PSID A national study of socioeconomics and health over lifetimes and across generations

PSID Main Interview User Manual: Release 2013

The PSID main interview user manual was prepared by Patricia Andreski, April Beaule, Mary Dascola, Denise Duffy, Eva Leissou, Katherine McGonagle, Jay Schlegel, and Robert Schoeni. The manual draws heavily from documentation from prior years written by numerous PSID staff members.

This document should be cited as follows:
PSID Main Interview User Manual: Release 2013. Institute for Social Research, University of Michigan, July, 2013.

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 2007 and subsequent waves of the main interview. In addition, this manual 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 if you have suggestions for enhancing the manual or if you find errors.

This user manual will be updated regularly to reflect changes in the data archive, dissemination tools created for users, or related developments.

Date	Update
07/23/13	Updated to include 2011 data
01/23/12	Table 7 on page 29, and Table 5 on page 22 have been updated. Release 3 – 2007 Family File - Variable ER36106 - A50 Why did you move? has been reconfigured. We have recoded cases originally coded as 'Other/ambiguous' into existing categories where possible. Release 2 – 2009 Individual File - Data and description updates for two sampling error estimation variables ER31996 and ER31997. Two outdated sampling error estimation variables have been dropped: ER31998 and ER31999. We have also reconfigured the indicators for the DUST supplement ER34040-ER34042.

TABLE OF CONTENTS

List of Acronyms

CDS: Child Development Supplement

CATI: Computer assisted telephone interviewing

DUST: Disability and Use of Time Study

EHC: Event history calendar

FIMS: Family Identification Mapping System

ISR: Institute for Social Research

NIA: National Institute on Aging

NICHD: National Institute for Child Health and Human Development

NSF: National Science Foundation

OEO: Office of Economic Opportunity

PSID: Panel Study of Income Dynamics

SEO sample: Survey of Economic Opportunity sample

SRC sample: Survey Research Center sample

TA: Transition into Adulthood Study

List of Figures and Tables

1. INTRODUCTION TO THE PSID

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).

Originally conceived as a five-year project, the PSID has now collected data for more than four decades. More than 70,000 people have participated in the PSID, over 6,000 have died, and as many as four generations within sample families are represented. The mission of the PSID has broadened well beyond its original more narrow 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

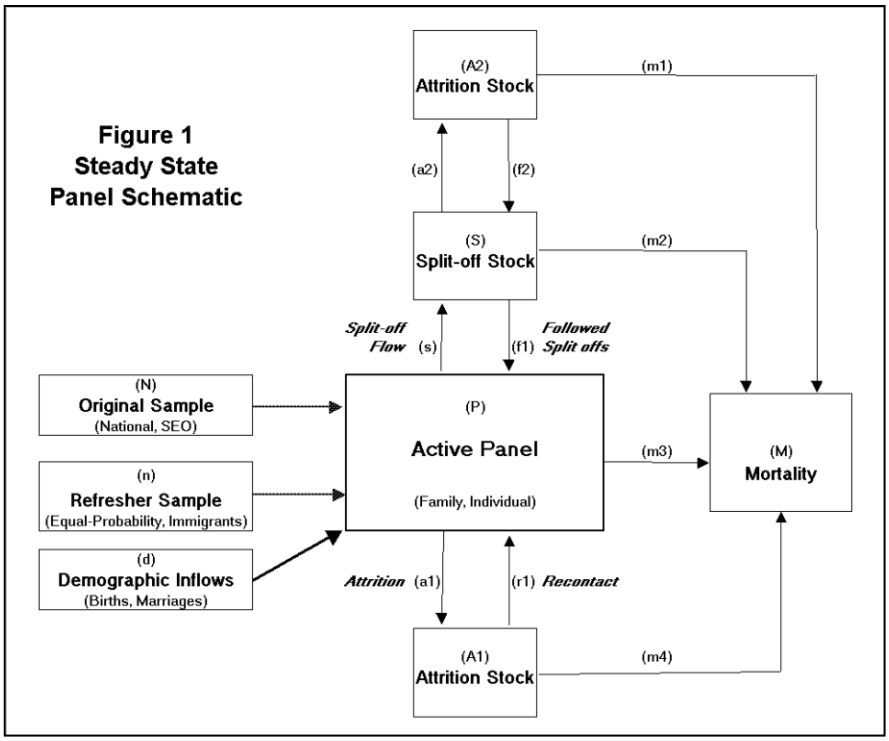
2.1 Core sample

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. This procedure replicates the population's family-building activity and produces a dynamic sample of families each year.

PSID families also include many "non-sample persons." The most common example is people who after 1968 marry sample persons. Information on non-sample 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 refresher sample of post-1968 immigrants (*n*), which is described below; and births and marriages in existing families (*d*). The intergenerational element is represented by children who split off 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 re-contacting (r_1) families refusing to be interviewed or not located (*A1* and *A2*) in previous waves, the PSID sample has grown despite losses due to mortality (*m1*, *m2*, *m3*, *m4*) and attrition from the active panel (*P*).



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 long time periods on a national population, the PSID is the only data set ever collected on life course and multigenerational economic conditions, well-being, and health in a long-term panel representative of the full U.S. population.

2.2 Immigrant refresher samples

While the design of the study augments the sample with new birth cohorts each year, it does not include individuals who arrived in the United States after 1968. (Individuals who arrived in the United States after 1968 and are living with individuals who were in the United States prior to 1968 are part of PSID family units and therefore are captured in the data collection. However, these individuals are not sample persons, i.e. they do not have the " PSID gene", and are therefore not followed unless a sample person continues to live with them.) To address this limitation, efforts have been made to add samples of immigrants who arrived in the United States after 1968. 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 include the full range of post-1968 immigrants, Asians in particular. Because of this crucial shortcoming, and a lack of sufficient funding, the Latino sample was dropped after 1995.

A total of roughly 500 immigrant families were added in 1997/1999 to update the PSID by adding a representative sample of recent immigrants to the United States: this new sample is called the 1997 immigrant refresher sample. A detailed description of the sampling 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 children born 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 roughly 500 immigrant families subsequently created split-off families, the number of post-1968 immigrant families continues to grow.

2.3 Sample reduction in 1997

Because the original sample of roughly 5,000 families had grown substantially 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 split-offs, 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

PSID follows sample members to whatever living arrangements they experience. Information is gathered about these sample members and their co-residents if they are living in a household situation. 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, or the military, 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, we still attempt 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 (e.g., ER42008 in 2009).

For those sample members still attached to PSID families that are active, the PSID keeps track of the location of sample members living in institutional housing, and it 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 time period individuals who could not be located or 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, several new approaches for recontacting former nonresponse sample cases were designed and implemented. First, interviews were attempted for all Latino and core 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. Non-response 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. The pool was then subset to 1,419 individuals who were recontacted in 1993, with the remaining individuals recontacted in 1994. In order to make the most of interviewer tracking efforts, the 1993 subset included all of the sample nonresponse individuals within a given original family identifier (i.e., 1968 interview number). The selection of original identifiers was geared toward recovering individuals who had not been followed when they initially became nonresponse because they were under age 18. If more than one individual was last present in the study in the same family, then PSID counted the group as one recontact family. If the recontact family was successfully interviewed in 1993, anyone who was no longer present was followed as a splitoff. 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 the following wave. Once a family was non-response for two consecutive waves, the family was no longer followed. PSID is currently reevaluating this design feature and is hoping to secure funding in order to follow sample into perpetuity regardless of how many waves they have been nonresponse in the past.

Following rules for sample individuals under age 18 were changed starting in 1993. In previous waves, PSID did not follow as split-offs sample members under 18 years of age if they left the family unit and their new family unit did not contain a sample person (usually with a nonsample parent as a consequence of a divorce, for example). 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, our family composition rules changed. Specifically, PSID families had always included a sample member as the Head or Wife/“Wife” 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 Head and the Wife/“Wife” may be nonsample. This could

occur if an underage sample member moves out with a nonsample parent who then remarries, for example.

2.5 Sample sizes

Reported in Table 1 are the number of individuals and families, in each of the main interview waves, separately by sample type, where sample type identifies SRC, SEO, Latino, and immigrant refresher. This table includes both sample and non-sample persons.

In Table 2 is reported the number of sample persons (non-sample persons are excluded) in each wave by whether they are in the SRC, SEO, or immigrant 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 Head or Wife are reported for each wave. While the number of families has increased substantially, the number of Heads and Wives who are sample persons has not changed appreciably in part because non-sample persons have become Heads and Wives after 1968. The final two columns report the number of Heads and Wives and the number of all individuals who have been in the PSID for at least five years and therefore can contribute to substantial panel analyses. These numbers were fairly steady until the sample was trimmed in 1997, 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 reason for 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. 5,617 individuals were dropped in 1997 when the sample was trimmed due to budget constraints. 300-400 children are born into PSID families annually, or roughly 600-800 between each post-1997 wave. In most years, new sample members who make their first entry into the PSID (or not born into the PSID) 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 and 1999 is associated with the addition of the new immigrant sample in those years. A substantial number of re-entrants were interviewed in 1993 and 1994 associated with the recontact efforts in those years, as described above. These re-entrants were living with individuals who had never participated in the PSID in the past, hence the large number of non-sample 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	9,481	8,749			18,230
1969	2,643	1,817			4,460	8,643	8,569			17,212
1970	2,754	1,891			4,645	8,752	8,597			17,349
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	9,191	9,045			18,236
1974	3,165	2,352			5,517	9,286	9,110			18,396
1975	3,252	2,473			5,725	9,437	9,186			18,623
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	2,107		10,765	12,576	11,936	7,034		31,546
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

Notes: Sample type is defined by ER30001: Core SRC=1-2930/Immigrant=3001-3511/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).

Table 2. Composition of individuals in the PSID: non-sample persons excluded							
	Individuals in the PSID who are sample persons						
Year	Total	SRC	Immigrant	Original Sample Member	Head or Wife/"Wife"	Heads or Wife/"Wife" in PSID>=5 Years	Individuals in PSID>=5 Years
1968	18,192	9,454		18,192	7,878		
1969	16,325	8,261		16,050	7,118		
1970	16,125	8,182		15,485	7,160		
1971	16,077	8,113		15,116	7,238		
1972	16,056	8,122		14,723	7,330	6,255	14,608
1973	15,955	8,068		14,305	7,445	6,299	14,503
1974	15,884	8,025		13,916	7,562	6,396	14,456
1975	15,846	8,014		13,555	7,668	6,463	14,405
1976	15,747	7,989		13,104	7,707	6,510	14,266
1977	15,705	8,006		12,714	7,769	6,610	14,159
1978	15,700	7,973		12,425	7,842	6,749	14,155
1979	15,687	7,987		12,063	7,995	6,848	14,070
1980	15,724	8,034		11,691	8,075	6,920	14,008
1981	15,753	8,037		11,391	8,098	6,986	14,037
1982	15,794	8,074		11,134	8,160	7,068	14,037
1983	15,838	8,101		10,840	8,217	7,194	14,094
1984	15,829	8,092		10,523	8,212	7,272	14,148
1985	15,808	8,091		10,192	8,258	7,267	14,085
1986	15,608	8,010		9,833	8,186	7,236	13,962
1987	15,544	8,022		9,511	8,164	7,222	13,882
1988	15,500	8,017		9,232	8,160	7,213	13,892
1989	15,396	7,971		8,937	8,106	7,268	13,881
1990	15,448	8,030		8,782	8,198	7,351	13,930
1991	15,421	8,033		8,524	8,178	7,352	13,932
1992	15,561	8,114		8,472	8,319	7,449	14,102
1993	15,927	8,340		8,236	8,376	7,372	14,217
1994	16,824	8,716		8,644	8,919	7,635	14,945
1995	16,506	8,575		8,294	8,762	7,600	14,878
1996	16,404	8,540		8,036	8,676	7,610	14,866
1997	14,081	8,501	1,659	7,377	7,122	5,743	11,312
1999	14,241	8,551	1,679	7,108	7,200	5,767	11,226
2001	14,540	8,750	1,633	6,804	7,370	5,703	11,163
2003	14,862	8,845	1,668	6,602	7,595	5,621	11,129
2005	15,394	9,139	1,681	6,391	7,895	6,126	12,289
2007	15,619	9,267	1,662	6,054	8,036	6,258	12,628
2009	16,163	9,518	1,710	5,876	8,311	6,503	13,041
2011	16,244	9,476	1,691	5,527	8,386	6,545	13,047

Notes: Cell values were determined by the following:

Total: Year sequence number 51-59 or sequence number 1-20 and person number (ER30002) < 170.

