

User Notes for the SPD Longitudinal File

1. General Cautionary Notes

The original SPD sample is five to six years old and has undergone a substantial amount of attrition (sample loss) from 1992 (or 1993) through 1998. As a result, estimates from this file are not as representative of the U.S. population as a cross-sectional survey or a longitudinal survey with less sample attrition would be. Therefore, results should be viewed with caution.

In addition, it should be noted that this file contains data collected using three different survey collection vehicles: the 1992/93 SIPP paper instruments that were used to produce data for calendar years (CY) 1992, 1993, and 1994, a modified March CPS CAPI instrument that was used to collect data for calendar year 1996, and the 1998 CAPI instrument that was used to collect data for calendar year 1997. Therefore, the different questions and modes of interview used to produce the estimates should be considered when analyzing changes over time. These effects are explained more fully in later notes.

a. 1992 Data

The SPD Longitudinal file does not contain a complete set of CY1992 data for 1993 SIPP Panel households because interviewing did not begin until 1993. Using weighted counts for 1992 data is not meaningful unless the necessary adjustments to the longitudinal weight are made. See the user note below titled *1992 Estimates* for further instructions.

b. 1995 Data

The SPD Longitudinal file does not contain any data for CY1995. The 1992 SIPP Panel did not collect any data for 1995 and the 1993 SIPP Panel did not collect data for the entire 1995 calendar year.

c. 1992, 1993, and 1994 Calendar Year Data - Recodes

Analysts interested in variable by variable mapping of SIPP to CPS need to review the SIPP and CPS Cross-Reference map on the SPD Website or Appendix D in the Technical Documentation.

The CY1992, CY1993, and CY1994 data were created by recoding the data collected in the SIPP 92/93 Panels to similar variables collected in the March 1994 CPS. Not all of the 1994 March CPS variables are recoded from the 1992/1993 SIPP data and no imputation flags were carried from the component SIPP data onto the 1992 - 1994 calendar year files. This recoding operation was done by the Urban Institute under contract with the Census Bureau, the Social Security Administration, the Office of the Assistant Secretary for Planning and Evaluation, and the Department of Health and Human Services. Collectively, the 1992, 1993, and 1994 data are referred to as the "CPS Look-Alike" file.

d. Creating Family and Household Structures

In the SIPP, family and household units evolve from month to month. Persons may leave a family by going away to school, joining the armed forces, establishing a new household, or dying. Others may join a family through birth of a child, marriage, or some other reason. There are no static family and household structures that exist for the duration of the survey, or even for a year. The monthly family and household information collected by the SIPP apply to the units that exist as of the month the data are collected. However, the CPS family and household sample units are defined as of a point in time--March of the survey year--and annual data for the preceding year, the reference year, are collected for units in existence as of the time of the survey.

In order to construct annual information such as is contained on a CPS file from information that varies by month, as is provided by SIPP, it is necessary to determine a fixed reference point for the construction of family and household units and the determination of demographic characteristics. December of the reference year has been chosen as that point on the SIPP file. Thus, the coding of all household and family relationship codes and demographic fields such as race, sex, marital status, and age and any other fields requiring a fixed reference point will be based on December values for the reference year.

The necessity of defining annual units from units that exist as of a given point in time (December of the reference year) also means that many of the household and family fields must be constructed from person fields, rather than the reverse coding scheme which would follow the coding logic of the CPS. For instance, on the CPS, if a household has any person who received social security benefits, the household field HSS-YN is coded with a 1 to indicate "Yes." Only then are questions asked that enable coding of person fields that refer to receipt of social security benefits. Thus, the universe for the person fields is $HSS-YN = 1$. However, a reverse process must be used for the CPS look-alike file. The CPS person fields are first coded from SIPP data. Then, the household and family fields are coded by cycling through persons in the family or household units that existed as of December. Thus, some of the universes on the SIPP to CPS look-alike file must differ from those on a CPS file. For example, DSS-YN, a CPS person field, has a universe of $HSS-YN = 1$. On the "look-alike" file, DSS-YN has a universe of all persons. This difference in universe is regarded as insignificant, but it is one that should be noted when frequencies on the SIPP to CPS look-alike file are compared with those from other CPS files.

