INSTITUTE FOR SOCIAL RESEARCH • SURVEY RESEARCH CENTER
PANEL STUDY OF INCOME DYNAMICS
UNIVERSITY OF MICHIGAN

PSID-2021 Main Interview User Manual: Release 2023

The PSID main interview user manual was prepared by April Beaule, Flannery Campbell, Noura Insolera, Paul Juska, Rose McAloon-Fernandez, Katherine McGonagle, Mohammad Mushtaq, Beth Simmert, and Jerry Warra. The manual draws heavily from documentation from prior years written by numerous PSID staff members.

This document should be cited as follows:
PSID-2021 Main Interview User Manual: Release 2023. Institute for Social Research, University of Michigan, June, 2023.

The Panel Study of Income Dynamics (PSID) is a household panel survey that began in 1968. This user manual serves as the primary source of documentation for the 2021 wave of the main interview. In addition, it provides critical information to users of the PSID such as the sample design, survey content, how to obtain the data, data quality, and much more. The manual includes important historical information about the survey, as well as information about the most recent data. For new users, this is the first document they should read before beginning to use the data.

Through the years, thousands of pages of PSID documentation, guides to using the data, and other such resources have been distributed to users. The current document does not replace these prior documents. Instead, this document serves as a starting place for understanding the PSID, with a focus on describing changes in the key features over the years. Within this document we point users to documentation from prior years where historical information is described in greater detail.

We expect the content of this manual to evolve and improve over time to meet the needs of the user community. Please contact us at psidhelp@umich.edu if you have suggestions for enhancing the manual or if you find errors

TABLE OF CONTENTS

List of Acronyms and Definitions

TOC

CDS: Child Development Supplement

CAI: Computer assisted interviewing

CATI: Computer assisted telephone interviewing

CRCS: Childhood Retrospective Circumstances Study

EHC: Event history calendar

FIMS: Family Identification Mapping

System FU: Family unit

ISR: Institute for Social Research

NIA: National Institute on Aging

NICHD: National Institute of Child

Health and Human Development

NSF: National Science Foundation

OEO: Office of Economic Opportunity

OFUM: Other Family Unit Member

PSID: Panel Study of Income Dynamics

RP: Reference Person—Head before 2017

SEO sample: Survey of Economic Opportunity sample

S/P: Spouse/Partner—Wife/"Wife" before 2015

SRC sample: Survey Research Center sample

TAS: Transition into Adulthood Supplement

WB: Wellbeing and Daily Life Supplement

List of Figures and Tables

TOC

1. INTRODUCTION TO THE PSID

TOC

The PSID was created to assess President Lyndon Johnson's War on Poverty. In 1966 and 1967, the Office of Economic Opportunity (OEO) directed the U.S. Bureau of the Census to conduct a study called the Survey of Economic Opportunity (SEO), which completed interviews with about 30,000 households. Interest in continuing this national study led OEO to approach the Survey Research Center (SRC) at the University of Michigan about interviewing a sub-sample of approximately 2,000 low- income SEO households. Professor James N. Morgan, who became the new study's director at SRC, argued successfully for adding a fresh cross-section of households from the SRC national sampling frame so that the study would be representative of the entire population of the United States, including non-poor as well as poor households. In addition, it was fortuitously decided to follow members of the families who moved away from their original households, such as children who came of age during the study. In this way, the sample could remain representative of the nation's families and individuals over time. This study became what is now called the Panel Study of Income Dynamics (Hill, 1992; McGonagle, Schoeni, Sastry & Freedman, 2012).

The PSID has now collected data for over 50 years. More than 84,000 people have participated in the PSID and as many as seven generations within sample families are represented. The mission of the PSID has broadened well beyond its original focus on income and poverty dynamics. As a result, the PSID has been used in thousands of peer-reviewed publications, and the user base has grown increasingly diverse, drawing in psychologists, medical researchers, public health scholars, geographers, and others. Numerous countries have developed their own PSID-like studies, facilitating cross national comparative research.

2. SAMPLE DESIGN AND FOLLOWING RULES TOC

2.1 Core Sample

The PSID was originally designed to study the dynamics of income and poverty. Thus, the original 1968 PSID sample was drawn from two independent samples: an over-sample of 1,872 low income families from the Survey of Economic Opportunity (the "SEO sample") and a nationally representative sample of 2,930 families designed by the Survey Research Center at the University of Michigan (the "SRC sample"). The oversampling of families who were poor in the late 1960s resulted in a sizable subsample of African Americans. These two samples combined constitute a national probability sample of U.S. families as of 1968.

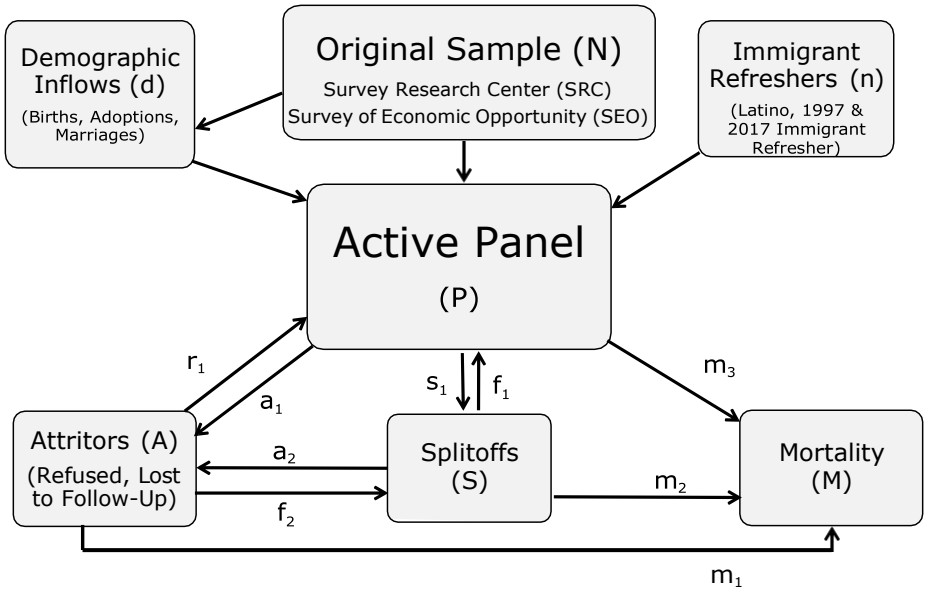
The rules for following individuals were designed to maintain a representative sample of families at any point in time as well as across time. To accomplish this, PSID "sample persons" include all persons living in the PSID families in 1968 plus anyone subsequently born to or adopted by a sample person. All sample members are followed even when leaving to establish separate family units (FUs). This procedure replicates the population's family-building activity and produces a dynamic sample of families each year.

PSID families also include many "nonsample persons." The most common example is people who after 1968 marry sample persons. Information on nonsample persons is collected while they are living in the same family unit as a sample person. However, once they stop living with a sample person, their household is not interviewed.

The steady-state panel design is depicted in Figure 1. Flows of people into the panel come from three sources: the original 1968 sample (*N*); the 1997 and 2017 refresher sample of post-1968 and post-1997 immigrants (*n*), which is described below; and births and marriages in existing families (*d*). The intergenerational element is represented by children who split off (s_1) as adults to form their own family units (*S*). Because of the follow-status rules (f_1, f_2), success in bringing in new families (i.e., boosting f_1, f_2), strategies to minimize attrition (a_1, a_2), and recontacting (r_1) families refusing to be interviewed or not located (*A*) in previous waves, the PSID active panel sample (*P*) has grown despite losses due to mortality (*m1, m2, m3*) and attrition from the active panel.

This self-replacing design implies that for many PSID families the data include self-reported information on multiple generations within the same family at various points in their lives. Through multiple waves collected over a long period on a national sample, the PSID is the only data set ever to provide information on life course and multigenerational economic conditions, well-being, and health in a long-term panel representative of the full U.S. population.

Figure 1. Steady State Panel Schematic



2.2 Immigrant Refresher Samples TOC

While the original design of the study augmented the sample with new birth cohorts each wave, it did not represent families who arrived in the United States after 1968 (post-1968 immigrants who co-resided with PSID sample members are captured in the data, but they are not followed when they are no longer co-residing with a sample member). To address this limitation, efforts have been made to add samples of immigrants who arrived in the United States after 1968.

2.2.1 Latino Sample TOC

In 1990, the PSID added roughly 2,000 Latino households, including families originally from Mexico, Puerto Rico, and Cuba. But while this sample did represent three major groups of immigrants, it did not fully represent all post-1968 immigrants. Because of this crucial shortcoming, and a lack of sufficient funding, the Latino sample was dropped after 1995.

2.2.2 1997 New Immigrant Sample TOC

A total of approximately 500 post-1968 immigrant families were added in 1997/1999 to update the PSID by adding a representative sample of recent immigrants to the United States: this sample is called the 1997 PSID New Immigrant Sample (NIS-1997). A detailed description of the sample design for the immigrant sample is provided here. Immigrants eligible for the supplemental sample were those who immigrated to the United States after 1968 or were children born in 1969 or later to people who were not living in the United States in 1968. These recent immigrants were not represented in the PSID sample before 1997. Because some of the immigrant refresher families subsequently created splitoff families, the number of families originating from the 1997 immigrant refresher sample continues to grow.

2.2.3 2017 New Immigrant Sample TOC

A total of 615 post-1997 immigrant families were added in 2017 to update the PSID by adding a representative sample of recent immigrants to the United States: this sample is called the 2017 PSID New Immigrant Sample (NIS-2017). 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. To be eligible for the PSID NIS-2017, the reference person and/or the spouse/partner must have been recent immigrants who were born abroad after 1997 (the year of our last cohort addition) or were born after 1997 to parents who were not living in the United States and moved to the U.S. after 1997. This includes reference persons and/or the spouse/partner who were born in U.S. territories and entered the U.S. after 1997. The selection criteria differed from the NIS-1997 in that reference persons and spouse/partners who were born in U.S. territories were not eligible. The purpose of the new immigrant refresher sample is to ensure that the PSID remains representative of the current U.S. population.

During the 2017 wave, 452 eligible families provided interviews, of which 421 agreed to be reinterviewed in 2019. That same year, an additional 29 families of the families who had been selected in 2017 provided and interview, for a total of 450 NIS-2017 families. As of 2021, 445 families have continued to participate.

2.2.4 Multiplicity Sample TOC

Additionally, in 2017, we asked our responding 2017 Immigrant families to provide information about 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 sample. We refer to this probability sample as the New Immigrant Multiplicity Sample (NIMS) in the documentation. We added 24 families of this type to the 2019 Family file. The baseline year for the New Immigrant Multiplicity Sample is 2019 and users should note that since these families are ‘related’ to our 2017 wave one Immigrant Samples they share the same 1968 ID (ER30001) as the family who recruited them.

The 2017/2019 Immigrant cohort were assigned 1968 Interview Numbers (ER30001) in the range of 4001-4462 if their baseline interview was 2017 (or 2019 if Multiplicity Sample) and 4701-4851 if their baseline interview was 2019. As of the 2021 Wave, there are a total of 445 families in the NIS-2017 cohort range 4001-4851.

2.2.5 Joint Inclusion Sample TOC

For couple headed family units to be eligible for the PSID NIS-1997, both members of the couple, reference person and spouse/partner, must have immigrated to the U.S. after 1968. In contrast, for couple headed family units in 2017, if either the reference person or spouse/partner (or both) of a screened family unit were recent immigrants, the family unit was eligible for the PSID NIS-2017 or NIMS. Under the 2017 eligibility rule, families who were eligible for the NIS or NIMS included: 1) single reference person – recent immigrant; 2) reference person – recent immigrant; spouse/partner – recent immigrant; 3)

reference person – recent immigrant; spouse/partner – not recent immigrant; and 4) reference person – not recent immigrant; spouse/partner – recent immigrant. Families in categories (1) and (2) could only enter the PSID through the NIS-2017 or NIMS. However, the families in categories (3) and (4) could have been selected for the PSID panel prior to 2017 if either the reference person or the spouse/partner were herself/himself Core-eligible and a sample person. 151 families from the NIS-2017 sample and 102 families from the PSID Core sample were identified as being in categories (3) and (4), hereafter referenced as Joint Inclusion Sample (JIS) families. An integration of weights to account for the dual chance of selection is needed for the Joint Inclusion Sample families.

2.3 Sample Reduction in 1997 TOC

Because the original sample of roughly 5,000 families had grown substantially due to splitoff families being added to the sample, and funding was not available to continue to interview the large number of families, the original core sample was reduced from roughly 8,500 families in 1996 to approximately 6,300 in 1997. The majority of the cuts were taken from the SEO sample. However, 43% of the SEO sample, or 1,714 families, remained in the active sample in 1997. Through natural sample growth generated by splitoffs, the SEO sample continues to grow. Because the SEO sample consists largely of low-income African-American families, it supports research on economic transitions, poverty, and disparities in health and other resources.

2.4 Sample Following Rules TOC

PSID follows sample members when they change households. Information is gathered about these sample members and others residing in the same household. A family member who moves out of a PSID family unit is eligible for interviewing as a separate family unit if they are a sample member and living in a different, independent household. If a sample member moves to an institution such as a prison, a college dormitory, the military, or a healthcare facility, the PSID records this fact and attaches an "institutional status" data record to the family they left. If the only sample person or the entire household is living in institutional housing other than prison, PSID still attempts to complete an interview for this family. We code the type of institution they are living in the variable on the yearly family file labeled "Type Institution" (ER78008 in 2021). For sample members still attached to active PSID families, the PSID keeps track of the location of sample members living in institutional housing, and attempts to interview them if and when they leave the institution.

Between 1968 and 1991, the sample following rules stated that individuals eligible for the next wave of interviewing would include only persons present in the prior wave. Therefore, during this period individuals who could not be located or who refused to be interviewed were not followed in later waves. Only if a nonresponse sample person subsequently moved into a currently responding family unit would they be followed in the future. While the "reappearance" of some formerly nonresponse sample individuals occurred in each wave, it was a relatively rare event.

In 1992, two new approaches for recontacting former nonresponse sample cases were adopted. First, in 1992, interviews were attempted for all sample persons who responded in 1990 but had become nonresponse in 1991, whether reinterview families or splitoffs. Second, interviews were attempted during 1992 for original sample individuals who had become nonresponse in any wave prior to 1991. These new approaches provided support for attempting a large-scale recontact effort for non-Latino cases, which occurred during the 1993 and 1994 waves. Nonresponse sample individuals who were last present in a PSID family in 1991 or earlier and who shared an original family identifier (1968 interview number) with someone who was still responding in 1992 were selected. More information about the recontact efforts and results are described in the 1992 and 1993 main interview documentation.

This work indicated that a significant portion of individuals who refused or were lost in one wave could in fact be successfully brought back into the study. Thus, starting in 1993, individuals who refused or were lost in a particular wave were designated as "recontact sample" and were followed in the subsequent wave. Once a family was nonresponse for two consecutive waves, the family was no longer followed.

Following rules for sample individuals under age 18 were changed starting in 1993. In previous waves, PSID did not follow as splitoff sample members under 18 years of age if they left the family unit and their new family unit did not contain a sample person (e.g., a sample child resided with a nonsample parent as a consequence of a divorce). Starting in 1993, the study began to follow these younger persons and attempt to interview an adult in the new family unit. As a corollary, the PSID family composition rules changed. Specifically, PSID families have always included a sample member as the Reference Person or the Spouse/Partner of the family unit, but this became impossible in some cases where we followed the underage sample member who moved out with a nonsample parent. Therefore, although all families contain at least one sample member, both the Reference Person and the Spouse/Partner may be nonsample. This could occur, for example, if an underage sample member moves out with a nonsample parent who then remarries.

2.5 New Terminology in 2021 TOC

PSID-2021 is the first wave in which a web instrument was available for completing the main PSID interview. This mixed mode design (WEB/CATI) has introduced web specifications in the questionnaire Box and Arrow and generated variables in the release file.

2.6 Sample Sizes TOC

Reported in Table 1 are the number of individuals and families in each of the main interview waves by sample type, where sample type identifies SRC, SEO, Latino, and Immigrant Refresher. This table includes both sample and nonsample persons.

Table 2 reports the number of sample persons (nonsample persons are excluded) in each wave by whether they are in the SRC, SEO, or Immigrant Refresher sample – the Latino sample is excluded. In

addition, the number of original sample persons – that is, individuals who were living in 1968 PSID family units – and the number of sample persons who are a Reference Person or Spouse/Partner are reported for each wave. In 2021, there are 2,927 individuals who were also present in the original sample in 1968. While the number of families has increased substantially, the number of Reference Persons and Spouse/Partners who are sample persons has not changed appreciably in part because non- sample persons have become Reference Persons and Spouse/Partners after 1968. The final two columns report the number of Reference Persons and Spouse/Partners and the number of all individuals who have been in the PSID for at least five waves and therefore can contribute to substantial panel analyses. These numbers were fairly steady until the sample was trimmed in 1997; they dropped substantially in 1997, but have increased subsequently.

Reported in Table 3 is the number of individuals entering and exiting the PSID in each wave, by reason for exit and entry. Although the exact number varies over time, roughly 100 individuals in the PSID die each year, or roughly 200 between each wave when interviewing became biennial in 1997. A total of 5,614 individuals were dropped in 1997 when the sample was trimmed. Each year, 300-400 children are born into PSID families, or roughly 600-800 between adjacent post-1997 waves. In most years, new sample members who make their first entry into the PSID (by moving into, rather than being born into, a sample household) are individuals whose parents are PSID sample members, but they themselves were born while their parent was not part of an interviewed family unit. These individuals were very rare prior to the recontact effort in 1993 and 1994. The large number of new sample members who first entered in 1997 is associated with the addition of the New Immigrant Sample. A substantial number of re-entrants were interviewed in 1993 and 1994 as part of the recontact efforts in those years.

These re-entrants were living with individuals who had never participated in the PSID in the past, hence the large number of nonsample individuals who first entered the PSID in 1993 and 1994. The number of reentrants after 1994 was higher than before 1993 because PSID began attempting interviews with individuals who were nonresponse in the prior wave.

Table 1. Number of individuals and families in each wave, by sample type: 1968 to present

	Families					Individuals				
Year	Core SRC	Core SEO	Latino	Immigrant	Total	Core SRC	Core SEO	Latino	Immigrant	Total
1968	2,930	1,872			4,802	9,461	8,772			18,233
1969	2,643	1,817			4,460	8,643	8,569			17,212
1970	2,754	1,891			4,645	8,751	8,597			17,348
1971	2,834	2,006			4,840	8,827	8,763			17,590
1972	2,947	2,113			5,060	9,109	8,942			18,051
1973	3,057	2,228			5,285	9,190	9,045			18,235
1974	3,165	2,352			5,517	9,285	9,109			18,394
1975	3,252	2,473			5,725	9,437	9,185			18,622
1976	3,318	2,544			5,862	9,556	9,212			18,768
1977	3,382	2,625			6,007	9,670	9,328			18,998
1978	3,416	2,738			6,154	9,697	9,443			19,140
1979	3,497	2,876			6,373	9,856	9,587			19,443
1980	3,589	2,944			6,533	10,034	9,713			19,747
1981	3,617	3,003			6,620	10,080	9,716			19,796
1982	3,673	3,069			6,742	10,232	9,880			20,112
1983	3,715	3,137			6,852	10,322	10,005			20,327
1984	3,729	3,189			6,918	10,349	10,044			20,393
1985	3,753	3,279			7,032	10,474	10,206			20,680
1986	3,750	3,268			7,018	10,400	10,037			20,437
1987	3,778	3,283			7,061	10,508	9,978			20,486
1988	3,809	3,305			7,114	10,555	9,951			20,506
1989	3,809	3,305			7,114	10,524	9,927			20,451
1990	3,935	3,393	2,043		9,371	10,677	10,068	7,452		28,197
1991	3,957	3,418	1,988		9,363	10,707	10,063	7,075		27,845
1992	4,051	3,510	2,268		9,829	10,934	10,211	8,130		29,275
1993	4,231	3,642	2,104		9,977	11,560	10,751	7,415		29,726
1994	4,624	4,034	2,106		10,764	12,576	11,936	7,033		31,545
1995	4,565	4,002	1,834		10,401	12,314	11,615	5,955		29,884
1996	4,547	3,964			8,511	12,294	11,516			23,810
1997	4,592	1,714		441	6,747	12,363	5,703		1,695	19,761
1999	4,740	1,787		470	6,997	12,787	5,937		1,791	20,515
2001	4,970	1,945		491	7,406	13,340	6,232		1,828	21,400
2003	5,159	2,126		537	7,822	13,684	6,661		1,945	22,290
2005	5,175	2,260		567	8,002	13,873	6,998		2,047	22,918
2007	5,295	2,412		582	8,289	14,150	7,263		2,088	23,501
2009	5,446	2,607		637	8,690	14,606	7,593		2,186	24,385
2011	5,495	2,767		645	8,907	14,607	7,844		2,210	24,661
2013	5,450	2,932		681	9,063	14,562	8,099		2,291	24,952
2015	5,318	3,037		693	9,048	14,151	8,247		2,239	24,637
2017	5,354	3,125		1,128	9,607	14,235	8,427		3,783	26,445
2019	5,255	3,172		1,142	9,569	13,925	8,401		3,758	26,084
2021	5,062	3,021		1,124	9,207	13,342	7,748		3,579	24,669

Notes: Sample type is defined by ER30001: Core SRC=1-2930/Immigrant 97/99=3001-3511/Immigrant 17/19=4001-4851/Core SEO=5001-6872/Latino=7000+. Sequence Number (SN) for each year can be used to determine family level (SN=1) and individuals in that year (SN>=1).

