

PANEL STUDY OF INCOME DYNAMICS  
1997 PUBLIC RELEASE I FILES  
DOCUMENTATION

---

TABLE OF CONTENTS

I. INTRODUCTION

- A. REASONS FOR AND LIMITATIONS OF THE PUBLIC RELEASE I FILES
- B. WHAT'S NEW FOR 1997
  - 1. NEW IMMIGRANT SAMPLE ADDITION
  - 2. REDUCTION IN CENSUS SAMPLE AND RESURRECTION OF BLACK CHILDREN
  - 3. BACKGROUND QUESTIONS
  - 4. CHILD SUPPORT QUESTIONS AND CHILD DEVELOPMENT SUPPLEMENT

II. CHARACTERISTICS

- A. FILES AND FORMAT
    - 1. THE .DAT FILES
    - 2. THE .SAS AND .SPS FILES
    - 3. THE .TXT FILES
  - B. VARIABLE NAMES, POSITIONS AND GENERATED VARIABLES
    - 1. FAMILY FILE VARIABLES
    - 2. CROSS-YEAR INDIVIDUAL FILE VARIABLES
  - C. DOCUMENTATION AND CODES
    - 1. 1997 QUESTIONNAIRE
    - 2. CODE CATEGORIES
  - D. BRINGING FORWARD BACKGROUND INFORMATION FOR HEAD AND WIFE/"WIFE"
  - E. PROBLEM VARIABLES, MISSING VARIABLES
    - 1. FAMILY FILE PROBLEM VARIABLES
    - 2. CROSS-YEAR INDIVIDUAL FILE PROBLEM VARIABLES
    - 3. CHANGE IN FAMILY-LEVEL VARIABLE LABELS FOR 1996
  - F. COMPARABILITY WITH 1993 PUBLIC RELEASE II FILE AND EARLIER PUBLIC RELEASE II FILES
  - G. NOTES ON NON-QUESTIONNAIRE VARIABLES IN 1997
- 

I. INTRODUCTION

A. Reasons for and Limitations of the Public Release I Files

The more than two-year interval between the completion of interviewing on a given PSID wave and the public release of a fully cleaned and documented data file has prompted demand for speedier release of an "Public Release I" version of PSID data files. In response to this demand, the PSID staff produced Public Release I versions of the 1990 through 1993 family and the 1968-1993 cross-year individual files; all these files are now available in Public Release II form.

These were followed by Public Release I versions of the 1994, 1995, 1996 and 1997 waves of data. The latest in this series is the Public Release I version of the 1997 wave of data -- a Public Release I version of the 1997 family file and of the 1968-1997 cross-year individual file. The Public Release I files are available both at the PSID's internet site.

The advantage of the Public Release I files is quicker access to recent waves of data; disadvantages include:

- \* more limited documentation -- for the 1997 Public Release I files, no documentation other than this readme file, univariate frequencies and statistics, and the 1997 questionnaire;
- \* missing variables -- relative to past waves, a more limited set of family and individual variables, most importantly in the family file

annualized work hours components and totals, generated variables such as the income to needs ratio and prorated poverty thresholds, and open-ended variables such as occupation and industry, and in the individual file the "summary variables" including variables about marital and fertility histories;

- \* minor problems with data values -- no item-by-item imputations for missing data, a few wild codes, and some inaccurate skip patterns on a handful of cases (due to post-data collection editing by staff without consistency checking); and

## B. What's New for 1997

We retained a very few questions from the 1996 module about home financing. We still ask what type of loan the family has, whether it is the original loan or refinanced, and what the current interest rate is. At the least, these will be helpful in imputations.

Otherwise, the core content of the questionnaire remained the same as in 1996, with the addition of both incoming and outgoing child support questions, the reasking of many background information items for heads and wives/"wives", and some new background questions relevant to the new immigrant sample.