Since longitudinal units are not identified for the SIPP, monthly data that pertain to monthly family and household units have little meaning for a longitudinal file, and, for the most part, such data are omitted from the longitudinal file. However, the CPS collects annual data that apply to family and household units that are identified as of the time of survey. To reproduce some of the family and household data from the SIPP that are provided by the CPS, it is necessary not only to define the family and household units as of a point in time (December of the reference year), it is also necessary to obtain information from the monthly data provided by SIPP wave files. Household information is taken from the record of the household reference person (identified by SIPP variable RRP) for December of the reference year; and family information is taken from the record of the family reference person (identified by SIPP variables FAMREL and RRP).

e. Modes of Data Collection and Data Creation

This SPD file covers a variety of subject areas which may be differentially affected by the mode of collection. For example, more frequent interviews (and the collection of contemporaneous

data) may result in more reports of part time work, one-time income sources, or part-year health insurance coverage. Be aware that observed changes over time may be the result of the different modes of collection over the lifetime of this survey.

The 1992 and 1993 SIPP Panels (CY1992- CY1994 data) collected data using a paper questionnaire. The data was collected three times per year and used a four-month recall. The 1992 and 1993 SIPP data was converted to a calendar year format to resemble the CPS March data. Users of the 1992 through 1994 data need to review Appendix D, “SIPP to CPS Cross-Reference Map,” to understand the conversion of SIPP variables to CPS variables.

The 1997 SPD collected data used a slightly modified 1997 CPS March automated instrument. The data was collected between May and June 1997 and used a “last year” (1996) recall for most questions.

The 1998 SPD collected data used an automated instrument specifically designed for the SPD. The data were collected between May and June 1998 and used a “last year” (1997) recall for most questions.

f. Variable Naming Convention

The variable name (also referred to as data item, item name, or data field name) is no longer than 8 characters, for example WORKYNE8, AGET8 or MARITLE8. The first 6 characters provide a descriptive name. Character 7 (or 2nd from right if the variable name is less than 8 characters) is the variable type code [unedited (U), edited (E), recoded (R), weights (W), imputation flag (I), or topcoded (T)]. The final character, character 8 (or the last character if the variable name is less than 8 characters) reflects the interview year, that is the year in which the data was collected (for example, 1998 = 8). In the case of 1992 through 1994, the interview year and the data year are the same. The 1997 and 1998 interview (data collection) years are not the same as the data year for most variables. For example WORKYNE8 refers to working during 1997 while MARITLE8 refers to marital status as of the interview date in 1998. Most variables collected during the 1997 interview refer to 1996 data year, while most variables collected in 1998 refer to 1997 data year. However, there are variables in 1997 and 1998 that refer to the interview year as shown above. The short descriptions of a variable specifies the year to which the variable refers.

g. Families

Like households, families are also defined each year. In order to look at everyone who is in the same family for a given year, concatenate the IHHKEY variable for that year with the variable indicating family number. For example, to look at families in 1993, concatenate IHHKEY93 with FAMNUM3 (e.g.,B1200107406901101).

h. Matching to the SIPP Data

When matching to the SIPP 1992 or 1993 panel, core, or topical module files note the field lengths differences in the following variables:

Field Length Differences

Short Variable Description	SIPP 1992 Panel Files			SIPP 1993 Panel Files			Longitudinal File
	Panel	Wave	Topical Modules	Panel	Wave	Topical Modules	
Person Number	4	3	3	3	3	3	4
Entry Address ID	3	2	2	2	2	2	3
Address ID	3	2	2	2	2	2	3

2. *Weighting and Estimation - Technical Notes*

The records on the First SPD Longitudinal File were weighted to the March 1993 population. Therefore, you should compare the population characteristics from this file with those from the CPS March 1993 survey universe.

The weights on this file (SPDLNWGT and ANNUALWT) are longitudinal; therefore, they are only valid for estimates of the characteristics of the cohorts of people who were represented by the original sample persons from the SIPP 1992 and 1993 panels. The variable SPDLNWGT is nonzero for only original sample people who were a self or proxy or imputed respondent in each and every reference month for SIPP, SPD 1997 Bridge, and SPD 1998. It should be used to calculate estimates covering multiple calendar years.