SRC: ER30001=1-2930; Immigrant: ER30001=3001-3511; Original sample: ER30002=1-29; Head/wife: Year Relation to Head=1,2 (1968-1982) 10,20,22 (1983+). Head/Wife in PSID > = 5 Yrs: Sum of waves for Head/Wife >=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

Table 3. Entry and exit of all individuals in the PSID, excluding Latino sample

						Entered the PSID				
		Exited the PSID					First Entry			
Year	Total	Total	Left	Died	Dropped	Total	Born into PSID	New Sample	Nonsample	Re-entrants
1968	18,230	0	0	0	0	0	0	0	0	0
1969	17,212	1,849	1,832	17	0	831	275	21	535	0
1970	17,349	725	634	83	0	862	375	8	479	0
1971	17,590	615	509	106	0	856	338	0	506	12
1972	18,051	569	442	127	0	1,030	392	1	607	30
1973	18,236	731	618	113	0	916	355	1	536	24
1974	18,396	751	647	104	0	911	368	0	519	24
1975	18,623	701	589	112	0	928	382	2	521	23
1976	18,768	850	748	102	0	995	423	1	525	46
1977	18,998	819	714	105	0	1,049	435	2	578	34
1978	19,140	784	708	76	0	926	369	0	505	52
1979	19,444	788	696	92	0	1,092	469	2	576	45
1980	19,748	910	804	106	0	1,214	522	1	632	59
1981	19,797	928	839	89	0	977	441	0	486	50
1982	20,113	746	633	113	0	1,062	427	1	584	50
1983	20,328	924	821	103	0	1,139	440	0	618	81
1984	20,394	1,000	890	110	0	1,066	459	0	537	70
1985	20,681	982	861	121	0	1,269	482	0	683	104
1986	20,438	1,235	1,128	107	0	992	369	0	553	70
1987	20,487	1,044	925	119	0	1,093	442	0	553	98
1988	20,507	1,030	912	118	0	1,050	419	0	532	99
1989	20,452	1,105	992	113	0	1,050	381	0	562	107
1990	20,746	898	794	104	0	1,192	373	0	597	222
1991	20,771	964	830	134	0	989	401	0	508	80
1992	21,146	1,090	969	121	0	1,465	355	0	720	390
1993	22,312	1,493	1,339	154	0	2,659	457	0	1,112	1,090
1994	24,513	1,550	1,398	152	0	3,751	419	507	1,066	1,759
1995	23,930	1,440	1,321	119	0	857	259	45	496	57
1996	23,811	1,247	1,126	121	0	1,128	351	44	612	121
1997	19,762	6,695	975	103	5,617	2,646	295	1,684	504	163
1999	20,516	1,534	1,317	217	0	2,288	604	334	900	450
2001	21,401	1,401	1,193	208	0	2,286	665	69	1,005	547
2003	22,291	1,499	1,294	205	0	2,389	695	72	1,113	509
2005	22,919	1,799	1,610	189	0	2,427	782	95	1,061	489
2007	23,502	1,761	1,505	256	0	2,344	822	88	1,082	352
2009	24,385	1,879	1,672	207	0	2,566	817	113	1,095	541
2011	24,661	2,044	1,802	242	0	2,320	827	81	1,060	352

Notes: Two adjacent years are compared to determine values. Categories are determined as follows using example code from Y1=1980 and

Total: Y2=Y1-Total Exits + Total Entrants

Exits:

- Total: Y1 sequence number (ER30314) > 0 and Y2 sequence number (er30344) =0
- Left: Total-Died
- Died: Y1 sequence number (ER30314) > 0 and Y2 reason for non-response=died (ER30371=41)
- Dropped: Y1 sequence number (ER30314) > 0 and Y2 reason for non-response=dropped (ER30371=80)

Entrants:

- Total: Born In+New Sample+Nonsample+Re-entrants
- Born In: Y2 sequence number (ER30344) > 0 and Y1 Type of Individual (ER30340)=9 and 1<=Person Number (ER30002) <=29
- New Sample*: Y2 sequence number (ER30344) > 0 and Y1 Type of Individual (ER30340)=9 and 30<=Person Number (ER30002) <=169
- Nonsample: Y2 sequence number (ER30344) > 0 and Y1 Type of Individual (ER30340)=9 and Person Number (ER30002) > 170
- Re-entrants: Y1 sequence number (ER30314)=0 and Year of first non-response (ER32008) < Y2

*Starting with the 1993/1994 comparison an additional requirement for this group is the variable Why Follow (ER33150) in Y2 =30 or 31

Starting in 1989/1990 must use sample ne 3 (Latinos out) - see SAS program.

3. SURVEY CONTENT

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.

Reported in Table 4 are the major topical areas contained in the main interview in recent waves, along with the average number of minutes it took to complete each section. For instance, in the 2011 wave, the average questionnaire length was 89.4 minutes. An additional 12.6 minutes were spent updating the family composition and collecting respondent contact and payment information, leading to an average total respondent burden of 102 minutes. The mean and median questionnaire length in each wave from 1968 to the present is reported in Table 5. (Note that the estimates in Tables 4 and 5 differ slightly because the estimates in Table 5 are based on data after the files have been processed.)

Table 4. Average interview length in minutes by section: 2005 to the present

Section of the questionnaire	2005	2007	2009	2011
Housing, utilities, computer usage (A)	4.9	5.1	5.9	6.7
Employment (B, C, D, E)	16.9	19.0	18.2	23.9
Housework, food expenses, food assistance, transportation, education expenses, other expenditures (F)	9.5	9.5	9.2	10.1
Income (G)	6.7	6.7	7.0	8.9
Health status, health behaviors, health expenditures (H)	11.9	16.8	14.9	18.6
Marriage/fertility (J)	1.1	1.0	1.1	1.1
New head/wife background (K,L)	2.3	1.8	4.4	1.8
Philanthropic giving and volunteering, religiosity, help received (M)	6.7	2.6	2.6	5.2
Pensions (P)	3.5	3.7	3.7	4.2
Off-year income and public assistance (R)	2.8	3.0	1.5	1.9
Wealth and active savings (W)	5.2	7.9	5.5	6.5
Total questionnaire length	71.5	77.1	74.0	89.4
Family unit composition update and split-off information, interview close, respondent address and payment information	8.3	9.2	9.7	12.7
Total respondent burden	79.8	86.3	83.7	102.0

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)	1980	29.0 (27.0)	1992	31.0 (29.0)	2011	88.9 (85.0)*
1969	61.8 (60.0)	1981	26.5 (25.0)	1993	34.5 (32.0)		
1970	60.5 (60.0)	1982	20.8 (20.0)	1994	42.8 (39.7)		
1971	59.1 (58.0)	1983	23.8 (22.0)	1995	34.4 (31.5)		
1972	66.2 (60.0)	1984	34.7 (32.0)	1996	29.1 (28.0)		
1973	20.1 (20.0)	1985	35.2 (33.0)	1997	36.9 (34.6)		
1974	23.1 (21.0)	1986	34.9 (33.0)	1999	59.5 (na)		
1975	26.9 (25.0)	1987	29.5 (28.0)	2001	60.2 (na)		
1976	25.3 (25.0)	1988	33.6 (31.0)	2003	69.6 (65.0)*		
1977	25.0 (24.0)	1989	33.9 (31.0)	2005	73.2 (69.0)*		
1978	26.9 (25.0)	1990	37.4 (33.0)	2007	80.0 (76.0)*		
1979	28.1 (26.0)	1991	29.5 (27.0)	2009	75.0 (70.0)*		

*Capped at 300 minutes; less than 0.3% of interviews in a given year were longer than 300 minutes. na=not available

The survey instrument for each data collection is available at the PSID website. A companion document to the survey instrument is prepared for each interviewing wave that describes the nuances of each question administered to respondents. This document is used by interviewers to assist them in addressing any questions raised by the respondent during the interview. These documents are called "question by questions objectives," or simply "QxQ," and they are available on the PSID website.

In the subsections below we describe the broad changes to the instrument between waves, with Appendix I containing a detailed list of all modifications.

3.1 Changes for 2007

There were three new questions in Section A. A25a1 indicates whether loan amounts had a fixed or variable interest rate. Questions A42_1 and A42a indicate if gas and electricity are paid on one bill and, if so, how much is spent on gas and electricity combined.

In the employment sections, BC for Head and DE for Wife/"Wife", questions were added about number of weeks actually worked on each of their jobs last year and the year before that (BC42a and BC47a for Head and DE42a and DE47a for Wife/"Wife").

In Section F, a question was added about which months the family was paying for child care in the previous year (F6d2). Question wording regarding stamp receipt (F16) changed from amount received last month to total amount received this year. New codes were added at F52 to capture some of the 'other specify' responses at this question.

Section W has new questions at the end (W129-W144) about intergenerational help.

The health section, Section H, underwent some major modifications. The most significant change was the addition of the childhood health calendar (H32-H45). For each condition and type of report about

Head or Wife/"Wife" there are yes/no indicators at each age. There are a total of six sets of Childhood Health Calendar reports. The series was asked about Head and Wife/"Wife" and depending on who the respondent was, the information is stored in the appropriate set.

(1) Self Report about Head (ER38344-ER38661)

(2) Spousal Report about Head (ER38662-ER38979)

(3) Other Report about Head (ER38980-ER39297)

(4) Self Report about Wife/"Wife" (ER39441-ER39758)

(5) Spousal Report about Wife/"Wife" (ER39759-ER40076)

(6) Other Report about Wife/"Wife" (ER40077-ER40394)

Since 2007 is the first wave of data collection for the childhood health calendar, only one report out of set 1-3 above and one report out of set 4-6 above may have data if any of the childhood health conditions were endorsed. These reports will be carried forward to future waves if Head or Wife/"Wife" remain the same. In subsequent waves, we may gather new types of reports (self report and spousal report, for example) about the same individual if the respondent changes. To help understand which reports are from which wave, we added variables at the start of each report to indicate which year this report originated. For example, for set 1: Self Report about Head, the year of report is stored at (ER38344).

Question H1d, about missing school for health reasons, is now in the childhood health sequence (H26). The depression scale that was used in 2003 was reinstated (H59-H59l).

The philanthropic giving section (Section M) is shorter as several questions from 2005 were dropped. Section R, which asks about income for the "off" year, or t-2, has some revisions. Separate questions for Head and Wife/"Wife" assets (R4-5 in 2005) are gone and a family total for assets is now at R16-19. New questions (R63-67) were added to capture any other t-2 income.

Section S, Katrina, is an entirely new series about health-related issues for respondents who lived in the Hurricane Katrina area in 2005. This information is available as a restricted use data file.

3.2 Changes for 2009

In Section A, a life satisfaction question was added (A3). Also added to this section was a series of questions asking homeowners whether they are behind on their mortgage payments and whether they have been faced with foreclosure (A27FOR1-6). Similar questions asking renters about previous foreclosures are found at (A37FOR1-7).

In the employment sections, BC for Head and DE for Wife/"Wife", summary confirmations about weeks and hours worked across all jobs were added at BC14a-f for Head and DE14a-f for Wife/"Wife." The flow of the questionnaire was altered so that only a fraction of the Heads and Wives/"Wives" were asked the job level questions BC/DE42. Only those cases for which we could not determine the weeks from the start and stop dates were asked these helper questions. These questions are

being used during editing to determine the value of weeks across all jobs at the BC/DE14 series, and therefore they are not released. Questions BC/DE49 and BC/DE57-60 were deleted.

In Section F, the frame of reference changed for F14 and F16. In 2007, these food stamp questions asked about 'last year;' in 2009 they were revised to say 'last month.' Some questions about car ownership were dropped. Deleted questions include those asking about who usually drives the car as well as some of the purchasing/leasing questions (F51, F52, F56 and F62-63).

Questions in Section W about intergenerational help (W129-W144) have been dropped because funding supporting the collection of these data was only available in 2007.

In the health section, Section H, the childhood health calendar (H32-H45) continues to be administered. For each condition and type of report about Head or Wife/"Wife" we have yes/no indicators for each condition at each age. There are a total of six sets of Childhood Health Calendar reports. The series was asked about Head and Wife/"Wife" and depending on who the respondent was, the information is stored in the appropriate set. These reports were carried forward from the 2007 wave if Head or Wife/"Wife" remained the same. In 2009, we may have gathered new types of reports (self report and spousal report, for example) about the same individual if the respondent changes. To help understand which reports are from which wave, we added variables – listed in parentheses below – at the start of each report to indicate which year this report originated.

(1) Self Report about Head (ER44317-ER44634)

(2) Spousal Report about Head (ER44635-ER44952)

(3) Other Report about Head (ER44953-ER45270)

(4) Self Report about Wife/"Wife" (ER45414-ER45731)

(5) Spousal Report about Wife/"Wife" (ER45732-ER46049)

(6) Other Report about Wife/"Wife" (ER46050-ER46367)

All Heads and Wives/"Wives" were asked the education series KL43-67 in 2009.

