SIPP User Notes

To: SIPP Users
From: Heather Boushey, economist, Center for Economic and Policy Research
RE: Set E: Child Care User Notes
Date: February 13, 2006

Where does the childcare data come from?
Data on the type of childcare arrangement, whether or not the parent/guardian paid for childcare, the amount paid for childcare, and assistance (government and other) for childcare payments are found in the following topical modules:

Panel	Topical Module	Calendar Year
1992	6	1993
1992	9	1994
1993	3	1993
1993	6	1994
1993	9*	1995
1996	4	1996/1997
1996	10	1998/1999
2001	4	2001/2002

*This topical module uses an entirely different coding system for the childcare variables—coding that is not consistent with the topical modules/panels before it or after. We do not use data from this topical module in what follows.

Construction of the childcare arrangement categories
We collapse the responses for the childcare questions into distinct categories or types of childcare arrangements:
Parent: parent/guardian respondent and other parent/guardian
Relative: sibling (over age 15), grandparent, other relative
Family: non-relative care outside the child's home; family day care in someone's home
Nanny/Sitter: non-relative care inside the child's home
Formal: formal day care center not in a home; head start; nursery or preschool
Self: child cares for his/her self (and sibling 15 years and younger)

Variables	92/93 Panels	96/01 Panels
Parent	tm8120, tm8122, tm8124	eckd01*, eckd02*
Relative	tm8120, tm8122, tm8124	eckd03*, eckd05*, eckd06*

Family	tm8120, tm8122, tm8124	eckd11*
Nanny	tm8120, tm8122, tm8124	eckd11*
Formal	tm8120, tm8122, tm8124	eckd08*, eckd09*, eckd10*
Self	tm8120, tm8122, tm8124	eckd04*, selfca*

Structure of the data

The structure of the raw data in the 1992/1993 panels is different from that in the 1996/2001 panels. In the earlier panels, the childcare arrangement and payment questions are asked for the youngest child, second youngest child, and third youngest child of each respondent. For example each respondent is asked what the primary childcare arrangement is for his/her youngest child; there are 13 categorical responses to this question. Therefore in the earlier panels there are only three childcare arrangement questions (*tm8120*, *tm8122*, and *tm8124*), each with 13 different responses. In the 1996/2001, the childcare responses are recorded as yes/no answers (dummy variables) to separate questions. For example, there is one set of questions for the "other parent care" (*eckd01*) and another set of questions for grandparent care (*eckd05*). In the 1996 and 2001 panels these eckd variables for each type of childcare (for example *eckd01* for "other parent care") are given by the suffixes *a* through *e* for the first five children under the age of six and by the suffixes *f* through *j* for the first five children six years or older. For example *eckd01a* is the variable name for the whether or not the first youngest child under the age of six received "other parent care;" while *eckd01b* is the variable name for whether or not the second youngest child under the age of six received "other parent care."

Pull program

In the pull program for the 1992/1993 panels, we then need to transform the categorical questions (*tm8120*, *tm8122*, and *tm8124*) into separate (dummy) variables akin to the *eckd* structure of the variables found in the 1996/2001 panels. To do this correctly, we need to determine the age of the children for each childcare question. If the youngest child (*tm8120*) over the age of six has parental care, we would not want to code that variable as "*eckd01a*" but rather "*eckd01f*". We merge information from Set C (family and household relationships) to get the age of each of the children and attach this information to the respondents (parents). We then create a dummy variable for whether or not each child is under the age of six or not. Each respondent then has the variables *kid1sub6*, *kid2sub6*,and *kid3sub6*, to let us know whether the children referred to in *tm8120*, *tm8122*, and *tm8124* are under the age of six.

For the 1992 and 1993 panels, the childcare questions are asked only about the primary childcare arrangement, that is, the childcare arrangement with the maximum number of hours. Comparisons of the 1992 and 1993 panels with childcare data from the 1996 and 2001 panels must therefore be among maximum hours/primary care arrangements.

Clean program

The clean program collapses the childcare arrangement and payment data into the final categories used for analysis, as described above. In doing this, we also add the childcare coverage and payment variables across children of the same parent. This allows us to create a variable that tells us if a parent had a certain type of childcare arrangement for any of her children.

The clean program for the 1996 and 2001 panels includes code that is necessary to make the data match that of the earlier panels.