For 2017: 452 new Immigrant families and 1,624 Individuals were added to the sample. Note however that 9 of them are mover out spouses as of 2017 - those 9 are non-sample (1624-9=1615 original sample Imm 17 listed on Table 2 cell G45)

For 2019: 53 new Immigrant families (of which N=24 are Multiplicity) were added and 173 individuals. Note however that 2 of them are mover out spouses as of 2019 - those 2 are non-sample (173-2=171 original sample Imm19 added in 2019)

Table 2. Composition of PSID Sample Individuals (Original, Born In, Moved In, or Joint Inclusion)

Year	Total	SRC	SEO	1997 Immigrant	2017 Immigrant	Original Sample Member*	RP or S/P**	RP or S/P in PSID 5+ Waves	Individuals in PSID 5+ Waves
1968	18,233	9,461	8,772			18,233	7,878		
1969	16,327	8,261	8,066			16,050	7,118		
1970	16,130	8,184	7,946			15,486	7,160		
1971	16,089	8,117	7,972			15,117	7,238		
1972	16,074	8,128	7,946			14,724	7,330	6,260	14,625
1973	15,978	8,074	7,904			14,306	7,445	6,299	14,513
1974	15,921	8,034	7,887			13,917	7,562	6,396	14,462
1975	15,902	8,031	7,871			13,556	7,668	6,463	14,417
1976	15,830	8,009	7,821			13,105	7,707	6,510	14,284
1977	15,810	8,032	7,778			12,715	7,769	6,610	14,178
1978	15,821	8,004	7,817			12,426	7,842	6,749	14,190
1979	15,830	8,022	7,808			12,064	7,995	6,848	14,122
1980	15,894	8,072	7,822			11,693	8,076	6,920	14,078
1981	15,933	8,074	7,859			11,393	8,099	6,986	14,126
1982	15,998	8,118	7,880			11,136	8,161	7,068	14,141
1983	16,704	8,147	7,927			10,842	8,218	7,194	14,217
1984	16,074	8,139	7,931			10,524	8,213	7,273	14,300
1985	16,081	8,140	7,941			10,193	8,259	7,268	14,249
1986	15,878	8,063	7,815			9,834	8,188	7,237	14,133
1987	15,822	8,081	7,741			9,512	8,168	7,223	14,069
1988	15,785	8,077	7,708			9,232	8,165	7,214	14,081
1989	15,676	8,029	7,647			8,938	8,112	7,269	14,071
1990	15,732	8,091	7,641			8,782	8,204	7,352	14,113
1991	15,701	8,093	7,608			8,524	8,184	7,355	14,124
1992	15,940	8,211	7,729			8,472	8,325	7,453	14,295
1993	16,516	8,482	8,034			8,236	8,382	7,376	14,442
1994	17,883	9,075	8,808			8,644	8,948	7,640	15,213
1995	17,557	8,924	8,633			8,294	8,807	7,605	15,161
1996	17,457	8,883	8,574			8,036	8,741	7,616	15,226
1997	14,867	8,845	4,363	1,659		5718+	7,179	5,742	11,699
1999	15,064	8,908	4,475	1,681		5,484	7,293	5,782	11,884
2001	15,408	9,110	4,653	1,645		5,297	7,513	5,723	11,842
2003	15,782	9,205	4,888	1,689		5,124	7,793	5,656	11,848
2005	16,413	9,535	5,167	1,711		4,982	8,180	6,175	13,032
2007	16,716	9,693	5,323	1,700		4,724	8,400	6,342	13,417
2009	17,350	9,967	5,624	1,759		4,565	8,744	6,635	13,889
2011	17,466	9,933	5,784	1,749		4,298	8,898	6,730	13,933
2013	17,635	9,873	5,959	1,803		4,063	9,029	6,874	14,186
2015	17,334	9,581	5,988	1,765		3,702	8,987	6,884	13,978
2017	19,091	9,652	6,125	1,699	1,615	3,507	9,823	7,054	14,186
2019	18,787	9,419	6,170	1,667	1,531	3,232	9,694	7,103	14,205
2021	17,665	9,005	5,701	1,554	1,405	2,927	9,306	6,962	13,502

* Original Sample Member from 1968 Wave (ER32006=1) and 1968 ID Number = SRC / SEO

** RP: Reference Person, S: Spouse, P: Partner

Notes: Cell values were determined by the following - Total: Year sequence number 51-59 or sequence number 1-20 and Whether Sample or NonSample (ER32006=1-4), SRC: ER30001=1-2930; SEO=5001-6862; Immigrant 97: ER30001=3001-3511; Immigrant 17: ER30001=4001-4851; Original sample: ER32006=1 (for SRC and SEO samples only; excludes 1997/2017 Immigrant samples); Head/Reference Person/Spouse: Year Relation to H/RP or S/P=1, 2 (1968-1982) 10, 20, 22 (1983+); H/RP or S/P in PSID > = 5 Yrs: Sum of waves for H/RP or S/P>=5 up to each year, starting in 1972; Individual in PSID >= 5 Yrs: Sum of waves for all individuals >=5 up to each year, starting in 1972; Note: 1997 immigrant sample persons reach up to 5 years in wave 2005

\+ Sample reduction of the SEO sample occurs in 1997

Joint inclusion sample persons added in 2019, but retroactively sample in all waves they are response in Table 2.

Table 3. Entry and Exit of All individuals in the PSID, Excluding Latino sample

		Exited the PSID				Entered the PSID						
						First Entry Sample Type						
Year	Total	Total Exit	Left	Died	Drop	Total Entry	Born in	Moved in	New Sample	Joint Inclusion	Non-sample	Re-entrants
68 Base	18,230											
1968	18,192	38	18	20	0	0	0	0	0	0	0	0
1969	16,862	2164	2080	84	0	834	275	2	22	0	535	0
1970	16,978	775	692	83	0	891	375	3	8	0	474	31
1971	17,252	612	506	106	0	886	338	6	0	0	501	41
1972	17,630	676	549	127	0	1,054	392	6	1	0	601	54
1973	17,777	795	682	113	0	942	352	8	1	0	530	51
1974	17,988	729	625	104	0	940	368	14	0	0	504	54
1975	18,173	775	663	112	0	960	383	20	2	0	501	54
1976	18,307	887	785	102	0	1,021	424	27	1	0	497	72
1977	18,524	857	752	105	0	1,074	435	27	2	0	551	59
1978	18,723	771	695	76	0	970	369	15	0	0	4,900	96
1979	18,914	927	835	92	0	1,118	469	26	2	0	549	72
1980	19,226	946	840	106	0	1,258	522	41	1	0	591	103
1981	19,288	957	868	89	0	1,019	441	15	0	0	471	92
1982	19,500	883	770	113	0	1,095	427	24	1	0	560	83
1983	19,749	943	840	103	0	1,192	440	38	0	0	580	134
1984	19,831	1,022	912	110	0	1,104	459	16	0	0	521	108
1985	19,998	1,150	1,029	121	0	1,317	482	43	0	0	640	152
1986	19,850	1,175	1,068	107	0	1,027	369	19	0	0	534	105
1987	19,852	1,145	1,026	119	0	1,147	442	31	0	0	522	152
1988	19,869	1,076	958	118	0	1,093	419	23	0	0	509	142
1989	19,862	1,107	994	113	0	1,100	381	25	0	0	537	157
1990	20,115	989	885	104	0	1,242	374	28	0	0	568	272
1991	20,124	1,038	904	134	0	1,047	401	18	0	0	490	138
1992	20,534	1,094	973	121	0	1,504	355	117	0	0	603	429
1993	22,631	1,555	1,401	154	0	2,652	455	216	0	0	887	1,094
1994	23,781	1,558	1,398	152	0	3,708	418	499	0	0	1,050	1,741
1995	23,406	1,273	1,321	119	0	898	259	45	0	0	496	98
1996	23,304	1,255	1,126	121	0	1,153	351	44	0	0	612	146
1997	19,315	6,639	978	103	5,614	2,650	295	25	1,659	0	472	199
1999	19,858	1,785	1,317	217	0	2,328	603	72	265	0	893	495
2001	20,727	1,510	1,193	208	0	2,447	665	69	69	0	1,004	640
2003	21,550	1,683	1,478	205	0	2,506	695	73	1	0	1,111	626
2005	22,034	2,044	1,855	189	0	2,528	780	99	0	0	1,059	590
2007	22,580	1,907	1,651	256	0	2,453	822	95	0	0	1,075	461
2009	23,427	1,775	1,568	207	0	2,622	813	120	0	0	1,092	597
2011	23,598	2,208	1,968	240	0	2,379	826	80	0	0	1,061	412
2013	23,912	2,335	2,100	235	0	2,649	749	103	0	0	1,149	648
2015	23,521	3,138	2,854	284	0	2,747	767	111	0	0	1,093	776
2017	25,389	2,534	2,306	228	0	4,402	704	111	1,615	0	1,006	966
2019	25,107	3,024	2,738	286	0	2,742	664	97	175	94	969	837
2021	23,667	3,879	3,539	340	0	2,439	678	77	1	0	792	891

Notes: Two adjacent years are compared to determine values. Total (Column D): Y2=Y1-Total Exits + Total Entrants; this number equals the total number of individuals whose SN = 1-59 for that year (excluding the Latino sample). For 1968 since SN is absent we have used Relation to Head (ER30003 GT 0) as a proxy for SN noting the above exclusions of Person Numbers 227/228. **1968 Base**: There are 18,233 individuals marked as original sample in the SRC cohort (ER32006=1 and ER30001 in (1-2390 or 5001-6872)) however only 18,192 of them are also response in 1968 (ER30003 gt 0) the rest (N=41) are individuals who are categorized as original sample but were listing errors in 1968 and are first response after the 1968 base year. They enter the study in subsequent waves starting in 1969 so are categorized as New Sample in this table; 18,230 includes the non-sample people (N=38) who are not original sample persons in the SRC cohort and are thus listed as exits from 1968. **1968**: 18,192 excludes non-sample people (N=38), specifically those who were Spouse of Head who were movers out in base year 1968 (ER30002=227) or Spouse of Head who died in base year 1968 (ER30002=228). **2017**: New sample in 2017 includes the addition of the Imm17 sample cohort individuals - it excludes 9 IMM17 non-sample individuals who were coded Spouse of Reference Person and were movers out in base year 2017 (ER30002=227) and (ER32052=2017) and SN 2017 (ER34502=71). **2019**: New sample in 2019 includes the addition of the Imm17 sample cohort individuals added in 2019 N=171- it excludes 2 IMM17/19 non-sample individuals who were coded Spouse of Reference Person and were movers out in base year 2019 (ER30002=227) and (ER32052=2019) and SN 2019 (ER34702=71). It also includes N=4 individuals who are original sample members belonging to the 2017 cohort addition (ER32052=2017) but were categorized as listing errors in 2017 (171+4=175). There are a total of N=104 Individuals designated as Joint Inclusion persons, however only N= 94 of them are response in 2019 when their follow status and why follow codes have been changed to sample persons. In waves prior to 2019, these individuals are not followable and categorized as non-sample in Table 3.

3. SURVEY CONTENT

TOC

Since its inception in 1968, the PSID has collected extensive information on employment, income and family demographics. With input from the PSID Board of Overseers and the broader scientific community, the content has evolved to allow the study of emerging scientific and policy interests.

Table 4 shows the major topical areas contained in the main interview since 1968, and new content that was added in 2021. Table 5 shows the mean and median length of the interview for each wave since 1968. In 2021, the mean questionnaire length was 94.4 minutes.

Table 4. Overview of PSID 2021 Core Survey Content

Economic & Health Content	Section	Initial wave
Housing	A	1968
Mortgage distress	A	2009
Employment	B,C,D,E	1968
Income	G,R	1968
Coivd-19 Financial Series; Stimulus Payments; Paycheck Protection Program	G	2021
Government program participation; Food security	F,R	1968
Family transfers	G	1968
Consumption expenditures	A,F,H	1968
Time use	F	2017
Wealth & active savings	W	1984
Pensions	P	1976
Health status, behavior, and insurance	H	1968
Covid-19 Health	H	2021
Philanthropic giving & volunteering	M	2001
Covid-19 Donations	M	2021
Housework & childcare	F	1968
Social & Demographic Content		
Family composition & Residential changes	Roster	1968
Deaths	Roster	1968
Marriage & fertility	J	1968
Educational attainment; New sample background & updates	K, L	1968

Table 5. Questionnaire length in each wave (minutes): 1968 to present

Year	Mean (Median)	Year	Mean (Median)	Year	Mean (Median)	Year	Mean (Median)
1968	63.1 (60.0)	1979	28.1 (26.0)	1990	37.4 (33.0)	2005	73.2 (69.0)*
1969	61.8 (60.0)	1980	29.0 (27.0)	1991	29.5 (27.0)	2007	80.0 (76.0)*
1970	60.5 (60.0)	1981	26.5 (25.0)	1992	31.0 (29.0)	2009	74.9 (70.0)*
1971	59.1 (58.0)	1982	20.8 (20.0)	1993	34.5 (32.0)	2011	90.4 (85.0)
1972	66.2 (60.0)	1983	23.8 (22.0)	1994	43.9 (41.0)	2013	82.0 (78.3)
1973	20.1 (20.0)	1984	34.7 (32.0)	1995	34.6 (32.0)	2015	77.5 (73.5)
1974	23.1 (21.0)	1985	35.2 (33.0)	1996	30.1 (27.0)	2017	85.8 (80.4)
1975	26.9 (25.0)	1986	34.9 (33.0)	1997	39.5 (36.0)	2019	80.3 (76.6)
1976	25.3 (25.0)	1987	29.5 (28.0)	1999	61.9 (59.0)	2021	94.4 (87.5)*+
1977	25.0 (24.0)	1988	33.6 (31.0)	2001	60.4 (57.0)		
1978	26.9 (25.0)	1989	33.9 (31.0)	2003	69.6 (65.0)*		

* Capped at 300 minutes; less than 0.3% of interviews in a given year were longer than 300 min.
+ This is the overall mean and median for the first mixed-mode PSID interview. This includes the sum of time spent on all fields in the instrument for cases completed by web and by telephone.

The questionnaire for each wave is available under the documentation tab on the PSID's website, psid.org. Also available is a companion document to the survey instrument that is prepared each wave to

assist interviewers in addressing any questions raised by the respondent during the interview. These documents are called “question by question objectives,” or simply “QxQ”.

3.1 Highlights of Changes in 2021 TOC

In 2021, PSID introduced a mixed mode design, offering some families the opportunity to complete their interview online. Approximately 40% of the sample completed their interview in the web mode in 2021. It should be noted that the design allowed respondents to switch between modes during the interview. The vast majority of respondents completed the interview in one mode. New variables were added to the family level release file that indicate the type of mode in which each section of the interview was conducted. These new variables are:

ER78024	SECTION A MODE
ER78166	SECTION BC MODE
ER78480	SECTION DE MODE
ER78793	SECTION F MODE
ER79050	SECTION G MODE
ER79918	SECTION W MODE
ER80136	SECTION P MODE
ER80549	SECTION H MODE
ER80975	SECTION K MODE
ER81102	SECTION L MODE
ER81229	SECTION NEW IMMIGRANT MODE
ER81337	SECTION M MODE

The variable for the full Mode of Interview (ER78006) is still available but please be aware that the code frame has changed to include the web mode. Additionally, there are several series of unfolding brackets for amounts that are included in only the CATI instrument (e.g. in Section A: A20AE_W) in the Box and Arrow questionnaire. These brackets were included in the web instrument as ranges that respondents could select. During data processing we folded these web questions into the corresponding CATI questions A20a-A20e for ease of use. In this wave the Event History Calendar was removed and stylized questions were added to collect the same information.

In 2021, questions were added throughout the instrument collecting information about the financial and health effects of the COVID-19 pandemic. A new series of questions asking about gig-work was added in Section G. Many sets of questions from the 2019 Wave continue to be part of the 2021 instrument. At the same time, there are notable additions, deletions, and revisions which are detailed below, by section. Users should consult the questionnaires and cross-year index for more specific details.

Table 6. Overview of Changes to the 2021 Main PSID Questionnaire

Section	New/Enhanced Content	Dependent interviewing	Other Streamlining (2019 item referenced for dropped items)
Housing, Utilities, Device and Internet Access (A)	• A50: code frame was modified. Code 7 was split into code (7) "Got divorced or had a job transfer", (11) "was evicted", and (12) "was foreclosed on". Also new code (10) "COVID-19 or coronavirus" was added.	--	• Deleted A42: The next few questions are about your utilities for your [apartment / mobile home / home]. Do you [and your family living there] get one bill that combines electricity, gas, and other fuel for your home? • A42-1 and A42-3: slightly modified and split into 2 questions to more accurately collect information on heating, gas and electricity.
	• BC70-BC75: questions about professional certificates were moved from section KL. These questions are now re-asked every wave. • BC10APTO, BC11APTO and BC12APTO: new questions about whether sick and vacation time off is paid.	--	• The EHC was removed and replaced with standard questions to facilitate the move to Web/CATI design.
Childcare, Food, Vehicles, Transportation and Expenditures (F)	--	--	--
Income and OFUM Education (G) and Off-Year Income (R2)	• G18GIGA-G18GIGCMEN3: this new series was added to collect information about gig jobs. • GCOVID1-GCOVID17: new series to capture information about financial difficulties encountered during the pandemic. • G44B2-G45B3: 2 new series asking about stimulus payments and paycheck protection program.	--	• Deleted G25eckpt: the $25,000 threshold filter for whether we collected information about TANF, SSI and other welfare was dropped.
Wealth (W)	--	--	--
Pensions (P)	--	--	--
Health (H)	• HCOVID1a-HCOVID17: new series about health exposure to COVID-19.	• H19a.G44B2-G45B3: new questions about vaping.	• H32, H33, H42 and H43 and their respective EHC dates were dropped.
Marriages and Children (J)	--	--	--
Background (New Reference Person & Spouse/Partner); Education Update (returning RP/S/P) (KL)	--	--	• K62/L62-K67-L67: dropped and moved to BCDE
New Immigrant (IMMIG)	• IMM13-IMM17: new series asking about green cards, residency status, visas and permits is available in the restricted data. • IMM18-IMM25: new series asking about residency, where parents live, and whether children and siblings live outside US.	--	--
Philanthropy (M)	• MCOVID1-MCOVID5: new series asking about donations made during COVID-19. • MGROUP1, MGROUP2: new questions that introduce groupings of philanthropic organizations.	--	• M1, M9 and M9a were dropped.
Interview Close	--	--	--

Important Note regarding values of Don't Know and NA/Refused:

Please note that for cases where the mode is computer-assisted telephone interview, the values of "Don't Know" and "NA/Refused" are indistinguishable and both mean that the respondent was asked the question but did not provide an answer. Users could choose to collapse these two values without loss of information. These values are not used when the interview is completed through self-administration on the web.

4. DATA COLLECTION

TOC

4.1 Questionnaire Development and Testing

Questionnaire development begins early in the year prior to data collection and continues through the end of that year. PSID staff and investigators review proposed new content and consider removing or modifying existing questions. Modifications may also be made in response to issues that arise during collection and processing of the prior wave of data. Significant content changes, such as adding a new content domain, are reviewed with the PSID Board of Overseers prior to implementation.

The full computer-assisted instrument is tested by means of a user interface that allows the tester to work through an interview and record programming bugs and revisions to the instrument on a per question basis. The interface collects and manages a database of tester comments, which the programmer and tester then use interactively to reprogram and retest the bugs and revisions.

4.2 Field Operations

TOC

Since the inception of the study in 1968, the Survey Research Operations (SRO), the data collection unit within SRC, has administered field activities for PSID. SRO staff members work closely with PSID faculty to incorporate the latest survey methods and technical systems into fieldwork activities. The technical systems used to collect PSID data over the past five decades have evolved with new technological developments. The majority of interviews in each wave since 1973 have been collected by professional field interviewers employed by SRO over the telephone, and a computer-assisted telephone collection technology was adopted starting in 1993. SRO has used industry-standard Blaise software since 2003 (developed by Statistics Netherlands) to program the PSID survey instrument and "SurveyTrak" software (developed internally by SRC) to manage the sample.

