April 2021

PANEL STUDY OF INCOME DYNAMICS
2019 PUBLIC RELEASE INDIVIDUAL FILE DOCUMENTATION
Release 2
_____

TABLE OF CONTENTS

_____

Release 2 Notes: For Release 2 we have updated Relation to Head/Reference Person and Marital indicators for several waves for a small fraction of individuals as part of our processing for the Family Relationship Matrix file.

I. INTRODUCTION: WHAT'S NEW FOR THE 2019 WAVE

   A. 2019 Questionnaire Changes

The most significant change for the 2019 Individual data file is the second wave of the new 2017 Immigrant cohort for 2019. As of 2019, there are a total of 1,627 individuals from the 2017/2019 Immigrant cohort. They consist of 1,348 individuals who were part of the 1968-2017 Individual file and continue to be response in 2019, an additional 97 individuals (96 of which are designated sample persons) who were non-response in 2017 but their family completed their first interview in 2019 were added and an additional 76 Multiplicity individuals (75 which are designated sample persons). The remaining 106 individuals added in 2019 are newly added family members who have moved into Immigrant families who were first response in 2017 and continue to be response in 2019.

These individuals were screened in a separate data collection effort in 2016 and those who met the screening criteria were included in the 2017/2019 main data collection effort. Specifically, at the time of the 2016 screening, the following criteria had to be met: the Reference Person and/or the Spouse/Partner were

born prior to 1997 (the year of our last cohort addition) and were not living in the United States at that time or were born after 1997 to parents who were not living in the United States in 1997.

The purpose of the new refresher sample is to ensure that the PSID remains representative of the current U.S. population. This new 2017 Immigrant cohort has been assigned 1968 Interview Numbers (ER30001) in the range of 4001-4851. Those in the range 4001-4462 completed their first interview in 2017 (with the exception of the Multiplicity sample described below) and those in the range 4700-4851 completed their first interview in 2019.

Additionally, in 2017, we asked our responding Immigrant 2017 families to provide information on non-co-resident parents, siblings, and children who met certain age and immigration criteria so that we may also pursue them and, if eligible, fold them into our Immigrant 2017/2019 cohort. We refer to this probability sample as the Multiplicity Sample in the documentation. We have added 76 individuals of this type to the 2019 file. The baseline year for the Multiplicity Sample is 2019 and users should note that since these families are 'related' to our 2017 Wave I Immigrant Sample, they must share the same 68ID (ER30001) as the families who recruited them.

Given that the 68ID range (ER30001) can no longer definitively capture the year this family's cohort began due to the addition of the Multiplicity Sample, for the reasons described above, we have provided a new summary variable, ER32052: YEAR THIS INDIVIDUAL'S COHORT BEGAN. This variable indicates which year this individual's original family was first added to the study. This variable contains values of 1968 for members of the SRC Cross-section sample and Census sample, 1990/1992 for the Latino sample, 1997/1999 for the first Immigrant sample, and 2017/2019 for the second Immigrant sample and Multiplicity sample.

Additionally, we have added new code value of 12 for the Multiplicity Sample at ER34854: WHY FOLLOWABLE to specifically identify these original sample members in this special cohort. Please note that one on the 76 Multiplicity individuals is non-sample and has a designated special person number of 227 *(please refer codebook for ER30002 for details)*

As of the 2019 wave, some previously non-sample individuals in the 1968 Census/SRC cross-section or the 1997 Immigrant cohort who met the same eligibility criteria as the newly Immigrant 2017 cohort have been designated as sample persons. These persons are identified in the documentation as joint inclusion individuals and they also have a new code value of 40 at ER34854 and have a value of 4 at WHETHER SAMPLE OR NONSAMPLE: ER32006. Across all waves there are a total of 104 joint inclusion individuals. As of the 2019 wave when they become sample, 94 of them are response.

In general, the majority of questionnaire changes in the 2019 Wave affected the Family data file. However, for the 2019 Individual data file, there are some new additions including ER34810: GS2 WTR AFFECTED BY GOVT SHUTDOWN. This variable indicates which individuals, if any, were affected by the U.S. federal government shutdown that began in December of 2018. The rest of the government shutdown series applies to the family and those variables appear on the Family file at ER72973-ER72991.