Section R, which asks about income for the "off" year, or t-2, was scaled back significantly. Remaining questions ask whether the family has received certain types of income, but the amounts and months of receipt were scaled back substantially. Please see the 2009 box and arrow version of the questionnaire for details.

Section S, which asked about the effects of Hurricane Katrina in 2007, was eliminated.

Some of the questions from the interviewer observation section are included on the public release file. These questions are asked of the interviewer after the interview is completed. Interviewer observation questions included on the family file include the respondents' reaction to requesting their Medicare number (ER46700-ER46705). Also included in this series are the sections of the interview that were completed with a helper (ER46706-ER46759).

3.3 Changes for 2011

The vast majority of survey questions from the 2009 wave continue to be part of the 2011 instrument, but a few sequences were dropped, some were added and others revised, as described below.

Section A: Housing, Utilities, Computer Use

In Section A, a new question (A6A) was added asking if the family unit lives in a place that requires someone to be 55 years old or older. The (A6) question asking if the R lives in a retirement community, and its follow-up question (A7) were dropped. A new follow-up question (A7A) was added regarding senior housing and whether the facility offered services that help with daily activities such as bathing, dressing and meals. Other new questions in Section A include a series (A20F-A20H) to collect information on lot rent and whether that amount includes water and sewer (A31A) for those living in mobile homes. The home air conditioning questions (A38) and (A39) were dropped in 2011. A new question (A47C) asking about internet access outside the home was added to the computing question series. The code frame at question (A50) was changed to accommodate respondents that answer 'homeless' to the, 'why did you move?' question.

Section BCDE: Head/Wife/"Wife" Employment

In the employment sections, BC for Head and DE for Wife/"Wife", a new question series (BCDE14B3-BCDE14B4PER) was added that asks if there was any overtime that wasn't included in the average hours per week reported by the Head or the Wife/"Wife." New employment questions were added that ask about the Head's job title (BC21a) for their current main job and how many minutes their commute time is (BC21b) for their current main job or most recent main job.

New skip rules were added so that only those cases where hours worked per week is not computable are asked question (BC43/DE43), and only those cases where overtime is not computable overtime are asked question (BC44/DE44).

Section F: Childcare, Food, Vehicles, Transportation and Expenditures

In Section F, the (F7) and (F7a) questions asking if anyone in the household was enrolled in adult day care, and how much they paid for adult day care, were dropped in 2011. Also in Section F, a new question was added (F49b2) which asks whether the vehicle in question is a hybrid.

Section G: Income

In parallel to Section BCDE, the same new questions about job title and commute time were added for the Unknown Job series. In addition, the job title (G77a) and round trip commute (G77b) questions were added for OFUM jobs as well. Additionally a multi-mention family level question asking about the Head's income from a Veteran's Administration pension was added at (G37a). These variables replace the single mention Veteran's Administration pension variable (G37) which was dropped. Question (G89) for OFUMs (aged 16-49) asking whether they are enrolled as a full-time or part-time student was dropped.

Section W: Wealth

In Section W, the (W38) question from 2009 asking about debts was disaggregated into separate questions for each of the five individual types of debt for the 2011 wave. This includes credit card debt (W38a), student loans (W39b1), medical bills (W39b2), legal bills (W39b3), and loans from relatives (W39b4). For each of these debt types, follow-up bracket questions are asked if the respondent is unable to provide an exact amount.

Section P: Pensions

A new question was added (P1a) asking if the Head/Wife/"Wife" is currently eligible for their employer's pension or retirement plan. Additional questions that were added this wave include (P22a-d), which collects information on the earliest age the Head/Wife/"Wife" could receive a full pension and the earliest age they could leave their employer and still receive pension benefits.

Question (P33 for current employer pension and P62 for previous employer pension), asking whether the Respondent can estimate their pension or retirement benefits was dropped as was the question asking about pension as a lump sum (P34LUMP/P62LUMP) The question text was revised for (P34AMT) and (P34PCT). Additional bracket questions (P34f/P62f) to (P34j/P62j) were added to capture how much the Head/Wife/"Wife" expects to receive as a percent of their pay at retirement if the exact amount was not given.

The (P40) and (P41) questions asking about when the Head/Wife/"Wife" (plans to/ thinks they will) stop working were revised. The new questions (P40NUM/P41NUM) ask in how many years (rather than in what year) Head/Wife/"Wife" (plans to/ thinks they will) stop working. The question text (P46) was revised to provide the Respondent with a clearer understanding of the code options.

Questions dropped from Section P include (P21) which asks how much the Head/Wife/"Wife" could take out of their account if they were to leave their employer today, the questions asking about when Head/Wife/"Wife" would receive full or partial benefits (P23) through (P31b), and the benefit value estimate as a fraction of pay questions (P33) through (P34e).

Section H: Health

The questionnaire text was slightly revised for questions (H6a2) and (H6b2). Text fills were added in 2011 to make the question intent clearer to the respondent depending on when they were last interviewed.

Several questions to the health condition series (H6c2-H6m2) asking if the condition became worse have been dropped. New follow-up questions (H7a2-H7m2) for each unique health condition were added that asks whether the Head/Wife/"Wife" are taking medication/treatment for endorsed conditions. New fields for Head/Wife/"Wife" height and weight (H23Meter/H22Kilo) were added to allow respondents the option of answering the questions in metric units. The cut-off age for checking on a state health insurance plan for uninsured children was changed from less than 18 years old to less than 19 years old at (H61a RULE).

New for 2011 are several edit variables for the chronic conditions section. First, codes for the open-ended variable response for other chronic condition at H6M1 are provided. As a part of coding process for this variable, a response which fit perfectly into a previous category may have been moved into that previous category. For those cases where data was moved from H6M1 as part of the coding process, the corresponding WTR EDITED variable was coded as '1' to indicate this data was moved from the open-specify field. An example of the WTR EDITED variable series is (ER49502).

The childhood health calendar questions (H32-H45) were included in 2011, marking the third wave of their inclusion in the instrument. For each condition and type of report about Head or Wife/"Wife," there are yes/no indicators for each condition at each age. There are a total of six sets of Childhood Health Calendar reports. The series was asked about Head and Wife/"Wife" and depending on who the respondent was, the information is stored in the appropriate set.

Reports about the childhood health conditions were carried forward from the 2007 and 2009 wave and not reasked in 2011 if Head or Wife/"Wife" remained the same. In 2011, new types of reports (self-report and spousal report, for example) about the same individual may have been collected if the respondent changed. To help understand which reports are from which wave, variables were added at the start of each report to indicate which year the report originated. For example, set one (1) Self Report about the Head and the year of report is stored at (ER49657).

(1) Self Report about Head (ER49657- ER49974)
(2) Spousal Report about Head (ER49975- ER50292)
(3) Other Report about Head (ER50293- ER50610)
(4) Self Report about Wife/"Wife" (ER50775- ER51092)
(5) Spousal Report about Wife/"Wife" (ER51093- ER51410)
(6) Other Report about Wife/"Wife" (ER51411- ER51728)

Section KL: Background of Head/Wife/"Wife"

In 2009, we asked all Heads/Wives/"Wives" about their current educational attainment regardless of whether they were new to the study or not. For 2011, the usual process of asking about the background of Head/Wife/"Wife" has been restored, with this section being asked only for new Head's/Wives/"Wives".

Section M: Philanthropy

The volunteering section (M14a-M57b) which was asked in prior waves (2003 and 2005) was included in the 2011 wave.

Section Interviewer Observations

Several questions from the interviewer observation section are included on the public release family file: respondents' reaction to requesting their Medicare number (ER52100-ER52105), and sections of the interview that were completed with a helper (ER52106-ER52159). New for 2011 are observation

questions (IO29/O29a) (ER52160-ER52167) asking what records and documents the Respondent had available during the interview.

4. DATA COLLECTION

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 principle investigators review proposed new content and consider removing or modifying existing questions. Changes are also made in response to issues that arise during processing and collection of the prior wave of data, including spelling errors, missing or new interviewer instructions, revisions to response category lower/upper limits, and reprogramming bad skips. Timings are estimated for proposed new and removed content in order to assess the impact of the changes for achieving the interview length goal for that wave.

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 Interviewer training and field operations

The PSID has gathered information about families and all individuals in those families by conducting interviews annually 1968-1997, and biennially thereafter. Since 1973 the majority of interviews have been conducted via telephone, with over 95% of interviews collected face-to-face 1968-1972. A single primary adult has typically served as the sole respondent and provides information about him/herself and about all other family members. (Exceptions were reports of retrospective information in 1976 and 1985 when separate interviews were completed with both the Head and Wives/"Wives.") The most detailed information is collected about the Heads and spouses of family units.

Since 1993, the survey has been administered using a computer-assisted telephone interview (CATI). Beginning in 2003, Blaise software was used to program the questions and SurveyTrak, a software developed at ISR, was used to manage sample and administrative information about the family. Other data collection systems that have been developed internally for the PSID include web-based systems (i.e., 'Webtrak' and 'Casetrak') used to communicate between interviewers and project staff on issues such as refusal conversion, requests for persuasion letters, and tracking of respondents. An Oracle-based respondent address payment system (RAPS) was developed by PSID staff in 2005 and it has several functionalities, including: the archiving of contact information for respondents, including the address at which they wish to receive their incentive, secondary contact information, and the address at which the interview occurred (for geocoding). Contact information is updated and archived by the RAPS in the year before the interview occurs with the update protocol based on results from a set of experiments

(McGonagle, Couper & Schoeni, 2011; McGonagle, Schoeni & Couper, 2013). RAPS also provides for the automated generation of tailored letters for respondents, including different refusal aversion/conversion templates based on the type of reluctance the respondent is exhibiting (e.g., Groves & McGonagle, 2001), as well as sympathy letters, and advance contact letters; and the automated generation of respondent incentives immediately after the interview is completed, with most respondents receiving their incentive payment within a few days after their interview has been completed.

A major innovation introduced in the 2003 wave was the Event History Calendar (EHC), which provides 2-year long timelines of employment, residence, and features of employment across job transitions. Having 2-year data in these content areas has helped fill the gap of data caused by moving the study to a biennial data collection. The fine-grained EHC timeline data can be used to support the construction of traditional measures – such as weeks of employment, unemployment, and time out of the labor force. Methodological research has shown that the EHC interviewing methodology leads to consistently higher quality retrospective reports in comparison to traditional standardized question-asking methods (Belli et al, 2001; Belli et al, 2004). In addition, these timeline data can be used to analyze interrelated events such as the timing of auto purchases, residential moves, and employment transitions. Additional details about the EHC in the PSID are available in Belli (2003).

Beginning in 2007, a training DVD containing a description of the study terminology, concepts, and the interview sections was created and distributed to all interviewers as part of their training. The training videos are available on the PSID website. Interviewers review the DVD training material prior to an in-depth training session held in the Ann Arbor, Michigan area.

A variety of strategies are used to minimize sample attrition including incentive payments, refusal conversion letters, off year address update mailings, tracking, respondent newsletters, and more. A full description of the procedures is provided in Schoeni, Stafford, McGonagle, and Andreski (2013).

Table 6 reports the beginning and end dates of the field period, the percent of interviews completed over the 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 Head, and the percent of family units for whom the interview was completed by a sample person.

4.3 Response rates

Response rates are calculated separately for the core (also known as original) sample families and for the families that were part of the immigrant refresher in 1997/1999. 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; split-offs, who are individuals who became economically independent creating their own family unit; and recontact split-offs, 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 7.

Table 6. Characteristics of field operations: 1968 to present

Year	Number of families	Field Period		% by telephone	Mean (median) # of calls to complete a case**	% calls 8+
		Start	End*			
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	10,765	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

* End date not exact for years 1969-1980; ** Calls 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.

Table 6, Continued. Characteristics of field operations: 1968 to present				
Year	Incentive ($)	% of interviews in Spanish	% of interviews provided by head	% 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
Notes: Type of respondent=1 (Head) 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, ER30002=1-169.				

Table 7. Response rate each wave by sample type and interview type: 1968 to present

		Main PSID					Latino (1990-95)/ Immigrant (1997-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
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

Notes: na= not available; deceased are included in base 1968-1972 and are excluded in all waves 1973-2009. Sample sizes for recontact split-offs for the Latino/Immigrant samples are quite small and therefore the response rates fluctuate substantially.