### 1. New Immigrant Sample Addition

The PSID's original sample was selected in 1968, but since that time the profile of the U.S. population has changed because of immigration. In order to represent the current population more accurately, we mounted a large screening effort for 1997 interviewing. We managed to locate and persuade 441 immigrant families to be interviewed. These families are included on the files along with the core PSID families.

Details of the immigrant sample and its sample weights are provided in a 2000 PSID technical working paper on our website.

Immigrant families and individuals were assigned a unique range of values so that they can easily be distinguished from the other PSID samples. The 1997 Interview (ID) Number is in the range 10001-10441, and the 1968 Interview (ID) Number is in the range 3001-3441. Person Numbers for immigrants were assigned in the usual manner, with all those present in the first (1997) wave receiving values from 1-20. Institutionals received values of 20-29, and movers-out were given nonsample Person Numbers (170 or greater).

All new heads and wives/"wives", both immigrant and core sample, were asked in what year they came to the U.S. to stay. These questions served as screening questions for the immigrant sample and were useful in development of 1997 sample weights for both the core and the immigrants. These variables are located on the 1997 portion of the cross-year individual file.

### 2. Reduction in the Census Sample and Resurrection of Black Children

In order to keep data collection costs down, we were forced to cut the PSID sample for 1997. Several scenarios were discussed, but in the end, the Census (SEO) subsample was selected for reduction by two thirds. The rule was that all current Census sample families sharing the same original 1968 Interview (ID) Number would either be kept or cut.

At the eleventh hour, we received funding that enabled us to reinstate some of the dropped families because of the intense interest in children and child development for the 1997 wave. The families to be reinstated were headed by a Black individual and contained at least one child aged 12 or under in 1996. If other related families (those sharing the same 1968 ID Number) had been dropped, they were not reinstated unless they, too, met the rule about race of head and presence of a child in the proper age group.

### 3. Background Questions

Because of the new immigrant sample for 1997, many of the PSID's traditional background items (from questionnaire Sections K and L) were reasked of all heads and wives/"wives", and some new items were added.

We collected information about where parents were born, where they grew up, their education, and their occupations. We asked race, ethnicity and occupation of the first full-time, regular job for head and wife/"wife". The information about where parents were born is completely new, but we have

always asked where the head's parents grew up, their education, and the father's occupation. These items had not been regularly asked of wives/"wives", however.

We did not reask heads and wives/"wives" who retained their head/wife/"wife" status from 1996 about the number of their siblings and whether those siblings were living or deceased, nor did we repeat the education series and religious preference questions for them.

Heads and wives/"wives" from the new immigrant sample were asked questions about their immigration experience (questionnaire Section M, ER11901-ER11982 for heads and ER11983-ER12064 for wives/"wives").

#### 4. Child Support Questions and Child Development Supplement

An extensive set of questions was asked about both incoming and outgoing child support, custody and visitation. These items were asked for all children in the family under age 18, and parallel information was collected for all parents in the family who had children living elsewhere. These data are not available as part of the PSID main Public Release I files (and will not be part of the main Public Release II files), but instead are destined to become a separate, supplemental file.

The Child Development Supplement is an intensive study of 3,500 PSID children aged 12 or under. It includes assessments of the children's cognitive, behavioral and health status; caregiver time inputs; time use in school; and other measures of such things as school and neighborhood resources and the home learning environment. The CDS has its own website:

<http://www.isr.umich.edu/src/child-development/home.html>

Data are available there for linking with PSID main data.

---

## II. CHARACTERISTICS

### A. Files and Format

The Public Release I package for the 1997 wave consists of two data files: the 1997 Public Release I family data file and the 1968-1997 Public Release I individual data file.

File name	Number of variables	Number of Records	LRECL	Contents
ER97F.DAT	2,084	6,748	3,558	1997 raw family data
ER68-97I.DAT	1,014	59,888	2,318	1968-1997 raw individual data

In addition to these two data files and the file you are reading now, README.TXT, the 1997 Public Release I packages include a number of other files listed in the tables immediately below. Their contents are described in the subsequent, more detailed paragraphs.

