NLSY Children Data Creation Programs – REV 4
(updated Jan. 2008, May 2008, May 2010)

Codes: **programs**, **data sets**, *variables*

1. **makemomkid.sas**

Pulls data from renamed data sets and creates the following data sets (no sample limitations):
mom: contains all info. on mothers and their backgrounds stacking it by *momid-year*
kid: contains all info. on kids and stacks it by *idchild-year* (can use *momid-year* to merge with **mom**)

Note: multiple variables for mother's highest grade completed:

higgrc: mother's highest grade completed as of May 1 that year (mother's file)
hgcmom: mother's highest grade completed at survey date (mother's file)
hgcm: mother's highest grade completed from children's file

Note: On July 25, 2008, updated hours worked last year (*hrswri*) to account for weeks unaccounted for (*wksuni*).

Note: It takes variables from all the "renamed" data sets.

2. **afqtreg2000.sas**

Creates a normalized (mean 0, std. dev. 1) afqt measure (*afqtn*) and a normalized measure that accounts for age effects (*afqtadj*). These are calculated for the entire sample, but are based on the distribution of mothers from the random NLSY sample of women. These are calculated once for each *momid* and saved in the data set **afqtadj**.

3. **normalizing depvars.sas**

Creates normalized dependent variable measures (*bpitosan, homtosn, motsosn, piamatsn, piarecsn, piarersn, ppvtosn, vermabn*) with mean 0 and std. dev. 1. Stores these variables (and original non-normalized variables as well as *momid*) by *idchild-year* in **normtestscores**.

4. numdepend3.sas

Generates variables for the number of dependents in different age categories (*young1*, *young5*, *young618, oldkids*) by *momid-year* based on summing children of the NLSY. Stores data in **numdep** (by *momid-year*). Also creates a few measures of the total number of children/dependents:

numdep: number of children of a mother that are in the children of NLSY survey
nchh: number of children of mother in the household (from Children of NLSY)
numofb: number of children of mother in the household (from NLSY mothers)
maxndep: max of *numdep* and *nchh*
maxndeply: *maxndep* for previous calendar year

Note that *nchh* and *numofb* are nearly identical and should be based on mother's responses. Category variables are also created to denote (0,1,2) children: *depcat* and *maxdepcat*.

Creates a second data set **imputedkidage** which contains imputed measures of child's age (in years) *agecsimp* by *idchild-year*.

5. imputeincomeREV4.sas

Uses mother's information to impute income components and total income measures for full sample of women. Creates **imputed** data set organized by *momid-year*. It includes/imputes data for non-survey years (1995, 1997, 1999).

Note: Also creates *ageofrimp*, which fills in missing ages with age in 1979 + (year-1979) and brings in *marrlyimp*, which is an indicator equal to one if the mother is married in the previous year—to match up with income measures.

6. impmomkidREV.sas

Merges data from mother (**mom**, **imputed**, **afqtadj**, **numdep**) and children (**kid**, **normtestscores**) data files for everyone and saves it in **impmomkid**. Pre-cleaned data ready for analysis. Also, makes some sample indicator variables for subsequent use:

samprandom: all random sample
sampnm: all non-military samples
samprandwhite: all random white sample
sampnmblack: all non-military black samples
sampnmhisp: all non-military hispanic samples

7. **incbyageyrREV.sas (not updated for some time)**

Uses mother's imputed total family income measure (*totincRSimp*) to compute an annual average of the discounted present value of future family income. Saves total imputed income and average discounted PV of income for every year. Then, creates variables for total income and average discounted PV of future income by the age of the child (relative to birth year of each child). These are saved in **incbyageyrTOT** by *idchild*.

Selected variables in 'impmomkid' data set

<u>Demographic variables:</u>

ageofc:	age of child in months
agecsimp:	age of child in years (missing values imputed based on earlier ages)
black:	dummy
datofby:	year of birth for child
fatstg:	dummy=1 if mother's father lives in household
hispanic:	dummy
idchild:	child's id
male:	dummy
maxndep:	number of dependents/children in household
maxndeply:	*maxndep* in previous calendar year
marrsepimp:	married or separated in previous calendar year (some imputations)
marrylimp:	married last calendar year (some imputations)
motstg:	dummy=1 if mother's mother lives in household
numofa:	number of adult family members living in household

samprandom:	all random sample
sampnm:	all non-military samples
samprandwhite:	all random white sample
sampnmblack:	all non-military black samples
sampnmhisp:	all non-military hispanic samples

