

July 2009

**PANEL STUDY OF INCOME DYNAMICS
2007 PUBLIC RELEASE INDIVIDUAL FILE DOCUMENTATION**

TABLE OF CONTENTS

I. INTRODUCTION: WHAT'S NEW FOR 2007

A. 2007 Questionnaire Changes

B. New Generated Variables

C. Other Additions

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

E. MISSING DATA

I. INTRODUCTION: WHAT'S NEW FOR 2007

A. 2007 Questionnaire Changes

The majority of questionnaire changes in 2007 affected the Family data file. Very few questionnaire changes affected the Individual data file. For specific information about questionnaire content, see the Introduction in the 2007 Public Release Family File documentation for details. Two new variables were added in 2007 that describe whether or not an individual was included in the multi-stage Attrition Project that was conducted in 2007: the outcome variable, ER33946, "WTR INCLUDED IN ATTRITOR PROJECT" and ER33947, "ATTRITOR PROJECT SOURCE IN 2007" which indicates from which source the information about this individual was obtained. For more detailed information about the Attrition Project, please see the following PSID technical paper The PSID Sample Leaver Tracking Project #08-03 at http://psidonline.isr.umich.edu/Publications/Papers/tsp/2008-03_PSID_Sample_Leaver_Tracking.pdf

B. New Generated Variables

Two new generated variables have been added to the 1968-2007 Individual data file: 1) ER33926, "WHETHER MEDICARE NUMBER GIVEN". This variable provides a summary for those who were eligible to be asked for a Medicare number; 2) In addition, an adjustment was made to summary variable ER32004, "WTR EVER MOVED OUT OF FU OR DIED" by adding a code of 3 to capture additional deaths that are not included in code 2. Beginning in 1999, we have been recording decedent information that has surfaced

anecdotally. These decedents had records in the restricted mortality data file but they were not being captured on the public file by this summary variable. This adjustment means that all decedents that have information in our restricted data file are now marked using codes 2 or 3.

C. Other Additions

For 2007, children were eligible for either the Child Development Supplement, Wave III (CDS-III) Interview or the Transition to Adulthood (TA-2007) interview if they had been part of the 1997 Child Development Supplement and were part of a PSID family interviewed in 2007. Variables indicating 2007 eligibility (ER33944) and the result of the interview attempt (ER33945) are part of the 1968-2007 Individual data file.

II. DATA CHARACTERISTICS

A. Files and Format

The 1968-2007 Individual data consists of one data file with 69,263 records and 1,397 variables. The data are merged across all waves of the study; that is, they include person information collected from 1968 through 2007. Each person ever in an interviewed family, even for just one wave, has his or her own data record. Consequently, the file contains records for both 2007 response and 2007 non-response individuals.

Using the web-based Data Center is the most efficient way to obtain the data. At this time, we continue to provide the data in a .zip package. If you download the data in ASCII form from the PSID website, you can choose whether to receive SAS, SPSS or STATA statements. 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, Positions and Generated Variables

The 1968-2007 Individual data file consists of yearly items (e.g., Sequence Number, Relationship to Head, Family Identification Number) and a set of summary or cross-year variables with up-to-date values (ER32001-ER32049 and Sex of Individual, ER32000).

Variables ER30001 through ER30866 remain the same as for all previous Public Releases (with the prefix change from "V" to "ER"). Weights for each wave of data are available through 2007 and are located at or near the end of each year's list of variables.

The order of variables is as follows:

- * RELEASE NUMBER, ER30000,
- * 1968 through 1993 Individual data arranged, as usual, by wave, ER30001-ER30866
- * the summary variables, ER31990-ER31999 and ER32001-ER32049,
- * the 1994 Public Release Individual data, ER33101-ER33150,
- * the 1995 Public Release Individual data, ER33201-ER33299B,
- * the 1996 Public Release Individual data, ER33301-ER33347,
- * the 1997 Public Release Individual data, ER33401-ER33438,
- * the 1999 Public Release Individual data, ER33501-ER33547,
- * the 2001 Public Release Individual data, ER33601-ER33639,
- * the 2003 Public Release Individual data, ER33701-ER33742,
- * the 2005 Public Release Individual data, ER33801-ER33848, and
- * the 2007 Public Release Individual data, ER33901-ER33950

C. Hand Coded and Generated Variables

The summary variables fall into three groups. ER31990-ER31999 are used for sampling error and weights calculations; ER32001-ER32008 are true cross-year indicators derived from values in each individual's data record; and ER32009-ER32049 are summaries from the Family History files (these latter files are part of the PSID supplemental file collection and available as .zip downloads from our website).

Other generated items include variables designating the originating family and move-out date for splitoffs for every wave from 1969 to 2007 (2007: ER33939-ER33941) and income summaries from 1968 to 1993.

D. Documentation and Codes

1. 2007 Questionnaire

Our website includes an HTML depiction of the computer-assisted interviewing (CAI) applications for 1993-2007. 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 2007 Individual file data items. The downloadable .zip package of the Individual data on our website contains a .pdf version of the entire codebook for 2007 Individual file.

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.

E. Missing Data

From 1968 through 1993, Heads and Wives/"Wives" have values for the equivalent employment status variables but, for 1994 through 2007, they are zero. Several mentions for employment status are available on the 1994 through 2007 family files.