Data collection for Wave 42 of PSID occurred between March 19, 2021 and December 31, 2021. Following several years of planned development and testing, the study implemented mixed-mode survey administration. It continued the use of Blaise software to program the survey instrument into both interviewer-administered (telephone) and self-administered (web) versions and adopted the new Michigan Sample Management System (MSMS; developed internally by SRC) to manage the sample in a mixed-mode environment. To evaluate potential effects of mode on survey response and fieldwork outcomes, two sequential mixed-mode protocols were introduced using an experimental design. One protocol randomized sample families to a "web-first" treatment, which encouraged response through an online interview, followed by an offer of telephone to complete the interview; a second protocol randomized sample families to a "CATI-first" treatment until the last phase of fieldwork when the option to complete a web interview was offered.

Table 7 reports the beginning and end dates of the field period, the percent of interviews completed by telephone, the average number of calls to complete a case, the amount of the incentive, and the percentage of interviews completed in Spanish. The table also reports the percent of Family Units for

whom the interview was completed by the Reference Person, and the percent of Family Units for whom the interview was completed by a sample person.

4.3 Response Rates

TOC

Response rates are calculated separately for the core (also known as original) sample families and for the families that were part of an immigrant refresher, in 1997/1999 or 2017. For each of these two samples, response rates are provided for four "interview types:" reinterview families, which are families that were interviewed in the prior wave; recontact families, which are families that were interviewed two waves prior to the current one but not the immediately prior wave; splitoffs, who are individuals who became economically independent creating their own Family Unit; and recontact splitoffs, which are families that have split off from recontact families within the current interviewing cycle. The wave-to-wave response rates - that is, the percentage of families who completed an interview in the current wave among those who completed an interview in the prior wave - by sample type and by interview type are reported in Table 8.

Table 7. Characteristics of Field Operations: 1968 to Present

Year	Number of families	Field Period: Start	Field Period: End*	% by telephone	Mean (median) # of calls to complete a case**	% calls 8+
1968	4,802	4-Mar	10-Jun	NA	2.5 (2.0)	1.9
1969	4,460	10-Mar	9-May	NA	2.3 (2.0)	1.6
1970	4,645	1-Mar	31-May	1.2	2.5 (2.0)	2.8
1971	4,840	1-Mar	1-Jul	2.4	2.2 (2.0)	1.7
1972	5,060	1-Mar	1-Jul	2.6	2.1 (1.0)	1.6
1973	5,285	1-Mar	1-Jul	76.6	2.6 (2.0)	3.7
1974	5,517	5-Mar	1-Jul	82.5	2.6 (2.0)	4.2
1975	5,725	1-Mar	1-Jul	84.5	2.7 (2.0)	4.6
1976	5,862	1-Mar	1-Jul	91.4	2.8 (2.0)	5.9
1977	6,007	1-Mar	1-Jul	83.9	2.7 (2.0)	5.4
1978	6,154	1-Mar	1-Jul	85.9	2.8 (2.0)	6.3
1979	6,373	1-Mar	1-Jul	88.4	3.0 (2.0)	8.0
1980	6,533	1-Mar	1-Jul	89.2	3.3 (3.0)	10.3
1981	6,620	1-Mar	29-Oct	91.9	3.4 (3.0)	12.0
1982	6,742	2-Mar	29-Sep	92.8	3.4 (3.0)	11.6
1983	6,852	21-Feb	11-Oct	93.4	3.4 (3.0)	12.3
1984	6,918	27-Feb	31-Oct	92.1	3.7 (3.0)	15.2
1985	7,032	4-Mar	31-Oct	91.2	14.4 (4.0)	19.6
1986	7,018	24-Feb	31-Oct	92.0	9.9 (3.0)	15.5
1987	7,061	3-Mar	25-Aug	91.8	11.5 (3.0)	14.6
1988	7,114	3-Mar	19-Sep	91.5	9.8 (3.0)	16.3
1989	7,114	2-Mar	16-Nov	91.7	7.3 (3.0)	18.1
1990	9,371	24-Feb	30-Nov	88.7	5.5 (3.0)	18.3
1991	9,363	18-Mar	24-Nov	93.9	6.4 (3.0)	22.4
1992	9,829	2-Mar	8-Dec	95.9	7.9 (4.0)	29.0
1993	9,977	20-Apr	22-Dec	97.3	6.7 (4.0)	26.4
1994	10,764	24-Feb	23-Dec	95.7	8.8 (5.0)	35.3
1995	10,401	20-Feb	20-Oct	97.9	5.9 (4.0)	24.1
1996	8,511	1-Feb	30-Jul	97.4	5.1 (3.0)	18.9
1997	6,747	13-Feb	13-Oct	97.5	5.9 (4.0)	22.6
1999	6,997	31-Jan	31-Oct	na	na	na
2001	7,406	3-Mar	17-Nov	97.0	na	na
2003	7,822	14-Mar	7-Nov	96.2	10.6 (6.0)	39.6
2005	8,002	14-Mar	8-Nov	96.6	10.7 (6.0)	37.4
2007	8,289	12-Mar	31-Dec	97.5	11.2 (6.0)	38.0
2009	8,690	19-Mar	27-Dec	97.4	12.6 (6.0)	41.2
2011	8,907	3-Mar	31-Dec	98.6	13.2 (6.0)	43.6
2013	9,063	10-Mar	31-Dec	97.3	14.2 (7.0)	44.8
2015	9,048	3-Mar	31-Dec	97.0	20.3 (8.0)	51.6
2017	9,607	1-Mar	31-Dec[++]	95.5	17.9 (9.0)[+]	56.6
2019	9,569	28-Feb	31-Dec	95.6	16.8 (9.0)	57.1
2021	9,244	14-Mar	31-Dec	66.9[+++]	21.8 (8.0)	50.5

* End date not exact for years 1969-1980;

** Attempts top coded 8+ for years 1968-1984; NA=not applicable. na=not available. Cell values were determined using relevant variables from the Data Center, with the variable names for 1990 as follows: Number of interviewers=v18044. Field dates=v18046. Telephone=v17709=1. Number of calls=v18857.Spanish interview=v18859=1.

[+]2017 forward uses total attempts=telephone calls, emails, text messages

[++]Core data collection ended 12/31/2017. Data collection for immigrant families was extended to Feb 15, 2019 to add bilingual interviewer staff to meet response rate goals.

[+++]First year of web collection as alternative mode to CATI.

Table 7, Continued. Characteristics of Field Operations: 1968 to Present

Year	Incentive	% of interviews in Spanish	% of interviews provided by Reference Person	% of interviews provided by a sample person
1968	5.00	NA	93.7	99.2
1969	5.00	NA	93.6	97.9
1970	5.00	NA	92.8	95.5
1971	5.00	NA	92.7	93.6
1972	5.00	NA	92.5	91.8
1973	7.50	NA	90.2	90.2
1974	7.50	NA	88.8	89.7
1975	7.50	NA	88.3	88.8
1976	7.50	NA	92.6	85.7
1977	7.50	NA	90.0	86.5
1978	7.50	NA	90.2	85.1
1979	7.50	NA	88.5	85.4
1980	9.00	NA	85.8	85.2
1981	10.00	NA	84.3	86.0
1982	10.00	NA	83.8	86.5
1983	10.00	NA	82.2	86.1
1984	10.00	NA	81.0	86.1
1985	10.00	NA	87.1	73.4
1986	10.00	NA	81.5	84.2
1987	12.50	NA	79.0	85.2
1988	12.50	NA	76.9	86.0
1989	12.50	NA	76.2	85.9
1990	15.00	13.5	74.1	87.3
1991	15.00	13.1	72.1	87.3
1992	15.00	13.5	70.7	86.8
1993	15.00	12.1	69.5	85.2
1994	15.00	11.9	69.3	81.6
1995	20.00	8.8	68.5	80.8
1996	20.00	0.2	69.6	78.8
1997	20.00	0.1	69.0	79.2
1999	40.00	4.9	68.2	80.6
2001	55.00	4.7	66.5	79.1
2003	55.00	4.4	67.2	78.3
2005	60.00	4.6	65.6	80.1
2007	60.00	4.4	66.4	79.9
2009	65.00	3.0	67.0	79.9
2011	65.00	2.9	68.7	77.9
2013	70.00	2.8	69.6	79.2
2015	70.00	2.7	69.9	84.5
2017	variable*	4.3	69.3	85.2
2019	variable*	4.6	70.0	86.2
2021	variable*	4.4	70.0	80.1

Notes: Type of respondent=1/10 (Reference Person) provided in family file. NA=not applicable.
For all years except 1968, 1985-1995 sample member was determined using Respondent=yes from individual file and ER30002=1-169. For years 1968, 1985-1995 sample member was determined using Who was Respondent from the family file, linking that with Relation to Head, and including only those individuals who were sample, ER32006=1-3.
*ranged from $75-$150 for Core sample and up to $300 for some immigrant families.

Table 8. Response rate each wave by sample type and interview type: 1968 to present

		Main PSID					Latino (1990-95)/ Immigrant (1997-present)					2017 Immigrant (2017-present)				
Year	Total	Re-interview	Re-contact	Split-off	Re-contact split-off	Total	Re-interview	Re-contact	Split-off	Re-contact split-off	Total	Re-interview	Re-contact	Split-off	Re-contact split-off	Total
1968	76.0															
1969	81.4	89.0		60.4		81.4										
1970	95.7	97.0		84.0		95.7										
1971	96.5	97.0		86.0		96.5										
1972	97.8	98.5		88.0		97.8										
1973	97.8	98.5		88.9		97.8										
1974	97.6	98.0		92.5		97.6										
1975	97.8	98.4		88.6		97.8										
1976	97.0	98.0		87.0		97.0										
1977	97.6	98.0		90.3		97.6										
1978	98.0	98.3		90.0		98.0										
1979	97.5	98.2		86.5		97.5										
1980	97.6	98.0		90.0		97.6										
1981	97.7	98.3		85.7		97.7										
1982	98.0	98.8		86.0		98.0										
1983	98.0	98.3		88.3		98.0										
1984	97.7	98.0		92.4		97.7										
1985	97.3	97.7		92.0		97.3										
1986	97.1	97.4		89.5		97.1										
1987	97.2	97.8		82.9		97.2										
1988	97.6	98.0		87.2		97.6										
1989	97.4	97.9		83.3		97.4										
1990	91.7	98.3		89.2		98.0					74.8					
1991	96.1	98.2		86.1		97.8	92.3		64.7		90.2					
1992	96.0	98.0		85.7		97.6	92.6		66.7		90.4					
1993	92.2	95.5	52.1	67.9	47.4	94.7	87.7	na	54.5	na	84.5					
1994	na	95.9	na	na	na	na	na	na	na	na	na					
1995	na	97.0	na	na	na	na	na	na	na	na	na					
1996	na	97.6	na	na	na	na										
1997	na	95.7	na	na	na	na					na					
1999	90.7	96.0	54.6	82.3	50.0	93.1	82.8	32.9	65.5	na	66.4					
2001	91.7	96.7	52.0	79.7	0.0	93.0	88.5	31.1	61.4	na	76.4					
2003	92.7	96.6	57.6	79.6	42.9	93.4	93.9	48.9	58.1	0.0	83.9					
2005	93.9	97.4	58.2	81.4	42.9	94.6	93.1	38.5	67.7	na	85.4					
2007	93.2	96.4	46.3	85.5	71.4	93.9	92.3	31.7	73.7	66.7	85.1					
2009	94.3	97.0	53.5	88.7	53.8	94.7	95.5	44.4	84.6	0.0	89.8					
2011	93.3	96.0	38.8	84.9	75.0	93.0	93.4	28.9	77.8	100.0	88.9					
2013	91.7	94.9	46.2	81.1	40.0	91.8	95.4	50.0	75.4	0.0	90.8					
2015	89.1	92.8	43.1	77.3	#	89.1	93.1	46.5	65.6	#	88.2					
2017	88.8	94.7	46.5	70.7	60.0	90.1	91.1	27.4	55.7	#	83.6	na	na	na	na	75.7
2019	88.3	93.7	45.8	65.4	#	89.9	90.5	53.1	72.2	100.0	85.5	83.6	21.3	48.8	na	68.7
2021	85.4	91.3	39.8	62.7	54.5	86.4	87.5	45.3	78.2	0.0	82.8	86.9	28.6	47.4	33.3	73.8

Notes: na= not available; deceased are included in base 1968-1972 and are excluded in all waves 1973-2021. Sample sizes for recontact split-offs for the Latino/Immigrant samples are quite small and therefore the response rates fluctuate substantially.

no sample in this category to calculate response rate

Note reinterview, splitoff, recontact, and recontact splitoff types are not applicable for the IMM17 sample in their wave one where year=2017

Note recontact splitoff types are not applicable for the IMM17 sample in their wave two where year=2019

4.4 Data File Organization TOC

Historically the PSID has released the main interview data in five different data files, and we continue to organize the data this way to facilitate use among established data users.

1. Family file
2. Cross-year Individual file
3. Childbirth and Adoption History (CAH) file
4. Marriage History (MH) file
5. Parent Identification (PID) file

Most variables are contained in what is called the family file, including all family level information as well as detailed the Reference Person and the Spouse/Partner. The cross-year individual file includes information on every individual who was ever in an interviewed family at any point since the study began in 1968. The information on this file is relatively limited; the vast majority of individual level information collected by the PSID is obtained only for the Reference Person and the Spouse/Partner, and this information is on the family file. The childbirth and adoption history file contains details about childbirth and adoption events of eligible people living in a PSID family at the time of the interview in any wave from 1985 through the most recent wave, including retrospective reports of such events. The marriage history file contains details about marriage events of eligible people living in a PSID family at the time of the interview in any wave between 1985 and the most recent wave, including retrospective reports collected in 1985 of all births and adoptions prior to that year. The parent identification file provides information collected about parent-child relationships from various sources since the 1983 wave, and the file consists of identifier variables that link children with their parents. For detailed information on the variables in these files, please see the PDF codebooks, located here.

5. FAMILY FILE

TOC

The family file contains one record for each family unit interviewed in a given year. It includes all family level variables collected in that year, as well as extensive information about the Reference Person and the Spouse/Partner. Therefore, the content of the family file is not restricted to family-level data. The Data Center is the most efficient way to obtain the family data, which creates a customized data extract and codebook for the user. The family data files are also available as zipped packages which include the codebook, the entire data file in ASCII format, and SAS, SPSS, and STATA data definition statements (which provide variable names, locations, and variable labels).

5.1 Format, Variable Names, And Positions

TOC

The 2021 Family Data File consists of one data file with 9,207 records and 3,958 variables. The variable names are in the range ER78001–ER81958.

5.2 Coding and Generated Variables

TOC

In this section we describe our coding scheme and the process for constructing generated variables. In general, code ending with “8” (98, 998, or similar depending on the specific variable) represents "don't know", a code ending with “9” (or 99, 999, etc.) represents other missing data or a refusal, and a code of “0” may represent “none” or a further defined inapplicable code. 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; this should be extremely rare.

Important Note regarding values of Don't Know and NA/Refused in PSID-2021

Please note that for cases where the mode is computer-assisted telephone interview, the values of "Don't Know" and "NA/Refused" are indistinguishable and both mean that the respondent was asked the question but did not provide an answer. Users could choose to collapse these two values without loss of information. These values are not used when the interview is completed through self-administration on the web.

The most complex generated variables are income, work hours, wages, and wealth. The next two subsections describe the construction of these variables, followed by a third subsection which describes the construction of all other generated variables.

5.2.1 Income, Work Hours, and Wages

TOC

A comprehensive technical paper was released in 2011 to provide users with an overview of the income and wage data in the 2007 PSID, as well as a detailed description of the methods used to impute missing and unreliable income and employment data. The procedures for 2011 were implemented in an identical fashion in 2021. Since 2015, the same labor and transfer income is collected for Spouse/Partners and Reference Persons. Imputations for these new variables were performed with the same methodology as for existing labor and transfer variables and a description of the imputation process can be found in Appendix A.

5.2.2 Wealth TOC

The wealth module was first included in 1984. This module was included again in 1989, 1994, 1999, and every wave since then. The question series includes unfolding brackets, and PSID staff members use this and other information to create variables representing the total value of wealth and its major subcomponents. Information from two sections -- the housing section (A) and the wealth section (W) -- were used to construct the 2021 net worth measures. PSID asks about nine broad wealth categories, including short-term debt:

1. Equity in business (also includes farm), now split into asset and debt components [W11A & W11B, ER79934 & ER79938].
2. Transaction accounts, now captured separately between (savings accounts and money market funds) [W28A, ER79970] and, (certificates of deposit, government savings bonds, and treasury bills) [W28, ER79976].
3. Value of debt aside from mortgage on the main home or vehicle loans, divided into sub-components: credit card debt [W39A, ER80002], student loan debt [W39B1, ER80012], medical bills [W39B2, ER80017], legal bills [W39B3, ER80022], loans from relatives [W39B4, ER80027], and unspecified other debt [W39B7, ER80033].
4. Equity in real estate (second home, land, rental real estate, or money owed on a land contract), now split into asset and debt components [W2A & W2B, ER79921 & ER79925].
5. Equity in stock (includes shares of stock in publicly held corporations, mutual funds, and investment trusts) [W16, ER79943].
6. Equity in vehicle (cars, trucks, a motor home, a trailer, or a boat) [W6, ER79929].
7. Other assets (includes bond funds, cash value in a life insurance policy, a valuable collection for investment purposes, and rights in a trust or estate) [W34, ER79997].
8. Equity in Individual Retirement Accounts [W22, ER79964].
9. Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2) [ER81834].

Questions about home equity (primary residence) are asked in section A.

Processing of the data includes three steps: a) imputation of the wealth components (1-8), b) computation of home equity (9), and c) construction of estimates for the total family wealth with and without housing equity. In the processing of the 2021 wealth data we followed the same approach as in prior waves. In particular, a hot-deck imputation technique was used for imputation of the missing data in each wealth component (1-8). Details on home equity imputation, including the numbers of cases imputed within each group, are given in Table 9. For the 257 cases missing at least one component of home equity, the mean value imputed was $149,117.

Table 9. Number of Cases by Missing Data Group and Imputation Category: Home Equity, 2021

Missing data group	Imputation category								
	A	B	C	D	E	F	G	H	Total
Group 1. House value is not missing; mortgage is missing	36	3	78	0	0	0	0	0	117
Group 2. House value is missing; mortgage is not missing	11	2	0	12	0	0	0	0	25
Group 3. House value is missing; there is no mortgage	30	4	0	0	26	12	0	0	72
Group 4. Both house value and mortgage are missing	5	2	0	0	0	0	21	15	43
Total	82	11	78	12	26	12	21	15	257

There are four rows (groups) in the table corresponding to four patterns of missing data. In Group 1, where mortgage is missing but not housing value, we first attempt to replace the missing mortgage value with a value from the last wave (imputation category A). If this is not possible, a value from the wave prior to the last is used (imputation category B). When this is also not possible, we use an assumed identity, Mortgage=0.6* House Value, to obtain an estimate for the mortgage.

When a housing value is missing but mortgage is not (Group 2) we proceed in the following way. First, using information on the "bracketed" or interval-censored responses, an estimate for the house value is calculated as an average of the lower and upper bracket values available. Then the missing mortgage is obtained as the maximum value of mortgage value reported in the last wave and the bracket estimate (imputation category A). If the last wave information is not available, we use information from the wave prior to the last wave (imputation category B). Finally, if mortgage value was not reported two waves ago then the missing value is imputed as maximum of house value divided by 0.4 and the house value estimate based on the bracket information (imputation category D).

In the case where house value is missing and there is no mortgage, i.e. mortgage value is zero (Group 3), we first try to use information on the house value from the last wave or, if needed, an earlier wave. When neither of these is available the missing home value is imputed in the same way as in Group 2, i.e., the missing value is assigned the maximum of the house value reported in a previous wave and the estimate obtained with help of the bracket questions (category A and B). When house value information is not available in the previous waves then the house value is assigned as an average of the upper bracket and the lower bracket values of the house value. Further, if the bracket information is not available then the median of the reported house values among those who have no mortgage is assigned.

Group 4 consists of cases with both the house value and mortgage missing. First, using information from two earlier waves, the missing house value is treated as in the same way as in Groups 2 and 3, and the missing mortgage is imputed as in the Group1 (imputation category A and B). If information in two preceding waves is not available but the bracket values for house value are available then house value is estimated as an average of the lower and the upper bracket values and home equity is

equal to 40% of this estimate, i.e. home equity is equal to 0.4 * (lower bracket + upper bracket)/2 (imputation category G). If the bracket information for the house value is also missing, then home equity is assigned the median value of home equity among all known cases (imputation category H).

Prior to the 2017 wave, question W27 asked about the ownership and value of many types of financial assets together, namely whether respondent families "*hold money in any of the following: checking or savings accounts, money market funds, certificates of deposits, government bonds, or treasury bills*." In 2015, 23% of respondents (weighted) replied "No" to this question and the value of these assets (W28) was set to zero for these cases. Comparisons to external, nationally representative data (such as estimates of checking account and savings account ownership in the Social Consumer Finances survey and estimates from the FDIC National Survey of Unbanked and Underbanked Households) raised the concern that W27 may undercount the number of respondents with basic banking services, in particular checking and savings accounts. One concern was that, as the list of financial products mentioned in W27 unfolds and becomes more complex, some respondents may have forgotten about the inclusion of checking and savings accounts at the beginning of the question.

