

June 2023

**PANEL STUDY OF INCOME DYNAMICS  
2021 PUBLIC RELEASE INDIVIDUAL FILE DOCUMENTATION  
Release 1**

---

**TABLE OF CONTENTS**

**I. INTRODUCTION: WHAT'S NEW FOR THE 2021 WAVE**

- A. 2021 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. 2021 Questionnaire**
    - 2. Codebook and Code Categories**
  - E. MISSING DATA**
- 

**I. INTRODUCTION: WHAT'S NEW FOR THE 2021 WAVE**

**A. 2021 Questionnaire Changes**

The most significant difference between the 2019 and the 2021 wave has been the introduction of the self-administered web mode for part of the sample. Approximately 40% of the sample was completed in the web mode in 2021

In general, the majority of questionnaire changes in the 2021 Wave affected the Family data file. However, for the 2021 Individual data file, there are two new additions regarding Covid-19 whether and an infection and whether vaccinated ER34999: WTR FU MEMBER HAD COVID-19 and ER35000: WTR FU MEMBER HAD COVID-19 VACCINE. These variables combine the questions asked of Reference Person/Spouse/Partner and those asked of OFUMS.

In the 2021 Wave questionnaire, we continue to ask the education series for eligible OFUMs (those aged 16-49 years at the time of the 2021 interview, who were in the Family Unit in the prior wave) that parallels the series asked of current Reference Persons/Spouses/Partners. Many of these variables have been asked for Reference Persons/Spouses/Partners in prior waves and also appear in the background sections

(Section K, for Spouses/Partners, and Section L, for the Reference Person, on the Family file). As of 2015, they continue to be included on the 2021 Individual file because they are asked of eligible OFUMs, as well.

As we did in 2019, users should note we have continue to use the CIP: NCES: Classification of Instructional Programs code frame for questions about college major. Given the granularity of the frame, we have provided two variables for each college major mention. The first variable is the first 2 digits of the CIP which is the main category (e.g., 01=Agriculture, Agricultural Operations, and Related Sciences). The 2-digit version of these variables will be publicly available. The second corresponding variable for each mention is the full CIP code XX.XXXX which are suppressed and only available for use under a restricted data use contract. The variables in question include KL55a/G88h3, KL55e/G88j, KL78f2/G88x2, and KL78g/G88y (up to 2 mentions each).

The PSID strives to keep education data current; thus, it has designed the questionnaire into two basic series. The first series is asked of all newly incoming Reference Persons/Spouses/Partners and eligible OFUMs: KL43/G88a through KL61B/G88M3 (ER34918-ER34951). The second series is asked of all returning Reference Persons/Spouses/Partners and eligible OFUMs, in order to obtain any additional educational attainment they may have received since their last interview. The update series includes questions KL74/G88N through KL84A/G88DD (ER345953-ER34987).

In addition, there are two generated variables for the education series. The first is “Years of Completed Education” (ER34952). For those returning Reference Persons/Spouses/Partners and eligible OFUMs, we calculate their “Years of Completed Education” from the update series. If the calculated level of completed education derived from the update series is higher than previously reported, then we update all the corresponding variables in the baseline series KL43/G88a through KL61B/G88M3 with this newly acquired information. In addition to updating the baseline series, we also update the variable “Year Highest Education Updated” (ER34917) with the year the baseline series was updated. As a result, the baseline series for Reference Persons/Spouses/Partners and eligible OFUMs always has the most up-to-date information.

## **B. Generated Variables**

For the 2021 Wave, we continue to provide the generated variable, ER35012: WHETHER MEDICARE NUMBER GIVEN. This variable provides a summary for those who were eligible to be asked for a Medicare number.

Longitudinal and cross-sectional weights continue to be included. For 2021, they are ER35064: CORE/IMM INDIVIDUAL LONGITUDINAL WT 21 and ER35065: CORE/IMM INDIVIDUAL CROSS-SECTION WT 21.

We continue to provide the generated variable in the education series ER34917: YEAR HIGHEST EDUCATION UPDATED. This variable indicates whether the baseline education was updated and when.

## **C. Other Additions**

For the 2021 Wave, children in the Family Unit were eligible for the Transition to Adulthood Study (TAS-2021) who met age requirements, were Followable, were not previously coded as “Unable to Cooperate” in a prior wave of TAS or CDS, and were part of a PSID family interviewed in 2021. Variables indicating 2021 eligibility (ER35056) and the result of the interview attempt (ER35057) are part of the 1968-2021 Individual data file.

