

March 2021

PANEL STUDY OF INCOME DYNAMICS
2007 PUBLIC RELEASE FAMILY FILE DOCUMENTATION

TABLE OF CONTENTS

I. INTRODUCTION -- WHAT'S NEW FOR 2007

- A. QUESTIONNAIRE CHANGES
- B. NEW GENERATED VARIABLES

II. DATA CHARACTERISTICS

- A. FILES AND FORMAT
- B. VARIABLE NAMES, POSITIONS AND GENERATED VARIABLES
- C. HAND CODED AND GENERATED VARIABLES
- D. DOCUMENTATION AND CODES
 - 1. 2007 Questionnaire
 - 2. Codebook and Code Categories

RELEASE 2 NOTES

October 2009: The imputed work hours and income variables (ER40873- ER41026) which were not completed for Release 1 are now available in Release 2.

RELEASES 3 NOTES

January 2012: We have reconfigured variable ER36106 - A50 Why did you move? We have recoded cases originally coded as 'Other/ambiguous' into existing categories where possible.

RELEASES 4 NOTES

December 2013: Correcting the following variable; Births to OFUMS only - 2005 (ER41067) and P20b Whether amount 10,000 (ER37762) In addition, we are suppressing two previously publicly released variables Beale Rural-Urban Code (ER41033) and Size of Largest City in County (ER41034).

RELEASE 5 NOTES

February 2014: Added one new coded variable in the philanthropy section M12code (ER40681A) and Collapsed Beale (ER41033A). Additionally, we have corrected extant T-2 variables for a handful of cases.

RELEASE 6 NOTES

January 2016: For Release 6 we have added 96 new variables in the T-2 (2005) Income section, 86 of which are about time unemployed for Head (ER40686A1-ER40686B5) and Wife/"Wife" (ER40686F4-ER40686G8), time out of the labor force for Head (ER40686B6-ER40686D1) and Wife/"Wife" (ER40686G9-ER40686J4), and time employed for Head (ER40686D2-ER40686E7) and Wife/"Wife" (ER40686J5-ER40686L1). The other 10 new variables are T-2 hours worked, labor income, and accuracy checks for Head (ER40686E8-ER40686F3) and for Wife/"Wife" (ER40686L2-ER40686L6).

RELEASE 7 NOTES

November 2017: For Release 7 we have added the 39 aggregated consumption variables to the family release file (ER41027A1-ER41027E3). These variables were previously released on a supplemental file. A correction was made to the imputation of missing data for vehicle gasoline costs, additional vehicle costs, and vehicle repair costs.

RELEASE 8 NOTES

June 2023: For Release 8, we have added 34 variables. 31 variables are accuracy codes and one is a generated variable TOTAL EXPENDITURE (ER41027E4) for consumption. 2 variables were added to chronic conditions. Users should also note that the range for HOME INSURANCE EXPENDITURE 2007 (ER41027A9) was revised.

I. INTRODUCTION – WHAT'S NEW FOR 2007

A. Questionnaire Changes

Most sets of questions from the 2005 wave continue to be part of the 2007 instrument, but a few sequences were dropped, some were added and others revised.

In Section A, there were three new questions. A25a1 indicates whether loan amounts had a fixed or variable interest rate. Questions A42_1 and A42a indicate if gas and electricity are paid on one bill and, if so, how much is spent on gas and electricity combined.

In the employment sections, BC for Head and DE for Wife/"Wife", questions were added about number of weeks actually worked on each of their jobs last year and the year before that (BC42a and BC47a for Head and DE42a and DE47a for Wife/"Wife").

In Section F, a question was added about which months the family was paying for child care in the previous year (F6d2). Question wording regarding stamp receipt (F16) changed from amount received last month to total amount received this year. New codes were added at F52 to capture some of the 'other specify' responses at this question.

Section W has new questions at the end (W129-W144) about intergenerational help.

The health section, Section H, underwent some major modifications; the most significant was the addition of the childhood health calendar (H32-H45). For each condition and type of report about Head or Wife/"Wife" we have yes/no indicators for each condition at each age. There are a total of six sets of Childhood Health Calendar reports. The series was asked about Head and Wife/"Wife" and depending on who the respondent was, the information is stored in the appropriate set.