First, the code drops the *eckd**d*, *eckd**e*, *eckd**i*, *eckd**j* variables from the data. These variables record childcare information for the fourth and fifth youngest children under 6 and fourth and fifth youngest children over 6. There is no information for these children in the 92/93 panels, and even in the 96/01 panels there are extremely few (less than 0.5% of the total child

responses) observations for these children. Thus we drop them from the data set. [[This part of the program drops information from 1996 and 2001 panels so that it matches, conceptually, the information in the 1993 panel. In the 1993 panel we have childcare information only for the 3 youngest children for each parent. In the 1996 and 2001 panels we have childcare information on the first five children under age 6 for each parent interviewed and for the first five children aged 6-17 for each parent interviewed. Thus we need to transform the 1996 and 2001 panels' childcare variables into the three youngest children for each parent interviewed. We do this by recoding information from the 1996 and 2001 panels on children that are not among the three youngest for each parent: we recode this information as missing.]]

Second the clean program includes code that uses the information from the "hours in type of childcare" variables to determine which type of childcare arrangement is the primary childcare arrangement. Again, this is necessary because the 92/93 childcare variables cover only the primary childcare arrangement.

In addition, the clean program creates childcare (*care**) and payment (*amt**, *paid**, and *tot**) variables that sum across children of the same parent. For example, the *amt***1 variable is the average amount paid (by a parent) for childcare for all children under the age of six; while *amt***2 is the average amount for childcare for all children six years and older.

Tables program
This program collects demographic, family, employment, and income data from other data sets (Sets A, B, C, D, and F) for the respondents to the childcare questions. We then merge this data into the childcare data so that we can look at childcare arrangements across demographic, family, and income characteristics. This also allows us to compute regression-adjusted means for the childcare arrangements, and to look at childcare arrangements for samples of interest—for example, working mothers with children under the age of six.

Topcoding Childcare Payment Variables
The childcare payment variables (*tamt** in the 1996 panel and *tm8152*, *tm8154*, and *tm8156* in the 1992/1993 panels) are topcoded. We have reason to believe that the 2001 panel childcare payment variables are not topcoded (see below).

The "cc_topcodes" program reports the maximum value and standard deviation of each childcare payment variable (*tamt**). It then calculates the mean of the variable for all observations greater than zero. Next it calculates the mean of the non-zero observations when we set the mean of the topcoded observations equal to maximum reported value plus one (2, 3, 4) standard deviation. For example, if there are 300 non-zero observations for a particular payment variable with a mean of 50 and a standard deviation of 50, and 6 of these observations are topcoded at 200, the program will calculate the mean of the 300 observations when we set the mean of the 6 topcoded observations equal to 250 (the maximum reported value plus one standard deviation), 300 (the maximum reported value plus two standard deviations), etc. The Topcode Tables (see Appendix) present the following information: the maximum value of each childcare payment variable (*tamt**), the % of the observations at the maximum value, whether or not the maximum value reported in the data is equal to the topcode for that variable listed in the codebook, and the means of all non-zero observations.

The 2001 panel (TM 4) childcare payment topcodes reported in the codebooks are all equal to maximum value of each variable reported in the data. Thus Table 1 reports the means when we assume that each reported value is actually topcoded. Because the maximum reported value in the data for the 2001 panel varies across variable type (e.g. the maximum reported value for *tamtdaya* is 240 while the maximum reported value for *tamtdayb* is 300), it is likely that there are no topcodes for these variables.

With the exception of the 2001 panel, the other panels show that very few observations in the childcare payment variables are topcoded. As should be expected, most of the topcoding takes place in variables for the youngest child under six (*tamt**a*) and the youngest child 6 and over (*tamt**f*). As can be seen in the topcode tables, when we increase the mean of the topcoded observations by one standard deviation, the mean of the new variable rarely increases by more than $1. Only when we set the mean of the topcoded values to the maximum reported value plus 3 or 4 times the standard deviation does the new mean increase by more than $2. So, although topcoding occurs for most of the payment variables, the incidence of topcoding is low. And this quick check shows that increasing the topcodes even by two standard deviations has little effect on the mean of the payment variables.

