

Tutorial 7

Young adult educational outcomes:

Matching grandparents and grandkids

Ngina Chiteji, Mohammad Mushtaq and Frank Stafford

Introduction

Is there a connection between young adults' educational levels and the level of education attained by their grandparents? Parent-child correlations have been observed in many domains (Behrman, Pollack and Taubman, 1995; Solon, 1992; for example). Researchers have argued that such associations exist partly because of demand side factors, such as parents imparting values to their children. For example, it is argued that highly-educated parents, whose own level of schooling reflects their valuing education, work to make sure that their offspring have similar attitudes toward educational attainment. Scholars also have noted that because post-secondary schooling requires substantial financial investment. In this context we can think of supply side factors, lowering the cost of education to young adults who come from well-to-do families (Becker 1967).

It seems natural to expect that grandparents also might exert some influence over young adults' educational outcomes and aspirations, or that they too might have an effect on a family's ability to make human capital investments in its youngest generation. This tutorial will show you how to obtain data that can be used to determine whether there is a correlation between individuals' early educational outcomes and the education level of their grandparents', focusing specifically on the grandchildren's college enrollment after age 18. In the process, the tutorial will show you how you might set up an analysis to examine other correlates of college enrollment as well. Or this tutorial can provide guidance on how to carry out studies of other intergenerational patterns. In an earlier tutorial ([Tutorial #6](#)) we show how to match labor market outcomes of adults (Baby Boom sons) and their parents (the dad's of the Baby Boomers) at a comparable life course point. <http://psidonline.isr.umich.edu/Guide/tutorials/IG/IG.aspx>

I. Using PSID for Generational Research

The PSID has supported unique scientific contributions to intergenerational research. The PSID bibliography includes over 250 papers on the topic of "intergenerational effects," with the full listing available at the PSID web site:

<http://psidonline.isr.umich.edu/Publications/Bibliography/default.aspx>

Here we provide a brief review of some of the highlights based on PSID and other sources.

Researchers have utilized the genealogical design of the PSID (<http://psidonline.isr.umich.edu/Guide/ug/stdydsgn.html#OVERALL DESIGN>) to examine the intergenerational transmission of income (Solon, 1992; Zimmerman, 1992) wealth (Charles and Hurst, 2003), and expenditures (Charles, Danziger, Li, and Schoeni, 2006). Solon in particular

has published a series of studies that concluded that the intergenerational transmission of economic status is substantially higher – with a correlation of at least 0.4 – than previously believed. And in recent work he has compared the U.S. with other developed countries finding that mobility is no greater in the United States. Intergenerational transmission has more recently been investigated using the new health data collected in the main survey instrument and the Child Development Supplement. This work finds strong intergenerational connections in obesity (Davis, McGonagle, and Schoeni, and Stafford 2006) and asthma (Andreski, McGonagle, and Schoeni, 2006).

PSID data have been used for some important work on intergenerational transfers. Altonji et al. (1996; 1997) examined theories of parental helping behavior and *inter vivos* assistance. They found that money transfers tend to reduce inequality in household incomes and that time transfers are only weakly related to income differences. Among parents and in-laws, the richer set of parent is more likely to give money and less likely to receive money. Richer siblings give more to parents and receive less. In contrast to the implications of simple exchange models of transfers, there is little evidence in the cross section or in the analysis using siblings that parental income or wealth raises time transfers from children or that time transfers are exchanged for money transfers. Work along this line includes that of Jayakody (1998), Wilhelm (1996), and Couch et al. (1999). Furstenberg et al. (1995) examined the effects of divorce on intergenerational transfers of money and time and found no evidence that divorced fathers who paid child support are more likely to be involved in intergenerational transfers than those who did not pay child support. These results support a growing body of evidence that marital disruption is altering the organization of kinship in American society. When men relinquish ties to their children during childhood, they rarely resume those ties in later life. Smock and Manning (1997) found that the characteristics of nonresident parents are central to understanding levels of child support.

There are several questions in the demographic literature that could not have been studied without PSID or other longitudinal data following parents and children over a long period of time. These questions include the effects of parental divorce on children (e.g., McLanahan and Sandefur, 1994), the relationship between teenage childbearing and the teens' subsequent social and economic outcomes (e.g., Hoffman et al., 1993; Geronimus and Korenman, 1992), and that effects that neighborhoods have on children's development independent of family characteristics (Brooks-Gunn et al., 1993).

II. PSID's Genealogical Sample Design

The authors of these and other intergenerational studies have chosen to use the PSID in large part because of its unique design. When the PSID began in 1968 a decision was made to follow the children of sample families as they grew up and left to form their own households. To follow such 'splitoffs', it was argued, would offset the problem of panel attrition and provide a way to refresh the sample with young families as the initial panel members aged. And by continually adding these young families, a panel study could provide a continuous, self-representing sample of the U.S. population. The PSID investigators realized that including the newly formed families of children who left to live on their own would provide continued representation of such young families, and it would support the study of early adult experiences of children from different

economic backgrounds and poverty exposures (Hill 1992). This genealogical-based design, which allows the study to maintain the representation of the young (with weights) and to study the effects of family background, has defined the PSID sample design ever since.

In effect, the PSID follows a bloodline, not a person. The fundamental identifier for every individual is the *1968 family ID* ("68 ID") in combination with their *person number* ("PN"). These variables uniquely identify any individual ever in the PSID and are the primary keys in the relational data archive. Use of this bloodline sample of individuals opened up the eventual study of intergenerational and life course connections. When played out over nearly 40 years, these design features enable the PSID to provide data on a national cross-section of families and individuals from 1968 forward, but excluding effects of immigration, which were corrected for in 1997/1999 as described in the discussion on <http://psidonline.isr.umich.edu/data/Documentation/ImmigSample.pdf>. Most importantly, the study design has led to the collection of panel data over what is now a major portion or the majority of the lives of many respondents and across generations.

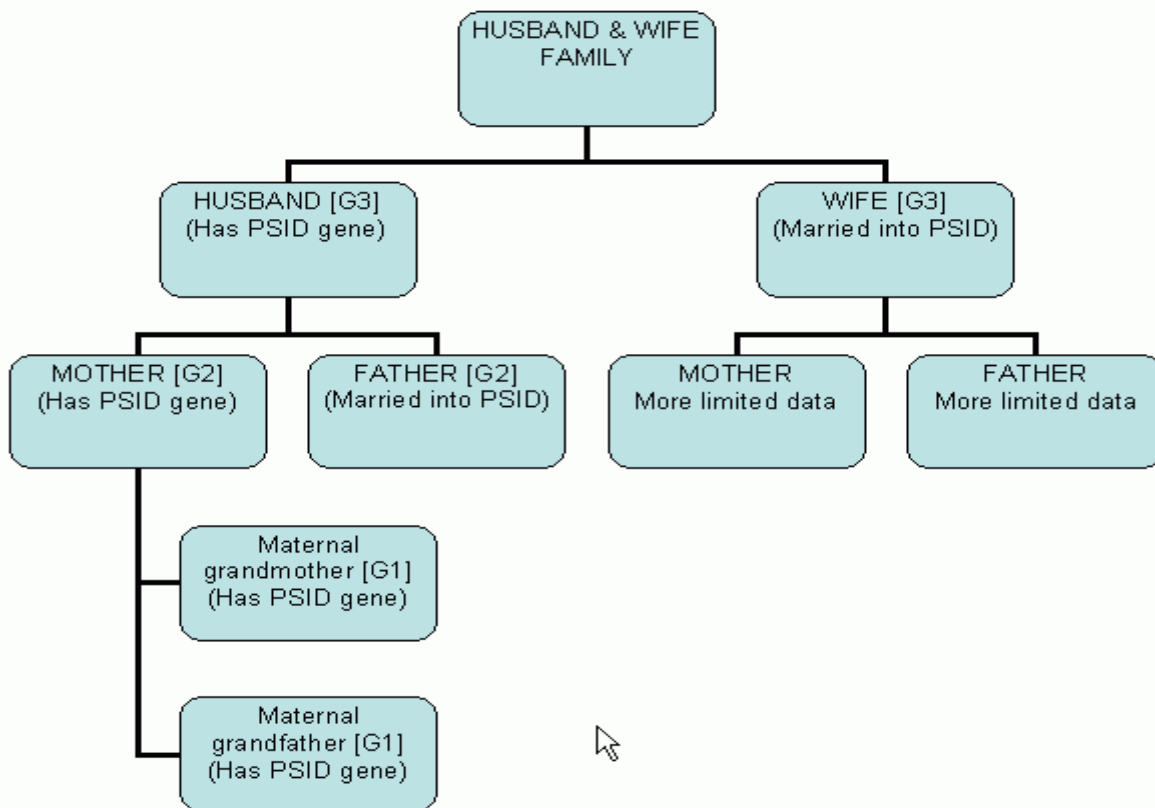
The PSID contains about 40,000 unique sibling pairs over the life of the study. In the 2003 wave alone, there are nearly 4,000 heads and wives who have a sibling who is also a head or wife. And the Child Development Supplement provides an additional 925 sibling pairs for children 5-18 as of 2002/2003. Moreover, 3,027 heads or wives have at least one parent who was also interviewed in 2003.

More recently there have become substantial increases in the availability of data on three generations within the same PSID families. This has arisen from the fact that many of the individuals who were children in the original 1968 interviews have had children themselves, and these children have become adults and left home. For example, 749 heads or wives in 2003 have at least one parent and one grandparent with data collected in 2003. In addition, the introduction of the Child Development Supplement in 1997, 2002/03 and 2007/008 accelerated the availability of three-generation information. Extensive data were collected on nearly 3,600 children ages 0-12 living in PSID families in 1997. These children were re-interviewed at ages 5-18 as of 2002/03, with reinterviews of these children in 2007/08. These children are in the age range where many of their grandparents are the early edge of the baby boom and making retirement plans or are in retirement already. Moreover, especially as the children in the CDS have aged, the investigators have spent substantial time making sure the CDS measures parallel the measures collected in the core PSID. As of 2005 and forward the CDS children who have become 18 or older have been followed in a project known as Transition to Adulthood (or TA, for short). With measures of the children's financial and health behavior, for example, one can study the extent to which their health and economic outcomes in young adulthood is related to not only their parents but also their grandparents characteristics. In the CDS sample alone, 3,025 children have at least one grandparent who has been part of the PSID.

The PSID follows people who have the 'PSID gene', and the only way to have the gene is to be part of a 1968 family or to be born or adopted into such a family. As a result, spouses who marry individuals with the PSID gene are not followed themselves (unless they live with someone with the gene). Moreover, the person who marries a PSID sample member is missing prospectively reported data on their family history. In the example in Figure 5.1, the wife has 'married in' to

the PSID. For both the wife and her husband, detailed information is available, and especially so for the husband since he has the PSID gene. Let's call this generation 3 (G3) to indicate that this couple is a third generation of PSID family. For the wife, there is limited recall-based information about her parents, which is collected at the point she became active in the PSID sample (i.e., when she married the person with the PSID gene). The husband, however, is the son of a couple who was interviewed in the PSID, and information on both his mother and father is available over a substantial portion of their lives. We will call his parents generation 2 (G2). His mother, and not his father, has the PSID gene. And in this illustrative example, his maternal grandmother and maternal grandfather were a married couple in 1968 and interviewed as part of the original 1968 PSID family. This couple we call generation 1 (G1).

