

May 17, 2007

PANEL STUDY OF INCOME DYNAMICS  
2005 PUBLIC RELEASE I INDIVIDUAL FILE DOCUMENTATION  
Release 2

---

TABLE OF CONTENTS

- I. INTRODUCTION
    - A. PUBLIC RELEASE I FILES VS. PUBLIC RELEASE II FILES
    - B. WHAT'S NEW FOR 2005
      - 1. 2005 Questionnaire Changes
      - 2. New Generated Variables
      - 3. 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. 2005 Questionnaire
      - 2. Codebook and Code Categories
    - E. PROBLEM VARIABLES, MISSING VARIABLES
- 

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.

Most of our usual inter-year consistency checking was performed for the Public Release I 1968-2005 cross-year individual file, so as we have added new waves of data and uncovered errors, the 1968-2005 file as it stands now is cleaner than the 1968-1993 Public Release II file.

## B. What's New for 2005

### 1. 2005 Questionnaire Changes

The changes in questionnaire content only slightly affected the individual data, as almost all substantive items are located on the family file. For information about content, see section II, part B, in the 2005 PR-I family documentation for details. Income for the "off" year (t-2, or 2003 income year) was asked in greater detail in 2005 than in the past. Indicators for those items were added to the cross-year individual file (ER33827-ER33837). The items from 2003 about receipt of meals for the elderly and school lunch and breakfast programs were dropped (2003: ER33730-ER33732).

The newest question for individuals is H61a, asking whether a child's health care coverage is through a state insurance plan (ER33823).

### 2. New Generated Variables

No additional generated variables were added for 2005, but we added values for completed education (ER33115, ER33215, ER33315, ER33415, ER33516, ER33616, ER33716, and ER33817) to Head and Wife/"Wife" records for 1994-2005 for Release 2. We also cleaned and regenerated the 2005 Medicare number permission variable (ER33826). We continue to create up-to-date summary variables (ER32001-ER32049).

### 3. Other Additions

The PSID's following rules were altered in 1990, when a special sample of elderly persons was selected for interview. Then, with the major recontact effort of 1994, we began following nonsample parents even if they were not living with their sample children. Additionally, children who otherwise fit the description of new sample members, i.e., were born to at least one sample member since the sample became part of the study (1968 for core sample and 1990 for the Latino sample), but who had not appeared in a panel family at birth, began to be designated as sample members and hence were to be followed.

For this reason, two variables indicating a person's status with regard to "followability" were added: follow status and reason for following. The follow status variable has been part of the yearly individual data record since 1990, but reasons for following were missing for 1994 and 1995. These have now been added (1994: ER33150; 1995: ER33299B).

An indicator for the philanthropy questions in Section M was included with the individual variables for 2005 and also for 2003, when it was first asked. The item designates the person who made most of the decisions regarding charitable giving. See ER33838 and ER33741 for details.

Another indicator, whether the individual was covered by TANF payments in the prior calendar year, was added for 2005 (ER33812).

Children who had been part of the 1997 Child Development Supplement and were aged 18 or older were selected for a follow-up in 2005. Variables indicating eligibility (ER33844) and the result of the interview attempt (ER33845) are part of the 2005 individual data.

## II. DATA CHARACTERISTICS

### A. Files and Format

The 1968-2005 individual data consist of one data file with 67,271 records and 1,342 variables. The data are merged across all waves of the study; that is, they include person information collected from 1968 through 2005. 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 2005 response and 2005 non-response individuals.

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, Positions and Generated Variables

All variable names for the 1968-2005 Public Release I individual file are prefaced with "ER" rather than "V" to assist both analysts and study staff in determining whether reference is to the Public Release I file or to its Public Release II version.

The cross-year individual 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 older Public Release II versions (with the prefix change from "ER" to "V"). Weights for each wave of

data are available through 2005, and are located at or near the end of each year's list of variables. A few more variables will eventually be added to the 1994 through 2005 portions of the individual data for Public Release II, most notably annual income components and totals for each individual.

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 I individual data, ER33101-ER33150,
- \* the 1995 Public Release I individual data, ER33201-ER33299B,
- \* the 1996 Public Release I individual data, ER33301-ER33347,
- \* the 1997 Public Release I individual data, ER33401-ER33437,
- \* the 1999 Public Release I individual data, ER33501-ER33546,
- \* the 2001 Public Release I individual data, ER33601-ER33637,
- \* the 2003 Public Release I individual data, ER33701-ER33741, and
- \* the 2005 Public Release I individual data, ER33801-ER33848.

#### 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 2005 (2005: ER33839-ER33841) and income summaries from 1968 to 1993.

#### 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 individual data items. The downloadable .zip package of the individual data on our website contains a .pdf version of the entire individual 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.

#### E. Problem Variables, Missing Variables

Please note that the current version of the 2005 Public Release I files have missing data for some variables. This will be corrected for the final version of the file, but in the meantime we want you to be informed of the following known problems with and idiosyncrasies of the data.

From 1968 through 1993, Heads and Wives/"Wives" have values for the equivalent employment status variables, but for 1994 through 2005 they are zero. However, values for Completed Education for Heads and Wives/"Wives", missing for Release 1, have been filled in for Release 2. In the meantime, several mentions for employment status are available on the 1994 through 2005 family files.