#### Other 1997 Public Release I Family Files

File name	Contents
ER97F.SAS	SAS statements for family variables
ER97F.SPS	SPSS statements for family variables
ER97FTAB.TXT	Frequencies and means for family variables
ER97FMD.SAS	SAS missing data statements

#### Other 1968-1997 Public Release I Individual Files

File name	Contents
ER68-97I.SAS	SAS statements for individual variables
ER68-97I.SPS	SPSS statements for individual variables
ER97ITAB.TXT	Frequencies and means for individual variables

### 1. The .DAT files

The data are in raw ASCII form. Refer to the corresponding .SAS or .SPS file for record format layout information, variable names and variable labels.

The 1997 Public Release I family data file contains one record for each family interviewed in 1997 and includes all family-level variables collected in 1997.

The 1968-1997 Public Release I individual data file is a merged cross-year file that contains a record for both 1997 response and 1997 nonresponse individuals; it has one record for each individual who was in an interviewed family for any wave of the study. The file includes individual-level variables collected from 1968 through 1997.

## 2. The .SAS and .SPS files

ER97FAM.SAS and ER97FAM.SPS, respectively, contain SAS and SPSS data definition statements that provide variable names, locations, and variable labels. The SPSS and SAS statements are NOT intended to represent complete and full programs for the respective statistical program packages to run extracts, analysis, etc. You must provide all other SPSS or SAS statements needed to complete a program.

Although traditional missing data information will be included with the Public Release II version of the data, missing data statements have not been provided as part of the .SAS and .SPSS files above. We have generated separate missing data statements for family variables that may be added to them (see the description for ER97FMD.SAS immediately below), but be sure to check the questionnaire and frequencies or means for each variable you intend to include in your analysis to determine whether you concur with these definitions.

ER97FMD.SAS contains quick-and-dirty missing data statements for all family variables. The statements were machine-generated and assign missing data to over-the-field amounts (codes of 97, 997, etc.) as well as to refused, don't-know and other missing responses. Zeroes are also recoded to missing data in these statements, even where zero is a substantive response meaning "none". We have not produced equivalent statements for individual variables.

## 3. The .TXT files

These ASCII text files provide additional information about the Public Release I files.

ER97FTAB.TXT contains univariate frequencies and statistics for the 1997 family variables. ER97ITAB.TXT provides the same for 1994-1997 individual variables. We produce frequencies for variables with less than 35 separate code categories; univariate statistics are calculated for variables with 35 or more codes. All frequencies and statistics are unweighted.

No annotation of differences in codes for the Public Release I files has been done, although this should not be a problem for most variables. The primary code scheme difference is the time unit codes for dollar amounts; a uniform code for all such was used from 1996 forward, but the 1994 and 1995 waves are idiosyncratic. Check questionnaires thoroughly!

## B. Variable Names, Positions and Generated Variables

All variable names for both 1997 Public Release I family and 1968-1997 Public Release I individual files are prefaced with "ER" rather than "V" to assist both analysts and study staff in determining whether reference is to the Public Release I (formerly called "Early Release") file or to its Public Release II version.

### 1. Family File Variables

The 1997 Public Release I family variables are in the range ER10001-ER12084. Most of these variables will eventually be incorporated into the Public Release II version of the 1997 data, but their variable names will change and the data will be cleaner.

Variable names and locations for the 1997 Public Release I family file are not the same as those we intend for the Public Release II version. The 1997 Public Release I family file includes neither variable names nor positions for so-called "imputed" and "generated" family-level variables.

By "imputed" variables we mean annualized totals and accuracy indicator codes from the first several hundred (almost 800 in 1993) variables usually present in each wave's final family-level data, beginning after the state of residence and ending with aggregated income detail for other family unit

members. By "generated" variables we mean those variables traditionally located at the end of the raw data after the Head's background information.

In short, imputation and accuracy variables equivalent to those in the 1993 variable range V21610-V22399 are absent, as are equivalents to 1993 V23322-V23363. This year, for the first time, the Public Release I family file includes some preliminary imputed income totals and a weight variable.