Figure 1: A PSID Family Across Generations



The point of Figure 1 is to demonstrate that even with the rich genealogical design of the PSID, the cross-generational information is selectively available. In terms of the full set of measures in the PSID adult core, the sample goes down one branch of an adult couple's genealogy one generation back, then down a single branch two generations back, and so on. As a result, the extensive set of characteristics of grandparents is oftentimes from a single pair or a single one of four biological grandparents. The impact of long generational processes is limited unless the marriage patterns are of similar individuals. At the same time, the PSID investigators have been mindful of the value of generational information from both sides of the family tree. A regular data collection in every wave is an extensive set of background questions for 'new heads' and 'new wives' who marry a person with the 'PSID gene'. These questions are part of the interview

each wave and include early work history, education, religious preference, experience growing up (urban rural, parental occupation industry and education) and number of brothers and sisters of the new head or wife.

III. Grandchildren Becoming Adults – How Do Grandparents and Early Childhood Matter?

The exercise outlined below presents an opportunity to use the PSID's new Transition to Adulthood (TA) datafile. The TA data were collected as part of a supplement to the Panel Study of Income Dynamics (PSID). This supplement contains observations on over 700 young adults who are the offspring of families in the PSID core. In 2005 these young adults were asked a series of questions, including questions about schooling, labor market participation, military service, religious affiliation, friendships and other relationships. For a more complete picture, visit <http://psidonline.isr.umich.edu/CDS/wavesdoc.html>.

Moreover, because the TA data can be linked back to data from the PSID's Child Development Supplement (CDS) <http://psidonline.isr.umich.edu/CDS/>, there are opportunities to examine aspects of the young adults' early life experiences that might affect their college attendance, viewing television instead of using one's free time for educational activities for example. Accordingly, this tutorial will make use of the CDS' time diary data to determine whether spending substantial time watching tv during one's formative years negatively influences the probability of attending college.

A. Goals of the exercise

The specific objective of this tutorial will be to assemble a dataset that allows one to conduct analyses designed to be able to fill in the following tables. As a starting reference point you may go to: <http://psidonline.isr.umich.edu/CDS/TA05-Overview.pdf>

Table 1. Correlations in educational outcomes across generations

	Young adults in college	Young adults with no college experience
% with grandparents who went to college		
% whose grandparents did not go to college		
average education level of grandparents		

Table 2. Influence of television on college attendance

	Young adults in college	Young adults with no college
Average time spent watching television as a child		

B. Before you get started

If you are a new PSID user, please note that the PSID requires users to register prior to downloading data in order to impart our Conditions of Use. Through your registration, we also are able to convey to our sponsors the size of our user community, allowing us to continue to collect these important data. Registration is very easy, and is required only once. If you do not already have a PSID login and password, you will want to register at the Data Center *before* you begin this tutorial. To do such, click on the word "login" that appears in the far right upper corner of the "Welcome to the Data Center" screen (under the day's date). Once you have registered, you can return to the Data Center to get going on the tutorial.

C. Assembling the dataset

To assemble our dataset, we will draw from several "files" that are available at the PSID online Data Center: (a) the Transition to Adulthood (TA) file <<http://psidonline.isr.umich.edu/CDS/TA05-Overview.pdf>>, (b) the Child Development Supplement's Time Diary Aggregate file <<http://psidonline.isr.umich.edu/CDS/timediary.html>>, and (c) the Family Identification Mapping System (FIMS) <<http://simba.isr.umich.edu/FIMS/>> and (d) the family-level data files for 2005.

1. Selecting variables from the TA module and from the Time Diary Aggregates file

We will start by obtaining the information about young adults, including that about their television-viewing habits during childhood. To get the data characterizing the young adults, we need to retrieve variables from the TA module and the Time Diary Aggregates <<http://psidonline.isr.umich.edu/CDS/timediary.html>> module at the same time while at the Data Center.

Step 1 Variables from the Transition to Adulthood (TA) module

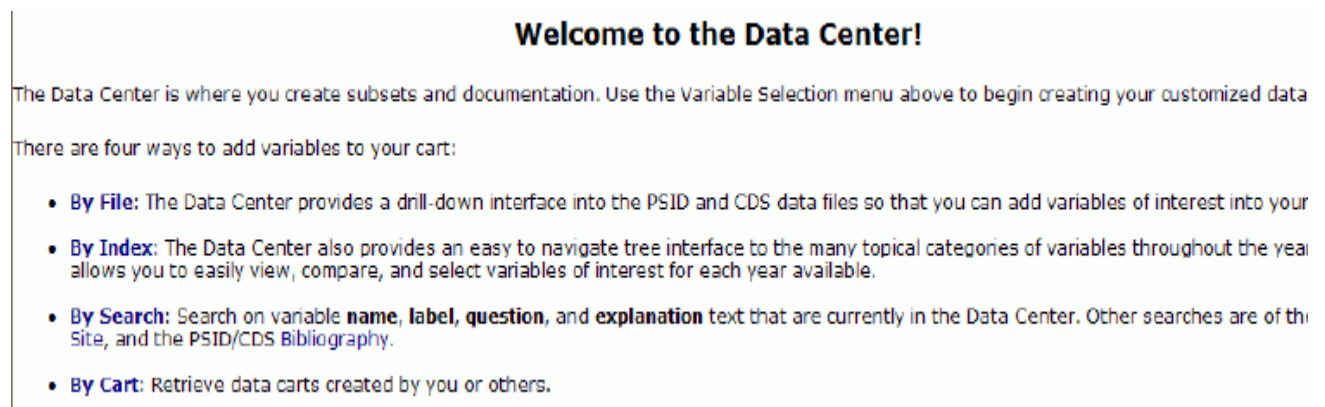
From the TA file we want seven variables. For young adults' educational outcomes, we will focus on college attendance since the modern economy places a great premium on higher education. The variables measuring young adults' college attendance are TA050594 and TA050595. The first variable records the young adult's answer to question G10, about whether the young adult ever attended college, including 2-year colleges. The second contains the response to question G11, about whether the young adult currently is enrolled in a post-secondary institution. To view the actual questions you can check out the TA questionnaire at <http://psidonline.isr.umich.edu/CDS/quescodetable.html>.

The TA file also includes measures of mothers' and fathers' years of schooling, TA050947 for moms and TA050949 for dads, and additional variables to indicate the year in which the parent's schooling was measured (TA050948 for moms and TA050950 for dads). We will select these variables as well in order to compare the association between young adults' college attendance and grandparents' schooling to that of the parent-child correlation in schooling outcomes.

Finally, we need the variable [TA050955](#). This is a statistical weight that ensures that analyses based on the data are nationally representative.

To get these 7 variables we go first to the Data Center's main page at <http://simba.isr.umich.edu/>. If you see a screen like the following, you'll know you're in the right place.

Screenshot 1. The Data Center's 4 options for getting data



Welcome to the Data Center!

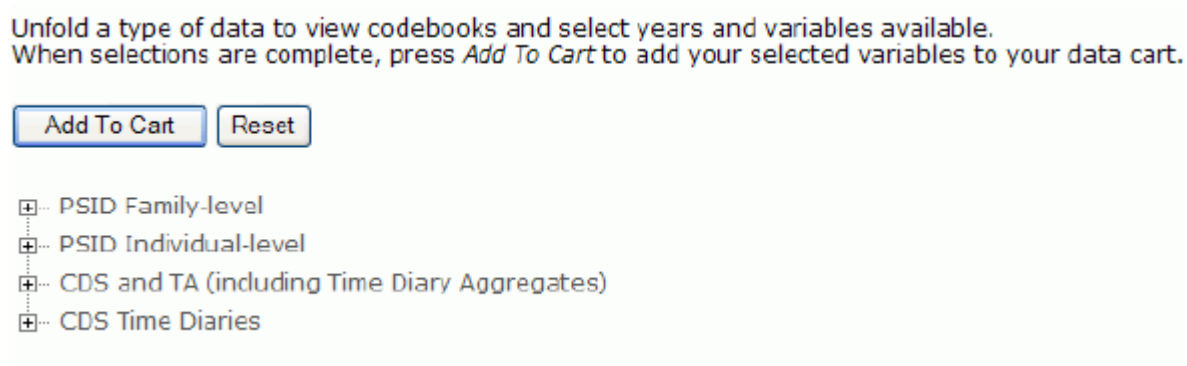
The Data Center is where you create subsets and documentation. Use the Variable Selection menu above to begin creating your customized data

There are four ways to add variables to your cart:

- **By File:** The Data Center provides a drill-down interface into the PSID and CDS data files so that you can add variables of interest into your cart.
- **By Index:** The Data Center also provides an easy to navigate tree interface to the many topical categories of variables throughout the year allows you to easily view, compare, and select variables of interest for each year available.
- **By Search:** Search on variable **name**, **label**, **question**, and **explanation** text that are currently in the Data Center. Other searches are of the Site, and the PSID/CDS Bibliography.
- **By Cart:** Retrieve data carts created by you or others.

Next click on "By File." This takes you to a screen with four different types of files that one can access, as shown in the screenshot below.

Screenshot 2. The "By File" data selection option



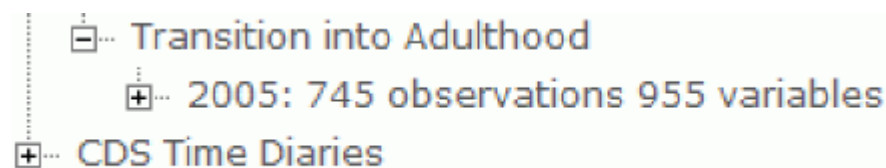
Unfold a type of data to view codebooks and select years and variables available. When selections are complete, press *Add To Cart* to add your selected variables to your data cart.

- PSID Family-level
- PSID Individual-level
- CDS and TA (including Time Diary Aggregates)
- CDS Time Diaries

This page provides a drill-down interface into the PSID and CDS data files so that you can add variables of interest.

If you click on the + sign next to any of the options listed, you will see an expanded list. (Note that clicking on the - sign that appears after one expands a list allows one to shrink the list back down.) For example, if we click on the + sign to the left of the phrase "CDS and TA" we are presented with several more choices. At the bottom of the choice list you'll see the phrase "Transition to Adulthood." These are the data that we want, so we need to click on the + sign next to the phrase, so that we will be able to select the 7 variables that we want from this data file. Clicking on the + sign next to 2005 then reveals a box that lists all of the variables that are part of the Transition to Adulthood file. This is shown in screenshot 3.

Screenshot 3. The variable selection box for the Transition to Adulthood module



Now scroll through the box and select the 7 variables that we need: (1) TA050594, (2) TA050595, (3) TA050947, (4) TA050949, (5) TA050948, (6) TA050950, and (7) TA050955. To these seven, we will add two more that are helpful for matching young adults with their families: (a) TA050003, the 2005 PSID family id number, and (b) TA050004, the young adult's "individual sequence number" for 2005. To select more than one variable hold down on the control key and use your cursor to highlight each variable you wish to select. (Note: Do not release the control key until you have finished selecting all your variables!!! If you do, you will lose all but the last variable that you highlighted.)

After you have selected all your variables click on the "Add to cart" box at the top of the page. This will instruct the Data Center to create a datacart with the variables you selected in it. This is like a shopping cart at the grocery store—one that's full of items you pulled off the shelves as you head to the cash register to pay.