5. PUBLIC USE DATA AND DATA PROCESSING: FAMILY FILE

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

1. Family file
2. Cross-year individual file
3. Birth history file
4. Marriage history file
5. Parent identification file

Most variables are contained in what is called the family file, including all family level information as well as detailed information about the Head and the Wife/"Wife." The cross-year individual file includes information on every person 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 for the Head and the Wife/"Wife", 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 marriages prior to 1985 as provided through retrospective reports. The parent identification file synopsizes 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.

The family file contains one record for each family interviewed in a given year. It includes all family level variables collected in that year, as well as extensive information about the Head and the Wife/"Wife". 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 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. These statements provide variable names, locations, and variable labels. The definition statement files are not intended to represent complete and full programs for the respective statistical program packages to run extracts, analyses, etc. Users must provide all other statements needed to complete a program. Missing data statements are not provided as part of the files.

5.1 Format, variable names, and positions

5.1.1 2007

The 2007 family data file consists of one data file with 8,289 records and 5,069 variables. The variable names are in the range ER36001-ER41069.

5.1.2 2009

The 2009 family data file consists of one data file with 8,690 records and 5,012 variables. The 2009 family variable names are in the range ER42001- ER47012.

5.1.3 2011

The 2011 family data file consists of one data file with 8,907 records and 5,136 variables. The variable names are in the range ER52344-ER52395.

5.2 Coding and generated variables

In this section we describe our coding schema and the process for constructing generated variables. In general, code 8 (or 98 or 998, etc.) represents "don't know" and code 9 (or 99 or 999, etc.) represents other missing data or a refusal. Inappropriate questions are padded with zeroes. If a variable contains a code value that is neither included in the codebook nor one of the "zero", "eight" or "nine" codes just mentioned, assume missing data for that value; this should be extremely rare.

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

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. Within this section of the user manual, we provide a summary of differences in the income and wage calculation and imputation methodology since 2007, as well as a collection of updated tables that reflect the frequency of 2009 and 2011 employment and income imputations.

5.2.1.1 2009

Income, work hours and wage calculation and imputation differences between 2009 and 2007 include the following:

- Weeks worked and hours worked
- Accuracy variables
- Minor changes to the imputation of some variables

As discussed in the technical paper on 2007 income and wages, the employment section of the PSID (BC/DE) asks respondents for detailed information for each of the Head's and Wife's jobs, including information on weeks worked, weeks not working work by time off category (illness, unemployment, etc.), and hours worked. In 2007, respondents were asked to report hours and weeks worked for the Head's and Wife's current main job, as well as for up to three additional jobs. In 2009, Head's and Wife's were asked to provide this information across all jobs, rather than by individual job. As a result, we will not have weeks worked and hours worked by job in 2009 (and therefore do not calculate total weeks work as discussed in section 9.1.1 of the 2007 technical paper). Instead, in the 2009

wave, each Head and Wife will have one weeks worked variable and one hours worked per week variable. The list below details the variable change:

Variable Description	2007 Variable	2009 Variable
Head Weeks Worked (Job 1)	ER36168	ER42146 (across all jobs)
Head Weeks Worked (Job 2)	ER36200	
Head Weeks Worked (Job 3)	ER36232	
Head Weeks Worked (Job 4+)	ER36264	
Wife Weeks Worked (Job 1)	ER36426	ER42398 (across all jobs)
Wife Weeks Worked (Job 2)	ER36458	
Wife Weeks Worked (Job 3)	ER36490	
Wife Weeks Worked (Job 4)	ER36522	
Head Hours Worked per Week (Job 1)	ER36170	ER42148 (across all jobs)
Head Hours Worked per Week (Job 2)	ER36203	
Head Hours Worked per Week (Job 3)	ER36234	
Head Hours Worked per Week (Job 4+)	ER36266	
Wife Hours Worked per Week (Job 1)	ER36428	ER42400 (across all jobs)
Wife Hours Worked per Week (Job 2)	ER36460	
Wife Hours Worked per Week (Job 3)	ER36492	
Wife Hours Worked per Week (Job 4)	ER36524	

For the 2009 PSID wave, new accuracy variables were added to provide users with additional details on imputed and/or edited variables. Specifically, accuracy variables have been created for Head/Wife weeks worked, hours worked, overtime and weeks not working. The new accuracy variables and their labels are provided in the list below:

Variable Description	Variable Name
ACCURACY HEAD WORK WEEKS-2008	ER46762
ACCURACY HEAD WEEKLY WORK HOURS-2008	ER46764
ACCURACY HD OVERTIME WORK HOURS-2008	ER46766
ACC HD WKS MISSED ILLNESS OF OTRS-2008	ER46769
ACC HD WKS MISSED FOR OWN ILLNESS-2008	ER46771
ACC HEAD WEEKS OFF FOR VACATION-2008	ER46773
ACCURACYHEAD STRIKE WEEKS-2008	ER46775
ACCURACY HEAD WEEKS LAID OFF-2008	ER46777

ACCURACY HEAD UNEMPLOYMENT WEEKS-2008	ER46779
ACC HEAD WEEKS OUT OF LABOR FORCE-2008	ER46781
ACCURACY WIFE WORK WEEKS-2008	ER46783
ACCURACY WIFE WEEKLY WORK HOURS-2008	ER46785
ACCURACY WF OVERTIME WORK HOURS-2008	ER46787
ACC WF WKS MISSED ILLNESS OF OTRS-2008	ER46790
ACC WF WKS MISSED FOR OWN ILLNESS-2008	ER46792
ACC WIFE WEEKS OFF FOR VACATION-2008	ER46794
ACCURACY WIFE STRIKE WEEKS-2008	ER46796
ACCURACY WIFE WEEKS LAID OFF-2008	ER46798
ACCURACY WIFE UNEMPLOYMENT WEEKS-2008	ER46800
ACC WIFE WEEKS OUT OF LABOR FORCE-2008	ER46802

Finally, for the 2009 wave, a number of small changes were made to imputation methodology. The imputation process was modified for the following variables:

- Head and Wife Labor Income
- Bonus Income

As detailed in the technical paper on 2007 income and wages, the steps in imputing Head and Wife labor income are as follows:

Step 1: Calculate wage and salary income using wage rate information collected on individual jobs in the employment section (sections BC and DE) of the 2007 family file. The employment section of the PSID asks respondents for wages, weeks worked, hours worked and time off detail for up to four jobs held by the Head (section BC) and Wife (section DE). If employment section data are not available, then go to Step 2.

Step 2: Impute using the wage rate from the prior wave, adjusted by the change in the average annual CPI (CPI-U) from the prior wave. If not available, then go to Step 3.

Step 3: Apply overall Head/Wife median wage and salary

In 2009, we are making a change to Step 3 in this process. Rather than applying the overall median Head/Wife wage to missing values, we are calculating the overall median Head/Wife wage rate, and applying that to total hours worked.

We are making a similar change with respect to bonus income. In the 2007 wave, bonus income was imputed first using the average bonus percent by occupation code (Step 1), and if average bonus percent by occupation wasn't available, we used an overall median bonus amount

(Step 2). For 2009, we are applying the mean bonus percent (mean of the ratio of bonus/wages) to Head wages with missing bonus income.

The technical paper on 2007 income and wage imputation contained a series of tables and figures describing the naming conventions and imputation statistics associated with the income and wage variables for the 2007 data. These same tables and figures are provided here in Appendix II, but based on the 2009 data. Please refer to the technical paper on 2007 income and wage imputations for additional details and descriptions of these tables.

5.2.1.2 2011

Income, work hours and wage calculation and imputation differences between 2011 and 2009 include the following:

- Wage Rate Calculation
- Social Security Type Variable

Since 1994, PSID has calculated and reported a Head and Wife wage which equals total labor income divided by total hours worked. Because of updates in the hours per week measure (as noted in section 3.3), we must modify the Head and Wife wage rate calculation. Specifically, beginning in 2011, we only ask respondents for hours worked at an individual job level when they are unable to provide average hours per week worked across all jobs. The average hours per week worked across all jobs includes hours worked in the respondent's unincorporated businesses. As a result, in order to ensure that income and hours are consistent in wage rate calculations, we need to include labor income from unincorporated businesses in total labor income. We had not included labor income from businesses historically in this calculation.

We also now provide six binary variables on the individual file to indicate the type of social security income a respondent received (if any). Individuals can receive more than one type of social security income, and if an individual received no social security income, then all six variables are coded 0. In 2009, social security type was a single variable. Values for the new social security type variables include:

Variable Description	2011 Code
Whether Social Security Disability	ER34137
Whether Social Security Retirement	ER34138
Whether Social Security Survivor's Benefits	ER34139
Whether Social Security Dependent of Disabled Recipient	ER34140
Whether Social Security Dependent of Retired Recipient	ER34141
Whether Social Security Other	ER34142

5.2.2 Wealth

The wealth module was first included in 1984. This module was included again in 1989, 1994, 1999, and every wave since 1999. 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.

5.2.2.1 1984-2001

Release 3 (1984, 1989, 1994, 2001): March, 2011

Release 4 (1999): March, 2011

Notes regarding the March, 2011 release: The wealth data were updated in March, 2011 to include accuracy codes for each of the components of total wealth as well as the totals themselves. These accuracy variables indicate whether the value was changed compared to the value as recorded in the original release of the family file, but do not capture whether such changes were due to editing, imputation, post-release updating, or some combination thereof. Additionally, the value of home equity was released as a separate component variable in each file. Also, the variables were renamed in the 1999 file to match the naming scheme used in each of the other wealth files (e.g. "Sxnn", where 'x' is a sequential number representing the wave, and 'nn' is a sequential number within the wave representing the individual variables). Lastly, descriptive labels were added for each of the variables.

The release of the PSID wealth data permits previously unavailable opportunities for researchers interested in long-term financial wealth dynamics as a part of modeling the behavior of households. For more detail, see Hurst, Erik, Ming Ching Luoh, and Frank Stafford. 1998. *The Wealth Dynamics of American Families, 1984-94.* Brookings Papers on Economic Activity 1: 267-337. A working paper, *Five Years Older: Much Richer or Deeper in Debt?*, by Joseph Lupton and Frank Stafford is illustrative of the uses of the wealth data.

The data from the wealth sequences come from an interviewing technique of 'unfolding brackets', pioneered in the PSID in 1984. When respondents cannot give an exact dollar value to the initial question on amount of such assets held, they are routed through a sequence of ranges. That is, if they initially report "don't know" or refuse, they are routed through questions with dollar ranges as answers. Within these ranges there are other subsequences, depending on their answers. To impute values for those not providing exact dollar amounts from this question structure requires special imputation techniques. We have developed and then improved such an imputation process and applied it to each wave of the PSID wealth data. This imputation process allows us to have a high-quality imputation procedure which applies consistently to all waves of PSID wealth data, and this is crucial for descriptive and modeling purposes. The components of wealth were asked as family-level variables. The variables included in the data files are noted below.

- The first variable for each wave (S100, S200, S300, S400, and S500) is the release number for that particular file. The value for this variable is updated each time the file is re-released for whatever reason. Codebook notes for this variable indicate which changes were made in the most-recent release.
- There are two sets of imputed variables (S102-S117, S202-S217, S302-317, S402-S419, and S502-S519) representing the processed components of net worth. There are lead-in variables for most of the components of wealth (except for value of vehicles), indicating whether the family has the particular component, followed by the corresponding value amount for the component. The un-imputed versions of each of these variables can be found in the corresponding family file. This sequence ends with the pair of total wealth variables, WEALTH1 (Sx16) and WEALTH2 (Sx17) -- WEALTH1 *does not include home equity* whereas WEALTH2 *includes home equity* (market value of the owner-occupied housing less outstanding mortgage balances).
- Accuracy variables (S102A-S117A, S202A-S217A, S302A-317A, S402A-S419A, and S502A-S519A) have been added for each component of wealth as well as for the two totals. These are dummy-indicators for whether the value differs from the value as released in the family file.
- There are two minor differences in how family wealth was ascertained during these five waves:
    - Home equity is calculated as the difference between home value and remaining mortgage amount. In 1984 and 1989, only one mortgage amount was asked for during the survey, whereas in 1994, 1999, and 2001 up to two mortgage amounts were asked for.
    - IRA amounts were included in the questions on stocks (W15) and bonds (W27) in 1984, 1989, and 1994. IRA amount was included as a separate question sequence beginning in 1999 (W21).