In 2017, the unchanged question W27 was therefore followed by a new question, W27a, asking "Do you [or anyone in your family living there] have a checking or a savings account?" We found that 1,913 respondents who had answered "No" to W27 (which included checking or savings accounts alongside the list of more complex financial products) went on to answer "Yes" to W27a (which re-asked only about ownership of checking or savings accounts). That is, more than half (57%) of those who reported that they did not "hold money in any of the following: checking or savings accounts, money market funds, certificates of deposits, government bonds, or treasury bills" (W27) did report that they owned a checking or savings account.[1]

For the purpose of imputing missing wealth information during the 2017 wave, respondents' answers at W27 and W27a were combined into a single variable, W27x, to match the structure of prior waves and following the assignment rules shown below. For cases where W27x resulted in a response of 'Don't Know' (DK) or 'Refusal' (RF), an imputed value for the 2017 wave family level variable ER71435 (W28) was calculated.

W27	YES	NO	NO	NO	NO	DK	DK	DK	DK	RF	RF	RF	RF
W27a	ZERO (skip)	YES	NO	DK	RF	YES	NO	DK	RF	YES	NO	DK	RF
Final W27x value:	YES	YES	NO	NO	NO	YES	NO	DK	RF	YES	NO	RF	RF

[1] One situation in which this combination of reports is *not* indicative of a question wording problem in W27 is, e.g., if a respondent holds only a checking account that carries a negative balance (overdraft). That respondent would correctly respond "No" to W27 (since it asks about "hold[ing] any money in any of the following…) and "Yes" to W27a (which asks about ownership of an account). However, few respondents are in this situation and made that fine distinction, as also indicated by the fact that only 298 of these 1,913 respondents were recorded as having "zero" money in the account in question W28.

In the 2019 wave, the W27 question asking about ownership and value of many types of financial assets was changed again. The question was split into two separate and distinct questions: W27a and W27. W27a asks the respondent if the family unit has a checking, savings or money market account. This is followed by question W27 which asks if the family unit has any money in certificates of deposit, government bonds, or treasury bills. In 2017, both W27 and W27a asked about checking and savings accounts, in 2019 only W27a asks about checking and savings accounts. In 2019, both of these questions have a set of unfolding bracket follow up questions which are asked if the respondent indicates that they don't know how much of these assets they have. Because we are now capturing both of these sets of assets as separate dollar amounts we no longer use the hybrid internal combined question indicator, W27x, to process and impute missing values for W27 and W27a. The imputed Wealth variables are located at ER81776-81839.

5.2.3 Consumption and Expenditures TOC

In response to a growing interest in understanding household consumption choices, the PSID began expanding the number of questions on consumption expenditures in 1999. Four questions for out-of-pocket spending for health care were added: hospital and nursing home care, doctor's visits, prescription drugs, and insurance premiums. Also included were: assessments of educational expenses, including payment for tuition, books, supplies, room and board; and transportation-related expenses (for up to three owned or leased vehicles) including outlays on vehicles, vehicle loan and lease payments, down payments on vehicles, vehicle insurance payments, gasoline, repairs and maintenance, parking, bus fares and taxicabs.

Estimates of expenditures on these items using the PSID have been compared with estimates from the Consumer Expenditure survey. In general, estimates from the two data sources align fairly closely, although some differences do exist for certain categories. Details are described in Li, G., R.F. Schoeni, S. Danziger, and K.K. Charles. 2010. New expenditures in the PSID: comparisons with the CE, *Monthly Labor Review*, March, pp. 20-30 and in Andreski, Li, Samancioglu and Schoeni 2014 "Estimates of Annual Consumption Expenditures and Its Major Components in the PSID in Comparison to the CE", American Economic Review: Papers & Proceedings 2014, 104. For information on how these measures compare across studies, please see the Data Comparisons page.

Consumption expenditure questions were further expanded in 2005 to include information on spending on home repairs and maintenance, household furnishings, clothing, trips, vacations and entertainment. In 2017 questions were added asking about computing expenses and the rental value of the family unit's home.

The purpose of the PSID Consumption Expenditure Data is to provide summary consumption expenditure data for families in the 1999-2021 Family Data Files. The summary variables were calculated

from responses to the consumption questions collected in Sections A and F. All Consumption Expenditure data from 1999-2021 are available on the family-level files.

Consumption expenditures are annualized. If an amount was reported for a period of less than a year, it was inflated by the reciprocal of the fraction of the year that the report covers. If the report was for more than one year, the amount was deflated.

The expenditure variables do not include the value of in-kind government transfers. For example, the value of food stamps received by family units is not included in estimates of food expenditures. Users who wish to include food stamps in calculating food expenditures or total expenditures will want to incorporate information contained in the food stamp variables contained in the family data files on the PSID website.

For some expenditure categories the PSID offers the respondents unfolding brackets when they cannot recall the exact amount of expenditures. The conditional mean expenditure for each bracket was estimated using the exact-number-responses that fall into the particular bracket. The mean estimates were then assigned to households who responded using the option of unfolding brackets. Even though PSID typically has a very low rate of nonresponse, an imputation strategy was used to eliminate missing values. Imputation models included a third-order polynomial in age and an unrestricted spline for family size and were fit separately for each expenditure category using ordinary least squares. Imputation may result in negative values due to linear regression model. These negative values are kept in order to preserve population mean consistent with the estimation.

Beginning in 2019, accuracy code variables for consumption questions that were imputed using the method noted above have been released. A value of one indicates that the missing value was imputed. This is similar to the accuracy codes released for PSID income and wealth variables. These accuracy codes can be found in the range from ER81842 to ER81913 in the 2021 PSID family level codebook. Prior wave variables can be found contained in the ranges listed below:

1999 – [ER16515A2A--ER16515D7]

2001 – [ER20456A2A--ER20456D7]

2003 – [ER24138A2A--ER24138D7]

2005 – [ER28037A2A--ER28037E4]

2007 – [ER38263A1--ER41027E4]

2009 – [ER44236A1--ER46971E4]

2011 – [ER52395A2A--ER52395E4]

2013 – [ER58212A2A--ER58212E4]

2015 – [ER65411A--ER65448B]

2017 – [ER71488A--ER71527C]

2019 – [ER77515-- ER77586]

Two aggregated variables were also added in 2019 to the imputed consumption data set. They aggregate all of the sub-categories into a total household consumption sum. The first summation variable added in 2019 is Total Expenditure, which includes: food at home, food delivered, food eaten out, hospital, doctor bills, prescriptions, health insurance, mortgage, rent, utilities, telephone & internet, homeowners insurance, property taxes, household repairs, household furnishings, vehicle loans, vehicle leases, vehicle down payments, auto insurance, additional vehicle expenses, vehicle repairs, gasoline, parking, bus, taxi, other transportation expenses, education, childcare, clothing, trips, other recreation, and computing expenses.

The second summation variable added in 2019 is Total Consumption with Rental Value Total. Consumption with Rental Value includes the following expenses: food at home, food delivered, food eaten out, hospital, doctor bills, prescriptions, health insurance, rent value, rent, utilities, telephone & internet, homeowners insurance, household repairs, household furnishings, vehicle loans, vehicle leases, vehicle down payments, auto insurance, additional vehicle expenses, vehicle repairs, gasoline, parking, bus, taxi, other transportation expenses, education, childcare, clothing, trips, other recreation, and computing expenses.

Put more simply, the Total Expenditure variable includes mortgage payments and property taxes, but excludes the value of the home if rented (ER81850). Conversely, the Total Consumption with Rental Value variable includes rental value (ER81850), but excludes the mortgage payments and property taxes. These new summation variables were included in the release of accuracy codes from the past waves from 1999 to 2017. However, of those past years, only the 2017 wave includes both total expenditures and total expenditures with rental value. The total expenditures with rental value variable is not available in waves 1999 to 2015. For the current 2021 wave, the consumption accuracy codes can be found in the variable range ER81840 to ER81915. The 2021 Total Expenditures variable is ER81914 and Total Consumption with Rental Value variable is ER81915.

For a listing of available variables in the Consumption Expenditure Module, please see the Consumption Expenditure Content Summary.

5.2.4 All Other Coded or Generated Variables TOC

In this subsection we describe all coded or generated variables other than income, work hours, wages, wealth, consumption, and expenditures, which are discussed above.

The PSID hand codes several data items for the Family File. For example, Reference Person's and Spouse's/Partner's occupations and industries, to a maximum of four jobs apiece, are coded for 2021, using the four-digit codes from the 2010 US Census Occupation and the 2012 version of the US Census Industry codes. 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 Reference Person's and Spouse's/Partner's non-academic degrees and certificates is coded for as many as three each. Any education received outside the U.S. is coded for Reference Person, Spouse/Partner, and

their parents.

Background items, such as education, are collected only for “new” Reference Persons and Spouses/Partners in a given wave. In 2019, we reclassified “new” to exclude Recontact individuals. During processing, we have traditionally "brought forward" background information from the preceding wave or the wave prior to that for Reference Persons or Spouses/Partners who are the same persons. Beginning in 2013, we asked for an education update from our returning Reference Persons and Spouses/Partners. If they reported additional education attainment, then we have updated their education variables in the background section and reported the year of the education update on the Family File (see variables “Year Highest Education Updated” for Reference Persons (ER81153) and Spouses/Partners (ER81026). 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 ER81931, for Reference Persons, and ER81932, for Spouses/Partners. Completed Education of Reference Person and Spouse/Partner variables (ER81026 and ER81153, respectively) are generated from the background information.

Family composition and change variables include Family Composition Change (ER78007), Splitoff Indicator (ER78005), Reference Person/Spouse/Partner Sample Status (ER81933), and variables about births to Reference Person, Spouse/Partner, and other family members during the prior calendar year, 2020 (ER81949– ER81952), as well as during the "off" year, 2019 (ER81953– ER81956). Please note that the Splitoff Indicator is only assigned to a family in the year that family first moves out of the main family and forms its own separate household; after that one wave of being a splitoff, these families receive code values that designate reinterviews. A new code with a value of 5 was added in 2019 to the Splitoff Indicator to designate a new 2017 Immigrant Family who was nonresponse in 2017 but response in 2019. This new code also indicates a Multiplicity Sample Family whose wave one interview is also 2019 (see, 2.2 Immigrant Refresher Samples, for a description of the Multiplicity Sample). Two other variables concerning Splitoffs are the number of Splitoffs arising from a main family (ER81934) and the Family Interview Number of the main family from which a Splitoff family originated (ER81935).

The PSID produces sets of variables about families sharing the same housing unit (HU): Family ID numbers, relationships, and sizes of up to four other PSID families sharing the same HU (ER81936–ER81947), the Household ID number (ER81948), and the number of persons not included in any PSID Family Unit who are sharing the Household Unit (ER78023). 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 consist of Reference Person’s current marital status (ER78025), the generated form of marital status comparable with years prior to 1977 (ER81928), change in marital status of Reference Person between waves (ER81929), and couple status of Reference Person (ER81930).

Location variables include PSID/GSA and FIPS state codes (ER78003 and ER78004), Current

Region, collapsed Beale Rural-Urban code, and Size of the Largest City in the County (ER81918–ER81921). We continue to provide the Beale Rural-Urban Code; however, we have updated the frame so that we are using the 2013 classification scheme as published by the USDA. The Beale Rural-Urban Code (ER81920) and Size of Largest City in the County (ER81921) are suppressed to protect the anonymity of our respondents and may be obtained under a restricted data contract. More information can be found on the Restricted Data webpage. A Metro/Non-Metro Indicator based on the Beale Rural-Urban Code (ER81919) is a public release variable we provide for users' convenience. Finally, we continue to provide two derived variables, from background information: Region where Reference Person (ER81922) or Spouse/Partner Grew Up (ER81924) and Reference Person's (ER81923) and Spouse's/Partner's (ER81925) Geographic Mobility. The codes for the FIPS and PSID/GSA codes are found on our website under State and Foreign Country Codes.

The Census Needs Standard was generated for the prior calendar year, 2020 (ER81916). Additionally, since the PSID has switched to biennial interviewing, a comparable needs standard has also been generated for the "off" year, 2019 (ER81917).

The variable indicating whether a PSID family lives in institutional housing is ER78008. The variable indicating the total number of data records from the Cross-Year Individual data file that are associated with a panel family is ER81575.

The weight variable included on the PSID-2021 Family file is ER81958: Core/Immigrant Family Weight.

The imputed work hours and income variables are found at ER81576–ER81775. Employment information for both 2020 and 2019, including whether Reference Person/Spouse/Partner was employed, unemployed, or out of the labor force, and the number of weeks is located at ER81426-ER81518).

6. CROSS-YEAR INDIVIDUAL FILE TOC

The cross-year individual file contains one record for each individual present in an interviewed family in any survey year.

6.1 What's New for 2021

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. Beyond this substantial mode change, 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 a family member had been infected by the COVID-19 virus (ER34999) and whether a family unit member had been vaccinated for COVID-19 (ER35000). 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 continued 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 two 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 (ER34953- 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.

6.1.2 Other Additions TOC

For PSID-2021, children in the Family Unit were eligible for the Transition into Adulthood Supplement (TAS-2021) if they met the 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 TAS-2021 eligibility (ER35056) and the result of the interview attempt (ER35056) are part of the 1968-2021 Individual data file. There are also eligibility (ER35058), selection status (ER35059), result (ER35060), and completeness (ER35061) variables for the 2021 Child Development Supplement (CDS-2021). Please note that the CDS-2021 selection status, result and completeness variables are at the time of PSID-2021 Release 1 filled with zeros, as the final files for the supplement have not yet been released.

6.2 Data Characteristics TOC

This section provides a brief overview of how the file is created, how the variables are generated and named, as well as the documentation and supporting information that coincides with the release of the data.

6.2.1 Files and Format

The 1968-2021 Individual Data File consists of one data file with 84,121records and 2,605 variables. The variable names are in the range ER34901-ER35065. 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 nonresponse individuals.

6.2.2 Variable Names, Positions, and Generated Variables TOC

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-ER32051 and Sex of Individual, ER32000). Summary variable ER32051: "Year Most Recent Pregnancy Intention Rec" was updated, as of 2021. It indicates the year of the most recent report on the Pregnancy Intentions File, which are a series of questions added in the 2013 wave and is provided to users as a separate file. Please see our packaged data page for more information.

Individual level sample weights continue to be included. For 2021, the individual level longitudinal sample weight is ER35064and individual level cross-sectional sample weight is ER35065.

6.2.3 Coded and Generated Variables TOC

The summary variables fall into three 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. 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 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.

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.

7. CHILDBIRTH AND ADOPTION HISTORY FILE, 1985 – 2021 TOC

7.1 Overview

The 1985-2021 Childbirth and Adoption History (CAH) file is designed to facilitate access to information collected in the 1985 through 2021 waves of the PSID regarding retrospective histories of childbirth and adoption. This file contains details about childbirth and adoption events of eligible people living in a PSID family at the time of the interview in any wave from 1985 through 2021.

Each set of records for a specified individual contains all known cumulative data about the timing and circumstances of his or her childbirth and adoption experience up to and including 2021, or those waves during that period when the individual was in a responding family unit. If an individual has never had any children, one record indicates that report. Similarly, if the individual never adopted any children, one record contains the denial.

Records contain identifiers for the individual and their children; dates of birth for both parent and child; geographic identifiers about the place of birth; the child's birth order, birth weight, birth length, race, and date of death; the year of most recent report and number of births or adoptions for the parent. Several significant changes have been made since 1985. Beginning with the 2005 Wave, the child's Hispanicity (CAH27) was also asked. In 2013, a much extended series of questions about the pregnancy, delivery, whether contraception had been used prior to the pregnancy, and whether the child was wanted by the mother/father was introduced. Most of these questions are asked of childbirth records only and many of them are only asked if the reported child was a newborn. Also, the ethnicity question series has been dropped, as of the 2013 Wave. The ethnicity variables on the cumulative file (CAH32-CAH34) have been maintained but they will only apply to records obtained between years 1985-2011.

In addition, the set of race variables (CAH28–CAH30) have had a series of changes that are important to note. In 2007, an important change was made to the code frame for these variables.

In 2005, the code frame for race was:

1) White
2) Black
3) Native American
4) Asian/Pacific Islander
5) Latino Origin or Descent
6) Color Besides Black or White
7) Other

In 2007, the code frame was changed to:

1) White
2) Black, African-American, or Negro
3) American Indian or Alaskan Native
4) Asian
5) Native Hawaiian or Pacific Islander
7) Other

That is, the "color besides black or white" code was dropped, and the "Asian/Pacific Islander" code was split into "Asian" and "Native Hawaiian or Pacific Islander".

In order to maintain the cumulative file we did two types of recoding:

(a) for those children in the 1997-2005 records who had a race mention of 5, Latino Origin or Descent, we:

- Recoded the Hispanicity question (CAH27) to a new value of 6 (Latino, no further information), and;
- Recoded their race questions (CAH28-30) to 7 (Other)

(b) in the 2013 CAI interviewing instrument, Native Hawaiian and Pacific Islander were two separate categories. To maintain consistency with the previously collected records, 2013 records coded as either Native Hawaiian or Pacific Islander were combined into Code 5, as in the 2007 code frame. The same approach was applied for 2015 forward.

The 1985-2021 Childbirth and Adoption File stores information in an efficient manner that allows a high degree of flexibility in linking with the PSID's Individual File. Linkages can be done from either the parent's or the child's standpoint. These linkages are more limited for children than for parents, since some children have never lived in a PSID study family and hence have no record on the 1968-2021 Individual File.

Data users who want only some of the detail of the childbirth data will find childbirth and adoption history information summarized on Public Release versions of the PSID Individual and Family files. Individual-file variables include number of births, birth dates of most children, identifiers of mother and father, whether the mother was married at the time the individual was born and birth order of the individual. Family variables include the number of births in the prior calendar year to the Reference Person, Spouse, Partner and Other Family Unit Members (OFUMs).

7.2 Individuals for Whom the Data are Available TOC

The childbirth and adoption history data were collected for individuals in responding families who were of childbearing age, i.e., individuals meeting the age requirements who had values in the range of 1-20 for the "Sequence Number" variable in a given wave. In waves when individuals were non-response or in an institution, no information was collected.

Starting in 2013, the way childbirth and adoption histories were collected changed significantly. Before 2013, information was collected in various ways for each eligible person about their biological and adopted children. In 2013, we started asking the first eligible person about their children and then asked who the other parent was of that child. During processing we copied the childbirth and adoption information reported for one individual to the other as per the ‘other parent’ report.

Adoption history data were gathered in a fashion similar to childbirth history, except that

information was collected for PSID family-unit Reference Person and Spouses/Partners, but not for OFUMs.

A number of complexities in the overall study design present special challenges for collecting and processing the childbirth and adoption history data:

(a) In any wave of the PSID, some family members appear in the study for the first time, although most are people who have been participating for years.

(b) From one wave to the next, a PSID individual can enter or leave eligibility for being asked marital or childbirth histories by passing the threshold ages for these questions. For reports from 1985-2011, the entry age for eligibility is 12 and the exit threshold for eligibility was 45 for many persons. As of 2013, the entry age for eligibility is 15 and the exit threshold for eligibility is 45.

(c) A PSID individual can change his or her relationship to Reference Person from one wave to the next and this can affect whether the childbirth and adoption history is self-reported or proxy-reported by a parent or by some other relative.

(d) From one wave to the next, the range of demographic events asked about a given individual can expand--information about adoptions is gathered for Reference Persons, Spouses, and Partners, but not for OFUMs.

(e) If a PSID individual becomes nonresponse the childbirth and adoption history is not updated.

(f) While both Reference Persons and Spouses/Partners were interviewed in 1985 (each giving a self-report), only one person (usually the Reference Person) has been the respondent in each wave since then.

(g) In the 1992 wave, the PSID undertook a pilot effort to recontact former respondents who had attrited from the study and persuade some of them to rejoin. Additional information regarding the 1992 recontact samples is discussed in the 1992 PSID Documentation.

7.3 Background for the Childbirth and Adoption History Files TOC

The 1985-2021 Childbirth and Adoption History File originated with the 1985 collection of comprehensive, retrospective questions about a number of demographic events, including childbirth, adoption, marriage, separation, divorce and substitute parenting. In each wave from 1986 through the present, these histories, with the exception of substitute parenting, were updated for eligible individuals.

Since 1986, the demographic history data are released annually as two separate files: the Childbirth and Adoption History File and the Marriage History File. These files are cumulative, and so their size increases each year as more events occur and additional people become eligible.

Data on childbirth and adoption are assembled into one file to facilitate analysis that may treat births and adoptions in the same framework. A primary function of the childbirth and adoption information is to clarify the relationships between individuals in the PSID. This information helps distinguish step relations from biological and adoptive ties.

7.4 How To Obtain A File And Whom To Contact About Questions TOC

The 1985-2021 Childbirth and Adoption File is available in the Data Center, as well as in .zip format. If you have questions that are not answered by this documentation, you can contact PSID staff through our website.

