PANEL STUDY OF INCOME DYNAMICS
2005 PUBLIC RELEASE FAMILY FILE DOCUMENTATION
RELEASE 4

_____

TABLE OF CONTENTS

RELEASE 3 NOTES

November 2013: Correcting the following four variables: Births to OFUMS in 2003 (ER28077), Number of businesses owned by FU in 2004 (ER27907), Number of businesses owned by Head (ER27912), and Number of businesses owned by Wife (ER27942). Due to further hand look-ups after Release 2 was made, there are 198 cases that have different values for Head and Wife Taxable Income - 2004 (ER27953) and Total Family Money Income - 2004 (ER28037). In addition, we are suppressing previously released variables Beale Rural-Urban Code (ER28043), and Size of Largest City the county (ER28044).

RELEASE 4 NOTES

February 2014: Added four new coded variables in the philanthropy section: M12code, M41code, M43code and M52code (ER27510A, ER27634A, ER27644A, ER27667A) and Collapsed Beale (ER28043A). In addition we have dropped fifteen previously released variables from the T-2 (Section R) range (ER27712-ER27726). These were originally released as family level variables for Head and Wife/"Wife" but since they also contain data for OFUMs we have moved them to the Cross Year Individual File instead.

RELEASE 5 NOTES

January 2016: For Release 5 we have added 96 new variables in the T-2 (2003) Income section, 86 of which are about time unemployed for Head (ER27711A1-ER27711B5) and Wife/"Wife" (ER27711F4-ER27711G8), time out of the labor force for Head (ER27711B6-ER27711D1) and Wife/"Wife" (ER27711G9-ER27711J4), and time employed for Head (ER27711D2-ER27711E7) and Wife/"Wife" (ER27711J5-ER27711L1). The other 10 new variables are T-2 hours worked, labor income, and accuracy checks for Head (ER27711E8-ER27711F3) and for Wife/"Wife" (ER27711L2-ER27711L6).

I. INTRODUCTION

A. Public Release I Files vs. Public Release II Files

Over the past several waves, PSID data, collected using Computer Assisted Telephone Interview (CATI) technology, have been suitable for a timely, single Public Release of data files. These data are of high quality and appropriate for research analysis. The recent waves of data, labeled Public Release I, differ from the historical, pre-CATI data labeled Public Release II in only two ways:

* Public Release I longitudinal data are subject to only minor revision as added waves of data are collected on individuals and families (as the result of family composition and economic editing), and

* Certain variables that were included in the historical, "pre-desktop" era Public Release II data are now readily computable by the user and so are no longer included.

The PSID data files, as with the data files from any complex longitudinal study, are subject to minor changes and subsequent updated releases, as new information is gathered in later waves in the process of economic and family composition editing. If such information necessitates some revisions in previously released data files, the revisions will be carried out shortly after or in conjunction with the release of the new wave data. For example, when the 2007 Public Release files are posted in 2008, we may also post minor updates to the 2007 Public Release files that have been made in light of the newest wave's data.

When the minor updates are posted, PSID staff will retain only the most recent data files for distribution. Therefore, we strongly urge users to retain and save all data files that they download from the PSID website and upon which individual research analysis is dependent. One exception to our general policy of posting only minor updates concerns the data files produced in the initial CATI-transition years, 1994-1996. Re-releases are planned soon.

The latest in this series of data files is the Public Release version of the 2005 wave of data, which includes the 2005 family file and the 1968-2005 cross-year individual file, available at the PSID's internet site.

B. What's New for 2005

1. 2005 Questionnaire Changes

Most of the sets of questions added in the 1999 wave continue to be part of the 2005 instrument, but a few sequences and sections were dropped, some were added and others revised. Questions about services for senior housing (2003 questions A8-A14) were removed, as were most of the questions about computer use in the home (2003 questions A47c-A47i), leaving only A47a and A47b. Questions A20a-e are new unfolding brackets that attempt to capture a ballpark amount for homeowners' house values (A20) if the initial response is "Don't know".

In the employment sections, BC for Head and DE for Wife/"Wife", the question from 2001 and earlier waves (2005 BC3 and DE3) about whether he or she is doing any work for money now, was reinstated. A question about the number of fellow employees working for Head's or Wife's/"Wife's" present employer was resurrected from 1999, with unfolding brackets for missing responses (2005 BC25a-f and DE25a-f). A brand-new question asks whether the current job requires the use of a computer (BC28a and DE28a).

In Section F, questions about free meals for family members in adult day care (2003 F7e-f) were dropped, as were the food stress measures (2003 F26-F46). Also omitted was 2003 F59, whether a vehicle was used for a disabled person's transportation. A module asking about consumption amounts was added in questions F87-F91c.