Variables not included in the Public Release I file for which component items are available include:

- \* imputed and annualized mortgage and rent payments,
- \* imputed and annualized food costs,
- \* poverty thresholds,
- \* imputed and annualized work hours,
- \* imputed and annualized unemployment, etc., hours, and
- \* education of Head and Wife/"Wife".

Since component items exist on the Public Release I file, you can generate these items. Needless to say, imputations have NOT been done for missing data. To create variables from the 1997 Public Release I data that resemble those on Public Release II files from 1993 and earlier waves, we suggest you consult the 1993 codebooks where we provide sufficient information about how the variables were created for 1993 to create them for 1997.

Most background items have not been asked about all Heads and Wives/"Wives" each and every year. We ask the questions for new Heads and new Wives/"Wives" only. During processing, we have traditionally "brought forward" background information from previous waves for Heads or Wives/"Wives" who are the same persons as in the prior year. In every wave, each set of background variables is preceded by a variable indicating whether data needed to be brought forward. The 1997 Public Release I file, in keeping with our practice for other Public Release I files, has not undergone this "bringing forward". See Section D below for a detailed description of how you can do this yourself.

Other variables are not generatable because income components of individuals other than Head and Wife/"Wife" are not included in the 1997 Public Release I data. Variables not included in the Public Release I which cannot be generated from available information include:

- \* detailed incomes for other family members (although collective transfer and taxable totals are included on the Public Release I family file),
- \* urbanicity,
- \* Head's geographic mobility,
- \* county unemployment rate, and
- \* variables linking related families who share the same dwelling.

## 2. Cross-Year Individual File Variables

Recent cross-year individual PSID files have consisted of annual measures and a set of "summary variables" that have appeared at the end of the individual data record. In the 1968-1997 Public Release I individual file, most of the annual measures (e.g., Sequence Number, Relationship to Head, Family Identification Numbers) are available. However, virtually NONE of the "summary variables" (i.e., V31996-V32049) are included; the single exception is V32000, Sex of Individual, which was too important to omit; it appears in the 1968-1997 Public Release I individual file as ER32000.

Variables ER30001 through ER30866 will remain the same for the Public Release II version (with the prefix change from "ER" to "V"). Now that weights for each wave of Public Release I data are available, they have been added to the Public Release I file at the end of each year's variables. A few more variables will be added to the 1994-1997 individual data. The order of variables in the 1968-1997 Public Release I individual file is as follows:

- \* RELEASE NUMBER, ER30000,
- \* 1968 through 1993 individual data arranged, as usual, by wave, ER30001-ER30866
- \* the lone summary variable, SEX OF INDIVIDUAL, ER32000,
- \* the 1994 Public Release I individual data, ER33101-ER33121,
- \* the 1995 Public Release I individual data, ER33201-ER33277,
- \* the 1996 Public Release I individual data, ER33301-ER33318, and
- \* the 1996 Public Release I individual data, ER33401-ER33430.

For the Public Release II version, the 1994, 1995, 1996 and 1997 variables will

be moved to follow to completed 1993 individual data and ER32000 will appear in its usual place among the summary variables as V32000.

Some 1994, 1995, 1996 and 1997 equivalents of traditional annual individual variables are not included in the Public Release I file:

- \* individual income components and totals,
- \* linking measures for splitoffs, and
- \* reason for nonresponse.

In the Public Release II individual files, these variables will be located near the end of the yearly data, just as in 1993 and earlier waves.

### C. Documentation and Questionnaire

We have not produced the traditional codebooks for the Public Release I files. Our website includes an HTML depiction of the computer-assisted interviewing application for 1997. This replaces the PDF-format questionnaires we have provided for the past several years. Use the SAS and SPSS data definition statements to match variables with questions in the CAI application. The HTM application contains codes for most data items.