After clicking "add to cart" you should see a new box just below that phrase that says "variables added to your cart." Clicking on this option allows you to doublecheck to make sure you have all the variables you need (the ones you think you selected). If you click on the phrase "your cart" you'll be taken to a screen that shows you the variables that you selected. When you "expand the list" you'll see that the Data Center gave you the 9 variables you selected, plus 5 others. The Data Center automatically adds the individual's 2005 relationship to head (ER33803), and ER30001 (the 1968 interview number) and ER30002 (person number-68). The latter two are important because they can be used to uniquely identify any individual that ever appears to the PSID. Another way of stating this is to say that they can be used to construct an identification variable for each young adult. And, each young adult's id will be unique. Moreover, because it is unique it allows one to use the id variable to trace the young adult throughout all the years that he or she has been part of the PSID.

Step 2 Adding the variables from the Time Diary Aggregates file

To get information about the time spent watching television during childhood, we need to turn to the Time Diary Aggregate file. It contains information from both weekday and weekend time diaries for two calendar years (1997 and 2002). Because every young adult respondent in the TA file was once a CDS participant, the time diary data that was collected as part of the CDS offers us useful information about each young adult's background, particularly early developmental experiences that the young adult would have had while he or she was a child growing up.

How do you get the data from the Time Diaries? Right now you are at a screen that shows you the variables you selected from the TA module. You now need to instruct the Data Center to let you add variables to this cart—variables that will come from the Time Diary Aggregates file.

To do this, look toward the top, left portion of your screen and move your cursor to the phrase "Data Center." A list of options should appear including "variable selection." Highlighting this phrase displays additional options, including "by file." If you click on this phrase, you're returned to the screen that shows the different files that the PSID has, from which variables can be selected. (This is basically what you saw in screenshot 2 above.)

Now expand the options under "CDS and TA (including Time Diary Aggregates)" by clicking on the + sign to the left of this phrase. Next, instead of scrolling down to the phrase "Transition into Adulthood" stop at the option just above it. That should be "Time Diary Activity—Aggregate file." Clicking on the plus sign at this point calls up two calendar years, 2002 and 1997. As you learned above, if you click on the + sign next to the year, a box will appear that includes all the variables collected for that particular year. For 2002, you want the variables WD02_919 and WE_02_919. From 1997, you want the variables WD97_919 and WE97_919. (If you're wondering about the similarity in the variable names, the prefix "wd" indicates a weekday time diary, while "we" represents a time diary recorded on the weekend. And, 919 is the activity code for watching tv.) Remember that you need to hold down on the "ctrl" key while highlighting your desired variables since you wish to select more than one.

After you have selected these 4 variables you need to hit "add to cart." This will add these 4 variables to your existing data cart (the one with our 9 analysis variables from the Transition to Adulthood file.) It also takes you to a screen that allows you to click on "variables added to your cart" so you can doublecheck to make sure you have everything you need.

Step 3 Checking out

You are now ready to check out. Hit the "check out" box toward the top of the screen, and the Data Center will take you to a screen that allows you to select output options. That screen should look like the one below (screenshot 4). Note that this screen reminds you how many variables you have selected at the top (there should be 24 variables in your dataset). Then, it allows you to specify the format in which you would like your codebook and your dataset. The codebook that the Data Center will send you is a booklet that contains the variables you selected with their codes and accompanying questions. (PDF is the recommended format). As you can see, there are

several options for how the dataset is delivered. You can have the data put into Excel format, or SAS, Stata or SPSS format. In this tutorial we will use SAS.

ALSO, note that you want to click on the box next to the phrase "CDS kids only." This will restrict the dataset to cases (observations) for individuals who were in the CDS. Since all of the young adults in the Transition to Adulthood module were once CDS respondents, we want to impose this restriction. (That way we don't get information for every single individual who ever appeared in the PSID.)

ADDITIONALLY, you want to enter some text in the subsetting command box. Because we do not want data for all CDS participants, but only those who have grown up and become part of the TA module, we need to restrict our dataset. To tell the Data Center to only send records for our TA young adults we need the following subsetting command:

```
TA050955 ^= .
```

(Here the sequence of 3 symbols " ^=. " represents a SAS designation for 'does not have a blank' value for the variable.)

This tells the Data Center to only send records for individuals who have a positive statistical weight. Anyone who is not in the TA module will have a missing value for this variable. (All TA respondents have a positive weight.)

Finally, NOTE that you can instruct the Data Center to send you a link that can be used to retrieve your codebook, dataset and SAS starter commands by clicking on the "send links to my e-mail" box. This would allow you to hit submit, and to leave your computer and return later to work on your analysis. If you do not select the e-mail option, the Data Center will simply assemble your dataset while you are here at the PSID website, and it will display the files so that you can download them from the webpage you are looking at right now (as shown in screenshot 5 below). Hit submit to see!

Screenshot 4. Output options screen

Your file will contain **24** variables

Codebook Type

HTML PDF XML No thanks, I don't need a codebook

Codebook only (no data will be output)

Data Output Type [?]

ASCII Data With SAS Statements Microsoft Excel Spreadsheet

ASCII Data With SPSS Statements dBase Data File (DBF) / Spreadsheet

ASCII Data With Stata Statements SAS V9 Transport Data File

Subsetting Criteria [?]

TA050955 ^-.

Subsetting of CDS and Family Members [?]

All Individual Data From CDS Families

CDS Kids Only

Compression Option [?]

Create files in compressed (.zip) format for faster downloading

Cart Options [?]

Make my cart available to the public

Why do you have 24 variables even though you only actively selected 13 (nine from the TA file and four from the Time Diary module)? You have additional variables because there are some variables that the Data Center automatically adds to your dataset. As noted earlier, when you select variables from the TA module, the Data Center automatically includes additional variables that can be used to construct unique identifiers for our young adults. Similarly, when we select data from the Time Diary file, the Data Center automatically includes some key variables that can be used to make sure that children who are in the CDS can be properly matched to the families from which they originate. The Data Center is organized this way because most users may need these variables to do their analysis and to check the matches made from different files to create their dataset. These are not variables users consider to be data, but are instead helpful if any problems arise concerning who is in the file from the Data Center. ("Better safe than sorry!" is the mantra of the Data Center.)

Screenshot 5. Downloading your data

Data Center >> Data Cart >> File Links

File Type	File Size (KB)
Customized Codebook Right-click and choose <i>Save As</i> to download.	138
SAS Statements	3
Rectangular ASCII Text Data	59

Your customized data contain 745 observations and 24 variables, created Monday, Jul 28 at 6:50 PM.

Output more than 7 days old will be deleted from this server

In addition to the customized codebook, we highly recommend that you also download the complete documentation:

- [Packaged Main Data, Documentation, and Questionnaires](#)
- [Packaged Supplemental Data and Documentation](#)
- [CDS Documentation and Questionnaires](#)

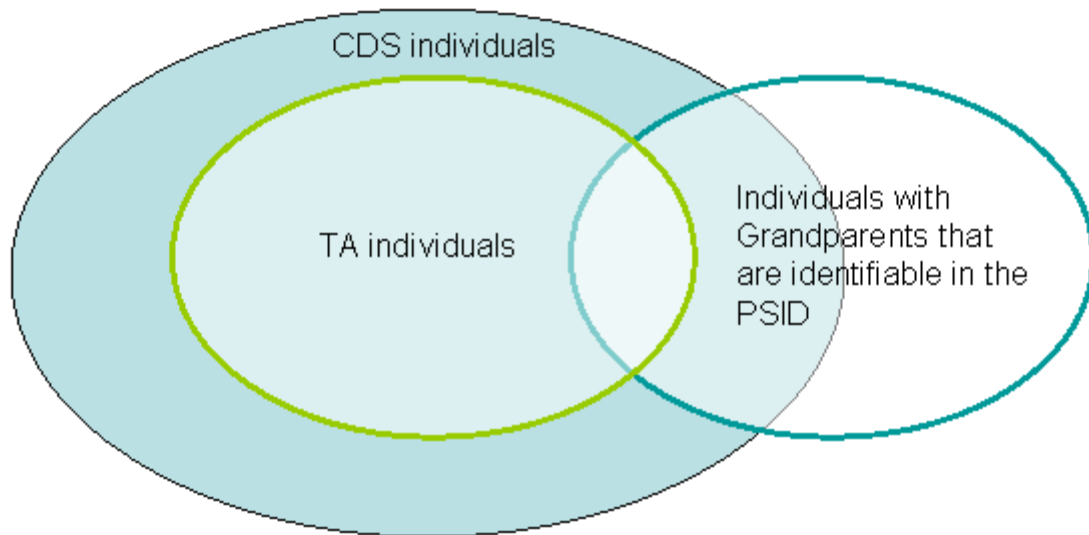
Right click on the phrase "customized codebook" to download your codebook. Rightclick on the "SAS Statements" phrase to download a starter program that will allow you to input your data into a SAS dataset. Then right click on the ascii file to download the data for SAS to read in. Make sure you put these files in a convenient location so you have access to them later. We will return to all of these files later. Before doing so, however, we need to turn to the challenge of determining who the grandparents of our young adults (TA respondents) are.

2. Retrieving the grandparent ids from FIMS

In addition to the variables obtained from the TA file and from the Time Diaries there is the issue of information about the grandparents and parents - possibly in their roles as PCG's and OCG's. We have an Excel file which provides some of this {[link to CDS and TA Map file: http://psidonline.isr.umich.edu/help/cds_ta_map.xls](http://psidonline.isr.umich.edu/help/cds_ta_map.xls)}, but for a more widely useable tool you will need to use the PSID's Family Identification Mapping System (FIMS) so that you can match young adults to their grandparents. The TA module does not contain any information about grandparents. Accordingly, to find out how much education the grandparents of our young adults received one has to look to data from the PSID core. The PSID "family files" for each year contain information about education for heads and wives for all households in the dataset, so we can use these data to assign information about grandparents of our young adults. First, however we need to determine which households correspond to grandparents of our young adult sample.

As Figure 2 shows, not all youth in the TA file will have grandparent information available for them. While all TA youth are the descendents of original sample members (discussed in Section II) and many are likely to be 3rd generation PSID members (G3 in Figure 1), some may be second generation PSID members, meaning that their parents are original sample families. (Consider, for example, the case of a TA respondent who was 22 in 2005, whose parents had children late. If the parents are in their 60s in 2005, meaning they had the child in their 40s, these parents would be the original family that the PSID sampled in 1968.) Other TA youth may have grandparents that are identifiable in the PSID but for whom no education data was collected (in rare instances in which this question was not answered by the respondent, for example).

Figure 2.
TA youth in relation to CDS and select PSID core individuals



Step 1 Extracting a FIMS dataset

To get your FIMS dataset you can start at the PSID's homepage and then click "Data and documentation," or you can go directly to <http://psidonline.isr.umich.edu/data/>. In the right column of this page, you will see the words "Family Identification Mapping System (FIMS)." If you click on that phrase, you'll be directed to the FIMS portion of the PSID website. FIMS will ask you for your login information before you can select a file from it. Once you have logged in, you'll see a screen like the one depicted below.

Screenshot 6. The Family Identification Mapping System (FIMS)

The genealogical sample design of the PSID means that the parents, grandparents, and siblings of respondents are respondents themselves in some waves of the data. This feature supports unique analyses of the inter- and intra-generational connections and relationships within families. To facilitate creation of data extracts that include information on parents, grandparents, and siblings of respondents, we have designed the tool on this web page.

To understand how the tool works, review the User Help examples.

Select Map Type... Select User Help...

