0 & resp12 == 0

replace cumprob4 = (1-phat08)*phat10_10 *phat12_101 ///
   if resp08 == 0 & resp10 == 1 & resp12 == 1
replace cumprob4 = (1-phat08)*phat10_10*(1-phat12_101) ///
   if resp08 == 0 & resp10 == 1 & resp12 == 0
replace cumprob4 = (1-phat08)*(1-phat10_10)*phat12_100 ///
   if resp08 == 0 & resp10 == 0 & resp12 == 1
replace cumprob4=(1-phat08)*(1-phat10_10)*(1-phat12_100) ///
   if resp08 == 0 & resp10 == 0 & resp12 == 0

* Compute overall probability of response pattern.
gen cumprob_case = cumprob4
sum cumprob_case

* Reshape data set, keeping variables of interest.
gen ln_inc1 = ln(H8ITOT + 1)
gen ln_inc2 = ln(H9ITOT + 1)
gen ln_inc3 = ln(H10ITOT + 1)
gen ln_inc4 = ln(H11ITOT + 1)

* Baseline sampling weight for individual from 2006.
gen basewgt = kwgtr

keep hhid pn gender marcat_06 diabetes_06 arthritis_06 ///
   racecat edcat ln_inc1-ln_inc4 secu stratum basewgt ///
   cumprob_case

reshape long ln_inc wgt, i(hhid pn) j(year)

* Form a unique ID variable.
egen newid = concat(hhid pn)
destring newid, gen(newid_num)

* Compute the case-specific weights for weighted GEE.
gen casewt = basewgt * (1 / cumprob_case)

* Compute measure of years since 2006, and squared version.
gen yrssince06 = 0 if year == 1
replace yrssince06 = 2 if year == 2
replace yrssince06 = 4 if year == 3
replace yrssince06 = 6 if year == 4

gen yrs06sq = yrssince06*yrssince06

* Fit weighted GEE model. Output in Text.
xtset newid_num year
xtgee ln_inc c.yrssince06##i.gender c.yrs06sq##i.gender ///
   i.stratum [pweight = casewt], corr(uns)
estat wcorrelation
```