An original sample person is a sample person who was a self or proxy respondent in Wave 1 of the SIPP 1992 or 1993 panel. Sample children born after the first interview (Wave 1) of the SIPP 1992 or 1993 panel and having an original sample person as their designated parent (blood-related or adopted) also receive longitudinal weights (ANNUALWT) so that their annual characteristics can be estimated.

a. Household weight

There is no weight for household analysis in the SPD Longitudinal file. Instead it has become acceptable in the research community to assign sample people the characteristics of their household or family. The analytic estimate then becomes the number of people in households (or families) with a particular characteristic.

b. Longitudinal weight with non-original sample children

SPD Longitudinal Annual Weight (ANNUALWT) - This weight is for calculating annual or calendar year estimates. This variable is nonzero for original sample people who provided a self or proxy, or whose data were imputed in each and every reference month for SIPP, SPD 1997 Bridge, and SPD 1998 and for non-original sample children (age six or less) born of or adopted by original sample people after the first interview. When calculating annual estimates, include only weighted non-original sample children born during or in prior calendar years. The non-

original sample children weights, for those age six or less, are derived by assigning the weight of child's designated parent (blood-related or adopted) to the child and is intended only to approximate growth in the universe. Some caution should be given when interpreting results for children 6 or less because this strategy has resulted in general overestimate of 2.2 percent based on 1998 benchmarks. By race, the results are a 3.6 percent overestimate for non-Black and a 5.4 underestimate for Black children six or less by 1998.

c. 1992 Estimates

To estimate for calendar year 1992 using the SPD longitudinal weight, analysts should multiply the weight of people present in 1992 by 2. This covers all situations, but gives only an approximation. Variance parameters will also need to be inflated by 2 for analysis of 1992 estimates. For more accurate 1992 calendar year estimates, use the 1992 SIPP Longitudinal file with its 1992 calendar year weight, FNLWGT92.

3. Demographic Variables - Technical Notes

The specific notes below address standards for editing relationship to reference person and cases that switch values in the unedited data for sex, vet, and race status. Note that in 1997, using the CPS instrument, Field Representatives could change demographic data that was input since the data provided was from SIPP paper control cards. If the information was incorrect during SIPP, they got a chance to change it in 1997, yet because demographic variables are edited forward from 1994, the edits force the responses to be consistent with the original SIPP data. That is why there is a higher number of imputations for 1997 than for other years.

a. Relationship to Reference Person (RRP)

An adult must be at least 12 years older than a child who points to them as their designated biological parent.

Same sex partners who report they are married are made unmarried partners if they report being spouses.

Parent pointers for adults age 18 and over were blanked out during the editing process for 1998 data.

People less than 15 years old must be 'never married.'

A person to whom a child points as their designated parent must be at least 15 years old.

A person under 18 years of age cannot be the reference person if their parent is present in the household.

Checks run on the data for 1999 and 2000 were the same as in 1998, except that in 1999, we decided to edit and output the designated parent pointer for people of all ages, so checks for the age gap between parents and children were done for people of all ages. For 1999, we also used the TYPMOM and TYPDAD variables which indicate whether a parent is a biological, step or adoptive parent. If the parent was identified as a step or adoptive parent, we allowed an otherwise illogical age gap between parent and child since remarriage sometimes creates situations where step parents are not old enough to be the biological parent of their spouse's

biological children. Users should note that the edit for RRP, MARITL, SPOUSE, and DESPAR is not truly longitudinal, but checks for consistency in each year cross sectionally. So the RRP values of two people shift over time, even though they are the same people. This happens at times when a later interview gives us more information and we are able to specify the relationship in greater detail-for example, making someone a grandchild rather than other relative.

b. Age

Some people who were recorded as children in 1992, 1993 or 1994 are recorded as adults in 1997 or 1998 in the unedited data, or vice versa. Since we are only editing forward, some of these people end up on the longitudinal file with information for variables like income and labor force status when they may no longer fit the universe for those variables. For example, if an individual reports being age 7 in 1992, 8 in 1993 and 9 in 1994, but in 1997 reports being 32 in the unedited data, they will appear as age 12 in the edited longitudinal file, even though they may have edited labor force information in 1997 which shows they work 40 hours a week and have an annual personal income of \$25,000. In the same way, the family structure variables which were edited in the cross sectional files for 1992, 1993, 1994 and 1997 may now be incorrect in some cases where age or sex switched over time.