Options

Output Options

- SAS V9 System File
- Excel
- ASCII With SAS Statements
- ASCII With SPSS Statements
- ASCII With Stata Statements

Submit

At this point you need to select a map type. We want an inter-generational map, since we want to match individuals from the PSID who are from different generations (grandparents and grandchildren). After you have selected "inter-generational" from the map type box, more options appear. For parent type select biological and adoptive. Under "generation map" choose individual to grandparents, under "map type" choose unbalanced. For file format, select "wide." This will give you a file that has one record per young adult, with the information for the young adult's grandparents appearing in the columns associated with the young adult's row. More specifically, it will be the grandparents' 68 ID and PN numbers that appear in the columns. You will then use these to create a unique identifier for each grandparent. If you're an avid tutorial user, you probably recall that FIMS was discussed extensively in [Tutorial 6](#).

Next choose SAS for the output option (as shown in the screenshot below—screenshot 7). Then check the box for "merge with CDS map," which will instruct FIMS to restrict the dataset to individuals who were in the CDS (thereby eliminating a lot of observations that we do not need). Then hit submit.

Screenshot 7. FIMS output options screen

The genealogical sample design of the PSID means that the parents, grandparents, and siblings of respondents are respondents themselves in some waves of the data. This feature supports unique analyses of the inter- and intra-generational connections and relationships within families. To facilitate creation of data extracts that include information on parents, grandparents, and siblings of respondents, we have designed the tool on this web page.

To understand how the tool works, review the User Help examples.

Inter-generation (GID) ▾	Select User Help... ▾
Options	
Parents: Biological and Adoptive ▾	Output Options <input type="radio"/> SAS V9 System File
Generation Map: Individual to Grandparents ▾	<input type="radio"/> Excel
Map Type: Unbalanced Map ▾	<input checked="" type="radio"/> ASCII With SAS Statements
File Format: Wide - 1 obs. per individual ▾	<input type="radio"/> ASCII With SPSS Statements
	<input type="radio"/> ASCII With Stata Statements
	<input type="button" value="Submit"/>
<input checked="" type="checkbox"/> Merge with CDS Map -- Wide Format	

The dataset that FIMS creates for you will contain 3563 records—one for each child that appears in the Child Development Supplement. Because the young adults in the TA file are a subset of the individuals who were once surveyed as dependent children—in the CDS—this dataset that FIMS created of CDS individuals mapped to their grandparents serves as a useful starting point for us.

Once you've hit the "submit" key it will take a few seconds for FIMS to create your dataset. Wait patiently! In a few seconds, links should appear that allow you to download your FIMS file and a codebook along with a SAS starter program to help you convert the ascii data into a SAS dataset. (You download by right-clicking like you did above to get the TA data.)

3. *Selecting the Grandparent education data*

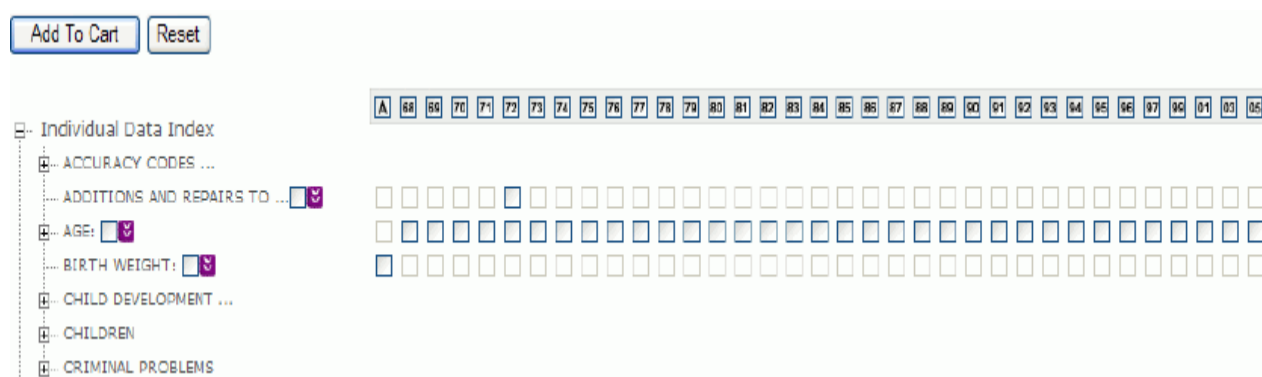
The PSID routinely collects data on heads' and wives' levels of educational attainment. For example, the 2005 PSID core contains a variable indicating what the education level of the head of household is (ER28047) and a variable indicating what the wife's level of education is (ER28048). We will use these data to construct a variable indicating whether the head of the household ever attended college, and a variable indicating whether the wife ever attended. These variables then will be used to assign education levels to the grandparents. Most individuals who are grandparents should show up as heads and wives of their own households in the PSID, so the head/wife educational variables offer the ability to assign information about educational attainment to our young adults' grandparents.

To obtain these data, we need to use the PSID's cross-year index. Where will we find this index? We go back to the Data Center home page at <http://simba.isr.umich.edu/>. (If you've forgotten what this page looks like, go back and take a second look at screenshot 1 above.) This time,

however, we need to choose the option of selecting "by index." A cross-year index exists for individual-level data and for family-level data. We will use both of these.

Start by clicking on the + sign next to the phrase "individual data index." A list of all the variables in this index will appear. (See screenshot 8 below, for example. While it does not display all of the variables that appear under the individual data index--the list actually runs from A to W--it does show the first few variables, so that you can determine if you have expanded the list correctly.)

Screenshot 8. Using the individual data index to select data



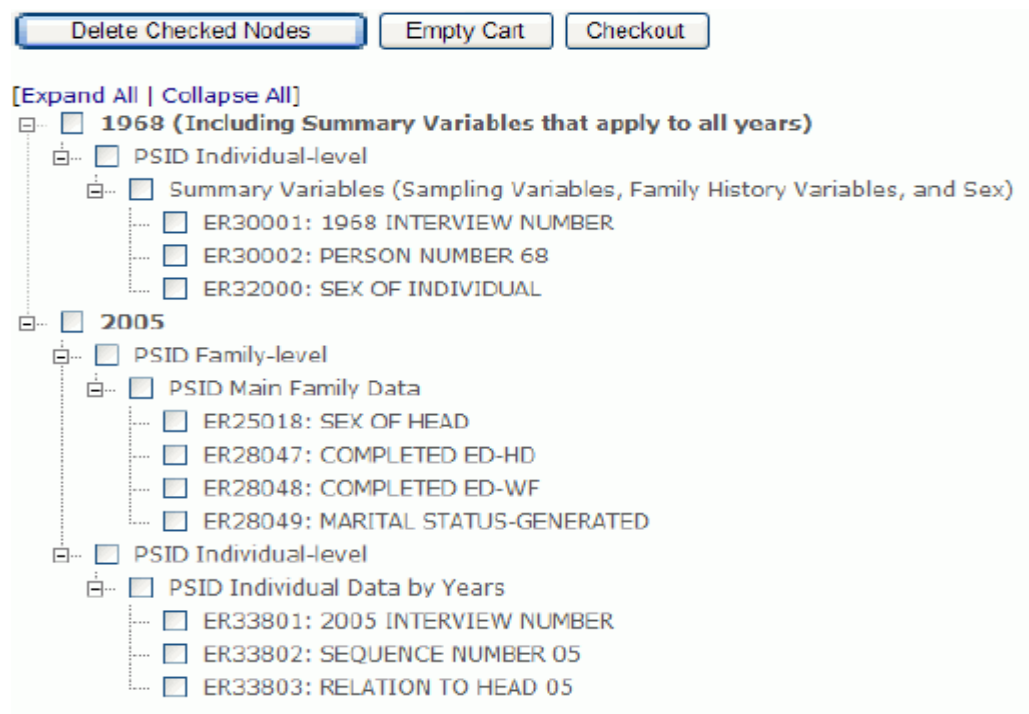
What do we need here? We need a series of variables that will allow us to determine whether the individuals whose records we will obtain were heads or wives in our year of interest (2005). As noted earlier, the PSID contains data for a substantial number of individuals. Most of the data are collected via family interviews however (not separately by individual). This means the interviewer sits down with the head (or wife) of the family and asks questions about the family. Accordingly, the information collected is recorded as "head's education" or "wife's education level," for example. So in order to properly assign the education information to specific individuals we have to know (a) which family a given individual was in during our year of interest, and (b) whether he/she was head or wife. Which variables do we want here? We want (1) the individual's PN or "*person number*," (2) a variable called "*relationship to head*," and (3) a variable called "*sequence number*." (See [Tutorial 1](#) for an extensive discussion of the concepts of head, wife, and relationship to head.) We also want (4) ER25018 so that we know what sex the household head is, and ER28049 to tell us whether the family represents a married household or not. (This will help us figure out who is a grandmother and who is a grandfather, and on which 'side' of the family.) You can find these variables by moving down the list of variables contained in the index, look for the section with words starting with the letter "P." When you find these variables, click on the box next to the phrase "person number" to select it. Then—for relationship to head and for sequence number—we want to select the variables from the calendar year 2005. This means you need to click on the box to the far right of each row. This will be the box corresponding to the 2005 wave of the PSID. (As you can see, the cross-year index contains information that on the different variables listed that spans a range of years—from 1968 to the present.) To get the gender variable you simply click on the box next to the word "sex."

Next we want to expand the list of options under the "family data index" option. Do this by clicking on the "+" sign next to that phrase. This should cause a list of several variables to appear. Next, scroll down to the word "education." Click on the "+" sign next to that word and several types of educational information will appear. Next click on the "+" next to the phrase "head and wife." Then click on the "+" sign next to the phrase "grades completed." Then click on the "+" sign next to "including college." Then, click on the boxes for the year 2005, for both heads and wives. This will add the variables ER28047 and ER28048 to your data cart. FINALLY, we want one last variable from this index. We need the variable that tells us whether the head of the household is male or female. This variable is named ER25018 in the 2005 wave of the PSID. (Note that this is a different sex variable from the one we selected above.) You can find it by clicking on the plus sign "sex" in the family data index. Then choose "head" and click on the box to the far right of the row—the box that corresponds to the year 2005.

NOW--a BIG STEP—hit "add to cart" and the Data Center will construct a data cart for you with these 2 education variables in it, along with sex of the head of the household variable, and the person number, sequence number and relationship to head and sex variables that you selected from the individual-index.

Screenshot 9 below shows all the variables that should be in your datacart.

Screenshot 9. Expanded variable list



We now want to proceed to the checkout point, so that we can tell the Data Center to create our customized dataset, and provide instructions on where to send it. The steps you take for this are similar to what you did above to retrieve data from the TA module and from the Time Diaries. Hitting "add to cart" should take you to a screen that allows you to review the list of variables in

your shopping cart. After checking that to make sure you have everything, you hit "checkout" at the top right.

At this point you should be directed to a screen displaying output options (as shown below in screenshot 10), and you have to tell the Data Center how you want your codebook, whether you want the dataset in SAS format (ascii data with a SAS starter program), and how you want your dataset delivered to you. You ALSO want to make sure the "all individuals" option is highlighted (as shown in the screenshot below). Additionally, you may want to instruct the Data Center to compress your files for faster delivery (by clicking on the box next to that option). You then hit submit, and the Data Center will make the dataset for you. Your dataset will have 10 variables and about 22,918 observations. We ultimately will merge this data with the grandparent information in our FIMS dataset.

Screenshot 10. Output options

Codebook Type

HTML PDF XML No thanks, I don't need a codebook

Codebook only (no data will be output)

Data Output Type [?]

