

PANEL STUDY OF INCOME DYNAMICS  
2001 PUBLIC RELEASE FAMILY FILE  
DOCUMENTATION

TABLE OF CONTENTS

I. INTRODUCTION

A. WHAT'S NEW FOR 2001

1. QUESTIONNAIRE CHANGES
2. NEW GENERATED VARIABLES
3. OCCUPATION, INDUSTRY AND OTHER CODES

II. DATA CHARACTERISTICS

A. FILES AND FORMAT

B. VARIABLE NAMES, POSITIONS AND GENERATED VARIABLES FOR THE FAMILY FILE

C. DOCUMENTATION AND CODES

1. 2001 QUESTIONNAIRE
2. CODE CATEGORIES

D. PROBLEM VARIABLES, MISSING VARIABLES

E. NOTES ON CODED VARIABLES IN 2001

F. NOTES ON GENERATED VARIABLES IN 2001

**Release 4 Notes:** February 2014: Contingency corrections for a handful of cases for the family level T-2 variables ER20167-ER20176, ER20183-ER20188, ER20191-ER20193 after further cleaning of corresponding individual level T-2 variables. Added a new variable that was recently coded ER20081A T7G MAIN PURPOSE/CAUSE OF ORGANIZATION from the Philanthropy section. In addition, we also added another new variable Collapsed Beale (ER20377C).

**Release 5 Notes:** January 2016: Corrected the bottom code for H14/H38 (ER19709/ER19817) and H17/H41 (ER19712/ER19820) for both Head and Wife/"Wife" where 1 means one cigarette or fewer per day.

**Release 6 Notes:** November 2017: Added 25 generated food security variables (ER18470A-ER18470Y). The food security variables were put through the same imputation and calculation process as was completed on the 2015 data to produce food security scale score and category. The 33 aggregated consumption variables were added to the family release file (ER20456A1-ER20456D6). These variables were previously released on a supplemental file.

**Release 7 Notes:** June 2023: We have added 26 variables to the consumption series. 25 variables are accuracy codes and one is a generated variable TOTAL EXPENDITURE 2001 (ER20456D7). Users should also note that the range for HOME INSURANCE EXPENDITURE (ER20456A9) was revised.

## **I. INTRODUCTION**

### **A. What's New for 2001**

#### **1. 2001 Questionnaire Changes**

Most of the sets of questions added for 1999 continue to be part of the 2001 instrument, but a few sequences and sections were dropped, while some were added.

Section M, containing the questions asked of post-1968 immigrants regarding their background, employment, and family connections, was dropped since no additional new sample was being drawn for 2001. For the same reason, most of the eligibility questions were dropped. A few of these questions, useful for weighting, were retained.

At the end of Section H, the questions on nutrition and healthy eating were dropped, and a 30-day depression scale (K-6) was added. The depression scale was asked of the respondent, but only if he or she was Head or Wife/"Wife". The K-6 scale is the same one that is used in the National Health Interview Survey.

The expenditure questions at the end of the wealth sequence (section W) were dropped.

Employer size was dropped from the Head and Wife/"Wife" pension sequence (Section P).

Unfolding brackets are now included for both income tax itemization questions in Section G to provide more information when the answer is "don't know". The bracket detail for charitable contributions is in ER19163-ER19166; itemized medical costs are probed in ER19168-ER19171.

The one major section addition for 2001 was a question series on philanthropic giving (Section T) developed by the Center for Philanthropy at Indiana University-Purdue University. This series includes questions about charitable contributions, including time spent by Heads and Wives/"Wives" doing volunteer work for nonprofit organizations.

A new core/immigrant cross-sectional weight (ER20459) was constructed for 2001 family data. See Part 2., Section b., below, for references for weights.

## **2. Generated Variables**

### **a. Family Composition variables**

Family Composition Change, a key variable for identifying change in family composition since the prior wave, is available this year.

Another important variable for identifying changes in family composition, Splitoff Indicator, is also available. It provides useful information about who is being interviewed or what changes occurred in the family. It is important to note that the Splitoff Indicator indicates a splitoff family only in the first year such a family forms; thereafter these families receive code values designating reinterviews.

See Section II, Part F below for codes for these two variables.

Please note that our Human Subjects Review Board, in response to new Federal policy, now prohibits us from interviewing anyone who is incarcerated. This means that the study has a 100% non-response rate for single-person family units in 1999 who were incarcerated during the 2001 interviewing period.

### **b. Family level income/employment variables**

Family income variables are now provided this file. Similarly, work hour, wage rate, and wealth composite variables are now also provided in this family.