- (1) Self Report about Head (ER38344-ER38661)
- (2) Spousal Report about Head (ER38662-ER38979)
- (3) Other Report about Head (ER38980-ER39297)
- (4) Self Report about Wife/"Wife" (ER39441-ER39758)

- (5) Spousal Report about Wife/"Wife" (ER39759-ER40076)
- (6) Other Report about Wife/"Wife" (ER40077-ER40394)

Since 2007 is the first wave of data collection for the childhood health calendar, only one report out of set (1-3) and one report out of set (4-6) may have data if any of the childhood health conditions were endorsed. These reports will be carried forward to future waves if Head or Wife/"Wife" remain the same. In subsequent waves, we may gather new types of reports (self report and spousal report, for example) about the same individual if the respondent changes. To help understand which reports are from which wave, we added variables at the start of each report to indicate which year this report originated. For set one for example, (1) Self Report about Head the year of report is stored at (ER38344).

Question H1d, about missing school for health reasons, is now in the childhood health sequence (H26). The depression scale that was used in 2003 was reinstated (H59-H59I).

The philanthropic giving section (Section M) is much shorter as several questions from 2005 were dropped.

Section R, which asks about income for the "off" year, or t-2, has some revisions. Separate questions for Head and Wife/"Wife" assets (R4-5 in 2005) are gone and a family total for assets is now at R16-19. New questions (R63-67) were added to capture any other t-2 income.

Section S, Katrina, is an entirely new series about health-related issues for respondents who lived in the Hurricane Katrina area in 2005. It is not included on the public use data file but will be available as a separate file under confidential contract.

B. New Generated Variables

Background information (Section K for Wives/"Wives" and Section L for Heads) is only asked when a person newly acquires a relationship of Head or Wife/"Wife", values are copied forward from prior data collections for those Heads and Wives/"Wives" who maintain their relationship classification from the preceding wave.

A new variable to indicate whether this family has a record on the restricted Katrina data file was added (ER41068).

In Release 7, we have added the 39 aggregated consumption variables to the Family file. For the 2007 wave, the variables are ER41027A1-ER41027E3. 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.

II. DATA CHARACTERISTICS

A. Files and Format

The 2007 family data file consists of one data file with 8,289 records and 5,206 variables. The family file contains one record for each family interviewed in 2007 and includes all family-level variables collected in 2007, including extensive information about Head and Wife/"Wife".

Using the web-based Data Center is the most efficient way to obtain the data. At this time, we also provide the data in a .zip package. If you download the data in ASCII form from the PSID website, you will receive SAS, SPSS or STATA statements, as you request. In the .zip package, we include the entire data file in ASCII format and SAS, SPSS and STATA data definition statements. These statements provide variable names, locations, and variable labels.

The definition statement 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 the files.

B. Variable Names and Positions

The 2007 family variable names are in the range ER36001-ER41069.

The 2007 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 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 annualized mortgage and rent payments and food costs.

C. Hand Coded and Generated Variables

The PSID hand codes several data items for the family file. Head's and Wife's/"Wife's" occupations and industries, to a maximum of four jobs apiece, are coded for 2007 using the three-digit codes from the 2000 Census Occupation and Industry Code. The information for each job also includes a one-digit reason for job termination where applicable. Family-owned businesses are coded using a two-digit industry code, and the field of endeavor for Head's and Wife's/"Wife's" non-academic degrees and certificates is coded for as many as three each. Any education received outside the U.S. is coded for Head, Wife/"Wife" and their parents.

Background items such as education are collected only for new Heads and new Wives/"Wives" in a given wave. 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 wave in which the section was most recently asked is indicated by ER40142 for Heads and ER40143 for Wives/"Wives".

Completed education of Head and Wife/"Wife" (ER41037 and ER41038) are generated from the background information.