ASCII Data With SAS Statements Microsoft Excel Spreadsheet

ASCII Data With SPSS Statements dBase Data File (DBF) / Spreadsheet

ASCII Data With Stata Statements SAS V9 Transport Data File

Subsetting Criteria [?]

Subsetting of Individuals [?]

All Individuals

Only heads current in *any* years selected

CDS: PCG

CDS: DCG

CDS: Kid

Compression Option [?]

Create files in compressed (.zip) format for faster downloading

Now that you have selected your data using the "by index" strategy that we identified above, you may be thinking, "Are there other options for selecting variables?" [Tutorial 1](#) discusses the different options for selecting variables at great length. Here, however, we do want to remind you that if you are wondering whether you could obtain the education data that we obtained above if (a) you did not happen to know the names of the variables that we want, or (b) if you did not know that they happen to be located under the family index, the answer is "yes." In addition

to offering the option of searching "by index," as described above, the Data Center also allows users to search by keyword. If you were to choose the "by search" option under "variable selection" (instead of by index as you did above), the Data Center would take you to a screen that allows you to type in words of interest and to select specific years for which to search for the data that interests you. Just for fun, select 2005 under the year choices and then type in "college." You'll see that you are then given a list containing a number of variables. ER28047 and ER28048 appear toward the middle of this list. With so many choices you might wonder how you would ever know that you should opt for the two variables that we chose. If you click on the purple box next to any variable name, a screen pops up to show you the actual question to which the variable corresponds. This will help you determine what variable you want--since it tells you what type of education information the variable contains. At this point it is also worth making a note about an issue that sometimes confuses users of PSID data. Near the top of the list of college-related variables you probably see "ER27321 K47 wtr recd college degree--wf." The notation in the variable name is shorthand for "whether received college degree--wife," indicating that this is a variable that tells us something about wives, and whether they have college degrees. The reason we are not using this variable in our analysis is because it represents a question that is asked of new wives only--that is to say wives who are new to the household in 2005. Any wives who already were in the PSID in previous years will have a missing value code for this variable, even though the PSID is actually likely to have education information for them. A similar caveat applies to "ER27417 L54 wtr recd college degree--hd" (a variable indicating whether the head has a college degree--if the head is new to the household in 2005). Experienced users of the PSID know that any questions in the K and L sections of the PSID questionnaire (such as "K47" and "L54" represent questions that are asked of new wives and heads for the given wave of data collection ONLY.) The ER28047 and ER28048 variables that we chose to use circumvent the problems posed by ER27417 and ER27321 because ER28047 and ER28048 represent the updated education data for all heads and wives (so they do not simply apply to heads or wives who are new to the survey in 2005).

A final note about searching options: You also could have located our two education variables of interest--ER28047 and ER28048--via a search "by file." To do this you would have needed to have chosen the "by file" option instead of the "by index" option, when prompted to indicate the strategy for "variable selection" that you wanted to use. You would then go directly to a file containing all of the family-level data contained in the 2005 wave of the PSID by clicking on the box next to the phrase "PSID family-level," and then clicking on the "+" sign to the left of the phrase "main family data," and then clicking on the "+" sign next to 2005, and scrolling through the list of variables that appears until you see ER28047 and ER28048.

D. Steps to take to conduct the analysis

To analyze the correlation between young adults' college attendance and attendance by their grandparents we have to merge our FIMS-based dataset with some of the other information that we requested at the Data Center. Several steps must be taken to accomplish this objective. In what follows we provide some simple SAS code that will help you to perform the merge and your analysis. The sample code below can be added to the starter program(s) that the Data Center sent you when you requested your dataset(s). Remember, the SAS program sent by the

Data Center contains enough code to allow you to read your data into SAS, but you always need to add additional code to it in order to analyze the data and to save it as a SAS dataset.

1. First, it makes sense to create a permanent SAS dataset from the FIMS file that we obtained above. (Let "grace" be the libname that you assigned for your SAS session.)

Basic code to create a permanent dataset from the FIMS file the user obtained:

```
data grace.TA_gp_FIMS_map;
set GID_MAP_wide_cds;
run;
```

2. Next we want to create a single variable that can be used to uniquely identify each young adult in the FIMS file. We will do this in a temporary SAS dataset, so that our original FIMS file remains intact (in case we need to use it at a later date). You can create a unique identifier variable in the PSID by using the "68 id number" and the "person number" for each individual in the PSID. Because we are working with young adults who appear in the PSID's Transition to Adulthood module, we will call our identifier variable "youngad_id." This stands for young adult's id.

```
data FIMS;
set grace.ta_gp_fims_map;
length youngad_id 8;
label youngad_id = "unique identifier for TA respondent";
youngad_id = er30001*1000 + er30002;
run;
```

So now we have a FIMS-based dataset that has records for all of our young adults, along with each young adult's unique id variable (stored in one column of the dataset), with information that will allow us to identify the young adults' grandparents in the other columns of the dataset. The next step is to use the information provided about the grandparents to construct a unique id variable for each grandparent.

3. Now, we need some simple SAS code to create grandparent identifier variables. Remember that we need to pay attention to the type of grandparent here. In the process, we will create id variables for each young adult's paternal grandfather ("grandpa_pat_id"), maternal grandfather, and the paternal and maternal grandmother.

```
data FIMS2 (keep = youngad_id grandpa_pat_id grandpa_mat_id grandma_pat_id
grandma_mat_id);
set FIMS;
length grandpa_pat_id grandpa_mat_id grandma_pat_id grandma_mat_id 8;

if ER30001_GP_AFF ^=. then grandpa_pat_id = ER30001_GP_AFF*1000 +
ER30002_GP_AFF; else if ER30001_GP_FAF ^=. then grandpa_pat_id =
```

```
ER30001_GP_FAF*1000 + ER30002_GP_FAF; else if ER30001_GP_FF ^=. then  
grandpa_pat_id = ER30001_GP_FF*1000 + ER30002_GP_FF; else grandpa_pat_id = .;
```

```
if ER30001_GP_AFM ^=. then grandpa_mat_id = ER30001_GP_AFM*1000 +  
ER30002_GP_AFM; else if ER30001_GP_FAM ^=. then grandpa_mat_id =  
ER30001_GP_FAM*1000 + ER30002_GP_FAM; else if ER30001_GP_FM ^=. then  
grandpa_mat_id = ER30001_GP_FM*1000 + ER30002_GP_FM; else grandpa_mat_id =  
.;
```

```
if ER30001_GP_AMAF ^=. then grandma_pat_id = ER30001_GP_AMAF*1000 +  
ER30002_GP_AMAF; else if ER30001_GP_AMF ^=. then grandma_pat_id =  
ER30001_GP_AMF*1000 + ER30002_GP_AMF; else if ER30001_GP_MAF ^=. then  
grandma_pat_id = ER30001_GP_MAF*1000 + ER30002_GP_MAF; else if  
ER30001_GP_MF ^=. then grandma_pat_id = ER30001_GP_MF*1000 +  
ER30002_GP_MF; else grandma_pat_id = .;
```

```
if ER30001_GP_AMAM ^=. then grandma_mat_id = ER30001_GP_AMAM*1000 +  
ER30002_GP_AMAM; else if ER30001_GP_AMM ^=. then grandma_mat_id =  
ER30001_GP_AMM*1000 + ER30002_GP_AMM; else if ER30001_GP_MAM ^=. then  
grandma_mat_id = ER30001_GP_MAM*1000 + ER30002_GP_MAM; else if  
ER30001_GP_MM ^=. then grandma_mat_id = ER30001_GP_MM*1000 +  
ER30002_GP_MM; else grandma_mat_id = .;
```

```
run;
```

What is the intuition here? We are using the information provided by FIMS to create variables that indicate whether any given young adult's grandparents were ever surveyed in the PSID. In U.S. society grandparents customarily are classified by their relationship to a child's parents. The parents of one's mother of are called maternal grandparents—more specifically, each young adult has a maternal grandfather and a maternal grandmother. The parents of a person's father would be paternal grandparents, so each young adult has a paternal grandfather and a paternal grandmother. While this social conventions implies that an individual can have 4 types of grandparents (maternal grandmother, maternal grandfather, paternal grandmother, paternal grandfather), FIMS records more detailed information about each young adult's grandparents. It tells not only whether the grandparent is paternal or maternal, but whether the grandparent happens to be a parent of a biological mom (or biological dad); or whether the grandparent is the parent of an adoptive mom (or dad); or whether the grandparent happens to be an adoptive parent of an adoptive mom or dad of the young adult. For our purposes, it is sufficient to work with the 4 major grandparent types, so we need to write SAS code to convert the detailed FIMS grandparent classification scheme into the 4 more general types. The code reproduced above accomplishes this task.

Note also that the PSID will not necessarily have data covering each grandparent. For example, some individuals might have maternal grandparents who have been surveyed by the PSID, while their paternal grandparents might not have ever been surveyed. Similarly, for some young adults the PSID may contain information about their paternal grandparents, but not the maternal side of

the family. Or, there may be information about one grandparent only, if the grandparent household that was surveyed by the PSID was a in a family consisting of a single head. Finally, in some cases, we might find that we have no information about any of the grandparents. The SAS code listed above allows for all these possibilities.

What does this step do to our dataset? As you probably recall, we started this step with a dataset that was organized so that each row represented a record for a young adult from the TA module. At the end of this step, we end up with a dataset where each row has a column containing a unique id for the young adult's paternal grandfather, and a column containing the unique id for the young adult's paternal grandmother, and a column containing the unique id for the young adult's maternal grandfather, and another column containing the unique id for the young adult's maternal grandmother.

4. Now we need to set our FIMS-based dataset aside momentarily. (You may want to save the dataset created above as a permanent SAS dataset at this point.) We are going to turn to our family level data about heads' and wives' education levels so that we can create a dataset that has individuals who ever appeared in the PSID, along with their level of education in 2005. The sample SAS code below starts with a dataset called "J83911" produced by the Data Center. (Remember, you would be adding the code listed below to the code already contained in the starter program that the Data Center sends you.)

Basic code to create a dataset that contains education records for all individuals in the PSID—using the 2005 education information available for heads and wives;

*we will call the resulting dataset "adult_ed_data";

```
data adult_ed_data (keep = indiv_id er28049 grandpa_ed grandma_ed);
set J83911;
length indiv_id grandpa_ed grandma_ed 8;
label indiv_id = " unique identifier for the individual";
label grandpa_ed = "grandpa ed level";
label grandma_ed = "grandma ed level";
indiv_id = er30001*1000 + er30002;
if er33803 = 10 and er33802 = 1 and er25018 = 1 then grandpa_ed =
er28047;
else grandpa_ed = .;
if er33803 = 10 and er33802 = 1 and er25018 = 2 then grandma_ed =
er28047;
else if (er33803 = 20 or er33803 = 22 ) and er28049 = 1 then grandma_ed
= ER28048;
else grandma_ed = .;
run;
```

*now it makes sense to recode our grandparent education variables to convert the PSID missing value code into a dot and to create a dummy variable for grandparent education;

```
data test;
set adult_ed_data;
length grandma_ed2 grandpa_ed2 8;
label grandma_ed2 = "whether grandma college degree";
label grandpa_ed2 = "whether grandpa college degree";
```

```

if grandpa_ed = 16 or grandpa_ed = 17 then grandpa_ed2 = 1;
else if 0 <= grandpa_ed < 16 then grandpa_ed2 = 0;
else grandpa_ed2 = .;
if grandma_ed = 16 or grandma_ed = 17 then grandma_ed2 = 1;
else if 0 <= grandma_ed < 16 then grandma_ed2 = 0;
else grandma_ed2 = .;

*the above text creates a dummy variable indicating whether a
grandparent has a college degree;
*it converts missing value codes from 9 to a dot and combines 0s and 5s
as no college;

run;

*now make this a permanent SAS dataset so we can use it again in the future;

data grace.adult_ed_data;
set test;
run;

```

What are we doing here? Here we are using the 2005 data covering heads' and wives' education to "determine" whether a grandfather or grandmother has a college degree or higher. At this point, we do not really know whether any given individual in the dataset called "test" is actually a grandparent; instead this is a dataset of all individuals who have ever appeared in the PSID throughout the history of the survey. However because we will eventually match the individual id variable created above to those for grandparents of TA young adults by using our FIMS-based dataset, it is ok to label our education variables as if they apply only to grandparents (since we ultimately will drop the individuals who are not grandparents from the analysis).