b.1. AGE User Notes for the Second Longitudinal File

We have not filled in DOBYRT for all cases, although users could easily create it using the year and the age. At this point it is filled in for interviews in 1998 and 1999 and 2000.

c. Sex

Demographic variables are edited forward in the SPD. This means if the sex recorded in the unedited data in 1997 or 1998 differs from the sex recorded in the 1992-1994 Urban file, the 1994 value is considered to be correct. The final value for sex in the longitudinal file is set to match the 1992-1994 value. The imputation variable SEXI will let you know which cases contained contradictory reports in the unedited 1997 or 1998 data. The numbers of cases affected are as follows:

Cases do not switch sex within 1992-1994.

1994 value different from 1997 value: 510 cases

1997 value different from 1998 value: 122 cases

1994 value different from 1997 and 1997 different from 1998: 44 cases

d. Veterans' Status

There are 101 cases which have an illogical change in VET status in the 1992, 1993, or 1994 data. We are not changing the 1992, 1993, or 1994 data, but are only editing forward.

- There are 28 cases that have a value of yes for vet for 1992 and not a vet in 1993.
- There are 72 cases that have a value of yes for vet for 1993 and not a vet in 1994.

- The remaining case has a value of yes for vet for 1992 and not a vet for 1994. The value is missing for 1993—they were probably not interviewed in that year.

Note that in 1993, two cases are marked as vets in the unedited data, and are then recoded to non-vets because they are outside the age universe, since they are 15 and 16 years old.

e. Race

Through 1998, there were 114 cases in which RACE was imputed. Nine of these are imputed using the hot deck, and the remaining 105 are assigned logically based on the person's mother, father, or household reference person.

Of the 114 cases imputed, 18 cases were imputed where there is no valid value for RACE in any of the years. The other 96 cases imputed have a value of "other" for 1998.

Of the total 407 cases who reported their race as "other" in 1998, 238 had one valid value for an earlier year which was used, 7 were imputed using the hot deck, 89 were imputed logically based on family members, and 73 switch their reported race over the years they are in the survey.

Overall, 692 cases switch their race from one year to another. Of these, 37 percent have a first valid value of American Indian and 40.2 percent have a first valid value of Asian and Pacific Islander.

The total 806 cases which are imputed or switch over time are 0.87 percent of the 92,158 cases, or 1.7 percent of the interviews in 1998 (47,587).

f. NATVTY

We have added a value of 4 for the imputation flag so that we can indicate cases whose value on the first longitudinal file differs from the value they report in 1999 or in 2000, if they entered in these years.

g. Educational Attainment Variables

To create the educational attainment variables for 1992, 1993 and 1994 (HGAR2, HGAR3, HGAR4) the response categories from the SIPP were edited to match those from the CPS. Unfortunately, the SIPP codes do not contain sufficient detail to exactly re-create the CPS field. In particular, coding of college attainment is deficient. The assumption is made that SIPP HIGRADE codes 21-24 refer to undergraduate study, 25 to master's level work, and 26 to the doctorate level. There is no information to determine associate or professional school degrees.

The variables that measure individual, family, and household receipt of educational assistance are not strictly comparable across time. The educational aid variables for 1992 (ED_YNE2), 1993 (ED_YNE3), and 1994 (ED_YNE4) depict higher levels of aid receipt than the variables for 1997 (ED_YNE7) and 1998 (ED_AIDE8). Some of this difference is likely attributable to differences in variable universes, question wording, and reference periods. The 1992-1994 variables were constructed from the 1992 and 1993 SIPP panels. The SIPP asked a very detailed question about receipt of financial aid during the last 4 months of all individuals age 15 and over. Data from multiple waves were then combined to construct the variable for financial aid receipt in the last year (because the reference months differ for the four SIPP rotation groups the

reference period for the financial assistance variable is not the same for all sample members). In 1997 a much less detailed question was asked of all individuals age 15 and older regarding financial aid receipt in the last calendar year. The 1998 question was similar to the 1997 question but was asked only of individuals age 18 and over who were enrolled in either high school or college and collected information on financial aid receipt during the last 9-month academic year, September to May, rather than the prior calendar year. For more detailed information on financial aid receipt in 1992, 1993, and 1994 users should go directly to the 1992 and 1993 SIPP panels.