In the 2019 Wave questionnaire, we continue to ask the education series for eligible OFUMs (those aged 16-49 years at the time of the 2019 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 2019 Individual file because they are asked of eligible OFUMs, as well.

With regard to the education sequence, users should note we have migrated to a new code frame for questions about college major. Furthermore, we have recoded all college majors reported in prior waves and pulled forward these data to the current wave, using the new code frame by NCES: Classification of Instructional Programs (CIP 2010 frame). 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 (ER34718-ER34751). 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 (ER345753-ER34787).

In addition, there are two generated variables for the education series. The first is "Years of Completed Education" (ER34752). 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" (ER34717) 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

As of the 2019 Wave, we have added two new summary variables. ER32052: YEAR THIS INDIVIDUAL'S COHORT BEGAN, described in detail in the previous section, and ER32053: WTR THIS INDIVIDUAL HAS CENSUS MATCH RECORD. This second variable indicates whether this individual has a corresponding record in the PSID-CENSUS 1940 data file. This file provides 1940 Census variables for PSID individuals who have been linked to their corresponding 1940 Census record. The PSID-CENSUS1940 file is available to qualified users under special contractual arrangements with the PSID. Please see our restricted data page, on the PSID website, for further details.

For the 2019 Wave, we continue to provide the generated variable, ER34811: 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 2019, they are ER34863: CORE/IMM INDIVIDUAL LONGITUDINAL WT 19 and ER346864: CORE/IMM INDIVIDUAL CROSS-SECTION WT 19.

We continue to provide the generated variable in the education series ER34717: YEAR HIGHEST EDUCATION UPDATED. This variable indicates whether the baseline education was updated and when.

Please note that an error was discovered with the CORE/IMM INDIVIDUAL CROSS-SECTION WT for waves 1997-2003. Each of these variables, ER33438, ER33547, ER33639, and ER33742, have updated values. These updated weights have a minor impact on the weighted distribution of some individual characteristics. The mean age, years of education and percent male are not statistically different for 1999-2003 and only percent male is slightly different in 1997. While mean age is not statistically different, the frequencies of some age groups are different. The largest differences occur in 2001 and 2003 where the revised weights yield slightly more people in the older age groups.

C. Other Additions

For the 2019 Wave, children in the Family Unit were eligible for the Transition to Adulthood Study (TAS-2019) 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 2019. Variables indicating 2019

eligibility (ER346855) and the result of the interview attempt (ER34856) are part of the 1968-2019 Individual data file.

Also, for the 2019 Wave, there are eligibility, Fall Follow-up 2020 status, result, and completeness variables for the Child Development Supplement (CDS-2019): ER34857, ER34858, ER34859, and ER348860, respectively. Please note that CDS-2019 Fall Follow-up 2020 status, result and completeness variables (ER34858, ER34859 and ER348860) are filled with zeros for Release 1 as final data collection for this study was not finalized at the time of this release.

II. DATA CHARACTERISTICS

A. Files and Format

The 1968-2019 Individual data file consists of one data file with 82,573 records and 2,440 variables. The data are merged across all waves of the study; that is, they include person-level information collected from 1968 through 2019. 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 2019 response and 2019 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 1968-2019 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 2019 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 2019 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, and
* 2019 Public Release Individual data, ER34701-ER34864

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 2019 (2019: ER346850-ER34852), as well as income summaries for OFUMS.

D. Documentation and Codes

1. 2019 Wave Questionnaire

Our website includes box-and-arrow questionnaires for all waves, 1968-2019, 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 2019 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 2019 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 the longitudinal weight variable CORE/IMM INDIVIDUAL LONGITUDINAL WT (ER34650) contains zero values for the new Immigrant 2017 sample as no historical data are available for these individuals.

Additionally CDS-2019 variables result and completeness variables (ER34859 and ER348860) are filled with zeros for Release 2. Final data collection for this study was not finalized at the time of this release.