<u>Parental characteristics:</u>

afqtadj:	AFQT measure normalized on a scale with mean zero, std. deviation of one, then adjusted by age
ageofrimp:	age of mother (missing ages filled in with age in 1979 + year - 1979)
forborn:	dummy=1 if mother was born outside the US
hgcbyf79:	highest grade completed by mother's father
hgcbym79:	highest grade completed by mother's mother
hgcbys:	highest grade completed by spouse
higgrc:	highest grade completed by mother
liveboth:	dummy=1 if mother lived with both parents at age 14
marrsep:	married or separated in previous calendar year (for matching with income)
momid:	mother's id
rural14:	dummy=1 if mother lived in a rural area at age 14
rural14miss	
hrswri:hrswrinew	hours worked in previous calendar year (adjusted for missing weeks)
wksuni:	% of weeks unaccounted for in calculating *hrwri*

wagehr: hourly wage computed using only *wage* income divided by *hrswri* (not deflated) –
Note: probably best to compute using *earnincR/hrswri* to ensure   that it's deflated
and takes both wage and military income into account

Child outcomes (normalized with mean zero, standard deviation one):

bpitosan: behavioral problem index (ages 4+)
homtosn: home score (all ages)
motsosn: motor and social skills (ages 0-3)
piamatsn: PIAT math (ages 5+)
piarecsn: PIAT reading comprehension (ages 5+)
piarersn: PIAT reading (ages 5+)
ppvtosn: PPVT (ages 3+)
vermabn: verbal memory (ages 3-6/ages 4-6 in 1990)

Income variables (imputed and deflated to 1979 dollars using CPI):

cpi: CPI deflator in 1979 dollars
cpily: CPI deflator in 1979 dollars for last calendar year

earnincR: Total earned income for respondent

earnincS: Total earned income for spouses only (not partners)

earnincRS: Total earned income for respondent + spouse (does not include partner income)

unearnincRS: Total unearned taxable income for respondent + spouse (not partners)

nontaxincRS: Non-taxable income for respondent + spouse (not partners)

taxincRS: Total taxable income for respondent + spouse (not partners)

totincRS: Total income for respondent + spouse (not partners)

totincP: Total partner income created of partner (non-spouse)

These income variables are derived in **imputeincomeREV4.sas** and are described further below.

Initial income components (revised version 4, May 2010)

Brief Notes:

- All amounts received in the previous calendar year.
- All income variables are in 1979 dollars (deflated by CPI for previous calendar year).
- Components of income>200,000 are set to missing.
- Measures for mothers only (also includes any women ages 18+ or those living away from parents)—fewer income measures available for non-mothers/younger women living at home
- Unless otherwise noted, variable is available for all survey years
- We look to marital status to determine whether *swaged*, *sbusd*, *smild*,…, apply to spouses or partners for 1994+
- For unmarried women with a partner (years 1994+), we assume *totinad* (alimony) and *totwead* (welfare income) are attributed fully to them, while *totinvd* (veterans benefits, worker comp., disability) and *totinrod* (income from savings, rental income, etc.) are attributed fully to the partner
- We assume *totinad* (alimony+child support) is all child support prior to 1983
- From 1983-93, we assume *totinad*=0 whenever *totinad*=*totincd* (double-counting assumed)
- All 'spousal' or 'respondent + spousal' income variables are revised to exclude income for partners of unmarried respondents (e.g. *swaged, sbusd, smild, totamusd, totedbsd, totincsd, totinrod, totinvd*)
- Uses NLSY versions for *Rtotwead*, *Rtotamud*, and *Rtotamusd* for all years including any non-survey years, these variables are imputed like all other components as well
- All cleaning and modifications are done in **imputeincomeREV4.sas**

a) Earned Income (Excludes business income)

waged: Total income received during the previous calendar year by the respondent from wages, salary, commissions, or tips from all jobs, before deductions for taxes.

swaged: Total income received during the previous calendar year by the respondent's spouse from wages, salary, commissions, or tips from all jobs, before deductions for taxes. (spouse only 1979-93, spouse or partner 1994+)

mild: Total income received during the previous calendar year by the respondent from service in the military before taxes and deductions. It includes special pays, allowances and bonuses.