7.5 Questionnaire Detail TOC

The flow of the PSID questionnaire is complex. In addition, the types of individuals asked for history information and the detail gathered about their history may change over time.

Data users may find it helpful to actually see the questionnaires. The 1986-2011 sequences are identical to each other, with the addition, beginning in 1997, of questions about birth location, race and ethnicity. And beginning in 2005, a question about the child's Hispanic ethnicity was added. In 2013, an additional set of 75 variables were added to the Childbirth and Adoption History (CAH) file, with the majority focusing on pregnancy, prenatal care, aspects of the delivery and infant care of the child. Most of these questions apply to childbirth records only and more specifically only to reports of newborns.

7.6 File Structure TOC

7.6.1 Number of Records

The 1985-2021 Childbirth and Adoption History File contains a total of 145,599 records, with 106,684 childbirth records and 38,915 adoption records . The file has a one-record-per-event general structure. Each record contains information for a childbirth or an adoption event. For example, if an individual has one biological child and one adopted child, the file contains one childbirth record and one adoption record for him or her. The same parent may appear on both childbirth and adoption records.

Multiple records for a given parent can result from an individual having (or adopting) more than one child. Although the parent remains the same in such circumstances, the child differs from one record to the next. An individual who has his or her own biological children may also be an adoptive parent. The maximum number of childbirth records for a specific individual is 18 for biological children, and nine for adoption records.

A given child can also have multiple records on the file, if the birth or adoption was reported for two or more individuals. Since a child has two biological parents, the same birth can produce two records on the file, one for the father as parent, and the other for the mother. The same applies to the adoption history data, and a child could have as many as four records on the file if he or she was both born and adopted within the study, i.e., was adopted by sample relatives.

7.7 Idiosyncrasies, Data Cleaning and Variable Detail TOC

Several aspects of the Childbirth and Adoption History merit particular attention. This section discusses what they are and how to handle them.

7.7.1 How to Identify Individuals Who Have Never Had or Adopted A Child

One caution, particularly relevant to event-history analysis, concerns the records for individuals who have never had any biological or adopted children. Such a person has a data record denying the event. If he or she has neither had nor adopted any children, the file has two records, one for each type of event. On these records, codes indicating "Inapplicable" (9s) are padded in the fields for details about the child, with the exception of Child's 1968 Interview Number and Child's Person Number (CAH10 and CAH11); both of these variables are padded with zeroes.

7.7.2 How to Identify Births/Adoptions That Were Not Ascertained TOC

Persons who may have had, may not have had, or adopted children but for whom the PSID has been unable to determine anything relating to that particular type of parental experience also have one record for each demographic phenomenon on the file. On these records, the "Number of Births/Adoptions" variable (CAH106) has a value of 98, although this value is not unique to such individuals. Missing data codes (8s or 98s) are padded in all the fields for that record, with the exception of Child 1968 ID (CAH10) and Child Person Number (CAH11). These two variables contain values of 9s, which do uniquely identify this sort of record.

7.7.3 Treatment of Individuals Who Become Nonresponse TOC

The Childbirth and Adoption History File is cumulative; that is, all individuals who have ever been eligible for the childbirth or adoption history question sequences since they were first begun in 1985 have at least one record on the file. Thus, each new version is current through the most recent wave for individuals in responding families, but the data are only up to date through the last year that nonresponse individuals were living in a responding family.

The childbirth and adoption history is current through 2021 for those in a responding PSID family at the time of the 2021 interview and who are otherwise eligible for the childbirth and/or adoption history questions. For those who were nonresponse in the 2021 wave or who are no longer eligible for the questions, the history is current through the last year they were in an interviewed family unit and eligible. For example, if an individual became nonresponse for the 1988 wave and has not returned to a PSID family, his or her childbirth history (and adoption history, if applicable) is current only through 1987.

Similarly, if an OFUM who has remained in a responding family is now 47 years old, his or her childbirth history has not been updated in the last few waves, since OFUMs' childbirth information is not collected once they achieve 45 years of age. The variable indicating recency of an individual's childbirth or adoption reports is CAH104.

7.7.4 Location Data about the Child's Place of Birth TOC

Beginning in the 1997 wave, information about a child's place of birth was added to the child-specific questions. This was asked only for biological children reported from 1997 forward if they were born since January 1st of the preceding wave, but the birth year restriction was waived for adoption

reports. The birth location data comprise two variables using FIPS state and county codes. Foreign births are coded with the PSID foreign country code scheme, in which the state variable contains values of zero and the county variable indicates the specific foreign country.

Because of the PSID's policy on respondent confidentiality, only the variable for the state in which the birth occurred is included on the public Childbirth and Adoption History File. The county variable is classified as restricted and suppressed in the public release file. Access to restricted data must be obtained by a special request and confidential data use contract.

7.7.5 Treatment of Incomplete or Inconsistent Information TOC

We have tried very hard to assure access to all available information while also recording occurrences of missing data or unclear identification of children. In some situations, however, a parent was reported to have had biological or adopted children, but details about some or all of the children were not reported. PSID staff can and do assign an identifier to such a child, as it is clear that the child has never been part of the study.

Sometimes the same child is reported in more than one wave. In such cases, the initial report was chosen as the source for the child's sex, birth date, birth weight, etc., unless the information was not ascertained. In that event, a succeeding year's report was used on the principle that known information is better than missing information. However, values for the child's current whereabouts (CAH24), and death date (CAH25-CAH26) if applicable, are always picked up from the most recent report. The variable indicating the most recent wave in which the child was specifically mentioned, (CAH103), shows in which wave that happened.

If a child had ever been in the study and his or her birth or adoption records contained missing information about sex or birth date, values from the cross-year Individual File were used. Additionally, parental reports of a child's sex or birth dates were cross-checked for discrepancies, and Individual File data were consulted to help us resolve the differences where possible. Once birth dates were known, we checked the spacing of births to the same mother. All cases where successive children were born less than ten months apart were checked for possible coding, reporting or transcription errors in birth dates.

Aside from multiple births, a few cases remain where the children are nearer than ten months apart in age. Most of these are legitimate, but in some cases, we know the dates are incorrect and we are unable to resolve them.

Parent and child birth dates were compared in order to check births occurring to a parent under 13 years old, and to mothers aged 50 or older. Twenty-two cases of very young birth parents and four cases of very old birth mothers remain on the file. Most of the children have never been in the study (their Person Numbers have values of 800-995), so we are unable to consult another source for satisfactory confirmation. The parental birth dates are consistent with the cross-year Individual File. However, birth dates of all children ever in the study were not universally checked against the Individual File. Some

discrepancies with the Individual File may still be present.

On a related note, a parent's birth date is copied from the current wave of the Individual File when his or her births first appear on the cumulative Childbirth and Adoption History File. This birth date is not updated in later years for the initial record(s). If a new child is subsequently reported for the parent, then his or her birth date from that later wave is used in construction of the new record. Thus, birth dates for the same parent may disagree across children if the reports were not collected in the same wave. These discrepancies were cleaned to some extent, but only as a by-product of other cleaning activities.

Attempting birth date consistency with the Individual File is made difficult because of parents' and children's varying years of participation in the study. In the longer term, we hope to clean these dates, but for the present we advise analysts to use the parent's birth date or age variables from the most recent year of Individual File data for which he or she is present in the study.

Other data cleaning steps ensured that each child had no more than one birth mother and father. In a few cases, birth parents also claimed that they had adopted the child. These reports were verified against information in interviewer thumbnail sketches and marginal notes for corroboration. Child identifiers were compared to cross-year individual data and to the 1985-2021 Marriage History File to confirm that no spurious cases of intergenerational incest occur. Our final checks assured that all individuals who had ever qualified for childbirth or adoption questions had records of the appropriate type on the file, and that individuals who had never qualified for a specific type did not.

7.7.6 Who Has Cross-Year Information? TOC

Please keep in mind another PSID intricacy when matching across files: while all parents were present in a PSID family, some children identified in the birth and adoption histories have never been present in a PSID family unit during the years the study has been in progress; these children have values for "Person Number of Child" in the range 800-995. Consequently, each parent has been in a PSID family unit and has a record on the 1968-2021 Individual File, but his or her child may or may not.

7.7.7 Birth Order and Number of Children TOC

Children of a specified parent are ordered from the oldest to the youngest based on their birth dates. If no birth dates contain missing data, then each child is rank ordered from the earliest to the most recent date. If one or more birth dates contain missing data, then missing data are assigned to the order variable (CAH9) for all births. The birth order variable applies only to childbirth records.

Occasionally it is possible to assign birth order to some of the children, even though others may have missing information for birth dates. This can happen if an individual's retrospective history contains non-missing information about the number of existing children, although their birth dates are missing, but in a subsequent year the individual reports an update about a new birth. For example, a new Reference Person moves into the study. At that time we receive a report about two children living with his ex-Spouse, but we do not obtain information about their birth dates. The children are assigned values of 98

for birth order (CAH9) because we don't know which one is older. In the next wave, the Reference Person and Spouse have a new baby. This brings the total number of children to three, and we know with certainty that the new baby is his third oldest child.

In cases with known birth years for all children, unknown birth months can cause order for a pair of children to be assigned missing data values if they are born in the same year but with no evidence of twinhood. Updating the number of an individual's children can have a negative effect on the number of children (CAH106). If the number and order of all prior children is known but we have not ascertained whether he or she has had any additional children, then the order values for the known children remain as they are but missing data values must be assigned to the variable for total number of children. The 2021 file has 82 parents who fall into this category.

7.7.8 Adoption Dates TOC

The adoption date was not collected as part of the adoption history data. The cross-year Individual Files do, however, record move-in dates for adopted children coming to live in responding PSID families.

7.8 Linking Records TOC

7.8.1 Using the Childbirth And Adoption History File With the Individual File

The 1985-2021 Childbirth and Adoption History file is designed to be linked with the 1968-2021 Individual File for analysis purposes. The CAH file has only a modest amount of information about the parent and his or her child. Data users will no doubt want to access the much greater volume of data available for these individuals on the Family and Individual Files. Those files can provide abundant information for many individuals dating back to 1968 when the PSID began, although for some, the span of available data is more limited or nonexistent.

Because of the corrections, do not attempt to match this file with any other PSID data file or a subset derived from it using these unique individual identifiers. Non-matches in individual records definitely will occur if any other file besides the above-named is used.

Data processing is required to link records between these two files. To achieve linkages, one must match on the parent's unique individual-specific identifier. This unique identifier is a combination of two variables: "1968 Family Interview Number" (CAH3) and "Person Number" (CAH4). The corresponding variables for these identifiers on the cross-year Individual File are ER30001 and ER30002. Care must be taken with regard to the proper files to use, the choice of individuals on the Childbirth and Adoption File for whom matches are attempted, and the years for which data are available.

Parents and children vary substantially in terms of which years they have been present in PSID family units over the course of the study. This affects the availability of data for them on the main files because valid information is obtainable on that file only in the years that an individual is present in a PSID family unit ("present" means living in the family unit or having left it to enter an institution). For more

details about PSID tracking procedures and classification of people into family units, see the discussion in Chapter 8 on "Family Composition Change," in The Panel Study of Income Dynamics: A User's Guide, by Martha Hill (Sage Publications, 1992). First there is the matter of whether any record exists for an individual. If a record does exist, then the question is in which years of the study are data available for that individual and his or her family.

If a person, either parent or child, has a record on the 1968-2021 Individual File, but less than the full range (42 waves) of data in that record, variables in the years when he or she was not present in a PSID family unit are, for the most part, filled with zeros. In addition, the annual individual-level variable "Type of Individual Data Record" in those years indicates that he or she is nonresponse.

All eligible individuals have records on the 1968-2021 Individual File because they were present in a PSID family unit during that time period. However, many children do not have records on that file, since indeed they have never been present in a PSID family at any time during the course of the study. Children who have been present in a PSID family unit at some time since the study began have records on the 1968-2021 Individual File. Values for their Person Numbers are in the range 001-399. Children who have never been present have Person Number values in the range 800-995.

Even though all parents have records on the 1968-2021 Individual File, that file does not necessarily contain data for all of them for all years of the study. Some parents first entered the study in, e.g., 1985, and thus only have data since that year. Others have logged more than thirty years of inclusion in the study. Similarly, there is considerable variation among children regarding which years, if any, they were included in the study.

For those persons with records on both the Individual File and the Childbirth and Adoption History File, linkages rely on a match of individual-specific identifier variables that appear on both files. As noted earlier, the unique identifier involves two variables: "1968 Family Interview Number" and "Person Number". The corresponding set of variables on the two files must match to properly link an individual's records from the two sources. The variable names for these two variables are ER30001 and ER30002 on the Individual File. On the Childbirth and Adoption File, they are CAH3 and CAH4 for the parent and CAH10 and CAH11 for the child. Note that such linkages involve a one-to-many-match. One record on the Individual File may have more than one matching record on the Childbirth and Adoption History File because the specified individual has multiple children or has both childbirth and adoption records.

7.8.2 Using the Childbirth And Adoption History File With Other Files TOC

Some analysts may be interested in linking information from different records on the Childbirth and Adoption History file or linking information from records on different demographic files. For instance, access to all childbirth records for an individual is needed to identify full and half siblings; that is, biological children of the same parent. To determine, for example, ages of children of single parents

based on marital spells data, one would need to merge records on the Marriage History and CAH files. To make links such as these, one must match on the unique individual-specific identifier, which is a combination of two variables -- "1968 Family Interview Number" and "Person Number" (MH2 and MH3 for the person designated as the individual on the 1985-2021 Marriage History File; CAH3 and CAH4 for the person designated as the parent on the 1985-2021 Childbirth and Adoption History File).

7.9 Childbirth Information Available On the Individual and Family Files TOC

Some of the information provided on the Childbirth and Adoption History Files is also available on the final release versions of the cross-year Individual File. In addition, the Individual Files contain some detail relating to fertility issues that the Childbirth and Adoption History Files do not, and they provide information involving a combination of detail about marriage and fertility that would otherwise require data management.

7.10 Codebook TOC

The item-by-item descriptions for all of the variables in the Childbirth and Adoption History File can be found in the codebook. Unweighted frequencies were calculated for each variable.

8. MARRIAGE HISTORY FILE TOC

8.1 Overview of the 1985-2021 Marriage History File

The marriage history data were collected about individuals of marriage-eligible age in responding PSID families, i.e., those with values of 1-20 for the "Sequence Number" variable in a given wave. In waves when individuals were nonresponse or in an institution, no information was collected about them. The types of individuals asked marriage history information and the detail gathered about their history differed over the waves from 1985 through 2021; 1985 followed one pattern and 1986 through 2021 followed another.

To keep respondent burden to a minimum and data quality high, different question sequences about these events have been used for PSID individuals depending on their circumstances. Marriage history information was gathered as described below:

(a) in the 1985 wave, a complete retrospective marriage history was asked of a Reference Person, Spouse, or Partner of any age;

(b) in all succeeding waves, marriage history was updated for changes since the beginning of the prior calendar year for a Reference Person, Spouse, or Partner of any age who was also a Reference Person, Spouse, or Partner in the prior wave's interview;

(c) in succeeding waves, details about first and current or most recent marriages were asked for a new Reference Person, new Spouse, or new Partner of any age; and

(d) in all waves from 1985 through the present, details about first and current or most recent marriages were asked for an Other Family Unit Member (OFUM) aged 12-44 at the time of the interview. In 2013, OFUM age eligibility was changed to those aged 15-44; however marital status for OFUMS was still obtained for those aged 12-44 therefore we generated denial records for those aged 12-14 for those OFUMs whose marital status was reported as never married in the Coverscreen portion of the CAI interview

These latter two groups, although initially asked about only first and last marriages, may have additional marriage records on the file if those marriages occurred while the individual was in a responding family.

A number of complexities in the overall study design present special challenges for collecting and processing the demographic history data:

(a) In any wave of the PSID, some family members appear in the study for the first time, whereas most are people who have been participating for years.

(b) From one wave to the next, a PSID individual can enter or leave eligibility for being asked marital or childbirth histories by passing the threshold ages for these questions. For example, the entry age for eligibility is 12 and, for family members other than Reference Person, Spouse, or Partner, the exit threshold for eligibility is 45.

(c) A PSID individual can change his or her relationship to Reference Person from one wave to the next and this can affect whether the demographic event-history information is self-reported or proxy-reported by a parent or by some other relative.

(d) From one wave to the next, the range of demographic events asked about a given individual can expand or contract. For example, information about adoptions is gathered for Reference Persons, Spouses, and Partners but not for other family members.

(e) A PSID individual can become nonresponse, after which time demographic event history information is not updated.

(f) While both Reference Person and Spouses/Partners were interviewed in 1985, only one person (usually the Reference Person) has been the respondent in each year since then.

The marital history data of the Latino sample are also included in this file.

For more information on the 1992-1997 interview year recontact efforts, please see the 1992 PSID Documentation.

8.2 Background for The Marriage History Files TOC

The 1985-2021 Marriage History File originated with the 1985 collection of comprehensive, retrospective questions about a number of demographic events, including childbirth, adoption, marriage, separation, divorce and substitute parenting. In each wave from 1986 through the present, these histories, excepting substitute parenting, were updated for eligible individuals.

All the retrospective data collected in 1985 on these demographic phenomena were included in the 1985 Ego-Alter File. This file was mostly of interest for substitute parenting information and for child-specific information on public program participation and health care surrounding a birth. These sets of questions were discontinued after the initial retrospective in 1985. Beginning in 1986, due to the inherent complexities of the Ego-Alter File collection and data dissemination, the Ego-Alter File was replaced annually as two separate files: the Childbirth and Adoption History File and the Marriage History File. These files are cumulative and so their size increases each year as more events happen and additional people become eligible.

8.3 How To Obtain The File And Whom To Contact About Questions TOC

The 1985-2021 Marriage History File is available in the Data Center, as well as in .zip format. If you have questions that are not answered by this documentation, you can contact PSID staff through our website or at PSIDhelp@umich.edu.

8.4 Questionnaire Detail TOC

The flow of the PSID questionnaire is complex. As described above in 8.1, the types of individuals asked history information and the detail gathered about their history have changed over time.

Data users may find it helpful to actually see the questionnaires. The 1986-2021 sequences about marriages are identical to each other. PDF format versions of the 1968-2021 main questionnaires are also available on our website. In the more recent waves, they contain the complete series of questions for Reference Person/Spouse/Partner/OFUM marriage and birth/adoption history however for some earlier years where supplementary forms used to collect details about childbirth and adoptions, or for OFUMs' marriages and childbirths, those supplements are not available on the website as of this writing.

8.5 File Structure TOC

8.5.1 Number of Records

The 1985-2021 Marriage History File contains a total of 63,599 records. This file has a one-record-per-marriage general structure. Each record contains information for a specified marriage for an individual or information that indicates the individual has no marriage data. Information for an individual is current as of the most recent wave that marriage history was collected for him or her.

Multiple records for a given individual result from an individual having more than one marriage. Although the individual remains the same in such circumstances, the spouse differs from one record to the next. The maximum number of marriage records for a given individual is eight on the 1985-2021 Marriage History File. A few multiple records for a given spouse also exist. This occurs when an individual has remarried the same person. In situations where two individuals were married to each other twice and were both present in a responding family unit, the file contains four records for this pair, two records for each of the two individuals.

8.5.2 Sort Order of the File TOC

The 1985-2021 Marriage History File is sorted, in ascending order, by "1968 Interview Number of Individual" (MH2), "Person Number of Individual" (MH3), and "Order of This Marriage" (MH9).

Details for an individual's first marriage are followed by those for his or her second, third, etc., or the most recent marriage. As noted previously, complete marriage histories were gathered for Reference Persons, Spouses and Partners in 1985, but information on only first and most recent marriages was initially collected for New Reference Persons, Spouses, and Partners in 1986 through 2021 and OFUMs in any wave from 1985 forward.

All records for the first eligible member of a 1968 PSID family are followed by those for the next eligible member in the same family of origin. When all of the records for all eligible members in the first family are exhausted, records for eligible members in the second family follow.

8.5.3 Variables on the File TOC

The 1985-2021 Marriage History File contains twenty variables, which can be categorized into those relevant to the individual, those specific to his or her spouse, those in regard to the marriage in question, and some aggregate information about the marriage history for the individual. Please see the codebook for the full listing of these variables, along with the code frame, and frequencies in the sample.

8.6 Idiosyncrasies, File Cleaning And Variable Detail TOC

Several aspects of the 1985-2021 Marriage History File merit particular attention. This section discusses what they are and how to handle them.

8.7 How to Identify Individuals Who Were Never Married TOC

One caution, particularly relevant to event-history analysis, concerns the records for individuals who have never married. The file has one record for each such person. On these records, the "Number of Marriages" variable (MH18) has a value of zero. Codes indicating "Inapplicable" (9s) are padded in the fields for marriage details, with the exception of Spouse ID (MH7) and Spouse Person Number (MH8); both of these variables have values of zero.