### 2. Code Categories

The HTM version of the 1997 interviewing application provides most codes for the family data. In general, codes follow our traditional structure, although "don't know" responses are now largely distinguished from other missing data responses. Generally, code 8 (or 98 or 998, etc.) represents "don't know" and code 9 (or 99 or 999, etc.) represents other missing data or a refusal. Inappropriate questions are padded with zeroes. If a variable contains a code value that is neither included in the application nor one of the "zero", "eight" or "nine" codes just mentioned, assume missing data for that value. We will correct such cases for Public Release II, but time constraints do not permit this sort of cleaning for Public Release I.

For individual data, use the codebook in our 1993 documentation, Section II, Part 2; similar variables for 1994, 1995, 1996 and 1997 are coded identically to those from earlier waves.

A few new variables are included on the Public Release I family and individual files but are not part of the application. See Part G for further help.

### D. Bringing forward Background Information for Head and Wife/"Wife"

As noted above, background information that was not reasked in 1997 for Head and Wife/"Wife" has not been "brought forward" for the Public Release I family file. Background information is complete for 1993 on the 1993 Public Release II family file, but as of this writing, the 1994-1997 family data are available only in Public Release I form and have not yet undergone the bringing-forward process. Only families with Heads and Wives/"Wives" who were new in 1997 have all the background data in the 1997 Public Release I family file. You must search, respectively, the 1996, 1995, 1994 and 1993 family data to complete 1997 background variables.

Carefully compare the background variables item for item and code for code in the 1993 Public Release II family file and the 1994-1997 Public Release I family files before you attempt to bring forward prior-wave background information. You should be aware that the 1994-1997 background variables are not necessarily completely identical to each other!

In addition, some 1993 background questions are not included at all in the 1994-1997 Public Release I family files' background data because they have NOT YET been coded; among these are questions about:

- \* Head's/wife's/"wife's" father's occupation, and
- \* number of states and regions in which Head has lived.

One more factor complicates bringing forward background data: the absence of the 1993, 1994, 1995 and 1996 family interview (ID) numbers on the 1997 family file. You must obtain these variables from the Head's record in the 1968-1997 Public Release I individual file in order to match with 1993-1997 family files to bring forward the background information.

Below is a suggested procedure for bringing forward the background information.

Step 1. First, add the 1997 Head's 1993, 1994, 1995 and 1996 interview (ID)

numbers from the 1968-1997 Public Release I individual file (ER68-97IND) to the 1997 family file (ER97FAM):

- a. sort ER97FAM by interview (ID) number (ER10002 "1997 INTERVIEW #")
- b. sort ER68-97IND by 1997 interview (ID) number (ER33401 "1997 INTERVIEW NUMBER") for 1997 Heads (ER33402 "INDIVIDUAL SEQUENCE NUMBER 97" = 01)
- c. merge ER97FAM and ER68-97IND by 1997 interview number and add Head's 1996, 1995, 1994, and 1993 interview numbers (ER33301 "1996 INTERVIEW NUMBER", ER33201 "1995 INTERVIEW NUMBER", ER33101 "1994 INTERVIEW NUMBER", and ER30806 "1993 INTERVIEW NUMBER"). Output is ER97FI.

Step 2. Obtain information from the 1997 family file; check to determine whether the 1997 family includes a wife/"wife" and whether new head and new wife/"wife" information is present in the 1997 Public Release I family file. If it is, then the appropriate background information is already part of the 1997 Public Release I family file, and this case needs no further processing.

- a. If there is no wife/"wife" in 1997 (ER10011 "AGE OF WIFE" = 0) then wfstatus=1 else wfstatus=0
- b. If there is a new wife/"wife" in 1997 (ER11731 "K1 CKPT: WTR NEW WIFE" = 1) then wfstatus=1
- c. If there is a new head in 1997 (ER11812 "L1 CKPT: WTR NEW HEAD" = 1) then hdstatus=1 else hdstatus=0

Step 3. If new Head or new Wife/"Wife" information was not present in the 1997 family file, check to determine whether it is present in 1996. If it is, then replace the values of the variables in the 1997 family file with values of the corresponding variables from the 1996 file. Remember that these variables differ slightly from year to year, and that much information was reasked in 1997--but not education and siblings, to name two primary sets of background info.