The health section, Section H, underwent quite a few changes. For health conditions, the duration questions (number of days, weeks, months and years) for each were replaced with age at first diagnosis and whether there was a second occurrence, or a worsening if the condition was chronic. The IADL sequence now includes questions about whether special equipment is needed to perform the activity. Questions from 2003 about whether Head or Wife/"Wife" felt rushed (H12b/H36b), as well as the items for the depression scale (H59a-n) were dropped. Heavy and light physical activity questions (2003: H11n-H12/H35n-H36; 2005: H12a-b/H36a-b) were reworded, and a new question about strength activities (H12c and H36c) was added. Drinking questions were updated to include reports on the frequency of alcohol consumption within the past year and the number of drinks consumed per occasion 2005: (H21a-c/H45a-c).

Sections K and L, about Wife's/"Wife's" and Head's backgrounds, respectively, included the revived question about Latino identity, which had been dropped in 1997.

The philanthropic giving section (Section M) includes new questions (M52b-M52l) about the December 2004 tsunami in Asia. Question M54, asking about help received from family and friends, is also new.

Section R, which asks about income for the "off" year, or t-2, has been revised to collect much more income detail, including asset income and a longer list of transfers. Other than income amounts, the questions about welfare receipt and program participation have been eliminated. A few of the questions, e.g., food stamp values, were moved to Sections F and M.

2. New Generated Variables

We now include full family income detail, with variables containing imputed amounts for all of the components of family income. Since 1994, such detail has been confined only to labor income components for Head and Wife/"Wife". Asset and transfer components for Head, Wife/"Wife" and Other FU Members are now present, along with accuracy indicators for each. Wealth composite variables remain in a separate data file within the Data Center.

Background information (Sections K and L, for Wives/"Wives" and Heads, respectively) is only asked when a person newly acquires a relationship of Head or Wife/"Wife" with the exception of the race [ER27393-ER27396 for Head, ER27297-ER27300 for Wife/"Wife"] and hispanicity questions [ER27392 for Head, ER27296 for Wife/"Wife"] which were asked of all Heads and Wives/"Wive"'s in 2005. For all other KL background variables values are copied forward from prior data collections for those Heads and Wives/"Wives" who maintain their relationship classification from the preceding wave. Added for 2005 are hand-coded values for parents' and own foreign education levels (questions K9, K19 and K54 [ER27271, ER27281, and ER27326] for Wife/"Wife" and questions L9, L19 and L61 [ER27360, ER27370 and ER27422] for Head).

II. DATA CHARACTERISTICS

A. Files and Format

The 2005 family data file consists of one data file with 8,002 records and 3,164 variables. The family file contains one record for each family interviewed in 2005 and includes all family-level variables collected in 2005, 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 2005 family variable names are in the range ER25001-ER28078.

The 2005 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 2005 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. As mentioned in section I, part A, item 2, any education received outside the U.S. is coded for Head, Wife/"Wife" and their parents.

Background items such as education and race 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 ER28052 for Heads and ER28053 for Wives/"Wives".

Completed education of Head and Wife/"Wife" (ER28047 and ER28048) are generated from the background information.

Family composition and change variables include Family Composition Change (ER25007), Splitoff Indicator (ER25005), Head-Spouse Sample Status (ER28054), and variables about births to Head, spouse and other family members during the prior calendar year, 2004 (ER28070-ER28073), and during the "off" year, 2003 (ER28074-ER28077). Note that the Splitoff Indicator indicates a splitoff family only in the first year such a family forms; thereafter, these families receive code values that designate reinterviews.

Two other variables concerning splitoffs are the number of splitoffs arising from a main family (ER28055) and the family interview number of the main family from which a splitoff family originated (ER28056).

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 (ER28057-ER28068), the household ID number (ER28069), and the number of persons not included in any PSID family who are sharing the household unit (ER25022). 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 (ER25023), the generated form of marital status comparable with years prior to 1977 (ER28049), change in marital status of Head between waves (ER28050), and couple status of Head (ER28051).

Location variables include PSID/GSA and FIPS state codes (ER25003 and ER25004); Current Region, Beale Rural-Urban code, and Size of the Largest City in the County (ER28042-ER28044); and two derived from background information, Region where Head Grew Up (ER28045) and Head's Geographic Mobility (ER28046). 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, 2004 (2005: ER28038 and ER28039). Additionally, since the PSID has switched to biennial interviewing, comparable needs standards were also been generated for the "off" year, 2003 (2005: ER28040 and ER28041).

The variable indicating whether a PSID family lives in institutional housing is present (ER25008), as is the variable indicating the total number of data records from the cross-year individual file that are associated with a panel family (ER27882).

Weights are available as part of the 2005 family file in ER28078.

D. Documentation and Codes

1. 2005 Questionnaire

Our website includes an HTML depiction of the computer-assisted interviewing (CAI) applications for 1993-2005. Box-and-arrow questionnaires are also available for those 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 2005 family data items. The downloadable .zip package of the family data on our website contains a .pdf version of the entire family codebook for 2005.

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.