How are we assigning the grandparent education information? We use the relationship to head variable (ER33803) and the sex of household head variable (ER25018) and the marital status variable (ER28049) to determine whether the individual in question is a head or wife of the household. For heads we also have to check to see that they were actually in the household in the year 2005. The sequence number restriction, ER33802 = 1, accomplishes this task. Then we look at the gender of the household head. If the head is male, then we know we are dealing with a prospective grandfather (since grandfathers would show up as male heads of households in their household unit when surveyed by the PSID). If the head is female, we know that the education information is for a prospective grandmother (since grandmothers are female). Finally, if the individual is a wife in the household, indicated by the ER33803 code of 20 or 22, we know that we have another case in which the individual is a prospective grandmother. In this latter case we also check to make sure the head is married when we assign the grandmother education information.

The last thing we do while working with this dataset is to take an opportunity to recode some of the education variables. By convention, the PSID uses a "99" to indicate cases of missing values. SAS prefers to work with a dot (a ".") in these instances, so some of the text above is designed to recode the grandparent education variables (grandpa_ed2 and grandma_ed2) in order to make this change. We also do additional recoding to make our desired education measures dummy variables for whether grandfathers and grandmothers received a college degree or not. The

variables that we get from the Data Center--ER28047 and ER28048-- a measure of years of schooling. They take on values ranging from 0 (no schooling) to 17, with a 12 indicating that an individual finished high school, a value of 16 indicating 4 years of college and 17 indicating that the individual has done some postgraduate work (beyond college). We need to combine values of 16 and 17 into a "yes" for whether the individual has a college degree, and to collapse all of the other numerical responses into a "no" in our new dummy variable, with "0" representing no. (You may be asking yourself, "Hmm...does 4 years really mean an individual has earned his or her college degree?" That's a very good question to ask in today's world, since research suggests time to completion of degree has increased for today's college students, with many students taking 5 or even 6 years to finish college [Fitzpatrick and Turner]. However, the numeric codes are based on survey questions which ask not for years of schooling but ask for high school or college completion rather than a simple question of how many years of school. Individuals who say that they've graduated from college are then assigned a "16" (so PSID respondents are reporting actual receipt of degrees here). Go to <ftp://ftp.isr.umich.edu/pub/src/psid/questionnaires/q2005.pdf> and scroll to Questions K37-K48.

At the very end of the above code, we included a command to make the adult education level dataset a permanent SAS dataset. We do this so we can return to it again in the future.

5. For the next step we need to merge the adult education information into our FIMS-based dataset that has grandparent ids for each of the young adult's grandparents (for all young adults in the TA module). To assign a grandparent's education level correctly we will have to merge in the information about each individual's college attendance using the grandparent id variable, and we will have to do separate merges for each grandparent type. Intuitively what we are doing is taking a dataset that tells us about the education level of every individual in the PSID, and then looking to see whether any given individual in that file is a grandparent of one of our young adults. (If the answer turns out to be yes, we assign the education info to the young adult—in a column that stores info about the education received by the grandparent, more specifically whether the grandparent earned a college degree).

Basic code to merge in the grandparent education information;

*The goal here is to end up with a master dataset called "merged4" that has grandparent education information assigned for each type of grandparent (each of the 4 types).

*first sort the FIMS-based dataset by the unique id variable for paternal grandfathers;

```
proc sort data=FIMS2;  
by grandpa_pat_id;  
run;
```

*next sort the dataset with education info for all adults by each adults unique id variable;

```
proc sort data=adult_ed_data;
```

```
by indiv_id;  
run;
```

*now merge the adult ed variable into the FIMS-based dataset;

*here we are merging in ed info for PATernal grandFATHERS;

```
data merged1 (rename (grandpa_ed2 = grandpa_pat_edu) );  
merge FIMS2 (in=f2) adult_ed_data (rename = (indiv_id = grandpa_pat_id ) );  
by grandpa_pat_id;  
if f2=1 then output;  
run;
```

*now sort the merged dataset by id variable for paternal grandMOMs;

```
proc sort data=merged1;  
by grandma_pat_id;  
run;
```

*now merge the adult ed info in for paternal grandmothers;

```
data merged2 (rename = (grandma_ed2 = grandma_pat_edu) ); merge merged1  
(in = m1) adult_ed_data (rename = (indiv_id = grandma_pat_id) );  
by grandma_pat_id;  
if m1 = 1 then output;  
run;
```

*now sort the merged dataset by the id variable for maternal grandFATHERS;

```
proc sort data=merged2;  
by grandpa_mat_id;  
run;
```

*now merge in the maternal grandfathers education info;

```
data merged3 (rename = (grandpa_ed2 = grandpa_mat_edu) ); merge merged2 (in  
= m2) adult_ed_data (rename = (indiv_id = grandpa_mat_id) );  
by grandpa_mat_id;  
if m2 = 1 then output;  
run;
```

*now sort the merged dataset by maternal grandmoms;

```
proc sort data=merged3;  
by grandma_mat_id;  
run;
```

```

data merged4 (rename = (grandma_ed2 = grandma_mat_edu) );
merge merged3 (in = m3) adult_ed_data (rename = (indiv_id = grandma_mat_id)
);
by grandma_mat_id;
if m3 = 1 then output;
run;

```

*now rename the dataset to add language indicating that it contains ed info of grandparents;

```

data gp_ed_data;
set merged4;
run;

```

Where are we now? How close are we to the end of the analysis? Well, we haven't actually done any analysis yet, but we are almost done constructing our master dataset to use for our analysis. Now that we have reached the end of our current step (step 5), we have a dataset that contains records for all of the young adults in the TA module, with each young adult's grandparent education information assigned to him or her. So, now we will be able to tell whether a given young adult's grandparents went to college.

6. What's next? We probably want to add two new variables in the dataset we created above. We can use one to measure the average rate of college attendance among a young adult's grandparents (like whether 2 out of 4 went to college). We can use the second to tell whether at least one grandparent went to college.

*the resulting dataset will be called gp_ed_complete;

```

data gp_ed_complete;
set gp_ed_data;
length gp_ed_avg gp_ed_max 8;
label gp_ed_avg = "grandparent college attendance average";
label gp_ed_max = "whether at least one gp attended college";
gp_ed_avg = mean (grandma_mat_edu grandma_pat_edu grandpa_mat_edu
grandpa_pat_edu);
gp_ed_max = max (grandma_mat_edu grandma_pat_edu grandpa_mat_edu
grandpa_pat_edu);
run;

```

7. Guess what else we need to do our analysis though? If you're thinking—"the information about whether the young adult has ever attended college, silly"—you are correct. Our final step in this master dataset creation phase involves merging in the information about the young adults themselves, into the modified FIMS-based file that we finalized in step 6. In order to merge these 2 separate datasets we're going to need to sort each one by the youngad_id variable. Here is some sample SAS code that should get you moving. It results in a merged dataset called "TA_and_gp_ed," where the name is designed to indicate that this is a master dataset containing

transition to adult data for our young adults along with their grandparents' education information.

```
proc sort data=gp_ed_complete;
by youngad_id;
run;

data TA_data (drop = er30001 er30002);
set Jxxxxx;
length youngad_id 8;
youngad_id = er30001*1000 + er30002;
run;

proc sort data=TA_data;
by youngad_id;
run;

data TA_and_gp_ed;
merge TA_data (in = ta) gp_ed_complete;
by youngad_id;
if ta=1 then output;
run;
```

8. Now we are ready to start the analysis.

Before you can perform any analysis however, you need to recode some variables. The PSID uses values of 98 and 99 to indicate instances in which there are missing values for the parental education variables (TA050947 and TA050949). Accordingly, we need to recode these to create dummy variables for moms' and dads' college attendance. The other values these two PSID variables can take on are (i) values from 1 to 16, with 16 indicating completion of college; (ii) a value of 17 for anyone who has postgraduate training (beyond college--such as a masters, a law degree, an MD or a Ph.D); and (iii) a value of zero if the young adult's parent was unidentifiable, or if the mom or dad was institutionalized, or if the parent actually has no education. Despite the fact that 0 represents a valid response if the parent has no education, because determining whether the 0 represents this particular situation (instead of a missing parent, for example) requires a great deal of additional work, we will treat all zeroes as missing values for the purposes of this tutorial. (In the case of the mother's education variable there are only 7 observations with a zero value, so we are not losing many observations by ignoring these cases.) This means we need to recode the two parental education variables to create two dummy variables where values of 13 to 17 are converted to a value of 1 for each dummy variable (indicating the parent has some college), and 1-12 years of schooling correspond to a value of 0 for the dummy variable (indicating a parent who did not attend college), with missing values indicated by a dot (".").

We also need to recode the young adult college attendance variable. This variable (TA050594) can take on 5 possible values: (a) 1 for yes, (b) 5 for no, (c) 8 for don't know, (d) 9 for refused to

answer, and (e) 0 for individuals who did not finish high school. We want to use this information to create a dummy variable that gauges young adults' college attendance. Accordingly, the "0" and "5" responses need to be grouped together, while the "8" and "9" responses need to be recoded with the SAS missing value symbol (a dot ".").

Finally, we also want to recode the variables from the time diary. For these variables (the ones pertaining to tv watching), the PSID assigns a value of 99999 for instances of missing data. We want to recode these into dots (".") for SAS. All other values are valid data, reported in seconds of tv time. (You may want to re-code to convert the seconds to minutes. If so, divide by 60 as we do below.)