4. Poverty Variables - Technical Notes

Data users should be aware that SPD income and poverty data are affected by the different survey instruments implemented in different years. For example, poverty rates have historically been lower when using the SIPP survey instrument, where there are detailed income questions every four months, than in the Current Population Survey (CPS) instrument, where income questions are asked once annually. The 1992-94 SPD data come from the SIPP instrument, while the 1996-97 income and poverty data come from annual, retrospective questions such as in the CPS. Data users should therefore use caution when examining temporal trends in poverty.

5. Income Variables - Technical Notes

Be aware that the dollar amount reported for the variable “other income received in 1997” (OIVALT8) is not comparable to the dollar amounts reported in the prior years, 1992 (OIVALT2), 1993 (OIVALT3), 1994 (OIVALT4) and 1996 (OIVALT7). For these prior years the dollar amounts reported reflect the summation of, at most, two sources of other income received. Starting with the 1997 data (OIVALT8), the dollar amount reported for OIVAL reflects the summation of income reported from the following sources: other income; income from National Guard duty; income from casual earnings, such as hobbies; farm income; charitable assistance; and foster child care payments. This holds true, as well, across the years for both household (HOIVAL) and family (FOIVAL) amounts.

a. Energy Assistance Variables

Starting with data collecting in 1998, the reference time frame for HENGASE (Energy assistance received Y/N) and HENGVL (Value of energy assistance received) was “in the past 12 months”. For 1992, 1993, 1994, and 1997, the reference period was the six month time frame of October 1 through March 31 prior to the survey.

b. Public Assistance

Be aware of the following discrepancy in the longitudinal file variables (PAWMONE7 and PAWMONE8) that record the number of months that an individual received public assistance, AFDC, and TANF. The discrepancy occurs when an individual has a response of “yes” to having received public assistance in the prior year (PAWYNE7=1 and PAWYNE8=1), but the reported value for the number of months in receipt of public assistance is coded as zero (PAWMONE7=0 and PAWMONE8=0). This limited situation is the result of editing the data for consistency in the dollar amount of public assistance, AFDC, or TANF that an individual received. In the interview, respondents are given the opportunity at the end of the interview to list any other sources of income that they might have received, in addition to those already reported. For those limited number of cases where other income sources are reported, the cases are manually reviewed by a

subject matter analyst to see if the reported source can be included into an existing income source. While reviewing the data several analysts reported income sources which fit into the category of public assistance. This amount reported was then added to the public assistance amount and the PAWYNE7 and/or PAWYNE8 was changed from “no” to “yes”. The other income source reported, however, did not provide information on the number of months having received this income source, so the value for this limited number of cases will still contain a value of “0” for PAWMONE7 and/or PAWMONE8.

c. Dividends Variables

The universe for the variables which identify receipt of dividend income, DIVYNE7=1 and/or DIVYNE8=1, include individuals who reported owning stocks, mutual funds or other financial instruments that could pay a dividend, but reported not being paid dividends on these assets, DIVNONE7=1 and/or DIVNONE8=1. In the prior years, individuals that were not paid dividends on the potential dividend bearing assets they owned, DIVNONE2=1, DIVNONE3=1 and DIVNONE4=1, were not included as part of the universe for those variables concerning receipt of dividend income; DIVYNE2=1, DIVYNE3=1 and DIVYNE4=1. A similar situation exists for the variables HDIVYNE2, HDIVYNE3 and HDIVYNE4 because these household variables are aggregates of the person level variables.

c.1. Dividends Variables User Notes for the Second Longitudinal File

For 1999 and 2000, the imputation procedure to estimate a dollar value for dividends. DIVVAL, was modified to take into account if information was collected on the average value of all investments held by the household. A review of the relevant research indicated that on average, the dividend return to an average investment portfolio was approximately 4.5% over the past fifty years. This average return value was then used to estimate a value for missing dividends if an average value for investments was reported and the imputation flag for the variable was set to a value of three indicating that a statistical imputation was performed. Otherwise, if it was not possible to statistically impute the missing dividend value, a suitable donor was used to impute the missing dividend value and the imputation flag was set to one.