### **c. Two types of weights are available as part of the 2001 family file:**

(1) a recalculated core/immigrant family longitudinal weight (for details, go to our internet site PSIDONLINE.ISR.UMICH.EDU: select 'Data & Documentation', then select 'Sample Weights', then 'PSID Longitudinal Weights', and 'PSID Revised Longitudinal Weights 1993-2005'); and

(2) a new core/immigrant family cross-sectional weight (for details about how this weight was constructed, again, go to our internet site, select 'Data & Documentation', 'Sample Weights', then 'PSID Cross-sectional Weights', and 'Cross-sectional Analysis of PSID Data and Cross-sectional Weights 1997-2005').

### **d. Aggregated consumption variables**

In Release 6, we have added the 33 aggregated consumption variables to the Family file. For the 2001 wave, the variables are ER20456A1-ER20456D6. These were previously included in the Consumption supplemental file, but they have now been added to the Family file for all waves in which they are generated.

### **e. Food Security variables**

In Release 6, we have generated the 25 food security variables. For the 2001 wave, the variables are ER18470A-ER18470Y. The food security variables were put through the same imputation and calculation process as was completed on the 2015 data to produce food security scale score and category.

## **3. Occupation, Industry and Other Codes**

As in prior waves, all occupation and industry codes are available on the file. This includes the three-digit codes for Head and Wife/"Wife" on main and extra jobs, one-digit reason for leaving a job, and the two-digit industry codes for family-owned businesses, and the field of endeavor for non-academic degrees and certificates.

## **II. DATA CHARACTERISTICS**

### **A. Files and Format**

The 2001 family data file consists of one data file with 3,533 variables and 7,406 records. The family file contains one record for each family interviewed in 2001 and includes all family-level variables collected in 2001.

In addition to the data file and this document, we include SAS, SPSS and STATA data definition statements and SAS missing data statements.

Using the web-based Data Center is the most efficient way to obtain the data. At this time, we are also still providing the data in a .zip package. The .zip data are in raw ASCII form, and the file is named FAM2001ER.TXT. Refer to the corresponding .SAS, .SPS, or .DO file for record layout information, variable names and variable labels. If you download the data in ASCII form from the PSID Website, you will receive SAS, SPSS, or STATA statements, as you request.

These files 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 statements needed to complete a program. Missing data statements are not provided as part of these data definition files. We have generated separate missing data statements that may be added to the files, however we leave it to the individual analyst to check the questionnaire and frequencies or means, and codebook for each variable you intend to include in your analysis to determine whether you concur with these definitions.

### **B. Variable Names, Positions and Generated Variables**

The 2001 family variables are in the range ER17001-ER20459.

The 2001 data file provides many component variables that analysts may use to construct summary variables. We leave it to the individual analyst to make decisions about imputation methods to use, if any, and how to treat missing data. Analysts who wish to construct the summary variables that are parallel to those provided in historical waves of the PSID may consult codebooks for those years to obtain information about rules for their creation. Some examples of variables that may be constructed using component variables provided in this data file include mortgage and rent payments, and food costs.

Background items such as education are collected 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. Beginning with the 1999 data file, we have done this "bringing forward" from the 1997 and 1999 data. All items asked afresh for all Heads and Wives/"Wives" in 1997 have been copied to equivalent variables for 1999 and 2001 if the Head or Wife/"Wife" was the same person. So, for example, race is available for all Heads and Wives/"Wives" with no further effort by the analyst. We have included the most recent report for years of education of all Heads and Wives/"Wives" in a separate data file in the Data Center ("Other Family Data").

### **C. Documentation and Codes**

## **1. 2001 Questionnaire**

Customized codebooks are available in the Data Center for the 2001 family variables, and most other variables.

The PSID website includes an HTML version of the computer-assisted interviewing application for 2001, at:

[http://www.isr.umich.edu/src/psid/cai\\_doc/2001\\_Interview\\_Year/2001\\_Interview\\_Year.htm](http://www.isr.umich.edu/src/psid/cai_doc/2001_Interview_Year/2001_Interview_Year.htm).

This replaces the PDF-format questionnaires we have provided for the past several years. Use the labels from the SAS and SPSS data definition statements to match variables with questions in the CAI application. There is also a cross year index for 1994-2001 in section III below.

## **2. Code Categories**

Refer to the codebook in the Data Center for code listings and descriptions. 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.

See Part F for additional information on generated variables.

## **D. Problem Variables and Missing Variables**

Values for variables referring to fourth extra job for working heads (ER17477- ER17502) and for third and fourth extra jobs for nonworking heads (ER17733- ER17784) contain only zeroes. Similarly, no fourth extra job for working for working wives/"wives" (ER18048-ER18073) and no second through fourth extra jobs (ER18278-ER18355) for nonworking wives/"wives" were reported.

## **E. Notes on Coded Variables in 2001**

The 2001 family file includes quite a number of variables that are not lifted directly from the questionnaire.