In what follows we give sample SAS code that will help you recode the variables. Once you have recoded them you are ready to do your analysis.

```
data test;
  set TA_and_gp_ed;

  length tv_wkday97 tv_wkend97 tv_wkday02 tv_wkend02 college_ya m_college
d_college 8;
  label college_ya = "young adult ever attended college";
  label m_college = "mom 13 ys of school or more";
  label d_college = "dad 13 ys of school or more";

  if wd97_919 ^= 99999 then tv_wkday97 = wd97_919/60;
  else tv_wkday97 = .;

  if we97_919 ^=99999 then tv_wkend97 = we97_919/60;
  else tv_wkend97 = .;

  if wd02_919 ^= 99999 then tv_wkday02 = wd02_919/60;
  else tv_wkday02 = .;

  if we02_919 ^= 99999 then tv_wkend02 = we02_919/60;
  else tv_wkend02 = .;

  if TA050594 = 1 then college_ya = 1;
  else if (TA050594 = 5 or TA050594 = 0) then college_ya = 0;
  else college_ya = .;

  if (TA050947 =98 or TA050947 = 99 or TA050947 = 0) then m_college = .;
  else if (12 < TA050947 < 18) then m_college = 1;
  else m_college = 0;

  if (TA050949 =98 or TA050949 = 99 or TA050949 = 0) then d_college = .;
  else if (12 < TA050949 < 18) then d_college = 1;
  else d_college = 0;

run;
```

Next, here is some simple language to get SAS to compute the rates of college attendance among young adults given whether or not their parents or grandparents attended college. We also have included commands that allow you to calculate the average level of education among the

grandparents of the young adults who have attended college so that it can be compared to the average rate of college attendance among the grandparents of young adults who did not receive college educations. Note that our calculations use the statistical weight provided by the PSID (TA050955). Remember, to generate nationally representative results, we need to weight the data!

```
proc sort data=test;
by m_college;
run;

proc means data=test;
class m_college;
var college_ya;
weight TA050955;
title1 "young ad college by mom ed";
title2 "weighted data";
run;

proc sort data=test;
by d_college;
run;

proc means data=test;
class d_college;
var college_ya;
weight TA050955;
title1 "young ad college by dad ed";
title2 "weighted data";
run;

proc sort data=test;
by gp_ed_max;
run;

proc means data=test;
class gp_ed_max;
var college_ya;
weight TA050955;
title1 "young ad college by max gp ed";
title2 "weighted data";
run;

proc means data=test;
class college_ya;
var gp_ed_avg;
weight TA050955;
title1 "avg ed among gp for different young ad";
title2 "weighted data";
run;

*now for the tv viewing analysis;
proc sort data=test;
by college_ya;
run;
```

```

proc means data=test;
class college_ya;
var tv_wkday97 tv_wkend97 tv_wkday02 tv_wkend02;
weight TA050955;
title1 "avg tv viewing time";
title2 "weighted data";
run;

```

IV. The answers for the intergenerational correlation in college exercise

Table 3. Percentage of young adults who have attended college--by college status of parents and grandparents

	Moms have at least 13 years of school	Moms never attended college
% of young adults ever attending college	80.9%	60.4%
	Dads have at least 13 years of school	Dads never attended college
% of young adults ever attending college	84.5%	59.9%
	At least one grandparent has a college degree	No grandparent has a college degree
% of young adults ever attending college	85.1%	66.0%

The table above shows that the percentage of young adults who have ever attended college is higher among young adults whose moms are college educated themselves. About 81% of kids of moms with at least 13 years of schooling went on to obtain some college themselves, while only 60% of young adults whose moms had no college went on to college. A similar observation holds for dads with college compared to fathers who do not have college degrees. Among young adults whose dads attended college, the proportion attending college is higher than it is for young adults whose dads did not attend college; in fact, there is almost a 25% gap in college attendance rates between these two groups of young adults. A similar association exists between grandparents having attended college and young adults' college attendance. The college attendance rate is highest among young adults with at least one grandparent who has a college degree (compared to young adults with no grandparent with a college degree).

The table below uses the composite measure of grandparents' education that we created. This is the variable that measures the average rate of college attendance in the grandparent generation, for each TA young adult. Remember, a grandparent who received a college degree has a code of 1, while any grandparent who did not receive a college degree has a value of 0. So, for example, a young adult who has 4 grandparent records in the PSID would get a 1 for the composite variable if all grandparents received college degrees, while a young adult who only had 2 out of

4 grandparents with a college degree would receive a 0.5, and a young adult for whom there is information about only 3 grandparents where 2 have college degrees would have a score of 0.67.

Table 4. Grandparents' education levels and grandchildren's college attendance

	Average education of grandparents
Young adults who have been to college	.22
Young adults with no college	.09

Examining the association between young adults' college attendance and the average education level of their grandparents tells a story that is similar to that told by looking at college attendance individually by type of grandparent: the young adults who have been to college have a higher average on the grandparent education variable than young adults who have not been to college. Hence, we have a positive association between grandparents' having college degrees and college attendance in the grandchild generation.

V. The answers for the tv viewing part of the exercise

As shown in the table that follows, the analysis of tv viewing time suggests a slightly negative association between viewing tv in one's childhood and eventual progress on to college during young adulthood. Average viewing times are slightly higher for those who did not attend college, compared to young adults who have attended college. For example, if one looks at time spent viewing tv during the week, on average, the young adults who ended up in college watched only about 1 hour and 48 minutes of tv on school days in 2002, compared to the 2 hours and 34 minutes that the average young adult who did not attend college watched. The year 2002 should correspond, roughly, to the time that our young adults from the 2005 TA module were in high school.

Table 5. Average daily TV viewing time during childhood and adolescence for young adults who have attended college and those who have not

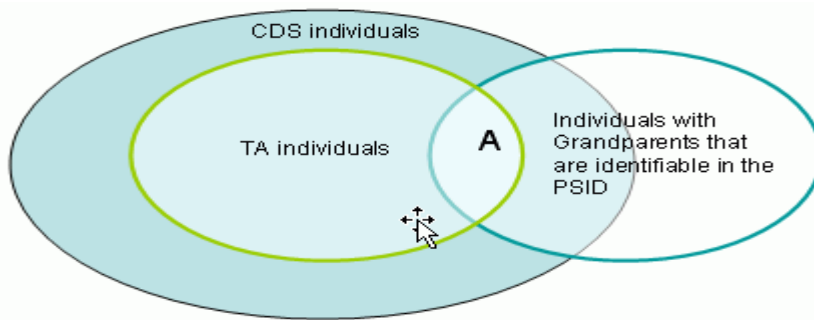
	College attendance yes	No college
TV viewing time on the weekday (from 1997 CDS)	101.98 minutes	107.11 minutes
TV viewing time on the weekend (from 1997 CDS)	155.27 minutes	174.20 minutes
TV viewing time on the weekday (from 2002 CDS)	107.84 minutes	154.16 minutes
TV viewing time on the weekend (from 2002 CDS)	153.33 minutes	177.07 minutes

VI. Thinking back on what you have done

Now that you have finished the tutorial exercise, some of you may want to take a moment to reflect upon what you have done (time and patience permitting). While you undoubtedly are pleased to have arrived at some answers, a good researcher will often find it useful to ask whether there are other ways that he or she might have gone about attempting to answer her research questions, or whether there are instances in which one might wish to refine or modify the research Qs. Let's ponder that a bit.

First, if you have done [Tutorial 4](#) you may be wondering how we managed to use CDS data in this tutorial without having any discussion of the primary caregivers (PCGs) of the CDS children. If you have worked with CDS data before, you know that the CDS data files provide information not just about children, but about the individuals who have primary responsibility for taking care of each child surveyed (the PCG), and about other caregivers (OCGs). The CDS does this so that researchers have information about the individuals who spend the most time taking care of the children about whom data are being collected. In about 95% of the cases, the PCG is the child's mom, so this means that any information recorded about the PCG is--de facto--mostly information about the child's mom. However, it's not always the case that PCG corresponds to mom. Similarly, the OCG need not be a child's dad. Why do we raise these points? It's possible that some researchers will care not so much about how a child's college attendance is related to the child's parents' education levels, but about whether a child is influenced by the person(s) who raised him. In this case, a researcher might want to relate college attendance of TA youth to the PCG's education level, so that we can correctly measure the influence of the person who actually raised the child). For example, consider the case of a child that was raised by his college educated grandparents (with the grandmom as PCG and the grandfather as OCG) because his mother and father were did not complete high school and were subsequently unable to afford to raise the child. (See Figure 3 below, which modifies Figure 2 to illustrate the type of case at issue here.) Suppose this child went on to college as a young adult. If we were to compare the association between the child's educational outcome and his parents we would get a different correlation than what we would get if we compare the child's educational attainment to his grandparents (the people who raised him). If the research question that we really have in mind is whether being raised by college educated people increases one's likelihood of attending college, we probably want to relate the young adult's outcome to his PCG's level of education (and to his OCGs)--not to his actual parents' outcomes (since the parents did not actually raise the child). However, we could hypothesize that the biological or adoptive parent could have an influence. For example, they may have had some caregiving role over the child's early life before the child was surveyed by the CDS. One could develop a statistical model which allows for influences of both the PCG and if the grandparents are not the PCG or OCG allow for that to be an influencing factor, too. Again, remember, in about 95% of the cases, the PCG happens to coincide with the child's mom, however in other cases one would be mismeasuring the educational influences of a child by using his mother's education instead of the PCG's. And, one would be making a similar mistake with dads and OCGs.

Figure 3. TA youth, grandparents, and grandparents who are PCGs or OCGs



Set "A" shows the set of TA individuals (who also happen to be part of the CDS set) for whom grandparent information is available in the PSID

Grandparents who are PCGs or OCGs would be a subset of set A.

Table 6. Grandparents in PCG and OCG positions

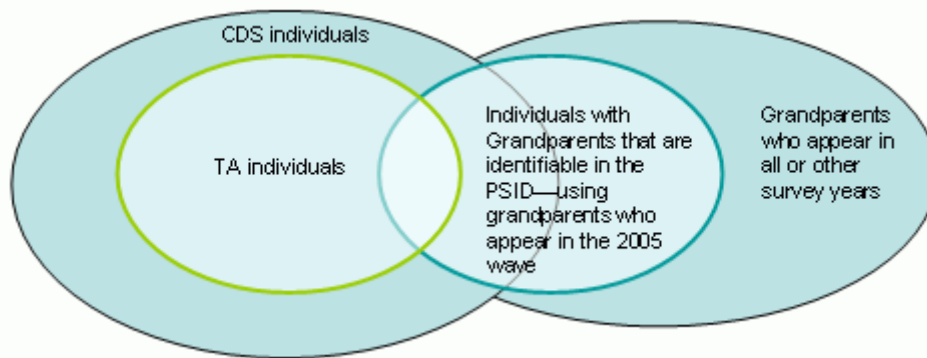
Grandparent type	Number of grandparents serving as PCG or OCG by year			
	CDS-1		CDS-2	
	PCG 97	OCG 97	PCG 02	OCG 02
Grandmothers				
Father's mother (FM)	19	6	16	7
Mother's mother (MM)	57	111	55	51
Grandfathers				
Father's father (FF)	2	3	2	4
Mother's father (MF)	3	16	1	23
Total number of CDS kids	3563	3563	2907	2907

Another reason to think about potential merits of linking young adults to their PCGs and OCGs, instead of mom and dad specifically, is that the CDS collected a great deal of information about the PCG's parenting style, the PCG's aspirations and expectations for the child, et cetera. Similar questions were asked of the OCG. Accordingly, rich analyses of the environments that young adults experienced as children can be done by connecting the TA young adults to the CDS-I and CDS-II data covering their PCGs and OCGs. Recall that Table 6 information can be from FIMS or the CDS/TA mapfile in Excel.

A second way that one could modify the analysis in this tutorial would be to use a broader range of years to obtain the grandparent education information. What is the issue here? In this tutorial we relied on the year 2005 to obtain the education information that we assigned to grandparents. This means that in any instances in which a young adult's grandparent(s) had died before 2005, we do not have the grandparent's education level, even though (i) the deceased grandparent

might very have been a college educated individual AND (ii) the grandparent's education level might have been observed in some earlier year in the PSID (when the grandparent was alive and serving as either head or wife of a household). (See Figure 4 below.) The challenge posed by the technique we adopted in the tutorial is that it creates the potential for having a lot of missing observations, and a smaller sample size than we would get if we swept across a range of years to identify grandparent education. In [Tutorial 6](#), we showed how one can "sweep" across a range of years to obtain information about the individuals' earnings. A similar technique could be applied here, to search over a number of years' worth of data in order to look for education information about any given young adult's grandparent, so that we have a greater chance of finding grandparents who are alive and active in the PSID (since we have to locate them when they are alive and active in order to obtain their education data).