The specific components of wealth are as follows:

- Net value of farm or business assets [Sx03]
- Value of checking and savings accounts, money market funds, certificates of deposit, savings bonds, Treasury bills, other IRAs (IRAs asked separately beginning in 1999) [Sx05]
- Value of debts other than mortgages, such as credit cards, student loans, medical or legal bills, personal loans [Sx07]
- Net value of real estate other than main home [Sx09]
- Value of shares of stock in publicly held corporations, mutual funds or investment trusts, including stocks in IRAs (IRAs asked separately beginning in 1999) [Sx11]
- Net value of vehicles or other assets 'on wheels' [Sx13]

- Value of other investments in trusts or estates, bond funds, life insurance policies, special collections [Sx15]
- Value of private annuities or IRAs (IRAs asked separately beginning in 1999) [Sx19]
- Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2 in each release of each of the five waves, but not released separately until March 2011) [Sx20]

Recall that WEALTH1 [Sx16] *does not include home equity* whereas WEALTH2 [Sx17] *includes home equity* (market value of the owner-occupied housing less outstanding mortgage balances). Note that the 'x' in the variable names referenced above refers to the wave of the corresponding wealth file: e.g., for 1984, x=1; for 1989, x=2; and so on.

Wealth Imputations. To work with the original wealth data it was necessary to process the values which were not provided in a dollar amount but rather as bracket ranges (the 'unfolding brackets') in a consistent fashion for the wealth components available in each of the five waves. The extent of item non-response was surprisingly low in the PSID, and this helps provide an extra measure of quality; we believe this is because the PSID respondents have confidence in the interviewers and have been interviewed on numerous prior occasions. Ferber (1959) has underscored the importance of reinterviews in gathering family wealth measures.

A standard series of unfolding bracket questions is as follows, using "real estate other than own home" as an example:

(1) "Do you (or your family living there) have any real estate other than your main home...?"

(2) If Yes in (1), "If you sold all that and paid off any debts on it, how much would you realize on it?"

If a dollar value is given in (2), then the questions skip to the next wealth component. Respondents who did not report an exact amount in (2) were asked a series of three or four questions (starting with "would it amount to $50,000 or more?). These questions ultimately yielded the following categories: $1-$999; $1,000-$24,999;$25,000-$99,999; over $100,000. Some respondents gave only partial brackets such as below $25,000, above $25,000, etc.; other respondents would not give any bracket value at all.

Assuming that respondents who could not or would not provide an amount have holdings distributed in the same way as respondents giving an exact amount, a hot deck method was used to impute the missing values. This imputation process consists of three levels. First, Respondents who answered "Don't know", "Refusal" or simply missing in question (1) are assigned to "Yes" or "No" with those who are imputed as "Yes" being treated as not having bracket information at the next level.

Second, respondents who only gave a partial bracket or no bracket are randomly assigned to one of the four brackets ($1-$999, $1,000-$24,999, $25,000-$99,999, and over $100,000) with probability in accordance with the distribution of brackets realized from respondents who gave an exact bracket. Note

that this does not include people who gave an exact amount despite the fact that they fall into a bracket. The assumption is that people who give partial bracket information or no bracket information at all are more like individuals who gave exact bracket information than those that gave exact amounts. Juster and Smith (1997) argue that the distribution of bracketed individuals is different from those who give an exact amount.

After the second level, every respondent has an index of the exact bracket to which the respondent belongs. (Note: the dollar values defining the bracket ranges in the questions differ some between 1984 and 1989, but are identical between 1989 and 1994 and 1999. The choice of these bracket values was informed by the empirical distribution of the values of the assets in question in combination with 'round' numbers *and* restricting the questions to a parsimonious set of bracket categories.

Third, respondents who did not give exact amount (including those with reported exact brackets and imputed brackets) are assigned a dollar value with a probability derived from the distribution of amounts from respondents who reported exact values and which fell within the range of the same bracket.

Because we were working with an early-release version of the 1994 and 1999 family files, some additional work was done. Prior to the processing described above, some pre-processing imputation was done for missing values in house mortgage amounts. In addition, subsequent to the wealth processing, a listing of 1994 data with wealth component values ending in 7, 8 and 9 was prepared. Based on interviewer notes and other information, editing judgments were applied, placing these as either "don't know", "refused" or over-the-field amounts.

Due to data collection and processing conventions, in 1984 and 1994 negative values for component wealth variables were coded as zero. The modest number and amount of negative values reported for component wealth variables in the 1989 leads us to believe that 1984 and 1994 net wealth values would not change significantly had negative values been recorded. Recall that WEALTH1 [Sx16] does not include home equity whereas WEALTH2 [Sx17] does include home equity (market value of the owner-occupied housing less outstanding mortgage balances). No inflation adjustments - the CPI-U for 1984=103.9, 1989=124.0, 1994=148.2, 1999=166.6, and 2001=177.1.

5.2.2.2 2003

Release 2: March, 2011

The second release of the 2003 wealth dataset contains 38 variables. Accuracy codes were added for each component of total wealth as well as for the two totals (S602A-S620A), to indentify records which have been imputed/changed. In addition, home equity value is now included as a separate release variable (S620).

Release 1: May, 2005

The 2003 wealth question sequence is essentially unchanged from the sequence asked in 2001. The only exception is an alteration in the numeration of the follow-up unfolding brackets for two wealth components: stocks (W15) and Individual Retirement Accounts (W21). For example, for stocks in 2001,

the bracket question W19 corresponds to a value of $100,000 while W20 corresponds to a value of $5,000. These brackets were switched in 2003, with W19 corresponding to a value of $5,000 and W20 corresponding to a value of $100,000. Note that the change affects only naming of the questions and does not affect the question sequence itself.

PSID asks about nine broad wealth categories, including debt:

1. Equity in business (also includes farm) [W11, S603]
2. Transaction accounts (includes savings accounts, money market funds, certificates of deposit, government savings bonds, and treasury bills) [W28, S605]
3. Value of debt aside from mortgage on the main home or vehicle loans [W39, S607]
4. Equity in real estate (second home, land, rental real estate, or money owed on a land contract) [W2, S609]
5. Equity in stock (includes shares of stock in publicly held corporations, mutual funds, and investment trusts) [W16, S611]
6. Equity in vehicle (cars, trucks, a motor home, a trailer, or a boat) [W6, S613]
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, S615]
8. Equity in Individual Retirement Accounts [W22, S619]
9. Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2, but not released separately until March 2011) [S620]. Questions about home equity (primary residence) are asked in section A.

Processing of the data include three steps: a) imputation of the wealth components (1-8), b) computation of home equity (9), and c) construction of estimates for total family wealth with and without home equity. In the processing of the 2003 wealth data we followed the same approach as in the prior waves. In particular, the hot-deck imputation technique was used for imputation of the missing data in the wealth components (1-8). Details on home equity imputation are given in Table 8.

Table 8. Number of cases by missing data group and imputation category: home equity, 2003

Missing data group	Imputation category A	B	C	D	E	Total
1. House value is not missing; mortgage is missing	0	0	37	12	81	130
2. House value is missing; mortgage is not missing	0	0	14	0	14	28
3. House value is missing; there is no mortgage	40	0	30	3	0	73
4. Both house value and mortgage are missing	0	33	9	1	0	43
Total	40	33	90	16	95	274

For the first two patterns of the missing data (Group 1 and Group 2), where either house value or mortgage value is missing but not both, we first tried to replace the missing value with the value from the previous wave (category C). If this was not possible, a value from the wave prior to the last (i.e., 1999) as used (category D). When this was also not possible, we used an assumed identity, Mortgage=0.6* HouseValue, to obtain an estimate for the mortgage or house value from the other known value.

When the house value was missing and there was no mortgage (Group 3) or when both house and mortgage values were missing (Group 4), we first attempted to impute based on the prior information (category C and D). If this was not possible, then the median home equity was used for imputation. For Group 3, the median was based on the sample of homeowners with no mortgage outstanding (median = $90,000). For Group 4, the median was calculated for the whole sample of homeowners (median = $13,000). Descriptive statistics of home equity values for the imputed cases are reported below.

Descriptive statistics of 2003 imputed home-equity values

N	Mean	Std Dev	Min	Max
274	64206.98	88317.7	-54900	968000

The specific components of wealth are as follows:

- Net value of farm or business assets [Sx03]

- Value of checking and savings accounts, money market funds, certificates of deposit, savings bonds, Treasury bills, other IRAs (IRAs asked separately beginning in 1999) [Sx05]
- Value of debts other than mortgages, such as credit cards, student loans, medical or legal bills, personal loans [Sx07]
- Net value of real estate other than main home [Sx09]
- Value of shares of stock in publicly held corporations, mutual funds or investment trusts, including stocks in IRAs (IRAs asked separately beginning in 1999) [Sx11]
- Net value of vehicles or other assets 'on wheels' [Sx13]
- Value of other investments in trusts or estates, bond funds, life insurance policies, special collections [Sx15]
- Value of private annuities or IRAs (IRAs asked separately beginning in 1999) [Sx19]
- Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2 in each release of each of the five waves, but not released separately until March 2011) [Sx20]

Recall that WEALTH1 [Sx16] *does not include home equity* whereas WEALTH2 [Sx17] *includes home equity* (market value of the owner-occupied housing less outstanding mortgage balances).
Note that the 'x' in the variable names referenced above refers to the wave of the corresponding wealth file: e.g., for 1984, x=1; for 1989, x=2; and so on.

Wealth Imputations. To work with the original wealth data it was necessary to process the values which were not provided in a dollar amount but rather as bracket ranges (the 'unfolding brackets') in a consistent fashion for the wealth components available in each of the five waves. The extent of item non-response was surprisingly low in the PSID, and this helps provide an extra measure of quality; we believe this is because the PSID respondents have confidence in the interviewers and have been interviewed on numerous prior occasions. Ferber (1959) has underscored the importance of reinterviews in gathering family wealth measures.

A standard series of unfolding bracket questions is as follows, using "real estate other than own home" as an example:

(1) "Do you (or your family living there) have any real estate other than your main home...?"

(2) If Yes in (1), "If you sold all that and paid off any debts on it, how much would you realize on it?"

If a dollar value is given in (2), then the questions skip to the next wealth component. Respondents who did not report an exact amount in (2) were asked a series of three or four questions (starting with "would it amount to $50,000 or more?). These questions ultimately yielded the following categories: $1-$999; $1,000-$24,999;$25,000-$99,999; over $100,000. Some respondents gave only partial brackets such as below $25,000, above $25,000, etc.; other respondents would not give any bracket value at all.

Assuming that respondents who could not or would not provide an amount have holdings distributed in the same way as respondents giving an exact amount, a hot deck method was used to impute the missing values. This imputation process consists of three levels. First, Respondents who answered "Don't know", "Refusal" or simply missing in question (1) are assigned to "Yes" or "No" with those who are imputed as "Yes" being treated as not having bracket information at the next level.

Second, respondents who only gave a partial bracket or no bracket are randomly assigned to one of the four brackets ($1-$999, $1,000-$24,999, $25,000-$99,999, and over $100,000) with probability in accordance with the distribution of brackets realized from respondents who gave an exact bracket. Note that this does not include people who gave an exact amount despite the fact that they fall into a bracket. The assumption is that people who give partial bracket information or no bracket information at all are more like individuals who gave exact bracket information than those that gave exact amounts. Juster and Smith (1997) argue that the distribution of bracketed individuals is different from those who give an exact amount.

After the second level, every respondent has an index of the exact bracket to which the respondent belongs. (Note: the dollar values defining the bracket ranges in the questions differ some between 1984 and 1989, but are identical between 1989 and 1994 and 1999. The choice of these bracket values was informed by the empirical distribution of the values of the assets in question in combination with 'round' numbers *and* restricting the questions to a parsimonious set of bracket categories.

Third, respondents who did not give exact amount (including those with reported exact brackets and imputed brackets) are assigned a dollar value with a probability derived from the distribution of amounts from respondents who reported exact values and which fell within the range of the same bracket.

Because we were working with an early-release version of the 1994 and 1999 family files, some additional work was done. Prior to the processing described above, some pre-processing imputation was done for missing values in house mortgage amounts. In addition, subsequent to the wealth processing, a listing of 1994 data with wealth component values ending in 7, 8 and 9 was prepared. Based on interviewer notes and other information, editing judgments were applied, placing these as either "don't know", "refused" or over-the-field amounts.

Due to data collection and processing conventions, in 1984 and 1994 negative values for component wealth variables were coded as zero. The modest number and amount of negative values reported for component wealth variables in the 1989 leads us to believe that 1984 and 1994 net wealth values would not change significantly had negative values been recorded. Recall that WEALTH1 [Sx16] does not include home equity whereas WEALTH2 [Sx17] does include home equity (market value of the owner-occupied housing less outstanding mortgage balances). No inflation adjustments - the CPI-U for 1984=103.9, 1989=124.0, 1994=148.2, 1999=166.6, and 2001=177.1.

