

2022 Parent-Child ID Matrix

Data Set Name: 2022_Parent-Child_ID_Matrix

Unit of observation: individual

Number of observations: 18,551

Number of variables: 5

Unique identifier: sampleid

Data collection date: 1996 to 2022

Date of last edits: 6/18/2026

2022 Parent-Child ID Matrix

sampleid	Respondent's identification number.		
	[001001003-151029004]	Respondent ID	18551
mother_id	Respondent's mother's identification number.		
	[001001001-151029003]	Mother ID	18009
		Missing	542
father_id	Respondent's father's identification number.		
	[001001002-151027004]	Father ID	15991
		Missing	2560
flag_agegap	Subjects with one or more parents who are less than 10 years older than the subject.		
	0	No parent <10 years older than subject	18534
	1	>=1 parent <10 years older than subject	17
flag_samesex	Subjects with one parent excluded from the data file because that parent's gender is the same as the other parent's gender.		
	0	No parent excluded for same gender	18548
	1	One parent excluded: same gender as other parent	3

2022 Parent-Child ID Matrix

APPENDIX

Purpose and Structure: The parent-child matrix is a supplemental file designed to be merged with any analytic file containing CVFS data. As shown above, it is a simple file with 5 variables: the subject ID, the ID(s) of the subject's parent(s), as well as flag variables that provide additional information about the parent-child links.

The purpose of this file is to facilitate identification of the mother and father IDs for each individual in an analytic dataset. Parent-specific characteristics can then be merged from other datasets, enabling multigenerational analysis. An illustration of the merging process is shown below, along with sample code for SAS.

File Construction: The parent-child matrix was constructed using the 1996, 2000, 2008, and 2017 Household Relationship Grid datasets. The file was subsequently updated to incorporate 305 months of household registry data collected February 1997 – June 2022.

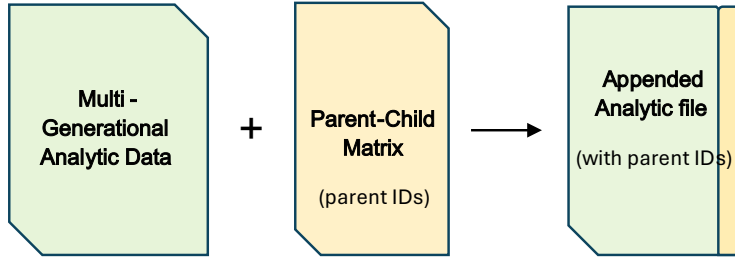
The relationship grid datasets included gender-specific parent ID variables for mothers and fathers. In contrast, the household registry data identified parents using non-gender-specific variables, “parent1” and “parent2”. To assign these parent IDs to mother and father variables, each listed parent ID was linked back to the corresponding respondent record for that parent, where available, in order to retrieve the parent's age and gender. Once parent gender was identified, the household registry parent records were merged with the relationship grid parent-child pairs, and “parent1” and “parent2” were assigned as mother or father IDs based on gender.

When inconsistencies were identified across data files, information from the most recent data file was prioritized. Some error in the household relationship and household registry data is expected, as the subject was often not the respondent providing the information. When parent-child relationship information is central to an analysis, we encourage data users to consult datasets in which subjects provided the information themselves, when available. Additionally, the parent ID variables do not necessarily indicate biological parents, which may help explain some discrepancies or unusual parent characteristics.

2022 Parent-Child ID Matrix

Illustration of the Merging Process:

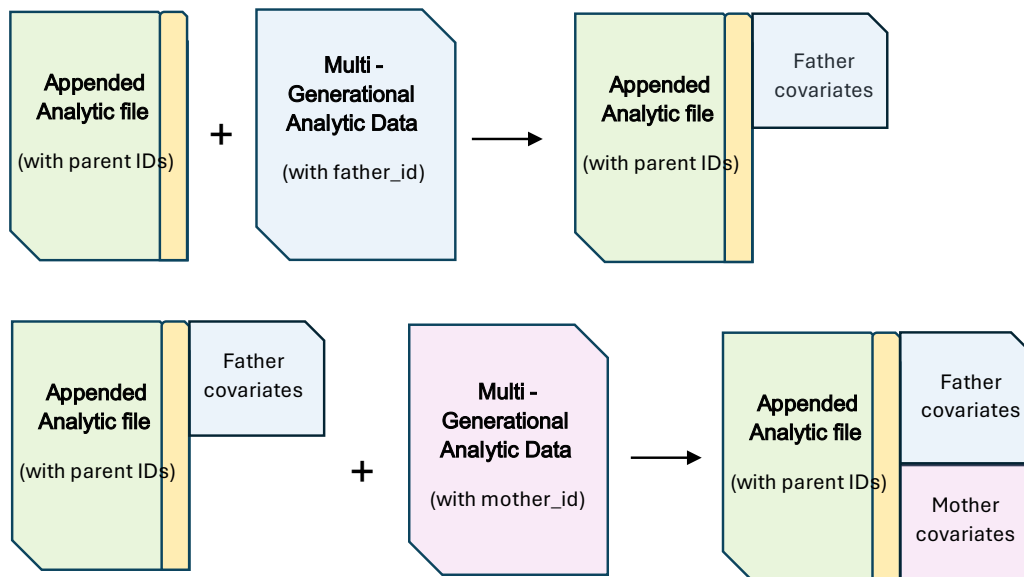
Step 1: Merge parent IDs to your analytic dataset (using SampleID variable)



Step 2: Create 2 copies of the analytic dataset, once renaming SampleID to "mother_id" and again renaming SampleID to "father_id"



Step 3: Merge the 2 copies of the analytic data with Appended Analytic File, once by each ID, to append father and mother-level variables.



2022 Parent-Child ID Matrix

Sample SAS Code for Merging:

```
*Sample code for appending parent covariates onto an existing analytic
dataset;
*Start with an analytic dataset of your choice, that you wish to add
parent-oriented variables onto (e.g. mother/father education,
mother/father mental health, etc.);

/*Prep step -- sort the parent-child reference file and your analytic
dataset by sampleid;*/
proc sort data=ParXCh.parentchild_omnibus; by sampleid; run;
proc sort data=analyticdata; by sampleid; run;

/*STEP 1: Merge Parent-Child Data onto your analytic dataset, which
adds mother ID and father ID to the "child" row*/
data merged_data;
  merge ParXCh.parchild_omnibus_HGQC_202605022 analyticdata(in=a);
  by sampleid;

  if a; /*this will filter back down to the casebase of your analytic
dataset*/
run;

/*STEP 2: Create 2 copies of the merged dataset one for mothers and one
for fathers (using childn as the example father-level var*/
data fatherdata; *drop mother_id and father_id from this dataset to
avoid collisions;
  set merged_data;
  drop father_id: mother_id;; *drop mother and father id from the input
dataset, this step creates a father-level file with no parent info;
  rename sampleid = father_id; *replace the old father_id variable with
the current sampleId, as this is the father-centric file, and these
will be appended back to the main file as father-centric variables;
  rename childn = father_childn; *rename all variables of interest with
"father_" prefix so as not to collide with the same variable at the
"child" level;
  keep sampleid childn; *keep variables of interest (the ID var and any
renamed variables -- in SAS since we are renaming these in this
datastep, the keep statement uses the original variable names, not the
new names;
run;

*create same dataset, but for potential mothers;
data motherdata; *drop mother_id and father_id from this dataset to
avoid collisions;
  set merged_data;
  drop father_id: mother_id;; *drop mother and father id from the
input dataset, this step creates a mother-level file with no parent
info;
  rename sampleid = mother_id; *replace the old mother_id variable with
the current sampleId, as this is the mother-centric file, and these
will be appended back to the main file as mother-centric variables;
  rename childn = mother_childn; *rename all variables of interest with
"father_" prefix so as not to collide with the same variable at the
"child" level;
  keep sampleid childn; *keep variables of interest (the ID var and any
renamed variables -- in SAS since we are renaming these in this
```

2022 Parent-Child ID Matrix

```
datastep, the keep statement uses the original variable names, not the
new names;
run;

/*STEP 3: Merge parent data back onto the original merged dataset*/
*3a: merge on father covariates;
proc sort data=merged_data; by father_id; run;
proc sort data=fatherdata; by father_id; run;
data merged_father; *this step adds father covariates back to the
original data;
  merge merged_data(in=a) fatherdata;
  by father_id;
  if a; *again, subset back down to original analytic casebase;
run;

*3b: merge on mother covariates to the output of 3a;
proc sort data=merged_father; by mother_id; run;
proc sort data=motherdata; by mother_id; run;
data Final_data; *this step adds mother covariates back to the original
data;
  merge merged_father(in=a) motherdata;
  by mother_id;
  if a; *again, subset back down to original analytic casebase;
run;
```