Also, for the 2021 Wave, there are eligibility, selection status, result, and completeness variables for the Child Development Supplement (CDS-2021): ER35058, ER35059, ER35060, and ER35061, respectively. Please note that CDS-2021 selection status, result and completeness variables (ER35059, ER35060 and ER35061) are filled with zeros for Release 1 as final reconciliation for this study was not finalized at the time of this release.

## **II. DATA CHARACTERISTICS**

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

The 1968-2021 Individual data file consists of one data file with 84,121 records and 2,605 variables. The data are merged across all waves of the study; that is, they include person-level information collected from 1968 through 2021. 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 2021 response and 2021 non-response individuals.

Using the web-based PSID 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 as well as 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, analyses, etc. You must provide all other statements needed to complete a program. Also, missing data statements are not provided.

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

The 1968-2021 Individual data file consists of yearly items (e.g., Sequence Number, Relationship to Reference Person, Family Identification Number) and a set of summary or cross-year variables with up-to-date values (ER32001-ER32053 and Sex of Individual, ER32000).

Summary variable, ER32051: Year Most Recent Pregnancy Intention Rec, was updated, as of the 2021 Wave. It indicates the year of the most recent report on the Pregnancy Intentions file which includes a relatively new series of questions added in the 2013 Wave. It is provided to users as a separate file. Please see our supplementary data page for more information: <http://simba.isr.umich.edu/Zips/zipSupp.aspx>.

Weights for each wave of data are available through 2021 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,
- \* summary variables, ER31987- ER32053,
- \* 1994 Public Release Individual data, ER33101-ER33150,
- \* 1995 Public Release Individual data, ER33201-ER33299B,
- \* 1996 Public Release Individual data, ER33301-ER33347,
- \* 1997 Public Release Individual data, ER33401-ER33438,
- \* 1999 Public Release Individual data, ER33501-ER33547,
- \* 2001 Public Release Individual data, ER33601-ER33639,
- \* 2003 Public Release Individual data, ER33701-ER33742,
- \* 2005 Public Release Individual data, ER33801-ER33849,
- \* 2007 Public Release Individual data, ER33901-ER33951,
- \* 2009 Public Release Individual data, ER34001-ER34046,
- \* 2011 Public Release Individual data, ER34101-ER34155,
- \* 2013 Public Release Individual data, ER34201-ER34269,
- \* 2015 Public Release Individual data, ER34301-ER34414,
- \* 2017 Public Release Individual data, ER34501-ER34651,
- \* 2019 Public Release Individual data, ER34701-ER34864, and
- \* 2021 Public Release Individual data, ER34901-ER35065

## **C. Hand Coded and Generated Variables**

The summary variables fall into four groups. ER31990-ER31997 are used for sampling error and weights calculations; ER32000-ER32008 are true cross-year indicators derived from values in each individual's data record; and ER32009-ER32051 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). Lastly, variable ER32052 helps users determine the start of this individual's cohort year while variable ER32053 indicates whether this individual has a PSID-CENSUS1940 record in the corresponding restricted release file.

Other generated items include variables designating the originating family and move-out date for Splitoffs, for every wave, from 1969 to 2021 (2021: ER35051-ER35053), as well as income summaries for OFUMS.

## **D. Documentation and Codes**

### **1. 2021 Wave Questionnaire**

Our website includes box-and-arrow questionnaires for all waves, 1968-2021, 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 PSID Data Center allows users to create and download customized codebooks that provide variable-specific documentation; variable descriptions are complete for all of the 2021 Individual file data items. The downloadable .zip package of the Individual data on the PSID website contains a .pdf version of the entire codebook for the 2021 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 (i.e., not ascertained) or a refusal. "Inappropriate" questions (that is, those that do not apply) are padded with zeroes. If a variable contains a code value that is neither included in the codebook nor one of the aforementioned "zero", "eight", or "nine" codes, assume missing data for that value.

## **E. Missing Data**

Users should note that CDS-2021 variables selection result and completeness variables (ER35059, ER35060 and ER35061) are filled with zeros for Release 1. Final data col reconciliation for this study was not finalized at the time of this release.