5.2.2.3 2005

Release 2: March, 2011

The second release of the 2005 wealth dataset contains 38 variables. Accuracy codes were added for each component of total wealth as well as for the two totals (S702A-S720A), to indentify records which have been imputed/changed. In addition, home equity value is now included as a separate release variable (S720).

Release 1: May, 2007

Information from two sections - the housing section (A) and the wealth section (W) - were used to construct the wealth file. In 2005, a few changes were introduced in the housing section. Relevant for the wealth data file construction is a new series of unfolding brackets for cases when house value (A20) is not available. The questions A20A-A20E ask whether the house value is greater or less than some amount. The bracket amounts are $25,000, $75,000, $100,000, $200,000, and $400,000. Wealth questions in section W are the same as in the previous wave.

PSID asks about nine broad wealth categories, including short-term debt:

1. Equity in business (also includes farm) [W11, S703]
2. Transaction accounts (includes savings accounts, money market funds, certificates of deposit, government savings bonds, and treasury bills) [W28, S705]
3. Value of debt aside from mortgage on the main home or vehicle loans [W39, S707]
4. Equity in real estate (second home, land, rental real estate, or money owed on a land contract) [W2, S709]
5. Equity in stock (includes shares of stock in publicly held corporations, mutual funds, and investment trusts) [W16, S711]
6. Equity in vehicle (cars, trucks, a motor home, a trailer, or a boat) [W6, S713]
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, S715]
8. Equity in Individual Retirement Accounts [W22, S719]
9. Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2, but not released separately until March 2011) [S720]. 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 2005 wealth data we followed the same approach as in the prior waves. In particular, the hot-deck imputation technique was used for imputation of the missing data in the wealth components (1-8). Details on home equity imputation are given in Table 9.

Table 9. Number of cases by missing data group and imputation category: home equity, 2005

Missing data group	Imputation category								Total
	A	B	C	D	E	F	G	H	
1. House value is not missing; mortgage is missing	51	12	91	0	0	0	0	0	154
2. House value is missing; mortgage is not missing	20	5	0	11	0	0	0	0	36
3. House value is missing; there is no mortgage	46	7	0	0	36	11	0	0	100
4. Both house value and mortgage are missing	4	3	0	0	0	0	24	15	46
Total	121	27	91	11	36	11	24	15	336

There are four rows in Table 9 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 (category A). If this is not possible, a value from the wave prior to the last (2001) is used (category B). When this is also not possible, we use an assumed identity, Mortgage=0.6* HouseValue, 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 (category A). If the last wave information is not available, we use information from the wave prior to the last wave (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 .4 and the house value estimate based on the bracket information (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 Group 2 and 3, and the missing mortgage is imputed as in the Group1 (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 (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 (category H).

Descriptive statistics of 2005 imputed home-equity values

N	Mean	Std Dev	Min	Max
336	89467.79	89921.46	-4000.00	613800.00

5.2.2.4 2007

Release 2: March, 2011.

The second release of the 2007 wealth dataset contains 38 variables. Accuracy codes were added for each component of total wealth as well as for the two totals (S802A-S820A), to indentify records which have been imputed/changed. In addition, home equity value is now included as a separate release variable (S820).

Release 1: July, 2009

The questions used to generate the wealth variables were not changed between the 2005 and 2007 waves. Information from two sections - the housing section (A) and the wealth section (W) - were used to construct the 2007 wealth file. PSID asks about nine broad wealth categories, including short-term debt:

1. Equity in business (also includes farm) [W11, S803]
2. Transaction accounts (includes savings accounts, money market funds, certificates of deposit, government savings bonds, and treasury bills) [W28, S805]
3. Value of debt aside from mortgage on the main home or vehicle loans [W39, S807]
4. Equity in real estate (second home, land, rental real estate, or money owed on a land contract) [W2, S809]
5. Equity in stock (includes shares of stock in publicly held corporations, mutual funds, and investment trusts) [W16, S811]

6. Equity in vehicle (cars, trucks, a motor home, a trailer, or a boat) [W6, S813]
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, S815]
8. Equity in Individual Retirement Accounts [W22, S819]
9. Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2, but not released separately until March 2011) [S820]. 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 2007 wealth data we followed the same approach as in the prior waves. In particular, the hot-deck imputation technique was used for imputation of the missing data in the wealth components (1-8). Details on home equity imputation are given in Table 10.

Table 10. Number of cases by missing data group and imputation category: home equity, 2007

Missing data group	Imputation category								Total
	A	B	C	D	E	F	G	H	
1. House value is not missing; mortgage is missing	40	17	94	0	0	0	0	0	151
2. House value is missing; mortgage is not missing	25	3	0	18	0	0	0	0	46
3. House value is missing; there is no mortgage	35	10	0	0	44	9	0	0	98
4. Both house value and mortgage are missing	5	2	0	0	0	0	31	20	58
Total	105	32	94	18	44	9	31	20	353

There are four rows in Table 10 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 (category A). If this is not possible, a value from the wave prior to the last (2003) is used (category B). When this is also not possible, we use an assumed identity, Mortgage=0.6* HouseValue, 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 (category A). If the last wave information is not available, we use information from the wave prior to the

last wave (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 .4 and the house value estimate based on the bracket information (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 Group 2 and 3, and the missing mortgage is imputed as in the Group1 (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 (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 (category H).

Descriptive statistics of 2007 imputed home-equity values

N	Mean	Std Dev	Min	Max
353	107697.66	107200.05	-30000.00	650000.00

5.2.2.5 2009

The questions used to generate the wealth variables were not changed between the 2007 and 2009 waves. Information from two sections - the housing section (A) and the wealth section (W) - were used to construct the 2009 wealth file. PSID asks about nine broad wealth categories, including short-term debt:

1. Equity in business (also includes farm) [W11, ER46938, corresponds to S803]
2. Transaction accounts (includes savings accounts, money market funds, certificates of deposit, government savings bonds, and treasury bills) [W28, ER46942, corresponds to S805]
3. Value of debt aside from mortgage on the main home or vehicle loans [W39, ER46946, corresponds to S807]
4. Equity in real estate (second home, land, rental real estate, or money owed on a land contract) [W2, ER46950, corresponds to S809]

5. Equity in stock (includes shares of stock in publicly held corporations, mutual funds, and investment trusts) [W16, ER46954, corresponds to S811]
6. Equity in vehicle (cars, trucks, a motor home, a trailer, or a boat) [W6, ER46956, corresponds to S813]
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, ER46960, corresponds to S815]
8. Equity in Individual Retirement Accounts [W22, ER46964, corresponds to S819]
9. Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2, but not released separately until March 2011) [ER46966, corresponds to S820]

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 2009 wealth data we followed the same approach as in prior waves. In particular, the hot-deck imputation technique was used for imputation of the missing data in the wealth components (1-8). Details on home equity imputation, including the numbers of cases imputed within each group, are given in Table 11.

Table 11. Number of cases by missing data group and imputation category: home equity, 2009

Missing data group	Imputation category								Total
	A	B	C	D	E	F	G	H	
1. House value is not missing; mortgage is missing	50	9	78	0	0	0	0	0	137
2. House value is missing; mortgage is not missing	40	3	0	16	0	0	0	0	59
3. House value is missing; there is no mortgage	38	7	0	0	26	9	0	0	80
4. Both house value and mortgage are missing	6	5	0	0	0	0	38	26	75
Total	134	24	78	16	26	9	38	26	351

There are four rows 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 (category A). If this is not possible, a value from the wave prior to the last (2005) is used (category B). When this is also not possible, we use an assumed identity, Mortgage=0.6* HouseValue, 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 (category A). If the last wave information is not available, we use information from the wave prior to the last wave (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 (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 (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 (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 (category H).

Descriptive statistics of 2009 imputed home-equity values

N	Mean	Std Dev	Min	Max
351	95103.29	121618.24	-280000.00	795750.00

5.2.2.6 2011

The questions used to generate the wealth variables were changed slightly between the 2009 and 2011 waves. The W38-W42 series on "other debt" was replaced with a series of questions for specific sub-components of debt: credit cards, student loans, medical expenses, legal expenses, and loans from relatives. Each of these sub-components was edited/imputed according to the same process as the other components of wealth.

Information from two sections - the housing section (A) and the wealth section (W) - were used to construct the 2011 wealth file. PSID asks about nine broad wealth categories, including short-term debt:

1. Equity in business (also includes farm) [W11, ER52346].
2. Transaction accounts (includes savings accounts, money market funds, certificates of deposit, government savings bonds, and treasury bills) [W28, ER52350].
3. Value of debt aside from mortgage on the main home or vehicle loans, now divided into sub-components: credit card debt [W39A, ER52372], student loan debt [W39B1, ER52376], medical bills [W39B2, ER52380], legal bills [W39B3, ER52384], and loans from relatives [W39B4, ER52388].
4. Equity in real estate (second home, land, rental real estate, or money owed on a land contract) [W2, ER52354].
5. Equity in stock (includes shares of stock in publicly held corporations, mutual funds, and investment trusts) [W16, ER52358].
6. Equity in vehicle (cars, trucks, a motor home, a trailer, or a boat) [W6, ER52360].
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, ER52364].
8. Equity in Individual Retirement Accounts [W22, ER52368].
9. Value of home equity (calculated as home value minus remaining mortgage; used in calculation of WEALTH2, but not released separately until March 2011) [ER52390].

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 2011 wealth data we followed the same approach as in prior waves. In particular, the hot-deck imputation technique was used for imputation of the missing data in the wealth components (1-8). Details on home equity imputation, including the numbers of cases imputed within each group, are given in Table 12.

Table 12. Number of cases by missing data group and imputation category: home equity, 2011

Missing data group	Imputation category								Total
	A	B	C	D	E	F	G	H	
1. House value is not missing; mortgage is missing	37	6	51	0	0	0	0	0	94
2. House value is missing; mortgage is not missing	27	6	0	8	0	0	0	0	41
3. House value is missing; there is no mortgage	34	8	0	0	31	15	0	0	88
4. Both house value and mortgage are missing	9	3	0	0	0	0	20	13	45
Total	107	23	51	8	31	15	20	13	268

There are four rows 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 (category A). If this is not possible, a value from the wave prior to the last (2007) is used (category B). When this is also not possible, we use an assumed identity, Mortgage=0.6* HouseValue, 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 (category A). If the last wave information is not available, we use information from the wave prior to the last wave (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 (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 (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 (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 (category H).

Descriptive statistics of 2011 imputed home-equity values

N	Mean	Std Dev	Min	Max
268	92004.13	132564.80	-384000.00	1024650.00

5.2.3 All other coded or generated variables

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

5.2.3.1 2007

Head's and Wife's/"Wife's" occupations and industries, to a maximum of four jobs apiece, are coded for 2007 using the three-digit codes from the 2000 Census Occupation and Industry Code. 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 Head's and Wife's/"Wife's" non-academic degrees and certificates is coded for as many as three each. Any education received outside the U.S. is coded for Head, Wife/"Wife" and their parents.

Background items such as education are collected only for new Heads and new Wives/"Wives" in a given wave. During processing, we have traditionally "brought forward" background information from previous waves for Heads or Wives/"Wives" who are the same persons as in the prior year. 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 ER40142 for Heads and ER40143 for Wives/"Wives".

Completed education of Head and Wife/"Wife" (ER41037 and ER41038) are generated from the background information.

Family composition and change variables include Family Composition Change (ER36007), Splitoff Indicator (ER36005), Head-Spouse Sample Status (ER41044), and variables about births to Head, spouse and other family members during the prior calendar year, 2006 (ER41060-ER41063), and during the "off" year, 2005 (ER41064-ER41067). Note that the Splitoff Indicator is only assigned to a family in the year they first move out of the main family and form their own separate household; after that one wave of being a splitoff, these families receive code values that designate reinterviews.

Two other variables concerning splitoffs are the number of splitoffs arising from a main family (ER41045) and the family interview number of the main family from which a splitoff family originated (ER41046).

The PSID produces sets of variables about families sharing the same household: family ID numbers, relationships, and sizes of up to four other PSID families sharing the HU (ER41047-ER41058), the household ID number (ER41059), and the number of persons not included in any PSID family who are sharing the household unit (ER36022). 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 Head's current marital status (ER36023), the generated form of marital status comparable with years prior to 1977 (ER41039), change in marital status of Head between waves (ER41040), and couple status of Head (ER41041).

Location variables include PSID/GSA and FIPS state codes (ER36003 and ER36004); Current Region, Beale Rural-Urban code, and Size of the Largest City in the County (ER41032-ER41034); and two derived from background information, Region where Head Grew Up (ER41035) and Head's Geographic Mobility (ER41036). The codes for the FIPS and PSID/GSA codes are included in Appendix 1 of the 1985 documentation, available in .pdf format on our website.