d. Free Lunch Variables

Users need to be aware that comparisons over time for the reported values in the variables HFLUNC and HFLUNN are not appropriate. In particular, the universe of children for these questions changed between the 1997 and 1998 data collections. For calendar year 1996 (HFLUNCE7), approximately 33% of households reported having one child receiving a free or reduced price meal at school, and for calendar year 1997 (HFLUNCE8), nearly 95% of respondents indicated that they had one child receiving a free or reduced price meal at school.

d.1. Free Lunch Variables User Notes for the Second Longitudinal File

For 1998, 1999 and 2000, the edited variable, FLUNC that determines whether a household has children that participate in the free or reduced price lunch and breakfast program at their school was modified to reflect the income tests required by this program. Specifically, households where the income to poverty threshold was greater than 185% should not be reporting children participating in this program.

e. Earnings Variables

There are 1,250 observations where the data reported in '92, '93 and '94 are a positive number of weeks worked (WKSWRK > 0), but there was no reported earning source from the longest job (ERNSRCE = 0) and no wage and salary data reported as well (WSVAL = 0). However, for these cases there is a reported value for longest job class of worker (CLSWKR > 0). Using this information, if the value of CLSWKR > 0, the value of ERNSRCE = 0 and the value of WKSWRK > 0, then the value of ERNSRCE was set to reflect the value of CLSWKR. Be aware that for these cases there will still be no wage and salary data (WSVAL = 0).

6. Employment Variables - Technical Notes

a. Number of Weeks Nonworker on Layoff

Use caution when using data for the number of weeks nonworker on layoff, ELAYWKE8, ELAYWKE9, and ELAYWKE0, as the weighted data is based on a very small number of observations. Data for this variable has been included for completeness.

b. Number of Stretches Looking for Work or on Layoff

The 1998, 1999, and 2000 SPD do not specifically ask the respondent whether the weeks looking for work or on layoff were in one stretch. Instead, LKSTRHE8, LKSTRHE9, and LKSTRHE0, the number of stretches looking for work, are constructed from the respondent's answers to questions 25 and 27 in the 'Employment and Earnings' section of the SPD questionnaire. In the 1998, 1999, and 2000 SPD, a stretch of looking for work or on layoff is considered to be interrupted when the respondent found employment or ceased looking for work for at least a one week period. For LKSTRHE7 (which is based on the CPS question design), a stretch of looking for work or on layoff is considered to be interrupted when the respondent found employment or ceased looking for work for at least a two week period. Hence, LKSTRHE7 is not perfectly comparable to LKSTRHE8, LKSTRHE9, and LKSTRHE0.

c. Weeks Looking for Work or On Layoff from a Job - Nonworker

The values for the variable NWLKWK, weeks looking for work or on layoff from a job (nonworker), vary considerably between the survey years 1994, 1997 and 1998. This is primarily due to the different survey designs associated with these years. Because information on the weeks spent looking for work or on layoff for the years 1992-1994 was collected monthly concerning current time spent looking for work or on layoff, while the 1997 information was collected from a fixed reference point about activities during the past year, the variables NWLKWKE4, NWLKWKE7 and NWLKWKE8 have considerable variation. To create annual information for the years 1992-1994, the time spent looking for work or on layoff was aggregated from data collected for the previous 4 months and then expanded to create an annual measure.

d. Industry of Longest Job by Major Industry Group

Individuals with a three-digit industry group code of 991 "Active Duty Armed Forces" are given a major industry group code 15 ("Never Worked") for the 'Bridge' data (WEMINDR7) and a major industry group code of '14' ("Public Administration") for the 1998, 1999, and 2000 SPD data (WEMINDR8, WEMINDR9, and WEMINDR0).