- a. sort ER97FI by 1996 interview number (ER33301 "1996 INTERVIEW NUMBER")
- b. sort ER96FAM by 1996 interview number (ER7002 "1996 INTERVIEW #") for new 1996 wife/"wife" (ER8979 "K1 CKPT: WTR WIFE" = 1) or for new 1996 head (ER9033 "L1 CKPT: WTR NEW HEAD" = 1)
- c. merge ER97FI and ER96FAM by 1996 interview number; if wfstatus=0 and ER7002=1, bring forward 1996 new wife/"wife" info and reset wfstatus to 1; if hdstatus=0 and ER9033=1, bring forward 1996 new head info and reset hdstatus to 1. Output is ER96-97FI.

Step 4. If new head or new wife/"wife" information was present in neither 1997 nor 1996, check to determine whether it exists for 1995. If it is, then replace the values of the variables in the 1996-1997 output file from the preceding step with values of the corresponding variables from the 1995 family file. Be sure to check for variables that differ and don't bring forward data items that were reasked in 1997!

- a. sort ER96-97FI by 1995 interview number (ER33201 "1995 INTERVIEW NUMBER")
- b. sort ER95FAM by 1995 interview number (ER5002 "1995 INTERVIEW #") for new 1995 wife/"wife" (ER6733 "K1 CKPT: WTR WIFE" = 1) or for new 1995 head (ER6787 "L1 CKPT: WTR NEW HEAD" = 1)
- c. merge ER96-97FI and ER95FAM by 1995 interview number; if wfstatus=0 and ER6733=1, bring forward 1995 new wife/"wife" info and reset wfstatus to 1; if hdstatus=0 and ER6787=1, bring forward 1995 new head info and reset hdstatus to 1. Output is ER95-97FI

Step 5. If new head or new wife/"wife" information was not present in 1997, 1996 and 1995, check to determine whether it exists in the 1994

family file. If it is, then replace the values of the variables in the 1995-1997 output file with values of the corresponding variables from the 1994 family file. Again, recall the slight differences and the 1997 reaskings.

- a. sort ER95-97FI by 1994 interview number (ER33101 "1994 INTERVIEW NUMBER")
- b. sort ER94FAM by 1994 interview number (ER2002 "1994 INTERVIEW #") for new 1994 wife/"wife" (ER3863 "K1 CKPT: WTR WIFE" = 1) or for new 1994 head (ER3917 "L1 CKPT: WTR NEW HEAD" = 1)
- c. merge ER95-97FI and ER94FAM by 1994 interview number; if wfstatus=0 and ER3863=1, bring forward 1994 new wife/"wife" info and set wfstatus=1; if hdstatus=0 and ER3917=1, bring forward 1994 new head info and set hdstatus=1. Output is ER94-97FI

Step 6. Last, if new head or new wife/"wife" information was not present in the 1997, 1996, 1995 or 1994 family files, obtain the information from the 1993 Public Release II family file. There is no need to check the value for the 1993 indicator, as all 1993 cases contain background information. Replace the values of the variables in the 1994-1997 output file with values of the corresponding variables from the 1993 Public Release II family file, again recalling that these variables do not match perfectly and that some background info was reasked in 1997. This is the final step.

- a. sort ER94-97FI by 1993 interview number (ER30806 "1993 INTERVIEW NUMBER")
- b. sort FAM93 by 1993 interview number (V21602 "1993 INTERVIEW NUMBER")
- c. merge ER94-97FI and FAM93 by 1993 interview number; if wfstatus=0, bring forward 1993 wife/"wife" info; if hdstatus=0, bring forward 1993 head info. Output has background information for all 1997 heads and wives/"wives".

#### E. Problem Variables, Missing Variables

Some variables included on the 1997 Public Release I files are known to include bad or completely missing data. These will be corrected for the Public Release II version of the file, but in the meantime we want you to be informed of the following known problems with and idiosyncracies of the data.