smild: Total income received during the previous calendar year by the respondent's spouse from service in the military before taxes and deductions. It includes special pays, allowances and bonuses. (1980+, amounts probably included in *swaged* for 1979) (spouse only 1979-93, spouse or partner 1994+)

b) Non-earned income contributing to AGI

busd: Total income received during the previous calendar year by the respondent from own farm or own non-farm business, partnership or professional practice (after expenses).

sbusd: Total income received during the previous calendar year by the respondent's spouse from own farm or own non-farm business, partnership or professional practice (after expenses). (spouse only 1979-93, spouse or partner 1994+)

totamud,*Rtotamud*: Total amount of unemployment compensation received during the previous calendar year by the respondent. (available all years including non-survey years; R version slightly revised by NLSY personnel)

totamusd, *Rtotamusd*: Total amount of unemployment compensation received during the previous calendar year by the respondent's spouse. (available all years including non-survey years) (spouse only 1979-93, spouse or partner 1994+; R version slightly revised by NLSY personnel)

totinrod: Total amount of money received by the respondent (or wife/husband) from any source excluding other categories, such as savings, payments from Social Security, net rental income or any other regular or periodic sources of income. (spouse only 1979-93, spouse or partner 1994+)

totinad:
> 1979-82: total amount received for alimony and child support in previous calendar year
> 1983-93: total amount received for alimony in previous calendar year (child support is reported in *totincd*)
> > • We do not included this in unearned income, since it's missing after 1993

c) Untaxed/other income sources

totinvd: Total amount of other veteran benefits, worker compensation or disability payments received by respondent (or spouse) during previous calendar year. (1980+) (spouse only 1980-93, spouse or partner 1994+)

totwead, *Rtotwead*: Total welfare amount received by the respondent (or spouse) from AFDC, Food Stamps, Supplemental Security Income or Other Public Assistance (welfare income from local, state or federal government). (available all years including non-survey years) (spouse only 1979-93, spouse or partner 1994+; R version slightly revised by NLSY personnel)

totedbd: Total educational benefits received during the previous calendar year by the respondent. It includes educational benefits for veterans under the G.I. Bill the V.E.A.P. any other kind of scholarships, fellowships or grants.

totedbsd: Total educational benefits received during the previous calendar year by the respondent's spouse. It includes educational benefits for veterans under the G.I. Bill the V.E.A.P. and any other kind of scholarships, fellowships or grants. (spouse only 1979-93, spouse or partner 1994+)

totincd: Total amount received for child support during the previous calendar year by the respondent from someone living outside the household. (1982+; before 1982, child support is included with alimony in *totinad* during early years)

totincsd: Total amount received for child support during the previous calendar year by the respondent's spouse/partner from someone living outside the household. (1993+; before 1993, this is included in *totincd*)

totinrd: Total amount of money received by the respondent (or wife/husband) from persons living outside household (if living in DU), outside her home in city of permanent residence (if living in Dorm, fraternity or sorority) or not living with respondent (if in military). (1979-84)

d) Income earned by other family members (non-spouse, non-partner) living in household

totinaod: Total income received by related adults that also live in the household (excluding spouse and children). It includes Aid to Families with Dependent Children, SSI, other public assistance or welfare from local, state or federal government, unemployment compensation, worker's compensation, or veteran benefits. (1980+)

totinard: Total income before taxes or deductions from all regular sources received by related adults that also live in the household (excluding spouse and children). It includes income from full or part-time job, net income from own farm, net income form non-farm own business, partnership or professional practice, income from Social Security or pensions and any other regular or periodic sources. (1980+)

e) Income from non-spousal partner

totinasd: Total income from all sources received by person of the opposite sex/partner living with respondent before taxes or deductions. It includes income from full or part-time job, net income from own farm, net income form non-farm own business, partnership or professional practice, payments from Aid to Families with Dependent Children, SSI, other public assistance or welfare from local, state or federal government, unemployment compensation, worker's compensation, income from Social Security or pensions and any other regular or periodic sources. (1979-89)

totinpmd: Total income received during the previous calendar year by the respondent's partner from service in the military before taxes and deductions. (1990-93)

totinpwd: Total income received during the previous calendar year by the respondent' s partner from wages, salary, commissions, or tips from all jobs, before deductions for taxes. (1990-93)