e. Number of Weeks Worked

These variables are a composite variable, composed of the number of weeks worked and the number of weeks on paid leave, such as vacation and sick leave. While most respondents reported their weeks worked, very few provided a response for weeks spent on paid leave (for example, there are 309 unweighted responses for the survey year 1998). Consequently, this led to a high imputation rate for these variables (WKSWRKE8, WKSWRKE9, and WKSWRKE0). However, for the majority of respondents, this high imputation rate reflects missing paid leave information, not missing weeks worked information. Note that for the years 1992, 1993 and 1994, WKSWRK (weeks worked and on paid leave) is an aggregate of monthly values, whereas for the survey years 1997, 1998, 1999, and 2000, the information was collected from a fixed reference point about the previous year's activities. Fluctuations between the survey years 1992 – 1994 and 1997 – 2000 may be due to these different reference periods.

NOTE: For further questions concerning variables comparability between survey years (particularly when comparing 1992 - 1994 variables to 1997 variables), please consult the "SIPP to CPS Cross-Reference Map" (Urban Institute, 1998) in Appendix D.

7. Health Insurance Variables

Several health insurance variables have inconsistent definitions over time, despite being named with the same base characters. Please read the data dictionary closely when using these variables, including MCAREE2-MCAREE8 and CH_MCE2-CH_MCE8.

a. HIOWNE2 - HIOWNE8: Did adult covered by private health insurance plan have policy in own name?

The universe for this variable is adults with a private health insurance plan. (HI_YNE_x=1) For 1992, 1993, and 1994, adults who had no private health insurance plan (HI_YNE_x=2) were assigned a value of 2 (no) to the HIOWNE_x. In 1997 and 1998 people without a private plan were assigned a 0, the correct code for "not in universe". One way to correct the problem is to use HI_YNE_x. If HI_YNE_x=2 then HIOWNE_x=0.

b. Household Variables On Children's Health Insurance Coverage

The household variables reflecting health insurance coverage of children include some inconsistencies in 1994 and 1997. CMCAREE_x and CHIEx have as their universes households with children under 15 years old. In 1994, there is some random noise with a handful of cases where people should be in universe but are not and vice versa. Double check the universe variable (UNDR15R_x) before completing your analysis. In 1997, most people living in households without children have a 2 meaning no children in the household have that coverage type. These people should instead have a 0 for "not in universe".

c. Notes as a result of 1999 Instrument Changes. These notes apply to the 1999 and 2000 instruments only.

MCAID - Was person covered by Medicaid at any time between January and December [of previous calendar year]?

NOTE(S): This variable also includes 1999 survey instrument items Q950_8 (Stateplan) and Q950_9 (Another government program). The 1999 survey instrument item Q950_9 is not capturing the same data as in the 1998 instrument (definition is not the same as

OTYP_5E8). The 1999 instrument question is a measurement of Medicaid. For consistency in the longitudinal file, MCAID combines these three categories (Q950_4, Q950_8, and Q950_9).

MON - Number of months covered by Medicaid [in previous calendar year] (values are 1-12)

NOTE(S): This variable counts months of coverage for either Medicaid or (Stateplan) or “another government program” (since all three are capturing Medicaid coverage). Since the three were combined in the MCAID variable, starting with the 1999 Instrument, it does not make sense to record occurrences (and sums) of each variable separately.

OTYP_5 - Was person covered by other (state plan) health plan at any time during [previous calendar year]?

NOTE(S): Based on inputs from 1998 Instrument variable (OTYP_5E8), it has been determined that this variable is NOT equivalent to Q950_9 in the 1999 instrument. The question was previously combined with CHAMP in the 1998 instrument, but when broken down as a separate question in the 1999 instrument, it lost its original intent, which was to collect “other types of *military* coverage.” The 1999 Instrument variable, Q950_9, will be included in the MCAID variable.

OTH - Excluding Medicare and Medicaid, was person covered by CHAMPUS, CHAMPVA, VA, military health, or Indian Health Services?

NOTE(S): The edit for OTHE9 is significantly changed from previous year (OTHE8), and only captures military types of coverage and Indian Health Services (CHAMP AND OTYP_4). Imputation flags for this variable are carried over from the imputations done on the corresponding coverage variables.

HI - Was adult household member a policyholder of a group health plan?