Figure 4. Expanding the range of years in which one searches for grandparent information



A final extension that a good researcher would want to contemplate is conducting multivariate analysis to determine whether the associations we have identified in Sections IV and V exist in the presence of other factors that influence young adults' educational attainment. Note that if you do decide to perform regression analysis you want to remember to adjust your standard errors in order to account for the fact that some young adults share the same PCG, OCG, and parents.

VII. References

- Altonji, Joseph G., Hayashi, Fumio, Kotlikoff, Laurence J., 1992. "Is the Extended Family Altruistically Linked? Direct Tests Using Micro Data," *American Economic Review* 82, 1177-1198.
- Altonji, Joseph G., Hayashi, Fumio, Kotlikoff, Laurence J., 1997. "Parental Altruism and Inter Vivos Transfers: Theory and Evidence," *Journal of Political Economy* 105 (6), 1121-1166.
- Amemiya, Takeshi, 1985. *Advanced Econometrics*. Harvard University Press: Cambridge, MA.
- Arrondel, Luc and Andre Masson (2001). "Family Transfers Involving Three Generations," *Scandinavian Journal of Economics*, Volume 103(3), pp. 415-443.
- Barro, Robert J., 1974. "Are Government Bonds Net Wealth?" *Journal of Political Economy* 82 (6), 1095-1117.
- Barro, Robert J., and Becker, Gary, S., 1989. "Fertility Choice in a Model of Economic Growth," *Econometrica* 57 (2), 481-501.
- Becker, Gary S., 1974. "A Theory of Social Interactions," *Journal of Political Economy* 82 (6), 1063-1093.
- Becker, Gary S., 1991. *A Treatise on the Family*. Harvard University Press, Cambridge, MA.
- Becker, Gary S., and Tomes, Nigel, 1979. "An Equilibrium Theory of the Distribution of Income and Intergenerational Mobility," *Journal of Political Economy* 87 (6), 1153-1189.
- Behrman, Jere; Robert Pollack and Paul Taubman (1995). *From Parent to Child: Intra-household Allocations and Intergenerational Relations in the United States*. University of Chicago Press: Chicago, IL.
- Bernheim, B.D., Schleifer, A., Summer, L.H., 1985. "The Strategic Bequest Motive," *Journal of Political Economy* 93 (6), 1045-1076.
- Blinder, A.B., 1974. *Toward an Economic Theory of Income Distribution*. MIT Press, Cambridge, MA.
- Carneiro, Pedro, Karsten Hansen and James J. Heckman (2003). *Estimating Distributions of Treatment Effects with an Application to Returns to Schooling and Measurement of*

the Effects of Uncertainty on College Choice. *International Economic Review* 44(2), 361-422.

Cavalli-Sforza, Luigi Luca and M.W. Feldman (1981). *Cultural Transmission and Evolution: A Quantitative Approach*. Princeton: University Press.

Charles, Kerwin and Erik Hurst (2003). "The Correlation of Wealth Across Generations," *Journal of Political Economy*. Volume 111(6), December, pp. 1155-1182.

Chiteji, Ngina and Frank Stafford (1999). "Portfolio Choices of Parents and Their Children as Young Adults," *American Economic Review*, Volume 89(2), May, pp. 377-380.

Corak, Miles (2004). "Generational Income Mobility in North America and Europe: An Introduction" in Corak, ed. *Generational Income Mobility in North America and Europe*. New York: Cambridge University Press.

Cox, Donald and Oded Stark (2005). "Bequests, Inheritances and Family Traditions," *Center for Retirement Research at Boston College CRR Working Paper*, No. WP 2005-09. Boston College.

Cox, Donald and Oded Stark (1994). "Intergenerational Transfers and the Demonstration Effect." Mimeo. Boston College and Harvard University.

Cox, D., 1987. "Motives for Private Income Transfers," *Journal of Political Economy* 95 (3), 508-546.

Davis, Matthew M., McGonagle, Katherine A., Schoeni, Robert F. and Stafford, Frank P., (2006). Relative contributions of parental and grandparental obesity to risk of child obesity: Three generations of American families. Mimeo. University of Michigan.

Diamond, Peter A., 1965. "National Debt in a Neoclassical Growth Model," *American Economic Review* 55 (5), 1126-1150.

Fitzpatrick, Maria and Sarah Turner, 2007. "Blurring the Boundary: Changes in the Transition from College Participation to Adulthood," in Rouse and Danziger (eds). *The Price of Independence: The Economics of Early Adulthood* (New York: Russell Sage Foundation).

French, Eric, 2005. "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *The Review of Economic Studies* 72 (2), 395-427.

Friedman, Benjamin M., and Warshawsky, Mark J., 1990. "The Cost of Annuities: Implications for Saving Behavior and Bequests," *The Quarterly Journal of Economics* 105 (1), 135-154.

Gustman, Alan L., and Steinmeier, Thomas L., 1986. "A Structural Retirement Model," *Econometrica* 54 (3), 555-84.

Gustman, Alan L., Mitchell, Olivia S., Samwick, Andrew A., and Steinmeier, Thomas L., 2000. "Evaluating Pension Entitlements," in Olivia Mitchell, P. Brett Hammond, and Anna M. Rappaport (ed.), *Forecasting Retirement Needs and Retirement Wealth*. Philadelphia: University of Pennsylvania Press.

Heckman, James J. and Salvador Navarro (2006). *Dynamic Discrete Choice and Dynamic Treatment Effects*. Forthcoming, *Journal of Econometrics*.

Heckman, James J. and James R. Walker (1989). Forecasting Aggregate Period-Specific Birth Rates: the Time Series Properties of a Microdynamic Neoclassical Model of Fertility. *Journal of the American Statistical Association* 84(408), 958-965.

Heckman, James J. and James R. Walker (1990). Estimating Fecundability from Data on Waiting Times to First Conception. *Journal of the American Statistical Association* 85(410), 283-294.

Heckman, James J. (1974). "Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle," *American Economic Review*, 54, (March 1974), p. 188-194;

Ryder, Stafford and Stephan, 1976 "Labor, Leisure, and Training Over the Life Cycle," *International Economic Review*, October, pp. 651-574.

Holtz-Eakin, Douglas, Joulfaian, David, and Rosen, Harvey S., 1993. "The Carnegie Conjecture: Some Empirical Evidence," *The Quarterly Journal of Economics* 108 (2), 413-436.

Hurd, M.D., 1989. "Mortality Risk and Bequests," *Econometrica* 57 (4), 779-814.

Hurst, Erik, Luoh, Ming Ching, Stafford, Frank P., and Gale, William G., 1998. "The Wealth Dynamics of American Families, 1984-94," *Brookings Papers on Economic Activity* 1998 (1), 267-337.

Ippolito, Richard A., 1997. *Pension Plans and Employee Performance*. Chicago: The University of Chicago Press.

Geronimus, Arline T. and Korenman, Sanders. The Socioeconomic Consequences of Teen Childbearing Reconsidered. *The Quarterly Journal of Economics*. 1992; 109(4):1187-1214.

Kotlikoff, Laurence J., 1988. "Intergenerational Transfers and Savings," *Journal of Economic Perspectives* 2 (2), 41-58.

- Kotlikoff, Laurence .J. and Spivak, A., 1981. "The family as an incomplete annuities market", *Journal of Political Economy* 89, 372-391.
- Laitner, John, 1988. "Bequests, Gifts, and Social Security," *Review of Economic Studies* 55 (2), 275–299.
- Laitner, John, 1992. "Random Earnings Differences, Lifetime Liquidity Constraints, and Altruistic Intergenerational Transfers," *Journal of Economic Theory* 58 (2), 135–170.
- Laitner, John, 1993. "Long–Run Equilibria with Borrowing Constraints and Altruism," *Journal of Economic Dynamics and Control* 17 (1-2), 65–96.
- Laitner, John, 1997. "Intergenerational and Interhousehold Economics Links," in *Handbook of Population and Family Economics, Vol. 1A*, North–Holland, 189-238.
- Laitner, John, 2001. "Simulating the Effects on Inequality and Wealth Accumulation of Eliminating the Federal Gift and Estate Tax," in Gale and Slemrod (eds.), *Rethinking Estate and Gift Taxation*. Brookings Institution Press: Washington, D.C., 258–298.
- Laitner, John, 2002. "Wealth Inequality and Altruistic Bequests," *American Economic Review* 92 (2), 270–273.
- Laitner, John, 2004. "Wealth Accumulation in the U.S.: Do Inheritances and Bequests Play a Significant Role?" University of Michigan mimeo.
- Laitner, John, Juster, F. Thomas, 1996. "New Evidence on Altruism: A Study of TIAA–CREF Retirees," *American Economic Review*, 86 (4), 893-908.
- Laitner, John, and Ohlsson, Henry, 2001. "Bequest Motives: A Comparison of Sweden and the United States," *Journal of Public Economics*, 79, 205-236.
- Laitner, John, and Silverman, Dan, 2005. "Estimating Life–Cycle Parameters from Consumption Behavior at Retirement," NBER working paper 11193.
- Loury, G.C., 1981. "Intergenerational Transfers and the Distribution of Earnings," *Econometrica* 49, 843-867.
- Lundberg, Shelley and Robert Pollak (1998). "An Intergenerational Model of Domestic Violence," manuscript, Department of Economics, University of Washington and Washington University in St. Louis, December 1998.
- Mayer, Susan; Greg Duncan and Ariel Kalil (2004). "Like Mother, Like Daughter?" SES and the Intergenerational Correlation of Traits, Behaviors and Attitudes," *Harris School Working Paper Series #04.15*. Harris School of Public Policy. University of Chicago.

- Modigliani, Franco, 1986. "Life Cycle, Individual Thrift, and the Wealth of Nations," *American Economic Review*, 76, 297-313.
- Modigliani, Franco, 1988. "The Role of Intergenerational Transfers and Life Cycle Saving in the Accumulation of Wealth," *Journal of Economic Perspectives* 2, 15-40.
- Rust, John, and Phelan, Christopher., 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets," *Econometrica* 65 (4), 781-831.
- Solon, G.R., 1992. "Intergenerational Income Mobility in the United States," *American Economic Review* 82 (3), 393-408.
- Solon, Gary (2004). "A Model of Intergenerational Mobility Variation over Time and Place," in Corak, ed. *Generational Income Mobility in North America and Europe*. New York: Cambridge University Press.
- Stark, Oded, 1989. "Altruism and the Quality of Life," *American Economic Review* 79, 86-90.
- Tabb, Rebecca M. 2004. "Intergenerational Labor Supply," Department of Economics, Stanford University, June.
- Toledo, Manuel. 2005. "On the Intergenerational Persistence of Work Hours", working paper University of Rochester, Department of Economics, December.
- Tomes, Nigel, 1981. "The family, inheritance and the intergenerational transmission of inequality", *Journal of Political Economy* 89, 928-958.
- Verba, Sidney; Kay Lehman Schlozman and Nancy Burns (2003). "Family Ties: Understanding the Intergenerational Transmission of Political Participation." Unpublished manuscript.