Analysis was conducted using Stata 10.0 Windows 64-bit, but the programs work in Stata 13.0 as well.

A. Analysis programs

1) Tab1_col1col2.do uses Couples2000.dta and Couples1960.dta to produce first two columns of Table 1.

2) Tabl1_col3col4col5.do uses CouplesNLSY.dta to produce remaining columns of Table 1.

3) Tab2.do uses CouplesNLSY.dta to produce Table 2.

4) Tab3_men_1960_col1col2.do uses Couples1960.dta to produce the results for men in 1960 in columns 1 and 2 of Table 3.

5) Tab3_men_1960_col3col4.do uses Couples1960.dta to produce the results for men in 1960 in columns 3 and 4 of Table 3.

6) Tab3_men_1970_col1col2.do uses Couples1970.dta to produce the results for men in 1960 in columns 1 and 2 of Table 3.

7) Tab3_men_1970_col3col4.do uses Couples1970.dta to produce the results for men in 1960 in columns 3 and 4 of Table 3.

8) Tab3_men_1980_col1col2.do uses Couples1980.dta to produce the results for men in 1960 in columns 1 and 2 of Table 3.

9) Tab3_men_1980_col3col4.do uses Couples1980.dta to produce the results for men in 1960 in columns 3 and 4 of Table 3.

10) Tab3_men_1960_col1col2.do uses Couples1960.dta to produce the results for men in 1960 in columns 1 and 2 of Table 3.

11) Tab3_women_1960_col3col4.do uses Couples1960.dta to produce the results for women in 1960 in columns 3 and 4 of Table 3.

12) Tab3_women_1970_col1col2.do uses Couples1970.dta to produce the results for women in 1960 in columns 1 and 2 of Table 3.

13) Tab3_women_1970_col3col4.do uses Couples1970.dta to produce the results for women in 1960 in columns 3 and 4 of Table 3.

14) Tab3_women_1980_col1col2.do uses Couples1980.dta to produce the results for women in 1960 in columns 1 and 2 of Table 3.

15) Tab3_women_1980_col3col4.do uses Couples1980.dta to produce the results for women in 1960 in columns 3 and 4 of Table 3.

16) Tab4 uses Couples_addhealth.dta to produce Table 4.

B. Analysis data sets from Census data

There are 4 analysis data sets using Census data: Couples1960.dta, Couples1970.dta, Couples1980.dta, Couples2000.dta that are called by the analysis programs listed in section A above.

1) Ipums downloads

The 4 analysis data sets are generated from the following provided ipums downloads: full1960.dta, full1970.dta, full1980.dta, full2000.dta and 1970fm1_fm2.dta. In all cases, the variable names and labels in Stata format correspond to the original variables on the ipums website. In all cases, the sample is limited to ages 25-60.

a) full1960.dta is downloaded from the 1960 1% sample on ipums.

b) full1970.dta is downloaded from the 1970 1% form 1 metro sample on ipums.

c) full1980.dta is downloaded from the 1980 5% state sample on ipums.

d) full2000.dta is downloaded from the 2000 5% sample on ipums.

e) 1970fm1_fm2.dta is a combined download from the 1970 1% form 1 metro sample and the 1970 1% form2 metro sample. This larger combined sample is used to obtain larger sample sizes when calculating wage characteristics by detailed occupation code.

2) Hours and Weeks of work imputation for 1960 and 1970.

The 1960 and 1970 ipums data only report hours/week and weeks of work in intervals. Therefore, a point estimate of hours/week and weeks of work is imputed based on 1980 data. The program 1980hours_weeks.do uses full1980.dta to calculate average hours of work by age group and sex, producing the output data sets avghours1980.dta and avgweeks1980.dta. These two output data sets are used to predict hours per week and weeks of work for the Couples1960.dta and Couples1970.dta analysis data sets. Therefore, the 1980hours_weeks.do program must be executed before running the other programs to generate Couples1960.dta and Couples1970.dta, described below.

3) Couples1960

Couples1960.dta is generated from full1960.dta by running, in order, the programs 1960_step1.do, 1960_step2.do, 1960step3.do. A number of intermediary data sets are produced in the process.

a) 1960_step1: merges husband records with wife records creating the couples sample.

b) 1960_step2: merges hours and weeks of work from the 1980 sample into the couples sample. This program calls the avghours1980.dta and avgweeks80.dta data sets described in section (B.2) above.

c) 1960_step3: calculates average occupational wages separately by sex and education and merges them into the couples sample, producing the final version of Couples1960.dta for analysis.

4) Couples1970

Couples1970.dta is generated from full1970.dta, 1970fm1_fm2.dta by running, in order, the programs 1970_step1.do, 1970_step2.do, 1970_step3.do.

a) 1970_step1: merges husband records with wife records creating the couples sample.

b) 1970_step2: merges hours and weeks of work from the 1980 sample into the couples sample. This program calls the avghours1980.dta and avgweeks1980.dta described in section (B.2) above.

c) 1970_step3: calculates average occupational wages separately by sex and education and merges them into the couples sample. This program calls the 1970fm1_fm2.dta sample to calculate average occupational wages.

5) Couples1980

Couples1980.dta is generated from full1980.dta by running, in order, 1980_step1.do and 1980_step2.do. There are only 2 programs because it is no longer necessary to merge in inputed hours and weeks of work information.

6) Couples2000

Couples2000.dta is generated from full2000.dta by running 2000_step1.dta. There is only one program because Couples2000 is only used to produce descriptive statistics in Table 1 of the paper, so additional variables created for the previous analysis data sets (such as occupational wages) are not necessary.

C. NLSY data

CouplesNLSY.dta is generated from NLSYfull.dta by running NLSY_couples.do. NLSYfull.dta was downloaded from the NLSY79 data. Variable names and labels in the NLSYfull.dta data set identify the relevant variables in the NLSY79 data.

D. Add Health Data

Add health analysis in Table 4 was produced using restricted access files. The process for obtaining access to the restricted access files can be found at http://www.cpc.unc.edu/projects/addhealth/data/restricteduse

The program file addhealth_data.do generates the analysis data set Couples_addhealth.dta using the following restricted-access data files:

Wave 1 in-home interview
Wave 3 in-home interview
Wave 4 in-home interview
Wave 4 Sect16b relationship file
Wave 4 Sect16c relatioship file
Wave 1 in-home weights

The addhealth_data.do file includes keep statements for each raw data file that lists all of the required variables from each file.

All of the variables used in our analysis are also available in the public use files (one variable, bmi, will need to be calculated from the height and weight variables). Therefore, by downloading the variables listed in the keep statements in addhealth_data.do, it is possible to run these programs on the public files. The sample sizes, however, will be smaller, so doing so will not replicate the results in the paper.