NOTE(S): There are new questions in the 1999 instrument that ask up front whether the respondent is the policy holder or dependent of the plan they recorded being covered by in Q950. This variable is created from the new questions (Q951A and Q951C). Previously in the 1998 instrument employer-provided and union provided were asked in the same question. The 1999 instrument separates them. To retain comparability over time, responses to Q951A and Q951C are combined in HI. Minimal changes to edit were made to accommodate for this change.

DEPHI - Was person a dependent on a household group health plan?

NOTE(S): This is the companion variable to HI and is derived from the same question (Q951A and Q951C). Those respondents less than 15 years of age (if they checked that they were covered by an employer or union-provided plan) auto-default to dependents (they skip question Q951A or Q951C and go right to the questions that ask about line number of policy holder – Q951B or Q951D). (See HI for changes to instrument that affected this variable.)

COV_GH - Was person covered by a group health insurance plan during [previous calendar year]?

NOTE(S): This variable is created from HI and DEPHI (if you were a dependent or a policyholder and marked employer-provided or union provided in Q950 (Q950_1 or Q950_2), you're considered covered by a group health plan). (See HI for changes to instrument that affected this variable.)

PAID - What part of health insurance premium did employer/union pay?

NOTE(S): As a result of changes to the 1999 Instrument, this variable combines data asked of those covered by either an employer-provided or union-provided plan (given that employer-provided coverage and union-provided coverage were asked independently of each other).

PRIV - Was adult household member a policyholder of a direct-purchase health insurance plan?

NOTE(S): Previously this variable had a universe of GE 15 years of age. The 1999 instrument, however, allowed for those under 15 years of age to answer as policyholders for a direct-purchase health plan. The unedited data file contains 29 occurrences of policyholders under age fifteen. 23 of these cases reported a dollar amount when asked how much they paid last year for this plan. Given that the longitudinal file does not contain any occurrences of a policyholder under the age of 15, it was decided to maintain consistency and edit those under-age-15 policyholders, placing them in the category of dependent. It is possible that a proxy (i.e., the child's parent) responded that their child was the policyholder of the plan, because the parent bought the policy, put it in the child's name, and paid for it. This is why a dollar amount could also be recorded when asked.

DEPRIV - Was person a dependent on a direct purchase private health plan?

NOTE(S): (see PRIV for policyholders under 15)

CH_HI - What is the child's private health insurance status? (Values are: Private coverage through someone in household, someone outside household, or not covered by private health plan.)

NOTE(S): Group health insurance (COV_GH) and dependent of a direct purchase (DEPRIV) are used for this variable. Universe is those people less than age 15.

CURR - Was person covered by any health plan last week?

NOTE(S): This variable is constructed from responses to a variety of questions in the health insurance section. As a result, the universe of the variable CURR (created from the 1999 instrument), takes into account responses to questions throughout the instrument that ask respondents about current coverage by the same plan as last year:

(q955 = '1' or q959 = '1' or q962 = '1' or q964 = '1' or q966 = '1' or q968 = '1' or q970 = '1' or q971C = '1' or q972 = '1' or q974 = '1')

So, if a respondent said "yes" to Q977 or they said "yes" to (Q955, Q959, Q962, Q964, Q966, Q968, Q970, Q971C, Q972, or Q974), then they were coded as "being covered by a health plan last week."

CTYP_(1-5) - What type of health plan was person covered by last week? (allow up to five types to be selected.)

NOTE(S): The 1999 instrument does not necessarily capture every type of “current coverage.” If respondents had one type of coverage last year and they were “currently covered” by that same plan at the time of interview, the instrument is not designed to capture new types of coverage that the respondent might have picked up between the reference period of “last year” and the time the survey questions were asked. Also, if respondents had more than one type of coverage and they were “currently covered” by one plan but not the other (lost a type of coverage), the instrument is not designed to capture any new types of coverage that the respondent might have picked up between the reference period of “last year” and the time the survey questions were asked.

This variable was created from the “current coverage” questions asked after responses were gathered for each type of health insurance. It also includes the current coverage question at the end of the section (Q978). Values for the edited variable were changed due to a change in the instrument from 1998 to 1999. All types of military health are included in one value. There is no dependent or policyholder separation. “Other government health care” was dropped.