8.8 How to Identify Individuals For Whom No Marriage Data Were Ascertained TOC

Persons who may or may not have married but for whom the PSID has been unable to determine anything relating to his or her marriage situation, not even marital status, also have one record on this file. On this type of record, the "Number of Marriages" variable (MH18) has a value of 98, although this value is not unique to such individuals. Missing data codes (8s or 98s) are padded in all the fields for that record, with the exception of Spouse 1968 ID (MH7) and Spouse Person Number (MH8). These two variables contain values of 9s, which do uniquely identify this sort of record.

8.9 Treatment of Individuals Who Become Nonresponse or Non-Eligible TOC

The Marriage History File is cumulative; that is, all individuals who have ever been eligible for the marriage history question sequences since it was first begun in 1985 have at least one record on the file. Thus, each new version is current through the most recent wave for individuals in responding families but the data are up to date only through the last year that nonresponse individuals were living in a responding family.

The Marriage History File is current through 2021 for those in a responding PSID family at the time of the 2021 interview and who are otherwise eligible for marriage history questions. For those who were nonresponse in the 2021 wave or who are no longer eligible for the questions, the history is current through the last year they were in an interviewed family unit and eligible. For example, if an individual became nonresponse for the 1988 wave and has not returned to a PSID family, his or her marriage history is current only through 1987. Similarly, if an OFUM who has remained in a responding family is now 49 years old, his or her marriage history has not been updated in the last few waves, since OFUMs' marriage information is not collected once they achieve 45 years of age. The variable indicating recency of an individual's marriage reports is MH17.

8.10 Treatment of Incomplete Or Inconsistent Information TOC

We have tried very hard to assure access to all available information while also recording occurrences of missing data or unclear identification of spouses. In some situations, the individual was

reported to have married, i.e., his or her marital status is known to be divorced, widowed, separated, or currently married, but details about the marriage or the identification of the spouse were not reported.

PSID staff can and do assign an identifier to such a spouse, as it is clear that the spouse has never been part of the study.

Often the same event (e.g., a divorce) is reported in more than one successive interview. In such cases, the initial report is chosen as the source for the associated date of the event (e.g., month and year of a divorce) unless the date was not ascertained. In that event, the succeeding year's reported date was used on the principle that known dates are better than missing dates.

Individuals from whom we have reports of their marriages to each other may disagree on the status. This seeming inconsistency can be legitimate if the timing of the spouses' reports differs. The variable indicating the wave in which the marriage history was most recently updated is MH17. As an example, in 1990 a female sample member marries and her new husband, the Reference Person, moves into the study. In that wave, we receive reports of their mutual marriage and each of them has a record on the Marriage History File registering the other as spouse. They each receive values of 1990 for MH17. In each successive wave their marriage is reconfirmed and values for MH17 are updated until 1994, when they separate and divorce. He, the nonsample husband and former Reference Person, leaves the PSID but the sample Spouse remains response. The record for her marriage to him is updated to indicate the revised status, the dates of separation and divorce are added, and MH17 receives a value of 1994. But his record is not updated; his status remains married, and MH17 retains a value of 1993.

The preparation of the 1985-2021 file involved a great effort to eliminate real inconsistencies. If a couple no longer living together disagreed on their marital status but both were responding at the same time, PSID staff attempted to reconcile the differences. Marriage and birth dates were cross-checked to ensure that marriages do not occur until an individual is at least 13 years old. Eleven marriages remain in which the individual reports a start date before that age. We are unable to resolve these cases. In most of them, marriage at age 11 or 12 is possible. Five of these persons are females from the 1990 Latino sample, and two are from the 1997 immigrant sample. For the latter, subsequent birth date information may clarify matters.

On a related note, an individual's birth date is copied from the current wave of the Individual File when his or her marriages first appear on the cumulative Marriage History File. This birth date is not updated in later years for the initial record(s). If a new marriage subsequently occurs for the individual, then his or her birth date from that later wave is used in the construction of the new record. Thus, birth dates for the same individual may disagree across marriages if the reports were not collected in the same wave. These discrepancies were cleaned to some extent but only as a by-product of other cleaning activities.

Attempting birth date consistency is made difficult because of individuals' varying years of participation in the study. In the longer term, we hope to clean these dates but, for the present, we advise

analysts to use the individual's birth date or age variables from the most recent year of Individual File data for which he or she is present in the study.

We checked the internal consistency of marriage dates: termination dates must not precede marriage dates and an earlier marriage must end by the time a later marriage begins. All cases in which divorce dates preceded separation dates were checked for coding/data entry errors and against corroborating sources (e.g., if the spouse had ever been in the study, his or her date of move out was compared to the separation date). In 153 cases, indeed the couples did not separate until after their divorces were finalized. In 27 cases, we were unable to resolve the final status of a person's earlier marriage. These individuals appear to be bigamists, and probably are. The earlier marriage records have values of 7 for the status variable (MH12).

Yearly fluctuations in status were reconciled as information from each successive wave was incorporated. For example, divorcees have been known to report themselves as widowed after their ex-spouses' deaths; their statuses were recoded to 'divorced'. Individuals who separate and then resume living together were recoded as married. The fact of their former separation must, of necessity, disappear from the file. Interruptions of this sort can be found by comparing the couple's data records on the cross-year individual file for co-residence; that is, comparing their yearly family interview numbers and sequence numbers (e.g., for 1990, V30642/ER30642 and V30643/ER30643). And every attempt was made to reconcile differing reports of status between couples if they were present in the same wave.

Spouse identifiers were checked against cross-year individual data and against the 1985-2021 Childbirth and Adoption History File to ensure that no spurious cases of intergenerational incest occur. If both spouses had been in the study, their sexes from the cross-year individual file were checked against each other. Our final checks assured that all individuals who had ever qualified for marriage history questions had records on the file and that individuals who had never qualified did not.

8.11 Who Has Cross-Year Information? TOC

Please keep in mind another PSID intricacy when matching across files: while all individuals were present in a PSID family, many spouses identified in the marriage histories have never been present in a PSID family unit during the years the study has been in progress; these spouses have values for "Person Number of Spouse" (MH8) in the range 800-995. Consequently, each individual has been in a PSID family unit and has a record on the 1968-2021 Individual File but his or her spouse may or may not.

8.12 What Cross-Year File to Use for Merging TOC

The 1985-2021 Marriage History File matches the 1968-2021 Individual File exactly. If you attempt to merge the 1985-2021 Marriage History File with earlier individual data releases, some cases on the merged file will not match and may result in the assignment of some erroneous spousal relationships. See the next section for details.

8.13 Marriage Order TOC

As described above, in the initial wave of demographic event history collection in 1985, all Reference Persons, Spouses and Partners were asked to provide details about all of their marriages. But in subsequent waves, the retrospective marriage history questions for new Reference Persons, Spouses and Partners permitted detailed information about only two marriages, the first and the current or most recent. OFUMs were never asked about all their marriages; even in 1985 we requested reports about only the first and current or most recent. Even so, the Marriage History File contains complete histories for most individuals, since more than two marriages is a relatively rare event. The number of individuals reporting more than two marriages is 4,530; 3,101 of them have reported all their marriages, but 1,429 have not.

The order of each marriage for an individual is indicated in MH9; values are assigned to each marriage in chronological order. The 1,429 individuals for whom we have received incomplete reports have gaps in the values for marriage order across their records. For example, if a person has been married three times but we have received detailed information about only the first and last marriages, the first marriage is assigned a value of 1 and the last marriage receives a value of 3. No record for marriage number 2 is on the file. These individuals can easily be identified, as values for the number of their marriages (MH18) are greater than values for the number of their records (MH20), of course excluding cases where the number of marriages contains missing data.

The relative order of marriages is always clear because of the way in which the questions are asked and updated. A missing beginning date for a marriage does not cause its order to be unknown. However, 353 individuals have a marriage of unknown order. This has happened in two different circumstances. If the interviewer did not obtain a complete marriage history when the individual entered the study but he or she was known then to be or have been married, we created a record for that current or most recent marriage with the information available. This situation accounts for the overwhelming majority of the cases (342). These persons have a known spouse at MH7-MH8, a known marital status at MH12, but the number of their marriages is not ascertained (MH18=98) and the number of records for them (MH20) equals 1.

The second, much rarer, circumstance in which a marriage is known to have taken place but its order is not ascertained occurs when the total number of marriages is not ascertained but the person reports a first and a last. Only eleven individuals on the file fit this profile. Their first marriage receives an order value of 1, of course, but the last must of necessity have a value of 98. At least through the 2021 wave, however, no person has more than one marriage with order unknown.

8.14 Linking Records TOC

8.14.1 Using the Marriage History File with the Cross-Year Individual File

The 1985-2021 Marriage History File is designed to be linked to the 1968-2021 PSID Individual data for analysis purposes. The Marriage History File has only a modest amount of information about the

individual and his or her spouse(s). The analyst will no doubt want to access the much greater volume of data available for these individuals on the Family and Individual Files. Those files can provide abundant information for many individuals dating back to 1968 when the PSID began, although, for some, the span of available data is more limited or nonexistent. Because of the corrections, do not attempt to match this file with any other PSID data file or a subset derived there from using these unique individual identifiers. Non-matches in individual records definitely will occur if any other file besides the above-named is used.

Data processing is required to link records between these two files. To achieve linkages, one must match on the unique individual-specific identifier. This unique identifier is a combination of two variables: "1968 Family Interview Number" (MH2) and "Person Number" (MH3). The corresponding variables for this unique identifier on the cross-year Individual File are ER30001 and ER30002. Care must be taken with regard to the proper files to use, the choice of individuals on the Marriage History File for whom matches are attempted, and the years for which data are available.

For those persons with records on both the Individual File and the Marriage History File, linkages rely on a match of individual-specific identifier variables that appear on both files. As noted earlier, the unique identifier involves two variables: "1968 Family Interview Number" and "Person Number". The corresponding set of variables on the two files must match to properly link an individual's records from the two sources. The variable names for these two variables are ER30001 and ER30002 on the Individual File. On the Marriage History File, they are MH2 and MH3, respectively, for the person designated "individual" and MH7 and MH8 for the person designated "spouse". Note that such linkages involve a one-to-many-match. One record on the Individual File may have more than one matching record on the Marriage History File because the specified individual has multiple marriages.

8.14.2 Using the Marriage History File with Other Demographic History Files

Some analysts may be interested in linking information from different records on the Marriage History File or linking information from records on different demographic files. For example, access to both marriage and childbirth records for an individual are needed to determine, via comparisons of marriage and childbirth dates, the number of biological children an individual has when he or she remarries. To make links such as these, one must match on the unique individual-specific identifier, which is a combination of two variables -- "1968 Family Interview Number" and “Person Number" (MH2 and MH3 for the person designated as the individual on the 1985-2021 Marriage History File; CAH3 and CAH4 for the person designated as the parent on the 1985-2021 Childbirth and Adoption History File)

8.15 Marriage Information Available On Individual Files TOC

Some of the information provided on Marriage History Files is also available on the cross-year Individual File. In addition, the Individual File contains some detail relating to marriage issues that the Marriage History File does not.

The following listing shows all of the marriage history-related variables included in Public Release

versions of the cross-year Individual File. All are individual-level variables.

ER32033	Year Marital Info Most Recently Updated
ER32034	# Marriages of This Individual
ER32035 ER32036	Month and Year First/Only Marriage Began
ER32037	Status of First/Only Marriage
ER32038 ER32039	Month and Year First/Only Marriage Ended
ER32040 ER32041	Month and Year Separated First/Only Marriage
ER32042 ER32043	Month and Year Most Recent Marriage Began
ER32044	Status of Most Recent Marriage
ER32045 ER32046	Month and Year Most Recent Marriage Ended
ER32047 ER32048	Month and Year Separated Most Recent Marriage
ER32049	Last Known Marital Status

These variables are compiled from marriage history information collected from the 1985 wave through the most current wave of cross-year individual information included on the file.

8.16 Codebook

TOC

The item-by-item descriptions for all of the variables in the Marriage History File can be found in the codebook. Unweighted frequencies were calculated for each variable.

9. PARENT IDENTIFICATION FILE (PID) TOC

The Parent Identification File synopsizes information collected about parent-child relationships from various sources since the 1983 wave of the PSID. This file consists of identifier variables that link children with their parents.

The data records are short. They contain relevant identifiers for the child, his or her birth and adoptive parents, and information source indicators. The file is intended to be used to facilitate linking children's and parents' data records from the 1968-2021 Individual File. Linkages can be done from either the child's or a parent's standpoint.

9.1 Sources of Parental Identifier Information TOC

Parent-child information has been collected in many different ways since the PSID began; this file exploits most of those sources.

For the 1983 and 1984 waves, interviewers were asked to indicate the names of birth mothers on the family listings for each person then associated with a responding family unit. The information was checked by PSID staff, and the mothers' identifiers were coded during family composition editing.

Beginning in 1985, and continuing through the present, retrospective childbirth and adoption histories have been asked for many individuals. See the 1968-2021 Childbirth and Adoption History File documentation for details about qualifying persons.

In the 1988 wave, a supplement about time and money help given to and received by the family was added to the usual questionnaire schedule. Part of this supplement included collecting the names of the Reference Person's and Spouse's/Partner's parents, regardless of whether any transfer of help had occurred. Some of the parents were or had been PSID family members, but others had never been part of the study. During family composition editing, individual identifiers were assigned to each of the parents. If a parent had ever been in the study, then his or her identifiers were coded; a "new" parent was given his or her own unique values. These unique identifiers for all parents of the 1988 Reference Persons and Spouse/Partner were never released by the PSID, although the 1988 Family File includes a lot of information about those parents and their assets and a concurrent linking identifier to the parents' family data if they were response in 1988.

In a related piece of the 1988 supplement, individual identifiers were coded for other people outside the family unit who had given or received help. Some of these individuals were children of the Reference Person or Spouse/Partner. Data for all givers and receivers, including records and identifiers for parents and children, were released as the 1988 Time and Money Transfers File.

Interviewers' thumbnail sketches and marginal notes can be precious sources of incidental information about family relationships. They have been remarkably useful to PSID staff during family composition editing for unusual cases in which youngsters, e.g., grandchildren, appear in a family unit but for whom no parental acknowledgement exists on the Childbirth and Adoption History File. Since PSID

samplehood depends on ancestral antecedents (staff have termed this "carrying the sample gene"), knowledge of parentage is crucial in determining sample membership and followability. Beginning in 1996, staff have been coding parental identifiers for some individuals whose progenitors were not established through the birth history reports but were named in interviewer notes.

In a procedure related to the PSID's Child Development Supplement (CDS) in 1997, and continued for 1999, interviewers were instructed to indicate mother and/or father for each child in the family unit if the parent was included in the same family unit. This information was used to determine whether a CDS absent parent interview was called for and, additionally, to invoke a set of questions about child support if one or both parents were not part of the same family unit as the child. In 1999, these questions were not used for additional data such as child support or the CDS, but they provided some verification of parenthood and are being continued in anticipation of future waves of CDS.

The 1997 child support sequences identified the person in the family who received support for a co-resident child, usually a parent. Additionally, they included asking whether any family member was responsible for children who were not currently co-residents in the family unit. If such a family member existed, then he or she was identified and the names of the children were collected. During 1997 family composition editing, individual identifiers were assigned to the children.

In the 2013 wave, a Family Rosters and Transfers Module was added to the PSID questionnaire. This module asked most respondents to provide the names of living parents and adult children for the Reference Person and Spouse/Partner. During data editing, individual identifiers were assigned to “new” parents and children.

In the 2017 Wave, we added four additional variables to allow for a second set of adoptive parent identifiers of the same sex.

9.2 How to Obtain the File TOC

The 2021 Parent Identification is available in the Data Center as well as in .zip format on the Packaged Data page of the PSID website.

9.3 Structure of the File TOC

The 2021 Parent Identification File contains a total of 102,074 records. Included are all individuals from the 1968-2021 Individual File and, additionally, records for known children from the Childbirth and Adoption File and the 1988 Time and Money Transfers File. Children in this latter group have never been included in a PSID family.

The file is sorted, in ascending order, by "1968 Interview Number of Individual" (PID2) and "Person Number of Individual" (PID3). These two variables, taken together, constitute a unique identifier or each person and record. The file contains 40 variables. Besides the pair of identifier variables for the child, six more sets of parental identifiers are present, one set each for birth and two sets for adoptive mothers and fathers (we may have 2 adoptive mothers or 2 adoptive fathers reported). Variables indicating

the source of the parental information are also included. Please see the codebook for the full listing of these variables, along with the code frame, and frequencies in the sample.

9.4 Idiosyncrasies, Data Cleaning, And Variable Detail TOC

Some people whose existence has been reported by PSID respondents do not have a record on this file. These "missing persons" are forever-absent spouses who have never been named as children by anyone in the study and some other individuals (e.g., siblings, nephews or nieces, grandchildren) listed in the 1988 Time and Money Transfers File.

Of the 102,074 records on the Parent Identification File, approximately two-thirds of the records contain identifiers for at least one natural or adoptive parent. Some of the remaining individuals, those with no identified parent, will acquire known parents in future waves.

Parent and child identifiers from the various sources were checked against each other for inconsistent parent reports. In addition, because the parent identifiers are sex-specific, they were checked against the Individual File's “Sex of Individual” variable, ER32000. However, we allow for changes in reporting of sex on the Childbirth and Adoption History File and the PID versus the latest report of sex ER32000 from the Individual File.

Parental reports from the Childbirth and Adoption History File, parent coding by PSID staff in 1983-1984, and, anecdotally, from 1996 onward differentiate between birth and adoptive parents, but none of the other sources specify whether a reported parent is biological or adoptive. For the purpose of file creation, all parents were assumed to be birth parents unless contradicted by one of the differentiated sources.

The 2021 Parent Identification File matches the 1968-2021 Public Release Individual File exactly. If you attempt to merge the 2021 PID file with an earlier release, some cases on the Individual File will not match and may ascribe erroneous parent information to a person. See the next section for details.

9.5 Linking Records TOC

The Parent Identification File is designed to be linked to PSID Individual data for analysis purposes. The Parent Identification File consists only of identifiers for child and parent, plus the dummy variables indicating sources of reports. The analyst most definitely must access the substantive data available for many of these individuals on the Family and Individual Files. Those files can provide abundant information for many individuals dating back to 1968 when the PSID began but, for others, the span of available data is more limited or nonexistent.

When matching the 2021 Parent Identification File to individual data, only the 1968-2021 Individual File should be used. During file merging and cleaning, a number of unique individual identifiers were corrected. Special care was taken to ensure perfect correspondence in individual identifiers between this file, the 1985-2021 Childbirth and Adoption History File, and the 1968-2021

Individual File.

Because of the corrections, do not attempt to match this file with any other PSID data file or a subset derived there from using these identifiers. Non-matches in individual records definitely will occur if any other file besides the above-named is used.

Children and parents vary substantially in terms of which years they have been present in PSID family units over the course of the study. Even though people have records on the 1968-2021 Individual File, that file does not necessarily contain data, for them, for all years of the study. For example, some first entered the study in 1990 and thus, only have data since that year while others have logged more than 30 years of inclusion in the study.

This affects the availability of data for them on the main files because valid information is available on those files only in the years that individuals are present in a PSID family unit. For more details about PSID tracking procedures and classification of people into family units, see the discussion in Section 8 on "Family Composition and Change", p. 55, in A PANEL STUDY OF INCOME DYNAMICS: A USER'S GUIDE, by Martha Hill (Sage Publications, 1992).

If a person, either child or parent, has a record on the 1968-2021 Individual File but less than the full 42-wave range of data in that record, variables in the waves when he or she was not present in a PSID family unit mostly contain zero values and the individual-level variable "Type of Individual Data Record" in that year indicates that he or she is nonresponse. As noted among the cautions in Section III, many children and parents do not have records on that file since, indeed, they have never been in a PSID family unit at any time during the course of the study. Parents or children who have been present in a PSID family at some time since the study began have Person Number values in the range 001-399. Those who have never been present have values of 800-995.

Data processing is required to link records between these two files. For persons with records on both the Parent Identification File and the Individual File, linkages rely on a match of individual-specific identifier variables that appear on both files. As noted earlier, the unique identifier involves two variables: "1968 Family Interview Number" and "Person Number". The corresponding set of variables on the two files must match for proper linkage to an individual's records from the two sources. The variable names for these two variables are ER30001 and ER30002 on the Individual File. On the Parent Identification File, they are PID2 and PID3, respectively, for the child and PID4 and PID5, PID6 and PID7, PID18 and PID19, or PID20 and PID21 for a parent.

9.6 Codebook TOC

The item-by-item descriptions for all of the variables in the Parent Identification File can be found in the codebook. Unweighted frequencies were calculated for each variable.

10. SAMPLE WEIGHTS

TOC

To account for differential probabilities of selection due to the original PSID sample design and subsequent attrition, the PSID data are provided with weights. The PSID's dynamic sample design and following rules are the building blocks for the strategy used in weight construction, the assignment of weights, and the use of weights in different types of analysis. The following rules are important for understanding how the weights are constructed, and how weights should be used in different types of analysis.

For the main interview in 2021, PSID creates longitudinal individual weights, longitudinal family weights, and cross-sectional individual weights. Further documentation describing the construction of the PSID weights are available on the questionnaires and supporting documentation page of the PSID website, as well as an overview in the video tutorial 'PSID Sample Weights.'