Both USDA and Census needs standards were generated for the prior calendar year, 2006 (ER41028 and ER41029). Additionally, since the PSID has switched to biennial interviewing, comparable needs standards were also been generated for the "off" year, 2005 (ER41030 and ER41031).

The variable indicating whether a PSID family lives in institutional housing is present (ER36008), as is the variable indicating the total number of data records from the cross-year individual file that are associated with a panel family (ER40872). Weights are available as part of the 2007 Family File in ER41069. The imputed work hours and income variables are at (ER40873- ER41026) .

Background information (Section K for Wives/"Wives" and Section L for Heads) is only asked when a person newly acquires a relationship of Head or Wife/"Wife", values are copied forward from prior data collections for those Heads and Wives/"Wives" who maintain their relationship classification from the preceding wave.

A new variable to indicate whether this family has a record on the restricted Katrina data file was added (ER41068).

5.2.3.2 2009

Head's and Wife's/"Wife's" occupations and industries, to a maximum of four jobs apiece, are coded for 2009 using the three-digit codes from the 2000 Census Occupation and Industry Code. 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 Head's and Wife's/"Wife's" non-academic degrees and certificates is coded for as many as three each. Any education received outside the U.S. is coded for Head, Wife/"Wife" and their parents.

Background items such as education are collected only for new Heads and new Wives/"Wives" in a given wave. During processing, we have traditionally "brought forward" background information from previous waves for Heads or Wives/"Wives" who are the same persons as in the prior year. 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 ER46986 for Heads and ER46987 for Wives/"Wives".

Completed education of Head and Wife/"Wife" (ER46981 and ER46982) are generated from the background information.

Family composition and change variables include Family Composition Change (ER42007), Splitoff Indicator (ER42005), Head-Spouse Sample Status (ER46988), and variables about births to Head, spouse and other family members during the prior calendar year, 2008 (ER47004- ER47007), and during the "off" year, 2007 (ER47008- ER47010). Note that the Splitoff Indicator is only assigned to a family in the year they first move out of the main family and form their own separate household; after that one wave of being a splitoff, these families receive code values that designate reinterviews.

Two other variables concerning splitoffs are the number of splitoffs arising from a main family (ER46989) and the family interview number of the main family from which a splitoff family originated (ER46990).

The PSID produces sets of variables about families sharing the same household: family ID numbers, relationships, and sizes of up to four other PSID families sharing the HU (ER46991- ER47002), the household ID number (ER47003), and the number of persons not included in any PSID family who are sharing the household unit (ER42020). 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 Head's current marital status (ER42023), the generated form of marital status comparable with years prior to 1977 (ER46983), change in marital status of Head between waves (ER46984), and couple status of Head (ER46985).

Location variables include PSID/GSA and FIPS state codes (ER42003 and ER42004); Current Region, Beale Rural-Urban code, and Size of the Largest City in the County (ER46974- ER46976); and two derived from background information, Region where Head (ER46977) or Wife Grew Up (ER46979)

and Head's (ER46978) and Wife's (ER46980) Geographic Mobility. The codes for the FIPS and PSID/GSA codes are included in Appendix 1 of the 1985 documentation, available in .pdf format on the PSID website.

The Census needs standard was generated for the prior calendar year, 2008 (ER46972). Additionally, since the PSID has switched to biennial interviewing, comparable needs standard was also been generated for the "off" year, 2007 (ER46973). The USDA needs standards variables have been dropped.

The variable indicating whether a PSID family lives in institutional housing is present (ER42008), as is the variable indicating the total number of data records from the cross-year individual file that are associated with a panel family (ER46760). Weights are available as part of the 2009 Family File in ER47012. The imputed work hours and income variables are found at (ER46761- ER46935).

Background information (Section K for Wives/"Wives" and Section L for Heads) is only asked when a person newly acquires a relationship of Head or Wife/"Wife;" values are copied forward from prior data collections for those Heads and Wives/"Wives" who maintain their relationship classification from the preceding wave.

A new variable to indicate whether this family has a record in the DUST supplement was added (ER46696).

Now that all Wives have been asked questions about where they grew up in the background section, we have generated Region where Wife Grew Up (ER46979) and Wife's Geographic Mobility (ER46980).

Also new for 2009 is a set of variables generated from the Employment EHC start and stop dates for the "off" year, 2007. These variables include whether Head or Wife were Employed, Unemployed, or Out of the Labor Force, and the number of weeks (ER4665-ER46682).

The wealth file contains imputed versions of items from Section W, which has been released in prior waves as a supplement is now part of the 2009 Family File (ER46936- ER46971).

5.2.3.3 2011

In this wave, background information (Section K for Wives/"Wives" and Section L for Heads) is only asked when a person newly acquires a relationship of Head or Wife/"Wife". Values are copied forward from prior data collections for those Heads and Wives/"Wives" who maintain their relationship classification from the preceding wave.

In 2011, a set of variables generated from the Employment EHC start and stop dates for the "off" year, 2009 are provided. These variables include whether Head or Wife were Employed, Unemployed, or Out of the Labor Force, and the number of weeks (ER52066-ER52084).

The wealth file contains imputed versions of items from Section W, which has been released in waves prior to 2009 as supplements continues be part of the 2011 Family File (ER52344-ER52395). Some

modifications were made to the imputed wealth variables that correspond to the changes in the wealth section at W38a- W39b4.

The PSID hand codes several data items for the family file. Head's and Wife's/"Wife's" occupations and industries, to a maximum of four jobs each, are coded for 2011 using the three-digit codes from the 2000 Census Occupation and Industry Code. 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 Head's and Wife's/"Wife's" non-academic degrees and certificates is coded for as many as three each. Any education received outside the U.S. is coded for Head, Wife/"Wife" and their parents.

Background items such as education are collected only for new Heads and new Wives/"Wives" in a given wave. During processing, we have traditionally "brought forward" background information from previous waves for Heads or Wives/"Wives" who are the same persons as in the prior year. 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 ER52410 for Heads and ER52411 for Wives/"Wives".

Completed education of Head and Wife/"Wife" (ER52405 and ER52406) are generated from the background information.

Family composition and change variables include Family Composition Change (ER47307), Splitoff Indicator (ER47305), Head-Spouse Sample Status (ER52412), and variables about births to Head, spouse and other family members during the prior calendar year, 2010 (ER52428- ER52431), and during the "off" year, 2009 (ER52432-ER52434). Note that the Splitoff Indicator is only assigned to a family in the year they first move out of the main family and form their own separate household; after that one wave of being a splitoff, these families receive code values that designate reinterviews.

Two other variables concerning splitoffs are the number of splitoffs arising from a main family (ER52413) and the family interview number of the main family from which a splitoff family originated (ER52414).

The PSID produces sets of variables about families sharing the same household: family ID numbers, relationships, and sizes of up to four other PSID families sharing the HU (ER52415-ER52426), the household ID number (ER52427), and the number of persons not included in any PSID family who are sharing the household unit (ER47320). 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 group consists of Head's current marital status (ER47323), the generated form of marital status comparable with years prior to 1977 (ER52407), change in marital status of Head between waves (ER52408), and couple status of Head (ER52409).

Location variables include PSID/GSA and FIPS state codes (ER47303 and ER47304); Current Region, Beale Rural-Urban code, and Size of the Largest City in the County (ER52398- ER52400); and two derived from background information, Region where Head (ER52401) or Wife Grew Up (ER52403) and Head's (ER52402) and Wife's (ER52404) Geographic Mobility. The codes for the FIPS and PSID/GSA codes are included in Appendix 1 of the 1985 documentation, available in .pdf format on our website.

The Census needs standard was generated for the prior calendar year, 2010 (ER52396). Additionally, since the PSID has switched to biennial interviewing, comparable needs standard was also been generated for the "off" year, 2009 (ER52397).

The variable indicating whether a PSID family lives in institutional housing is present (ER47308), as is the variable indicating the total number of data records from the cross-year individual file that are associated with a panel family (ER52168).

Weights are available as part of the 2011 Family File in ER52436.

The imputed work hours and income variables are found at (ER52169-ER52343).

6. PUBLIC USE DATA AND DATA PROCESSING: CROSS-YEAR INDIVIDUAL FILE

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 2009

6.1.1 2009 questionnaire changes

The majority of questionnaire changes in 2009 affected the Family data file. Very few questionnaire changes affected the Individual data file.

In 2009, we have added individual level social security variables, ER34030 "G33 TYPE OF SOC SEC RCD 09", ER34031 "G34 AMT SOC SEC RCD 09" and ER34032 "G34 ACC SOC SEC AMT 09". These variables describe the type of social security received by the individual, the amount received, and an accuracy variable if it was imputed. These three social security variables were last released in 1992.

Another new set of variables include three questions from the interviewer observation section of the questionnaire. We ask the interviewer if any other individuals helped the respondent with completing the interview. We identify the sequence number of the Family Unit Member who helped the respondent up to three mentions: ER34012-ER34014. These helper variables are attached to the individual who was the respondent in 2009.

Three new variables were added in 2009 that describe whether or not an individual was included in the follow-up project Disability and Use of Time (DUST) that was conducted in 2009: the outcome variables, ER34040, "WTR ELIGIBLE FOR DUST 09", ER34041, "WTR SELECTED FOR DUST 09"

and ER34042, “RESULT OF DUST IW 09” which indicates which Heads and Wives from the main PSID project were eligible, of those which were selected, and the outcome of the DUST interview.

6.1.2 New generated variables

In 2009 we continue to include the generated ER34029, “WHETHER MEDICARE NUMBER GIVEN”. This variable provides a summary for those who were eligible to be asked for a Medicare number.

A new summary variable for the 1968-2009 Individual File is ER32050, “YEAR OF DEATH”. The source of this variable is the more detailed restricted use Cumulative Death File. For some analysis that does not require the more detailed mortality information and a confidential contract, we have provided year of death on the Public Release File.

For 2009, we also have the addition of three cross-sectional weights. We have provided cross-sectional weights for 2005: ER33849, 2007: ER33951, and 2009: ER34046 “CORE/IMM INDIVIDUAL CROSS-SECTION WT”.

6.1.3 Other additions

For 2009, children were eligible for the Transition to Adulthood (TA-2009) interview if they had been part of the 1997 Child Development Supplement, met age requirements and were part of a PSID family interviewed in 2009. Variables indicating 2009 eligibility (ER34038) and the result of the interview attempt (ER34039) are part of the 1968-2009 Individual data file.

6.2 What’s new for 2011

6.2.1 2011 questionnaire changes

The majority of questionnaire changes in 2011 affected the Family data file. Very few questionnaire changes affected the Individual data file. For specific information about questionnaire content, see the Introduction in the 2011 Public Release Family File documentation for details.

In 2011, multi-mention individual level social security type variables were added: ER34137 “G33A WTR SOC SEC TYPE DISABILITY 11”; ER34138 “G33A WTR SOC SEC TYPE RETIREMENT 11”; ER34139 “G33A WTR SOC SEC TYPE SURVIVOR 11”; ER34140 “G33A WTR SOC SEC TYPE DEP OF DISABLED 11”; ER34141 “G33A WTR SOC SEC TYPE DEP OF RETIRED 11”; ER34142 “G33A WTR SOC SEC TYPE OTHER 11”. These replace the single mention social security variable G33 that were included in prior waves. Variables describing the types of social security received by the individual, the amount, and an accuracy variable if it was imputed are continued: ER34143 “G34 AMT SOC SEC RCD 11” and ER34111 “G34 ACC SOC SEC AMT 11”.

Another new set of variables include a series of questions about health insurance coverage. For each individual in the Family Unit, a question is asked about whether they have health care insurance or coverage now ER34128 “H61D WTR COVERED BY INSURANCE NOW 11”. For those individuals covered, up to three mentions of the type of insurance or coverage are asked: ER34129-ER34131 “H61E TYPE CURRENT HEALTH INS MEN 1-3”. For those individuals who are covered by employer

provided health insurance, we then ask whose employer provides the insurance ER34132 “H61F SN 1ST PERSON EMP PROVIDES INS 11” and ER34133 "H61F SN 2ND PERSON EMP PROVIDES INS 11". Finally for those individuals who are covered by private health insurance, information is collected on the the policy holder for that private insurance: ER34134 “H61G SN 1ST PERSON POLICY HOLDER 11” and ER34135 "H61G SN 2ND PERSON POLICY HOLDER 11".

6.2.2 New generated variables

In 2011, the generated variable providing information about whether those who were eligible to be asked for a Medicare number provided one, continues to be generated: ER34136, “WHETHER MEDICARE NUMBER GIVEN”.