##### 1. Family File Problem Variables

In general, we are very pleased with our Public Release I data quality for 1997. We have conquered past problems with data extraction that had heretofore caused unfortunate gaps in the data items.

Since for the most part these data have not yet been edited, wild values given by respondents or entered by interviewers remain. Some family variables contain wildly improbable or incorrect values:

- |         |  |
|---------|--|
| ER10032 | A16 ACTUAL 3 ROOMS has values of 50 and 97.  |
| ER10084 | B2 YEAR RETIRED (HD-R). Almost all cases that should have been asked this question have values of 9999. This was due to an error in the data collection instrument, so unfortunately the information is simply missing for 1997. |
| ER10566 | D2 YEAR RETIRED (WF-R). See ER10084 above.   |

Values for variables referring to nonworking heads' third and fourth extra jobs (ER10519-ER10562) contain only zeroes because in this wave no such head had more than two extra jobs. Similarly, working wives/"wives" reported no fourth extra job (ER10785-ER10806), and nonworking wives/"wives" reported no second through fourth extra jobs (ER10979-ER11044).

All ending wage rates and time units for second main jobs of nonworking heads/wives/"wives" are included in ER10419-ER10420 and ER10901-ER10902, not ER10417-ER10418 and ER10899-ER10900. These latter two sets of variables should not have been included in the data at all and contain zeroes for everyone.

##### 2. Cross-Year Individual File Problem Variables

In the 1968-1996 Public Release I individual file the EMPLOYMENT STATUS

variables, ER33111, ER33211, ER33311 and ER33411, for 1994, 1995, 1996 and 1997, respectively contain zeros for every person in the file.

### 3. Change in Family-Level Variable Labels for 1996

In processing the 1996 Public Release I file, we discovered that the labels provided in earlier years for some Public Release I variables were incorrect, and new labels were also provided for a few other variables. All variables affected are listed in ER96READ.ME, part of the 1996 Public Release I documentation. We did not provide new SAS or SPSS statements for the pre-1996 files. However, if you use these "older" variables in your analysis, you should be aware of the changes.

### F. Comparability with 1993 Public Release II File and Earlier Public Release II Files

Beginning with the 1993 wave, the data were collected using CATI (Computer Assisted Telephone Interviewing). This meant that information about each question was collected electronically by the interviewer and, in effect, was coded at the time of data collection. Conversion to standardized units of measurement, formerly performed as part of our coding operation, has not yet been done. As a result, the data in the 1994-1997 Public Release I and the 1993 Public Release II family files much more directly resemble the answers to questionnaire questions than the 1992 and early years' data did. For example, instead of one variable indicating monthly rental expense, rent costs now exist as two variables: one for the dollar amount and one for the time unit. Typical responses are \$500 per month, \$100 per week, or don't know how much per month. In addition, the 1993 Public Release II file includes an annualized rent amount, free of missing data, and an accuracy variable indicating whether an imputation was necessary to arrive at the annualization. This is more in line with 1992 and earlier procedures. But the Public Release I files have not yet been processed through the imputation algorithm. Therefore, many of the 1994 through 1997 Public Release I variables are not directly equivalent to variables from 1992 and earlier waves.

Unlike data collected through 1992, the Public Release I family data have NOT been cleaned with our manual economic edit process (nor have imputations been made), so you must convert these kinds of amounts into some sort of consistent unit for inter-case comparison and make decisions about handling missing data. In addition, we expect that values for quite a few cases will change when we do perform economic edit operations. For instance, time spent working, being laid off, unemployed, out of the labor force, etc., does not sum to 52 weeks per year in about 10 per cent of the cases in the Public Release I file. In addition, all time unit questions include an "other" code, as well as options for missing data; amounts associated with these "other" codes will be recoded from interviewers' notes when the data are cleaned.