11. SUPPLEMENTAL STUDIES TOC

In addition to the main PSID interview, numerous supplemental studies (outside the main interview) have been conducted throughout the years. This section provides an overview of ongoing supplemental studies and point users to relevant documentation.

11.1 Child Development Supplement TOC

The Child Development Supplement (CDS) is a research component of the Panel Study of Income Dynamics. The CDS provides researchers with extensive data on children and their extended families with which to study the dynamic process of early human and social capital formation. The Original CDS included up to two children per household who were 0 to 12 years old in 1997 and followed those children over three waves, ending in 2007-08. In 2014, CDS was redesigned as a steady state panel study and interviews all eligible PSID youth aged 0-17 every 5-6 years. The study design and questionnaire content are consistent with earlier waves of CDS to permit cross-cohort analysis of children's development.

Data collection for CDS-2019 was interrupted in March 2020 when the COVID-19 pandemic ended the in-home visit component of the interview for about two-thirds of the sample. A Fall-2020 follow-up effort—called CDS-2020—was undertaken to cover the cases from CDS-2019 that had not received the home interview. A subsequent round of CDS was undertaken in 2021 for those families who had participated in CDS-2019. CDS-2021 data are available in an early release file. The latest wave of collection for CDS-2023 will be collected starting later in 2023.

11.2 Transition into Adulthood Supplement TOC

The Transition into Adulthood Supplement (TAS), initiated in 2005 and collected biennially, captures data on the developmental pathways and outcomes of children who participated in the 1997 CDS through 2015, and young adults in all PSID families from 2017 forward as they transition into adulthood. The TAS fills a gap between information collected in the CDS and information on adulthood collected from panel members who enter the main PSID study once they have assumed economic independence as Reference Persons and spouse/partners. TAS-2021 is available in an early release file. TAS is planned to continue biennially and the next round of TAS-2023 will be collected later in 2023.

The TAS interview domains are coordinated with the CDS adolescent measures and the PSID employment and health measures, and also include measures unique to this transitional stage. Together with 50 years of longitudinal data in the PSID, the CDS and TAS support a wide range of research on the ways in which time, money, social resources, parenting practices, and characteristics of caregivers and family members are linked to the cognitive and behavioral development of children, adolescents, and young adults. All CDS and TAS documentation and data are available on the PSID website.

12. DATA DISTRIBUTION TOC

The long period over which data have been collected, the extensive range of measures captured in the instrument, and the genealogical design make the PSID a valuable data source. At the same time, these qualities combine to make the PSID increasingly complex for users. Therefore, tools have been developed to allow users to take full advantage of the many aspects of the unique data. In this section we describe some of these tools.

12.1 Internet-Based Data Center TOC

The PSID began distributing data through an online Data Center in 1996. The Data Center allows users to create customized longitudinal datasets from all waves of the main interview as well as some supplemental data collections by choosing various options, creating customized codebooks specific to the data that has been downloaded, searching and browsing for variables, adding multiple variables simultaneously via the variable list function, and archiving data downloads for shared and future use.

Users add variables to their data carts, and when they are ready to download their data, or "check out," they first view the contents of their data cart. Users also see an information icon next to each variable, and pressing on this icon takes them to a window that contains the full codebook documentation for that variable. Users can then choose to edit their cart by removing any unwanted variables, or they can add variables by returning to the "data aisle" for more items. They may also choose to completely empty their data cart, or to proceed to download their data, i.e., "check out."

Data carts may also be saved, named, and made public to other data users, allowing users to easily track specific data downloads. Users can choose from a range of output types including SAS, SPSS, STATA, dBase, and ASCII. Moreover, users can specify the data subset in a selection phase. For more information, see the video tutorial 'Accessing and Downloading PSID Data.'

12.2 Online Cross-Year Variable Index TOC

In the cross-year index, users view a given domain of variables – income, health, or wealth, for example – and then "unfold" this category to see all of the variables related to that domain. For a given specific variable -- for example, current employment status -- the index will list the years that the variable is available. The user can then click on the year to view the codebook for that specific year, and they can click on the year indicator for the given variable to add that variable for that year to their data cart. In sum, the cross-year index integrated with the Data Center allows users the option of "browsing" the entire PSID archive sorted by variable domains.

12.3 Family Identification Mapping System TOC

Because of its genealogical design, the PSID is one of the few nationally representative US datasets that can be used for intra- and intergenerational analyses. As described above, from its beginning the PSID has followed all 1968 family members and their descendants. When family members splitoff and

create their own separate family, the PSID interviews these new families as well as the original family. The numbers of sibling pairs, child-parent pairs, and grandparent-grandchild pairs are substantial. For example, tens of thousands of sibling pairs exist in the data archive. Of course, not all siblings are alive or reporting data in each and every wave of the PSID, therefore the number of siblings available for any particular analysis will be smaller. But in the more recent waves, there are thousands of sibling pairs who are Reference Persons or Spouses/Partners, which means that the full set of PSID data is collected on these siblings. These large samples support a wide range of analyses, but the creation of these files is complex, and can be prohibitively so for some users.

With the goal of facilitating the use of these data to support complex models of family and life course development, the PSID offers a Family Identification Mapping System (FIMS). FIMS creates a customized file – i.e., "map file" - that contains the identification variables of the relatives an analyst wishes to examine. FIMS also supplies code (for SAS, SPSS, and Stata) that uses the map file and a file containing the variables the analyst wishes to examine to create a new data file that includes the variables for the relatives of interest. For example, an analyst may be interested in "biological grandparents" as the relative of interest. FIMS would then generate a file that contains the IDs of all PSID sample members and each of their four biological grandparents if they were ever observed in the PSID. FIMS provides code that will create a data file in the shape desired by the analyst. The shapes available are "wide" – i.e., one observation per grandchild – and "long" – i.e., one observation per grandchild-grandparent pair. The user would then create an individual level data file that contains all of the variables they want and merge it with the FIMS file to match them to grandchildren and their grandparents.

FIMS offers three distinct types of maps. The intra-generational (SIB) map allows the identification of various types of siblings (full siblings, half siblings). The inter-generational (GID) map matches PSID individuals to their predecessors, going back up to three generations, i.e. parents, grandparents, and great-grandparents. This intergenerational map is thus retrospective in nature. That is, it starts with an individual and goes back along in his or her family lineage. In 2014, a prospective inter-generational map (GID PRO) was added to FIMS. Here, the starting generation (G1) is the original sample from 1968 (person numbers between 1 and 19). Descendants of original PSID households form subsequent generations, again up to three generations down (G2 - child, G3 – grandchild, G4 – great- grandchild). The prospective generation map format is long, i.e., each row is a distinct set of individuals observed in the PSID. The map also provides biological relation (father or mother) between individuals and generation position of all individuals listed on each row. For more information, see the video tutorial 'The Family Identification Mapping System' or the FIMS User Guide.

12.4 Video Tutorials TOC

Online video tutorials have been created to help users learn about the PSID. T are available on the PSID website, as well as on the PSID's YouTube Channel, with closed captioning available.

12.5 Cross National Equivalent File TOC

The Cross-National Equivalent File contains equivalently defined variables for the PSID and PSID- like studies in several other countries. The data and a description of this project, which is led by researchers at Cornell, is currently being housed at Ohio State University.

12.6 Tax Information TOC

For estimates using TAXSIM from 1999-2015 see Kimberlin, Kim, & Shaefer (2015) who provide an updated method for calculating income and payroll taxes from PSID data 1999-2015. The PSID website also includes Stata programs describing a method to calculate income and payroll taxes using TAXSIM. Data and programs for 2013 and 2015 are forthcoming.

12.7 Restricted Data TOC

In order to safeguard the confidentiality of respondents at the highest level, some data are provided only under conditions of a restricted use contract between the researcher and the University of Michigan. A description of the available data, the documentation, the procedures for obtaining the data, and the requirements for those who gain access to such data is provided on the restricted data section of the PSID website, as well as the video tutorial 'Restricted Data in the PSID.'

13. DATA QUALITY TOC

The U-M team regularly monitors PSID data quality and assesses the representativeness of the sample. In particular, the weighted, cross-sectional data from each wave are compared to five external data sources on income, wealth, expenditures, time use, and health (with results shared on the PSID website on the Data Comparison page, and all published papers archived in the PSID online bibliography). For income, comparisons between PSID and the Annual Social and Economic Supplement of the Current Population Survey (CPS) suggest similar distributions (with PSID estimates of mean, median and various percentiles a few percentage points higher than CPS). Comparisons of wealth in PSID with the Survey of Consumer Finances (SCF) show that the SCF captures more wealth than PSID, but there is high concordance for all but the top 2 percent of the wealth distribution and across most asset components (Pfeffer et al. 2016), and including 401k plans (Cooper & Dynan 2019). Recent analyses of the PSID consumption expenditure data show close alignment with those reported in the Consumer Expenditure (CE) Survey (PSID 2020). PSID data show strong alignment for time spent on various activities with estimates from the American Time Use Survey (Insolera et al. 2019) and very similar rates of specific health conditions and health behaviors compared to the National Health Interview Survey (Insolera & Freedman 2017). Taken together, these findings indicate that the sample, with weights, remains cross-sectionally representative of the U.S. population.

In any long-term panel, cumulative attrition is a potential concern because it can interfere with panel representativeness. Heeringa et al. (2018) examine cumulative attrition and find that attrition has been higher for certain groups, such as males and low-income individuals; each wave's nonresponse weights and regular post-stratification adjustments serve to ameliorate this problem. Other investigations suggest that attrition is random within families (Wiemers et al. 2010), does not substantially bias estimates of intergenerational correlations in income (Schoeni & Wiemers 2015) and does not consistently bias sibling models (Fitzgerald 2011).

Item nonresponse in PSID is historically low compared to other surveys (Moffitt & Zhang 2020). Regular analyses by PSID staff undertaken to examine item nonresponse in each wave shows that in 2017, very few key questions had rates of missing values of more than 3% - 4%. An updated item nonresponse table can be found in Appendix B.

13.1 Getting Help TOC

If you have questions about the PSID that are not answered in the user's manual, the first place to check is the list of frequently asked questions. If you cannot find the answer to your question after reviewing the documentation and FAQs, contact us via the PSID Help Desk at PSIDhelp@umich.edu.

14. FUNDING AND ADMINISTRATION TOC

The PSID has been funded from a variety of sources through the years. Over the past decade, the National Science Foundation (NSF), the National Institute on Aging (NIA), and the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development (NICHD) have been the primary sponsors. During the current funding cycle 2020-2024, support has also been received from the following agencies: the Indiana University Lilly Family School of Philanthropy and the Economic Research Service of the United States Department of Agriculture.

Since 1982, the study has been advised by a Board of Overseers, created by NSF to foster input from the national community of scholars, researchers, and policymakers. The members of the Board are listed on the PSID website.

From its beginning in 1968 until 1989, the PSID was founded and directed at the Survey Research Center, University of Michigan by James Morgan. Frank Stafford was the Project Manager beginning in 1968. From 1982-1989 responsibility for running the study was also shared by Greg Duncan (as co-Director), Daniel Hill, and Martha Hill. Between 1989 and 1995 Greg Duncan directed the study, with Martha Hill and James Lepkowski as co-Directors. Frank Stafford became Director of the study in 1995, with Sandra Hofferth as co-Director until 2001, and Wei-Jean Yeung as co-investigator. Between 2001 and 2010 responsibility for running the study was additionally shared by Robert F. Schoeni and Katherine McGonagle. Jacqueline Eccles and Robert Wallace were co-investigators starting in 2007.

Narayan Sastry and Vicki Freedman joined the team as co-Principal investigators in 2008 and 2010, respectively. During 2010 – 2011, Robert Schoeni and Charles Brown were co-Directors, Katherine McGonagle was Assistant Director, and the co-investigators included Vicki Freedman, Narayan Sastry, and Frank Stafford. From 2012-2016, Charles Brown was Director, Narayan Sastry and Vicki Freedman were Associate Directors, and Katherine McGonagle was Assistant Director. Co-investigators included Robert Schoeni, Frank Stafford, and Fabian Pfeffer. From 2017-2022, David S. Johnson was the Director of PSID, with Associate Directors Katherine McGonagle, and Narayan Sastry. Charles Brown, Paula Fomby, Fabian Pfeffer, and Robert Schoeni were Co-Investigators. Frank Stafford was Professor Emeritus starting in 2019. Currently, Katherine McGonagle and Narayan Sastry are Co-Directors of PSID and Esther Friedman is an Associate Director. Charles Brown, Noura Insolera, Fabian Pfeffer and Wei Zhao are Co-Investigators, with Paula Fomby as an external affiliate of PSID. Current PSID team leaders included Flannery Campbell of the Data Processing Team, Rose McAloon-Fernando of the Instrument Development & Production Management Team, Noura Insolera of the Education, Outreach, and Data Promotion Team, and Mohammad Mushtaq of the Applications Development Team.

15. INDICATORS OF SCIENTIFIC IMPACT TOC

In this section we report on several indicators of the usefulness of the data, including: number of published articles using the PSID, grants awarded by NIH and NSF to support research using the PSID, registered users, hits to the PSID website, and data downloads.

15.1 Peer-Reviewed Publications Using the PSID TOC

As of April 2023, PSID staff have located 7,266 peer-reviewed publications based on PSID data. This total includes 5,135 journal articles, 886 books and book chapters, and 1,245 dissertations.

Articles based on the PSID appear in top journals. PSID staff has ranked academic journals based on numbers of PSID manuscripts ever published, and the top 10 in rank order beginning with the top are: American Economic Review, The Review of Economics and Statistics, The Journal of Human Resources, Demography, Journal of Labor Economics, Journal of Marriage and Family, Journal of Monetary Economics, Social Science Research, Journal of Political Economy, Review of Economic Dynamics. Articles have appeared in many journals from a variety of scientific disciplines, including economics, sociology, demography, public health, medicine, child development, geography, and psychology.

15.2 Grants Awarded By NSF and NIH Using the PSID TOC

Although the vast majority of social science research in the U.S. is conducted without grant support, a substantial share is. Another indication of the value of PSID data is its use in grant supported research. As of April 1, 2023, there have been at least 425 awards made by NSF and NIH to support the collection and secondary analysis of PSID data, with NSF making about 40.7% and NIH making about 59.3% of all awards. The vast majority of all awards made have supported secondary data analysis. These totals are likely to be an underestimate of the total number of awards made by these agencies as the database searches abstracts only; thus an award that did not use ‘PSID’ or ‘Panel Study of Income Dynamics’ in its abstract could not be identified.

15.3 Website Activity, Data Downloads, and Numbers of Users TOC

Information on general website activity is monitored by PSID with summary statistics for each year reported. In calendar year 2022, there were over 4.8 million total page hits to the PSID website made by 137,320 unique visitors (i.e., IP addresses). User activity specific to the actual downloading of data is also assessed. There are two ways to download data, both through the PSID Data Center. The first way is to create customized datasets directly from the Data Center by selecting various types and years of data and variables. The second way is to download complete data files that are compressed in zip packages. In total, there were 33,757 data downloads during this period. Across both types of downloading, these datasets were created by 2,540 unique registered users.

Effective September 1, 2006, individuals wishing to download PSID data are required to provide basic information including their email address, name of institution or organization, affiliation (academic,

governmental, private, other), and scientific field or discipline. In calendar year 2022, there were nearly 40,000 registered users, which is an increase of over 2,500 registered users (6%) over the past year: about 68% have identified their major field as economics, 8% sociology, with the remainder distributed across education, psychology, demography, child development, medicine, geography and "other", with the remainder distributed across education, psychology, demography, child development, medicine, geography and "other."

16. REFERENCES

TOC

Belli, Robert F., William Shay, Frank P. Stafford. 2001. Event History Calendar and Question List Survey Interviewing Methods: A Direct Comparison. *Public Opinion Quarterly* 65.

Belli, Robert F, Eun Ha Lee, Frank P. Stafford, and Chia-Hung Chou. 2004. Calendar and Question-List Survey Methods: Association between Interview Behaviors and Data Quality. *Journal of Official Statistics* 20 (2).

Belli, Robert F. 2003. The Integration of a Computer Assisted Interviewing Event History Calendar in the Panel Study of Income Dynamics. PSID Technical Series Paper #03-01.

Butrica, Barbara A., and Richard V. Burkhauser. 1997. Estimating Federal Income Tax Burdens for Panel Study of Income Dynamics (PSID) Families Using the National Bureau of Economic Research TAXSIM Model. Aging Studies Program Paper No. 12. Maxwell School. Syracuse University.

Ferber, Robert. 1959. Collecting Financial Data by Consumer Panel Techniques, Urbana, Bureau of Economic and Business Research, University of Illinois.

Groves, Robert M., and Katherine McGonagle. 2001. A Theory Guided Interviewer Training Protocol regarding Survey Participation. *Journal of Official Statistics* 17: 249-265.

Hill, Martha. 1991. The Panel Study of Income Dynamics: A User's Guide Sage Press.

McGonagle, K.A., Schoeni, R.F., Sastry, N., and Freedman, V.A. 2012. The Panel Study of Income Dynamics: Overview, Recent Innovations, and Potential for Life Course Research. *Longitudinal and Life Course Studies* 3(2): 268-284. PMCID: PMC3591471

McGonagle, K.A., Schoeni, R.F., and Couper, M.P. 2013. The Effects of a Between-Wave Incentive Experiment on Contact Update and Production Outcomes. *Journal of Official Statistics* 29 (2), 1- 17.

McGonagle, K.A., Couper, M.P., and Schoeni, R.F. 2011. Keeping Track of Panel Members: An Experimental Test of a Between-Wave Contact Strategy. *Journal of Official Statistics* 27(2): 319-338. PMCID: PMC3253355.

Schoeni, Robert F., Stafford, Frank, McGonagle, Katherine A. and Andreski, Patricia. 2013. Response Rates in National Panel Surveys. *The Annals of the American Academy of Political and Social Science* 645 (1): 60-87.

Appendix A. Tables and Figures Describing Income and Wage Imputation

A 2011 technical paper describes in detail the calculation of income and wages for the 2007 wave, including the imputation procedures used. The technical paper also provides a series of tables describing that process. In this appendix the same tables are reported, but based on the 2021 wave of data.