Important items to note

- For the calendar year 1994 (92tm9 and 93tm6) the Census Bureau did not ask the childcare questions (*tm8120*, *tm8122*, and *tm8124*) about children six years or older. For the calendar year 1993 (92tm6 and 93tm3) the childcare questions were asked of all children under the age of 15. This explains why the response categories **"kindergarten"** and **"child day/group care"** differ so dramatically across calendar years. After we tabulate only for children under the age of six, these differences are no longer present in the data. (See "Tabs 92, 93 childcare raw variables")
- The 1992/1993 panels do not have certain types of responses that are covered in the latter panels: family day care for younger children and sports, clubs, and lessons for older children. This does not dramatically affect the childcare arrangement variables for children under six. But for children 6 years and older, there are very few observations for the childcare variables in the 1992, 1993 panels. This makes sense because the categories added in the 96/01 panels mostly deal with the type of activities for older children— lessons, sports, and clubs.
- In the 1992/1993 panels the head start question is a separate question (*tm8103* in calendar year 1993 and *tm7079*, *tm7080*, and *tm7081* in calendar year 1994) from the childcare arrangement questions, but the head start question is still coded into the *eckd* format as *eckd10*.
- The 1992/1993 panels do not include variables with information about assistance with childcare payments, government assistance or otherwise

Variable list

epppnum	person number of child
id	id
wave	wave
ccuniv1	had childcare for a child under 6
ccuniv2	had childcare for a child 6 or older
care_par1	used parental childcare for child under 6 (parent-variable)
care_par2	used parental childcare for child 6 or older (parent-variable)
care_rel1	used relative childcare for child under 6 (parent-variable)
care_rel2	used relative childcare for child 6 or older (parent-variable)
paid_rel1	paid for relative childcare for child under 6 (parent-variable)
paid_rel2	paid for relative childcare for child 6 or older (parent-variable)
amt_rel1	average amount paid for relative childcare for child under 6 (parent-variable)
amt_rel2	average amount paid for relative childcare for child 6 or older (parent-variable)
tot_rel1	total amount paid for relative childcare for child under 6 (parent-variable)
tot_rel2	total amount paid for relative childcare for child 6 or older (parent-variable)
care_fam1	used family childcare for child under 6 (parent-variable)
care_fam2	used family childcare for child 6 or older (parent-variable)
paid_fam1	paid for family childcare for child under 6 (parent-variable)
paid_fam2	paid for family childcare for child 6 or older (parent-variable)
amt_fam1	average amount paid for family childcare for child under 6 (parent-variable)
amt_fam2	average amount paid for family childcare for child 6 or older (parent-variable)
tot_fam1	total amount paid for family childcare for child under 6 (parent-variable)
tot_fam2	total amount paid for family childcare for child 6 or older (parent-variable)
care_nan1	used nanny childcare for child under 6 (parent-variable)
care_nan2	used nanny childcare for child 6 or older (parent-variable)
paid_nan1	paid for nanny childcare for child under 6 (parent-variable)
paid_nan2	paid for nanny childcare for child 6 or older (parent-variable)
amt_nan1	average amount paid for nanny childcare for child under 6 (parent-variable)
amt_nan2	average amount paid for nanny childcare for child 6 or older (parent-variable)
tot_nan1	total amount paid for nanny childcare for child under 6 (parent-variable)
tot_nan2	total amount paid for nanny childcare for child 6 or older (parent-variable)
care_for1	used formal childcare for child under 6 (parent-variable)
care_for2	used formal childcare for child 6 or older (parent-variable)
paid_for1	paid for formal childcare for child under 6 (parent-variable)
paid_for2	paid for formal childcare for child 6 or older (parent-variable)
amt_for1	average amount paid for formal childcare for child under 6 (parent-variable)
amt_for2	average amount paid for formal childcare for child 6 or older (parent-variable)
tot_for1	total amount paid for formal childcare for child under 6 (parent-variable)
tot_for2	total amount paid for formal childcare for child 6 or older (parent-variable)
care_sel1	self childcare for child under 6 (parent-variable, 1996/2001 only)
care_sel2	self childcare for child 6 or older (parent-variable, 1996/2001 only)
yr	calendar year

These variables appear in the 1996/2001 panels only:

help_cc	received any type of outside aid for childcare payment
hourslost	hours lost from work/school due to childcare
hr_sch1	hours child 4-5 in school
hr_sch2	hours child 6-15 in school
govt_help_cc	used government aid for childcare payment

Note: variables are available in all panels except where noted