For example, our traditional hand coded items--all occupations and industries, reasons for leaving employers, and fields of endeavor for non-academic certificates received--are included on the family file. Please see the codebook for codes.

Values for the ethnicity questions, asked of all Heads and Wives/"Wives" in 1997 and of new Heads and Wives/"Wives" since 1999, have been coded and are included. (The 1997 values were brought forward for same Heads and Wives/"Wives" in 1999, and likewise for 1999 and 2001, along with other background information, so all 2001 Heads and Wives/"Wives" have information here.)

One-digit ethnicity (ER19901 for Wife/"Wife"; ER19993 for Head):

1. "American"
2. Hyphenated American (e.g., African-American, Mexican-American)
3. National origin (e.g., French, German, Dutch, Iranian, Scots-Irish)
4. Nonspecific Hispanic identity (e.g., Chicano, Latino)
5. Racial (e.g., white or Caucasian, black)

6. Religious (e.g., Jewish, Roman Catholic, Baptist)
7. Other (including responses such as "a mutt", "Heinz 57", "good old boy", "hillbilly")
9. NA; DK

Two-digit ethnicity (ER19902-ER19903 for Wife/"Wife"; ER19994-ER19995 for Head):

If a national origin was given for the one-digit variable (code 2 or 3), then up to two mentions are coded here.

1. American (meaning U.S.)
2. American Indian, Eskimo, Aleut
3. British: English, Scottish, Irish, Welsh
4. Western European: French, German, Italian, Portuguese, Spanish, Swiss, Netherlander/Dutch, Belgian, Luxembourger, Liechtensteiner, Sicilian, Corsican, Sardinian, Maltese
5. Eastern European: Romanian, Polish, Russian and former Soviet republics, Greek, Bulgarian, Yugoslavian, Bosnian, Herzegovinian, Croatian, Slovenian, Slovakian, Czech, Albanian
6. Northern European/Scandinavian: Norwegian, Swedish, Finnish, Dane
7. Middle Eastern: Iranian, Turkish, Iraqi, Syrian, Lebanese, Egyptian, Afghani, Pakistani, Omani, Yemenite, Jordanian, Georgian, Armenian, Azerbaijani, Uzbekistani, Tajikistani, Turkmenistani, Saudi Arabian, Israeli
8. East Asian: Chinese, Japanese, Korean
9. South or Southeast Asian: Indian, Vietnamese, Cambodian, Laotian, Thai
10. Pacific Islander: Filipino, Indonesian
11. Canadian
12. Central American: Nicaraguan, Mexican, etc.
13. Caribbean: Cuban, Haitian, etc.
14. South American: Peruvian, Chilean, etc.
15. African
16. Oceania: Australian, New Zealander, New Guinean
97. Other
99. NA, DK
0. INAP.: no second mention; one-digit ethnicity=1,4-7,9

Codes are available in the Data Center for most variables. The exception is vehicle manufacturer, make, and type codes.

Details were asked for as many as three vehicles. Codes for manufacturer and make (ER17112-ER17113, ER17142-ER17143, and ER17172- ER17173)-the first two digits are the values for the manufacturer variable, and the second pair identify the make within manufacturer:

- 10 01. AMC
- 10 02. AMC Eagle (1980-1988)
- 10 03. AMC Jeep (1970-1988)
- 10 05. AMC Renault
- 10 06. AMC Studebaker
- 10 98. Other AMC; AMC, DK make
- 11 01. Chrysler
- 11 02. Chrysler Dodge
- 11 03. Chrysler Eagle (1987-1998)
- 11 04. Chrysler Jeep (1989 and later)
- 11 05. Chrysler Plymouth
- 11 98. Chrysler, DK make

12 01.	Ford
12 02.	Ford Lincoln
12 03.	Ford Mercury
12 04.	Ford Jaguar
12 98.	Ford, DK make
13 01.	GM Buick
13 02.	GM Cadillac
13 03.	GM Chevrolet
13 04.	GM GMC
13 05.	GM Oldsmobile
13 06.	GM Pontiac
13 07.	GM Saturn
13 08.	GM Saab
13 98.	GM, DK make
19 98.	Other domestic manufacturer (Delorean, Hudson, Nash, etc.); domestic, DK manufacturer
21 98.	Motor home (Winnebago, Coachman, etc.)
30 01.	Honda
30 02.	Honda Acura
30 98.	Honda, DK make
31 01.	Nissan
31 02.	Nissan Infiniti
31 03.	Nissan Datsun
31 98.	Nissan, DK make
32 01.	Toyota
32 02.	Toyota Lexus
32 98.	Toyota, DK make
33 01.	Hyundai
34 01.	Isuzu
35 01.	Kia
36 01.	Mazda
37 01.	Mitsubishi
38 01.	Subaru
39 01.	Suzuki
49 05.	Daihatsu
49 06.	Daewoo
49 98.	Other Asian manufacturer; Asian, DK manufacturer
80 01.	BMW
81 01.	Mercedes-Benz
82 01.	Porsche
83 01.	Volkswagen
83 02.	Volkswagen Audi
83 98.	Volkswagen, DK make
84 01.	Volvo
97 35.	Fiat
97 46.	Land Rover
97 60.	Peugeot
97 98.	Other European manufacturer; European, DK manufacturer
98 98.	DK manufacturer and make
99 99.	NA manufacturer and make
00 00.	Inap.: no first vehicle; no second vehicle; no third vehicle