You can create cross-year files with variables needed for your particular analysis by merging the necessary information from the appropriate family files. The needed family identification numbers appear both in the Public Release II and Public Release I cross-year individual files. Since they do not appear on any family data files from 1993 forward, you can obtain them from the Head's record in the 1968-1997 Public Release I cross-year individual file (see Section G, step 1, above). Detailed instructions for the process of creating the cross-year files are included in the 1993 family documentation volume and are also available at our Internet site; thus, they are not repeated here.

Much of our usual inter-year consistency checking was performed for the Public Release I 1968-1997 cross-year individual file, so we expect the records in this file to remain relatively stable for Public Release II.

### G. Notes on Non-Questionnaire Variables in 1997

The 1997 Public Release I family file includes some new variables that are not lifted directly from the questionnaire. Most of them are income summaries and are located at the end of the data file. Documentation about their calculation and imputation can be found in the package for the Family 'Income Plus' Files: 1994-1997, available with other supplemental files on our website. These income amounts do not quite match the 1997 'Income Plus' files, however, as we recalculated for the Public Release I data and included new immigrant cases (which were excluded from the 'Income Plus' file).

Each of these income components has an associated accuracy code that indicates whether imputations were made in order to calculate totals. Accuracy variable codes are 0 for no imputation and 1 if an imputation was

needed.

The state of residence is included on the 1997 Public Release I family file. This variable contains FIPS values. The code for FIPS states is located in Appendix 1 of the codebook for the 'Income Plus' files. See our Supplemental Files web page.

For the first time, we include a preliminary weight variable as part of the Public Release I family data. This variable is located at the very end of the data. The weight variable takes account of the new immigrant sample. It is scaled so that it may be used for the core sample, the immigrant sample, or the two samples together.

On the 1968-1997 individual file, we have included three variables related to the Child Development Study (CDS): whether eligible for the CDS (ER33418), whether selected as a CDS subject (ER33419), and the outcome of the interview attempt (ER33420).

All sample persons in the 1997 family aged 12 or under were considered eligible for the CDS and are coded 1 on ER33418. All other cases contain values of 5 or 0. The next two paragraphs explain code 5.

The maximum number of CDS interviews per family was limited to two. Thus, in a family with three or more eligible children, two were randomly selected for interview. ER33419 contains values of 1 for those children who were selected, and values of 5 for those who were eligible but not selected. All other individuals have values of 0 here.

Note that around 80 children are coded 5 on ER33418, whether eligible. All these children have values of 1 for ER33419, whether selected. These individuals were thought to be sample members during main PSID interviewing and were selected and interviewed for the CDS, but during the family composition editing process we discovered that they indeed were not sample. Their interviews were kept by the CDS, but ER33418 indicates their retroactively discovered ineligibility.

The third variable in this series, ER33420, contains codes indicating the reason for CDS nonresponse for those who were selected. We list the codes below.

- 1 Interview with sample child
- 2 Interview with nonsample child
- 3 Nonresponse--refused
- 4 Nonresponse--lost
- 5 Nonresponse--incapacitated, permanent condition
- 6 Nonresponse--deceased
- 7 Nonresponse--could not contact primary caregiver for actual interview
- 8 Nonresponse--area too dangerous for interviewer
- 9 Nonresponse--language barrier
  
- 20 Nonresponse--study ended before interview could be completed
  
- 96 Office Error--interviewer reports interview taken, but no primary caregiver interview logged
- 97 Office Error--eligible child who should have been selected for interview
- 98 Office Error--outside continental U.S., should have been called
- 0 Inap.; ineligible and no interview taken; eligible but not selected

The 1968-1997 Public Release I cross-year individual file also includes preliminary weights for all waves. We have added preliminary weights for 1994 through 1997 to the file. The weights for 1994 and 1995 include separate core, Latino and combined variables, very similar to those for 1990 through 1993. The Latino sample was dropped for 1996, so we only include a core weight. In 1997, a new immigrant sample was added, but the weights were scaled so that only one variable is needed for the core sample alone, the new immigrant sample alone, or the two combined.