totinpfd: Total net income received during the previous calendar year by the respondent's partner from own farm or own non-farm business, partnership or professional practice. (1990-93)

totinapd: Total welfare amount received by the respondent's partner from AFDC, Food Stamps, Supplemental Security Income or Other Public Assistance (welfare income from local, state or federal government). (1990-93)

f) Other measures

totinpd: Total income from person (and all family members that live with him/her) that pays at least half of living expenses (other than spouse) and does not live with respondent. (1979-88)

totmkve, ttlmkve: Total market value of any property or money from any estates, trusts, inheritances or gifts from relatives or friends. (1988+: 1988-92 *totmkve*, 1993-2000 *ttlmkve*)

<u>Created income components (revised May 2010)</u>

wagemild: all wage and military income *(waged + mild)*

swagemild: all wage and military income for spouses (not partners) *(swaged + smild)*

childcare *childsuppRS:* Total child support by respondent and spouse (not including partners) combining *totincd* and *totincsd* (and *totinad* for 1979-82)

totincP: Total partner income created from *waged, sbusd, smild, totamusd, totedbsd,, totinrod,* and *totinvd* for 1994+ and from *totinasd, totinpmd, totinpwd, totinpfd,* and *totinapd* for earlier years.

earnincR: Total earned income *(waged+mild)*

earnincS: Total earned income for spouses only *(swaged+smild)*

earnincRS: Total earned income for respondent + spouse (does not include partner income)

unearnincRS: Total unearned taxable income for respondent + spouse (does not include partner income) from *busd, sbusd, Rtotamud, Rtotamusd,* and *totinrod* (does not include alimony income for any years since its unreported after 1993)

nontaxincRS: Non-taxable income for respondent + spouse (does not include partner income) from *totinvd, Rtotwead,* and *childsuppRS*

taxincRS: Total taxable income for respondent + spouse (does not include partner income) from *earnincRS* and *unearnincRS*

totincRS: Total income for respondent + spouse (does not include partner income) from *nontaxincRS* and *taxincRS*

Notes on income imputations (imputeincomeREV4.sas)

I. All measures with a valid skip are set to zero before any other imputations take place. Any values greater than 200,000 are set to missing.

II. Imputations for all survey years are made based on other non-missing observations using three methods:

a. An individual-specific regression is run on an intercept, age and age-squared. When at least 8 non-missing values are available for an individual and the true value is missing for a year, the predicted value is used from the regression. Only observations when the individual is age 22 or older are used. This method is used for *waged, mild, wagemild, swaged, smild* and *swagemild*.

b. An individual-specific regression is run on an intercept and age. When at least 6 non-missing values are available for an individual and the true value is missing for a year, the predicted value is used from the regression. Only observations when the individual is age 22 or older are used. This method is used for *busd, sbusd, totinrod,* and *totincP*.

c. An individual-specific mean is computed for all available years. When at least 4 non-missing values are available for an individual and the true value is missing for a year, the average is used. Only observations when the individual is age 22 or older are used. This method is used for all variables not listed in a or b above.

All variables end in "imp" after this imputation.

III. All missing income values for 1995, 1997, and 1999 are imputed using the average of adjacent years (if there is at least one reported value).

All non-aggregated variables end in "imp2" after this imputation (any versions with "imp" ending that are updated here are dropped from **impmomkid** data file). All aggregated measures use these imputations (note: only affects 1995, 1997, 1999 relative to "imp" measures) even though they end in "imp":

earnincRimp, earnincSimp, earnincRSimp, unearnincRSimp, nontaxincRSimp, taxincRSimp, and *totincRSimp.*

VI. We set any income measures (imputed or not) to missing if they are either zero or greater than 200,000.

Imputing marital status (marrsepimp,marrlyimp)

This is done in a preliminary program **vars4incimputing (marrsepimp** is described but **marrlyimp** is similarly generated, however separation is coded as a zero):

1. First, we set *marrsepimp=1* if marital status in the previous survey is married, separated, or remarried. (In 1979, we set *marrsepimp=0* if first marriage was after 1978 and set *marrsepimp=1* if first marriage ended in 1978.)

2. Next, we check whether current marital status is 'never married'. If so, we set *marrsepimp=0*.

3. Next, we check whether the person has been married in past and has never been divorced as of the current period. If so, we set *marrsepimp=1*.