Family composition and change variables include Family Composition Change (ER36007), Splitoff Indicator (ER36005), Head-Spouse Sample Status (ER41044), and variables about births to Head, spouse and other family members during the prior calendar year, 2006 (ER41060-ER41063), and during the "off" year, 2005 (ER41064-ER41067). Note that the Splitoff Indicator is only assigned to a family in the year they first move out of the main family and form their own separate household; after that one wave of being a splitoff, these families receive code values that designate reinterviews.

Two other variables concerning splitoffs are the number of splitoffs arising from a main family (ER41045) and the family interview number of the main family from which a splitoff family originated (ER41046).

The PSID produces sets of variables about families sharing the same household: family ID numbers, relationships, and sizes of up to four other PSID families sharing the HU (ER41047-ER41058), the household ID number (ER41059), and the number of persons not included in any PSID family who are sharing the household unit (ER36022). The PSID documentation for 1993 and earlier waves has additional information about multiple PSID families sharing the same household (see "Linking Data" in Section I, Part 5 of the 1993 Guide).

The marital status variables group consists of Head's current marital status (ER36023), the generated form of marital status comparable with years prior to 1977 (ER41039), change in marital status of Head between waves (ER41040), and couple status of Head (ER41041).

Location variables include PSID/GSA and FIPS state codes (ER36003 and ER36004); Current Region, Beale Rural-Urban code, and Size of the Largest City in the County (ER41032-ER41034); and two derived from background information, Region where Head Grew Up (ER41035) and Head's Geographic Mobility (ER41036). The codes for the FIPS and PSID/GSA codes are included in Appendix 1 of the 1985 documentation, available in .pdf format on our website.

Both USDA and Census needs standards were generated for the prior calendar year, 2006 (ER41028 and ER41029). Additionally, since the PSID has switched to biennial interviewing, comparable needs standards were also been generated for the "off" year, 2005 (ER41030 and ER41031).

The variable indicating whether a PSID family lives in institutional housing is present (ER36008), as is the variable indicating the total number of data records from the cross-year individual file that are associated with a panel family (ER40872).

Weights are available as part of the 2007 Family File in ER41069.

D. Documentation and Codes

1. 2007 Questionnaire

Our website includes an HTML depiction of the computer-assisted interviewing (CAI) applications for 1993-2001. Box-and-arrow questionnaires are available for all waves in PDF format. Use the SAS, SPSS and STATA data definition statements to match variables with questions in the CAI application.

2. Codebook and Code Categories

The Data Center allows users to create and download customized codebooks that provide variable-specific documentation; variable descriptions are complete for all of the 2007 family data items. The downloadable .zip package of the family data on our website contains a .pdf version of the entire family codebook for 2007.

In general, codes follow our traditional schema, 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 codebook nor one of the "zero", "eight" or "nine" codes just mentioned, assume missing data for that value.

RELEASE 8 NOTES

March 2021:

A)

For Release 8 we have added 32 new consumption variables to the family release file. They are found in the variable range (ER41027A2A-ER41027E4). Thirty-one of these variables are accuracy codes which indicate if the consumption value was imputed (1) or not (0). The thirty second newly released consumption variable is an aggregate that indicates the total consumption spending (ER41027E4) for the entire family unit.

B)

Two new health section variables were added ER38263A1 and ER39360A1. These are chronic health conditions that have been coded from the respondent's response at the string variable H6M1. Previously this code value had been released with the 2011 to 2019 waves. We are now adding the H6M1 two-digit code to the 2007 and 2009 waves. The first variable is the Head's reported other chronic health condition, the second is the Wife's/"Wife's". As part of the coding process some Other conditions are found to actually belong in one of the previous question series from H5A ("had a stroke?" ER38209 for Head) to H5L ("Had any emotional, nervous, or psychiatric problems?" ER38256 for Head). In these cases, the respondent's chronic health condition mention was moved to the appropriate variable in the H5A to H5L series. In some cases, the respondent mentions both an Other chronic condition and a condition from the H5A to H5L series within their text response at H6M1. For these types of scenarios both health conditions are coded.

C)

Mention Mushtaq's GEO code fixes.

???