Codes for type of vehicle (ER17115, ER117145, and ER17175):

1. Car
2. Pickup or truck
3. Van
4. Utility vehicle
5. Motor home
8. DK
9. NA
0. Inap.: no first vehicle; no second vehicle; no third vehicle

## **F. Generated Variables**

For 2001, a number of generated variables are included in the family file.

1. One such group is location data. We include ER17004 and ER17005, PSID/GSA and FIPS state codes. We also include the PSID's traditional variables for region, Beale rural-urban code, and size of the largest city in the county (ER20376- ER20378, respectively). The codes for FIPS and PSID/GSA codes are included in Appendix 1 of the 1985 documentation, available in .pdf format on our website.
2. We also offer a group of marital status variables, including: marital status of Head (ER17024), the generated form of marital status comparable with years prior to 1977 (ER20369), change in marital status of Head between waves (ER20370), and couple status of Head (ER20371)
3. The variable indicating whether a PSID family is living in institutional housing is present (ER17020, equivalent to 1993 V21609), as is the variable indicating the total number of data records from the cross-year individual file that are associated with a panel family (ER20368, equivalent to 1993 V23321).
4. Both USDA and Census needs standards have been generated for the prior calendar year, 2000 (ER20372 and ER20373; see 1993 V23325 and V23326). Additionally, since the PSID has switched to biennial interviewing, comparable needs standards have also been generated for the "off" year, 1999 (ER20374 and ER20375).
5. Two traditional PSID family-level generated variables concerning splitoffs are: for a main family, the number of splitoffs arising from it; for a splitoff family, the family interview number of the main family from which it split (ER20379 and ER20380, the same as 1993 variables V23342 and V23343).
6. The PSID documentation for 1993 and earlier waves has information about multiple PSID families sharing the same household (See Section I, Part 5 of the front matter). The PSID produces sets of variables about these sharing families: family ID numbers, relationships, and sizes of up to four other PSID families sharing the HU (ER20381- ER20392, comparable to 1993 V23344-V23355), the household ID number (ER20393, the 2001 version of 1993 V23356), and the number of persons not included in any PSID family who are sharing the household (ER17018, comparable to 1993 V22411).
7. Finally, Family Composition Change and Splitoff Indicator are included.

Please note that PSID following rules changed somewhat in 1994 with the advent of followability of non-sample parents and of sample children under age 18. And beginning in 1994, children who did not appear in a response family at birth but who were born to sample members during the study period were designated as followable sample members. These two types of changes necessitated a slight rewording of the Family Composition Change codes, so we give the revised versions here (as well as in the codebook):

Family Composition Change, ER17007: (Codes V2 through 8 take priority over codes 0 and 1):

0. No change; no movers-in or movers-out of the family.
1. Change in members other than Head or Wife/"Wife" only.
2. Head is the same person as in 1999 but Wife/"Wife" left or died; Head has new Wife/"Wife"; used also when cohabiting, nonrelative female becomes "Wife."
3. Wife/"Wife" from 1999 is now Head.
4. 1999 female Head got married--husband (usually a nonsample member) is now Head. Used also when cohabiting, nonrelative male becomes Head.
5. Some followable individual other than 1999 Head or Wife/"Wife" has become Head of this FU. (Used primarily for male and unmarried female splitoffs.)
6. Some followable female other than 1999 Head got married and her husband is now Head. (Used primarily for married female splitoffs.)
7. Female Head in 1999 with husband in institution--husband in FU in 2001 and is now Head.
8. Other (used for recontacts and recombined families--these latter are usually Heads and Wives who have parted for a wave or more, been interviewed separately, and who have reconciled at some time since the 1999 interview but prior to the 2001 interview).
9. Neither Head nor Wife/"Wife" (if there is one) is followable and neither of them was a Head or Wife/"Wife" last year. (Used primarily for underage splitoff children.)

Codes for Splitoff Indicator have varied from wave to wave beginning in 1990 because of various sample additions and recontact scenarios. See the codebooks for details.