Table A1.1: Components of Reference Person and Spouse Taxable Income

Reference Person Wage and Salary Income	Spouse Wage and Salary Income	Reference Person Income From Assets	Spouse Income From Assets	Net Profit from Farm or Business
Reference Person Wages ER81626	Spouse Wages ER81654	Reference Person Interest Income ER81647	Spouse Interest Income ER81675	Reference Person Net Business Income ER81623, ER81624
Reference Person Bonus ER81628	Spouse Bonus ER81656	Reference Person Dividend Income ER81645	Spouse Dividend Income ER81673	Spouse Net Business Income ER81651, ER81652
Reference Person Overtime ER81630	Spouse Overtime ER81658	Reference Person Rental Income ER81643	Spouse Rental Income ER81671	Net Income from Farm ER81621
Reference Person Tips ER81632	Spouse Tips ER81660	Reference Person Trust Funds ER81649	Spouse Trust Funds ER81677	
Reference Person Commissions ER81634	Spouse Commissions ER81662			
Reference Person Professional Practice ER81636	Spouse Professional Practice ER81664			
Reference Person Additional Job Income ER81638	Spouse Additional Job Income ER81666			
Reference Person Miscellaneous Labor Income ER81640	Spouse Miscellaneous Labor Income ER81668			

Table A1.2: Reference Person Wage and Salary Income Imputation Process in the 2021 PSID

Table A1.3: Spouse Wage and Salary Income Imputation Process in the 2021 PSID

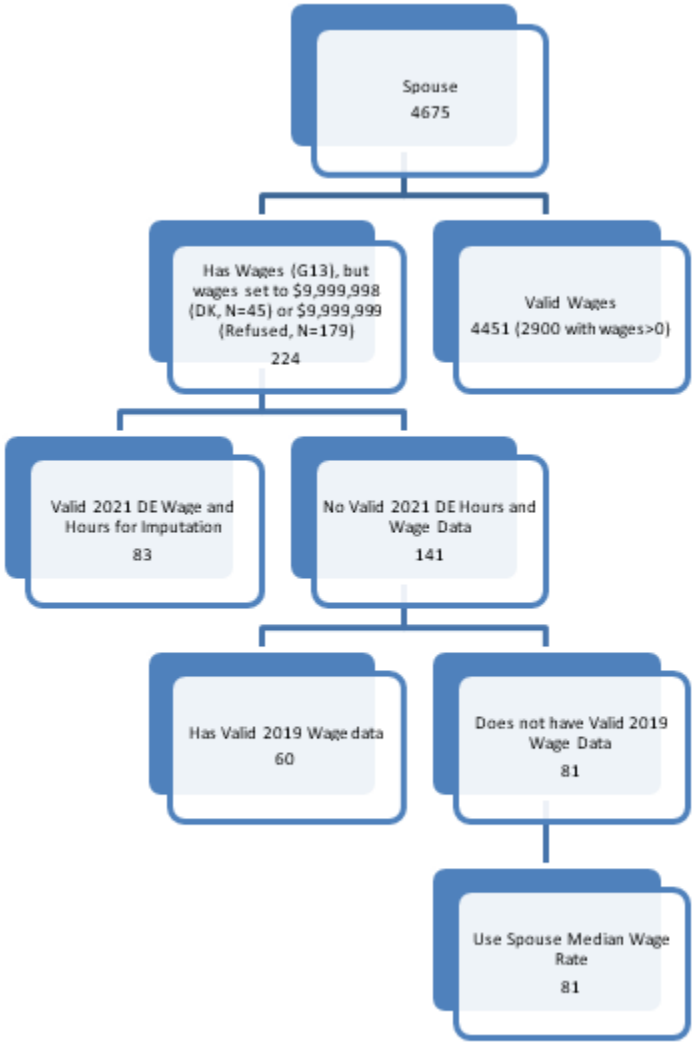


Table A1.4a: 2021 Overtime, Tips and Commission Imputation – Reference Person

Income Source	# Reference Person with non-zero income	Impute Using Individual Jobs Data (Step 1)	Impute Using Average Income as a % of Wages by Occupation (Step 2)	Impute Using Overall Median Income (Step 3)
Overtime	384	29	25	9
Tips	83	1	3	2
Commission	81	2	5	3

Table A1.4b: 2021 Overtime, Tips and Commission Imputation - Spouse

Income Source	# Spouses with non-zero income	Impute Using Individual Jobs Data (Step 1)	Impute Using Average Income as a % of Wages by Occupation (Step 2)	Impute Using Overall Median Income (Step 3)
Overtime	105	8	6	13
Tips	35	1	2	2
Commission	18	1	0	2

Table A1.5a: 2021 Reference Person Bonus Imputation

Reference Person with non-zero Bonus income	592
Reference Person requiring Bonus income imputation	27
Use average bonus % by occupation (Step 1)	21
Use overall median bonus amount (Step 2)	6

Table A1.5b: 2021 Spouse Bonus Imputation

Spouse with non-zero Bonus income	210
Spouse requiring Bonus income imputation	12
Use average bonus % by occupation (Step 1)	7
Use overall median bonus amount (Step 2)	5

Table A1.6a: 2021 Professional Practice Imputation – Reference Person

	Imputation Condition	# Reference Person with >0 income in 2021	Step 1: Use Prior Wave Income for Reference Person	Step 2: Use Median hourly rate. If hours NA, use 500 (done for 0 observations)
Professional Practice (ER81636)	Income outside the range of $0 and $9,999,996, or Don't Know (9,999,998) /Refused (9,999,999)	65	0	5

Table A1.6b: 2021 Professional Practice Imputation - Spouse

	Imputation Condition	# Spouses with >0 income in 2021	Step 1: Use Prior Wave Income for Reference Person	Step 2: Use Median hourly rate. If hours NA, use 500 (done for 0 observations)
Professional Practice (ER81664)	Income outside the range of $0 and $9,999,996, or Don't Know (9,999,998) /Refused (9,999,999)	22	0	2

Table A1.7: 2021 Asset Income Imputation

	# Reference Person/Spouse with Income>0	# Reference Person/Spouse with Imputed Values
Reference Person		
Rent	419	17
Dividends	1045	203
Interest	3210	996
Trust	61	9
Rent	297	11
Dividends	649	115
Interest	1998	592
Trust	24	0

Table A1.8: 2021 Net Business Income Imputation Prevalence by Family Business Ownership Type

Ownership	# Businesses with Non- missing Net Income	# Businesses with Imputation Required
Reference Person	369	75
Spouse	99	15
Reference Person & Spouse	53	3
OFUM Only	21	9
Reference Person & OFUM	4	0
Spouse & OFUM	0	0
Reference Person, Spouse & OFUM	1	0

Table A1.9: 2021 Net Business Income Imputation Methodology

Methodology	Number of Businesses Imputed
Using Self Employment data from Jobs section (Step 1)	27
Use Prior Wave's Net Business Income (Step 2)	5
Hot Deck Methodology (Step 3)	40

Table A1.10: 2021 Net Business Income Hot Deck Imputation Methodology

Cases for Which we Impute	Hot Deck Method	# Cases
Don't Know Loss (-999,998)	Assign Random Negative Income	0
N/A, Refused Loss (-999,999)	Assign Random Negative Income	1
Don't Know Gain (9,999,998)	Assign Random Positive Income	11
N/A, Refused Gain (9,999,999)	Assign Random Positive Income	28

Table A1.11: 2021 Transfer Income Imputation

Transfer Income Source	Who	Number where income amount>0	Number Imputed
Alimony (ER81702)	Reference Person	20	1
Annuity (ER81690)	Reference Person	103	10
Child Support (ER81700)	Reference Person	296	20
Help Non-Relatives (ER81706)	Reference Person	197	28
Help Relatives (ER81704)	Reference Person	733	75
IRA (ER81692)	Reference Person	196	20
Other Pension (ER81694)	Reference Person	17	6
Other Transfer Income (ER81708)	Reference Person	21	1
Retirement (ER81688)	Reference Person	789	49
SSI (ER81682)	Reference Person	244	27
TANF (ER81680)	Reference Person	65	4
Unemployment (ER81696)	Reference Person	1134	80
VA Pension (ER81686)	Reference Person	309	33
Welfare (ER81684)	Reference Person	46	3
Workers Comp (ER81698)	Reference Person	57	3
Alimony (ER81732)	Spouse	3	0
Annuity (ER81720)	Spouse	29	2
Child Support (ER81730)	Spouse	147	7
Help Non Relatives (ER81736)	Spouse	19	4
Help Relatives (ER81734)	Spouse	99	5
IRA (ER81722)	Spouse	62	3
Other Pension (ER81724)	Spouse	6	1
Other Transfer Income (ER81738)	Spouse	17	1
Retirement (ER81718)	Spouse	282	14
SSI (ER81712)	Spouse	47	1
TANF (ER81710)	Spouse	13	2
Unemployment (ER81726)	Spouse	451	24
VA Pension (ER81716)	Spouse	32	2
Welfare (ER81715)	Spouse	15	0
Workers Comp (ER81728)	Spouse	15	1

Table A1.12: 2021 Labor Income Imputation for Other Family Members

	# Jobs with income>0	# Imputations
Job 1	1593	522
Job 2	302	96
Job 3	44	10
Job 4	11	5

Table A1.13: 2021 Transfer Income Imputation for Other Family Members

	# OFUMS with Income Source>0	# Imputations
ADC (ER81748)	2	0
SSI (ER81750)	49	5
Welfare (ER81752)	11	6
VA Pension (ER81754)	9	6
Pension (ER81756)	47	12
Unemployment (ER81758)	37	6
Workers Comp (ER81760)	0	0
Child Support (ER81762)	10	0
Support from Relatives (ER81764)	13	1
Other (ER81766)	19	3

Table A1.14: 2021 Social Security

	Number with Social Security Income> $0	# Imputations
Reference Person (ER81769)	1874	161
Spouse (ER81771)	772	55
OFUM (ER81773)	409	83

Table A1.15: 2021 Weeks and Hours Variables

Variable Name	Variable Description	Variable Group
ER81576	REFERENCE PERSON WORK WEEKS	Weeks
ER81578	REFERENCE PERSON WEEKLY WORK HOURS	Hours Worked
ER81580	REFERENCE PERSON OVERTIME WORK HOURS	Hours Worked
ER81582	REFERENCE PERSON TOTAL HOURS OF WORK	Hours Worked
ER81583	REFERENCE PERSON WEEKS MISSED FOR ILLNESS OF OTRS	Weeks
ER81585	REFERENCE PERSON WEEKS MISSED FOR OWN ILLNESS	Weeks
ER81587	REFERENCE PERSON WEEKS OFF FOR VACATION	Weeks
ER81589	REFERENCE PERSON STRIKE WEEKS	Weeks
ER81591	REFERENCE PERSON WEEKS LAID OFF	Weeks
ER81593	REFERENCE PERSON UNEMPLOYMENT WEEKS	Weeks
ER81595	REFERENCE PERSON WEEKS OUT OF LABOR FORCE	Weeks
ER81597	SPOUSE WORK WEEKS	Weeks
ER81599	SPOUSE WEEKLY WORK HOURS	Hours Worked
ER81601	SPOUSE OVERTIME WORK HOURS	Hours Worked
ER81603	SPOUSE TOTAL HOURS OF WORK	Hours Worked
ER81604	SPOUSE WEEKS MISSED FOR ILLNESS OF OTRS	Weeks
ER81606	SPOUSE WEEKS MISSED FOR OWN ILLNESS	Weeks
ER81608	SPOUSE WEEKS OFF FOR VACATION	Weeks
ER81610	SPOUSE STRIKE WEEKS	Weeks
ER81612	SPOUSE WEEKS LAID OFF	Weeks
ER81614	SPOUSE UNEMPLOYMENT WEEKS	Weeks
ER81616	SPOUSE WEEKS OUT OF LABOR FORCE	Weeks

Table A1.16: Weeks Worked Imputation

	2021
Number of Reference Person/Spouses with Jobs	10066
Number of Reference Person/Spouses with Jobs with Weeks Worked Edits	130

Table A1.17c: 2021 Imputation Values for Time-Off Categories

Variable	Variable Description	Edits Made by Data Processing Staff	Imputations Made Using Constant Substitution	Constant Value Used for Imputation
ER81587	Reference Person Weeks not working due to vacation	104	9	1 Week
ER81585	Reference Person Weeks not working due to illness- self	10	13	0.4 Weeks
ER81583	Reference Person Weeks not working due to illness – other	8	17	0.4 Weeks
ER81589	Reference Person Weeks not working due to strike	0	2	3 Weeks
ER81591	Reference Person Weeks laid off	35	5	2.5 Weeks
ER81608	Spouse Weeks not working due to vacation	44	6	1 Week
ER81606	Spouse Weeks not working due to illness- self	7	4	0.4 Weeks
ER81604	Spouse Weeks not working due to illness – other	5	5	0.4 Weeks
ER81610	Spouse Weeks not working due to strike	0	0	3 Weeks
ER81612	Spouse Weeks laid off	18	2	2.5 Weeks

Table A1.18: 2021 Time Not Working Imputations

	# Reference Persons/Spouses with Weeks Manually Edited	# Reference Persons/Spouses with Weeks Adjusted	# Reference Person/Spouse Non-zero Weeks
Reference Person			
Unemployment	57	0	775
OOLF	58	0	2946
Spouse			
Unemployment	18	0	235
OOLF	20	0	1849

Table A1.19: 2021 Work Hours Imputation Summary

	2021
Number of Reference Persons/Spouses with Jobs	10066
Number of Reference Persons/Spouses with jobs for which Hours Worked has had Pre-imputation Manual Adjustments Applied	347
Number of Reference Persons/Spouses with jobs for which we Impute Using a Value of 40 Hours per Week	7

Table A1.20: 2021 Overtime Hours Imputation Summary

	Manual Edits	Imputed	# of Observations with Non-Zero Overtime
Reference Person Overtime (ER81580)	38	159	1554
Spouse Overtime (ER81601)	7	41	458

Table A1.21c: 2021 Number of Families by Number Income Sources Imputed

Number of Income Sources Imputed	Number of 2021 PSID Families	% of Families
0	6265	68.05
1	1630	17.70
2	885	9.61
3	232	2.52
4	121	1.31
5	48	0.52
6	14	0.15
7	9	0.10
8	1	0.01
9	1	0.01
10	1	0.01

Table A1.22: Income Imputation PSID Codes

Imputation Method	Imputation Code
Data Processing Edit	1
Imputed from Other Information in the Interview	2
Imputed from Last Wave's Report	3
Imputed from Subgroup Means	4
Imputed Using Median Value of all Non-Zero Cases	5
Hotdeck Replacement	6

Table A1.23c: 2021 Income Imputation Summary Table

Source	# Observations with >0 income (including imputed cases)	Number Observations requiring imputations	Methodology	Accuracy Variable
Reference Person Labor Income	6627	1008	Step 1, Use PSID Employment Section BC/DE Wages/Hours/Weeks Worked to impute (268), else Step 2, Use prior year income (158), else overall median wage rate (149)	ER81627
Spouse Labor Income	3124	358	Step 1, Use PSID Employment Section BC/DE Wages/Hours/Weeks Worked to impute (83), else Step 2, Use prior year income (60), else overall median wage rate (81)	ER81655
Reference Person Bonus	592	30	Step 1: Use average bonus percent by OCC code, apply to wages (21), else Step 2, Use overall median bonus percent (6)	ER81629
Reference Person OT	384	73	Step 1, Use BC jobs info (29), else Step 2, avg OT as pct of wages by occ code (25), else median OT amount (9)	ER81631
Reference Person Tips	83	7	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (3), else median tips amount (2)	ER81633
Reference Person Commission	81	11	Step 1, Use BC jobs info (2), else Step 2, avg tips as pct of wages by occ code (5), else median commission amount (3)	ER81635
Reference Person Professional Practice	65	9	Step 1, Use Prior year (0), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (5)	ER81637
Spouse Bonus	210	12	Step 1: Use average bonus percent by OCC code, apply to wages (7), else Step 2, Use overall median bonus percent (5)	ER81657
Spouse OT	105	29	Step 1, Use BC jobs info (8), else Step 2, avg OT as pct of wages by occ code (6), else median OT amount (13)	ER81659
Spouse Tips	35	5	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (2), else median tips amount (2)	ER81661
Spouse Commission	18	3	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (0), else median tips amount (2)	ER81663
Spouse Prof. Practice	22	2	Step 1, Use Prior year (0), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (2)	ER81665

Source	# Observations with >0 income (including imputed cases)	Number Observations requiring imputations	Methodology	Accuracy Variable
Reference Person Farm Income	27	5	Step 1, Farm Receipts - Farm Expenses (0), else Step 2, Prior Year Income (0) else Step 3, overall median farm income (5)	ER81622
Reference Person Business	369	75	Step 1, Use BC/DE Self Employment Income (22), else Step 2, Prior year's income if same industry (2), else Step 3, hot deck within industry (25)	ER81619
Spouse Business	99	15	Step 1, Use BC/DE Self Employment Income (3), else Step 2, Prior year's income if same industry (3), else Step 3, hot deck within industry (7)	ER81619
Reference Person/ Spouse Business	53	3	Step 1, Use BC/DE Self Employment Income (2), else Step 2, Prior year's income if same industry (3), else Step 3, hot deck within industry (0)	ER81619
OFUM Only Business	21	9	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (8)	ER81619
Reference Person & OFUM Business	4	0	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (0)	ER81619
Spouse & OFUM Business	0	0	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (0)	ER81619
Reference Person, Spouse & OFUM Business	1	0	Step 1, Use BC/DE Self Employment Income (0), else Step 2, Prior year's income if same industry (0), else Step 3, hot deck within industry (0)	ER81619
Rent Reference Person	419	32	Overall Median	ER81644
Dividend Reference Person	1045	208	Overall Median	ER81646
Interest Reference Person	3210	1000	Overall Median	ER81648
Trust Reference Person	61	9	Overall Median	ER81650
Rent Spouse	297	17	Overall Median	ER81672
Dividends Spouse	649	115	Overall Median	ER81674
Interest Spouse	1998	592	Overall Median	ER81676

Source	# Observations with >0 income (including imputed cases)	Number Observations requiring imputations	Methodology	Accuracy Variable
Trust Spouse	24	0	Overall Median	ER81678
OFUM Labor Income	1604	543	Overall Median (within Job #)	ER81744
OFUM Interest	29	9	Overall Median	ER81746
OFUM ADC	2	0	Overall Median	ER81749
OFUM SSI	49	5	Overall Median	ER81751
OFUM Welfare	11	6	Overall Median	ER81753
OFUM Veterans	9	3	Overall Median	ER81755
OFUM Pension	47	12	Overall Median	ER81757
OFUM Unemployment	37	6	Overall Median	ER81759
OFUM Workers Comp	0	0	Overall Median	ER81761
OFUM Child Support	10	0	Overall Median	ER81763
OFUM Relatives	13	1	Overall Median	ER81765
OFUM Other	19	4	Overall Median	ER81767
Social Security	3142	312	Overall Median	ER81770, ER81772, ER81774
Reference Person Alimony	20	1	Overall Median	ER81703
Reference Person Annuity	103	11	Overall Median	ER81691
Reference Person Child Support	296	21	Overall Median	ER81701
Reference Person Help Non Rel	197	29	Overall Median	ER81707
Reference Person Help Rel	733	75	Overall Median	ER81705
Reference Person IRA	196	22	Overall Median	ER81693
Reference Person Other	21	1	Overall Median	ER81709
Reference Person Other Retirement	17	7	Overall Median	ER81695
Reference Person Retirement	789	53	Overall Median	ER81689
Reference Person SSI	244	28	Overall Median	ER81683
Reference Person TANF	65	5	Overall Median	ER81681
Reference Person Unemp	1134	84	Overall Median	ER81697

Source	# Obs. with >0 income (including imputed cases)	Number Obs. requiring imputations	Methodology	Accuracy Variable
Reference Person VA Pension	309	36	Overall Median	ER81687
Reference Person Welfare	46	3	Overall Median	ER81685
Reference Person Workers Comp	57	4	Overall Median	ER81699
Spouse Alimony	3	0	Overall Median	ER81733
Spouse Annuity	29	2	Overall Median	ER81721
Spouse Child Support	147	7	Overall Median	ER81731
Spouse Help Non Relative	19	4	Overall Median	ER81737
Spouse Help Relative	99	5	Overall Median	ER81735
Spouse IRA	62	4	Overall Median	ER81723
Spouse Other	17	1	Overall Median	ER81739
Spouse Other Retirement	6	1	Overall Median	ER81725
Spouse Retirement	282	15	Overall Median	ER81719
Spouse SSI	47	1	Overall Median	ER81713
Spouse TANF	13	3	Overall Median	ER81711
Spouse Unemployment	451	28	Overall Median	ER81727
Spouse VA Pension	32	2	Overall Median	ER81717
Spouse Welfare	15	0	Overall Median	ER81715
Spouse Workers Comp	15	2	Overall Median	ER81729

Appendix B. Item Nonresponse Calculations 2009-2021

Selected Item Nonresponse Rates by Year*								
Topic	Who	2009	2011	2013	2015	2017	2019	2021
Health Status								
	Reference Person	0.23%	0.35%	0.38%	0.18%	0.28%	0.29%	0.82%
	Spouse/Partner	0.41%	0.43%	0.50%	0.55%	0.52%	0.64%	1.50%
Hours Worked								
	Reference Person	1.94%	1.48%	1.67%	1.04%	1.07%	1.21%	1.43%
	Spouse/Partner	1.67%	0.96%	0.94%	0.82%	1.05%	0.84%	1.30%
Industry								
	Reference Person	0.45%	0.44%	0.47%	0.31%	0.38%	0.22%	1.93%
	Spouse/Partner	0.56%	0.49%	0.62%	0.51%	0.33%	0.34%	1.87%
Occupation								
	Reference Person	0.44%	0.28%	0.36%	0.26%	0.33%	0.25%	1.67%
	Spouse/Partner	0.71%	0.41%	0.53%	0.54%	0.28%	0.25%	2.16%
Whether Union Job								
	Reference Person	2.62%	2.85%	2.01%	2.60%	2.96%	2.49%	2.71%
	Spouse/Partner	1.55%	1.35%	1.43%	1.82%	2.44%	1.75%	2.12%
Wages								
	Reference Person	8.42%	5.69%	4.91%	5.38%	6.47%	6.79%	9.50%
	Spouse/Partner	7.01%	4.69%	4.80%	5.05%	5.27%	5.18%	7.72%
Housework Hours								
	Reference Person	0.82%	0.89%	0.95%	0.67%	0.58%	0.65%	1.75%
	Spouse/Partner	1.08%	1.24%	1.11%	1.08%	0.86%	0.86%	1.97%
TANF Receipt								
	Reference Person	0.12%	0.15%	0.91%	0.85%	2.34%	2.11%	0.58%
	Spouse/Partner	0.11%	0.22%	5.02%	--	1.15%	1.83%	1.06%
TANF Amount								
	Reference Person	4.86%	0.69%	0.00%	4.26%	5.56%	4.92%	6.56%
	Spouse/Partner	11.54%	0.00%	0.00%	12.50%	0.00%	25.00%	18.18%
Dwelling Unit Type	Family Level	0.06%	0.10%	0.00%	0.01%	0.00%	0.02%	0.05%
Whether Food Stamps	Family Level	0.09%	0.12%	0.13%	0.08%	0.10%	0.09%	0.24%
Food Expenditures (No Food Stamps)	Family Level	2.09%	2.20%	1.74%	1.66%	1.89%	1.97%	2.79%
Food Expenditures (Yes Food Stamps)	Family Level	3.01%	2.25%	1.60%	2.02%	1.67%	1.81%	7.14%
* Replication and updates to PSID Technical Paper #11-02 'Trends in Item Nonresponse in the PSID, 1968-2009'								