Longitudinal and cross-sectional weights are provided in 2011: ER34154 “CORE/IMM INDIVIDUAL LONGITUDINAL WT 11” and ER34155 CORE/IMM INDIVIDUAL CROSS-SECTION WT 11".

6.2.3 Other additions

Variables indicating 2011 eligibility for the 2011 Transition into Adulthood study (ER34150) and the result of the interview attempt (ER34151) are part of the 1968-2011 Individual data file. Children were eligible for the 2011 Transition into Adulthood study if they had been part of the 1997 Child Development Supplement, were at least 18 years of age, had completed or left high school, and were part of a PSID family interviewed in 2011.

In previous versions of the Individual File, the generated employment status variable for OFUMs was complete but was missing for all Heads and Wives for the years 1994-2009. For this release, we have filled in the prioritized employment status variable for Heads and Wives for all missing years. Specifically those variables are:

1994-ER33111

1995-ER33211

1996-ER33311

1997-ER33411

1999-ER33512

2001-ER33612

2003-ER33712

2005-ER33813

2007-ER33913

2009-ER34016

6.3 Data characteristics

6.3.1 Files and format

The 1968-2011 Individual data consists of one data file with 73,251 records and 1,499 variables. The data are merged across all waves of the study; that is, they include person information collected from

1968 through 2011. 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 2011 response and 2011 non-response individuals.

Using the web-based Data Center is the most efficient way to obtain the data. PSID continues to provide the data in a .zip package for the convenience of long-term users. If you download the data in ASCII form from the PSID website, you can choose whether to receive SAS, SPSS or STATA statements. In the .zip package, we include the entire data file in ASCII format and SAS, SPSS and STATA data definition statements. These statements provide variable names, locations, and variable labels.

The definition statement files are not intended to represent complete and full programs for the respective statistical program packages to run extracts, analysis, etc. You must provide all other statements needed to complete a program. Missing data statements are not provided as part of the files.

6.3.2 Variable names, positions, and generated variables

The 1968-2011 Individual data file consists of yearly items (e.g., Sequence Number, Relationship to Head, Family Identification Number) and a set of summary or cross-year variables with up-to-date values (ER32001-ER32050, Sex of Individual, ER32000).

Variables ER30001 through ER30866 remain the same as for all previous Public Releases (with the prefix change from "V" to "ER"). Weights for each wave of data are available through 2011 and are located at or near the end of each year's list of variables.

The order of variables is as follows:

* RELEASE NUMBER, ER30000,
* 1968 through 1993 Individual data arranged, as usual, by wave, ER30001-ER30866
* the summary variables, ER31990-ER31997 and ER32001-ER32050,
* the 1994 Public Release Individual data, ER33101-ER33150,
* the 1995 Public Release Individual data, ER33201-ER33299B,
* the 1996 Public Release Individual data, ER33301-ER33347,
* the 1997 Public Release Individual data, ER33401-ER33438,
* the 1999 Public Release Individual data, ER33501-ER33547,
* the 2001 Public Release Individual data, ER33601-ER33639,
* the 2003 Public Release Individual data, ER33701-ER33742,
* the 2005 Public Release Individual data, ER33801-ER33849,
* the 2007 Public Release Individual data, ER33901-ER33951,
* the 2009 Public Release Individual data, ER34001-ER34046
* the 2011 Public Release Individual data, ER34101-ER34155

6.3.3 Hand coded and generated variables

The summary variables fall into three groups: ER31990-ER31997 are used for sampling error and weights calculations; ER32001-ER32008 are true cross-year indicators derived from values in each individual's data record; and ER32009-ER32050 are summaries from the Family History files (these files are described below).

Other generated items include variables designating the originating family and move-out date for splitoffs for every wave from 1969 to 2011 (2011: ER34145-ER34146) and income summaries from 1968 to 1993.

For the 2011 release, we have provided generated employment status for Heads and Wives/"Wives" as well as OFUMs for all the years it had missing data 1994-2009 (ER33111, ER33211, ER33311, ER33411, ER33512, ER33612, ER33712, ER33813, ER33913, ER34016 and the 2011 employment status variable: ER34116). See the codebook for details about employment status variable construction.

6.3.4 Documentation, codes, and missing data

Box-and-arrow questionnaires are available on the PSID website. The Data Center allows users to create and download customized codebooks that provide variable-specific documentation. The downloadable .zip package of the Individual data on our website contains a .pdf version of the entire codebook for 2011 Individual file.

In general, codes follow our traditional schema, although "don't know" responses are now largely distinguished from other missing data responses. Generally, code 8 (or 98 or 998, etc.) represents "don't know" and code 9 (or 99 or 999, etc.) represents other missing data or a refusal. Inappropriate questions are padded with zeroes. If a variable contains a code value that is neither included in the codebook nor one of the "zero", "eight" or "nine" codes just mentioned, assume missing data for that value; this happens rarely.

From 1968 through 1993, Heads and Wives/"Wives" have values for the equivalent employment status variables but, for 1994 through 2009, they are zero. Several mentions for employment status are available on the 1994 through 2011 family files.

7. PUBLIC USE DATA AND DATA PROCESSING: CHILDBIRTH AND ADOPTION HISTORY FILE, 1985 – 2011

7.1 Overview

The Childbirth and Adoption History File is designed to facilitate access to information collected in the 1985 through 2011 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 2011.

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 2011, 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.

The number of variables are few; they contain relevant identifiers for the individual and his or her children; dates of birth for both parent and child; geographic identifiers about the place of birth; the child's birth order, birth weight, race, ethnicity and date of death; and year of most recent report and number of births or adoptions for the parent. Beginning with the 2005 wave, we also asked the child's Hispanicity (CAH19).

In 2007, an important change took place with the race question (CAH20 –CAH22). The code frame for race 2005 was as follows:

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

In order to maintain the cumulative file, we recoded records from 1997-2005 into the new 2007 frame.

Specifically, for those children who had a race mention of 5, Latino Origin or Descent, the following recodes were made:

a) the Hispanicity question (CAH19) was recoded to a new value of 6 (Latino, no further information), and

b) the race questions (CAH20-22) were recoded to 7 (Other)

The 1985-2011 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. This file is designed to link to the 1968-2011 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 been part of the PSID and hence have no record on the 1968-2011 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 Head, Wife, "Wife" and Other Family Unit Members (OFUMs). For particulars, see Section **(7.7)** Idiosyncrasies, Data Cleaning and Variable Detail below.

7.2 Individuals for whom the data are available

The childbirth and adoption history data were collected about individuals of childbearing age in responding families, i.e., those with values in the range 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 for whom childbirth and adoption information was asked and the detail gathered about their history differed somewhat during the collection period from 1985 through 2011; 1985 followed one pattern, and 1986-2011 followed another. In addition, beginning in 1997, some new questions were added for births and adoptions reported during that wave. In 2005 Hispanicity question was added and in 2007 the code frame for the race questions were altered.

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. Childbirth history information was gathered as described below:

(a) in the 1985 wave, a complete retrospective birth history was asked for a Head, Wife, or "Wife" of any age;

(b) in all succeeding waves, birth history was updated for changes since the prior calendar year for a female Head, Wife, or "Wife" aged 44 or younger who also had been either a Head, Wife, or "Wife" in the prior wave's interview;

(c) in succeeding waves, birth history was updated for changes since the prior calendar year for a male Head of any age who was also a Head in the prior wave's interview, unless he was married to a Wife aged 45 or older who had also been his Wife in the prior wave's interview;

(d) in succeeding waves, a complete retrospective birth history was reported for a New Head, New Wife, or New "Wife" of any age; and

(e) in all waves from 1985 through the present, a complete retrospective birth history was reported for an Other Family Unit Member (OFUM) aged 12-44 at the time of the interview.

Adoption history data were gathered in a fashion similar to childbirth history, except that information was collected for PSID family-unit Heads, Wives and "Wives", but not for OFUMs (type e

above). 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, although most are people who have been 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. The entry age for eligibility is 12 and the exit threshold for eligibility is 45 for many persons.

(c) A PSID individual can change his or her relationship to Head 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--information about adoptions is gathered for Heads, Wives, and "Wives", but not for OFUMs.

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

(f) While both Heads and Wives/"Wives" were interviewed in 1985, only one person (usually the Head) has been the respondent in each year since then.

In the 1990 wave, a sample of 2,043 Latino households was added to the PSID sample to provide the representative information about Latinos that is now available for blacks and non-Latino whites in the original PSID. The Latino sample size was increased to 2,308 households in 1992. This sample was dropped from the PSID in 1996. The childbirth and adoption history data of the Latino sample are also included in this file.

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, pp. 1-3; but briefly, four types of recontact samples were selected for inclusion in this round of interviewing. They differed in the detail gathered about their childbirth and adoption history.

(a) 1992 Reinterview recontact families who were last interviewed during the 1990 wave, but attrited in 1991.

(b) 1992 Splitoff recontacts who were sample members who moved out of a responding 1991 family, but who were not interviewed themselves in 1991.

(c) 1992 Sample recontacts who were designated as members of original 1968 Panel families who were last interviewed at some time between 1969 and 1989.

(d) 1992 Latino recontacts who were families in the original Latino National Political Survey sample, but either refused or were never contacted by the PSID during the 1990 wave when the Latino sample was added to the study.

Splitoff, Sample, and Latino recontacts (the last three types described above), like all other New Heads/Wives/"Wives", were asked complete retrospective birth and adoption histories. The birth and adoption histories of Reinterview recontacts (the first type above) were updated for changes since January 1991. Since this group of people was interviewed in 1990, but not 1991, information was lost for births or adoptions occurring between the 1990 interview and January of 1991. The unaccounted-for time varies from a month or two up to a maximum of nine to ten months. Reinterview recontacts can be identified on the cross-year Individual Files by data values of 2 or 3 for "1992 Follow Status" (V30799/ER30799).

In 1993, 1994, and from 1997 on, the PSID also recontacted attritors, but these individuals were asked to complete retrospective birth and adoption histories.

7.3 Background for the childbirth and adoption history files

The 1985-2011 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, excepting substitute parenting, were updated for eligible individuals.

All the retrospective data collected in 1985 on these demographic phenomena are included in the 1985 Ego-Alter File. This file is mostly of interest for the substitute parenting information and for child-specific information on public program participation and health care surrounding a birth. The program participation and health care questions include whether prenatal care was received and where it was received, whether the mother had private health insurance, and whether she received Medicaid, WIC, food stamps, free government food, ADC/AFDC, or other public assistance. These questions were collected about children reported in the 1985 interview if the parent was a female Head, Wife, or "Wife" in 1985 and the child was born between January, 1979 and the time of the 1985 interview.

Both the substitute parenting and the program participation/health care sequences were discontinued after the initial retrospective in 1985. For more details about the 1985 Ego-Alter File, refer to A Panel Study of Income Dynamics: 1985 Ego-Alter File Documentation.

Beginning in 1986, we decided to release the demographic history data 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.

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

The 1985-2011 Childbirth and Adoption File is available on the Data page of the PSID website.

7.5 Questionnaire detail

The flow of the PSID questionnaire is complex, and the types of individuals asked for history information and the detail gathered about their history changed. The analyst may find it helpful to actually see the questionnaires. The 1985 sequence is available in the PDF format version of the main 1985 documentation volume on pp. 23-31 and pp. 58-68, located on our website. 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 Hispanicity was added. PDF format versions of the 1991-1996 main questionnaires are also available on our website. They contain the questions for Head/Wife/"Wife" marriage history but do not include the supplementary forms used for details about childbirth and adoptions, or for OFUMs' marriages and childbirths. These supplements are not available on the website as of this writing.

7.6 Structure of the file

The 1985-2011 Childbirth and Adoption History File contains a total of 124,325 records, with 91,763 childbirth records and 32,562 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 children may also be an adoptive parent. The maximum number of childbirth records for a specific individual is 18 and, for adoption records, it is five.

Multiple records for a given child also exist 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.

The 1985-2011 Childbirth and Adoption File is sorted, in ascending order, by "1968 Interview Number of Parent" (CAH2) and "Person Number of Parent" (CAH3), "Type of Record" (CAH1), birth order (CAH8), birth year and month for the child (CAH13 and CAH12), and "Person Number of Child" (CAH10). The childbirth records for an individual are followed by his or her adoption records. The record for an individual's oldest child is followed by that for the second oldest child, etc. All records for the first eligible member of a 1968 PSID family are followed by all records for the next eligible member in the same family of origin. When all of the childbirth and adoption records for all eligible members in the first family are exhausted, records for eligible members in the second family follow.

The 1985-2011 Childbirth and Adoption History File contains 31 variables, which can be categorized into those relevant to the parent, those specific to the child, and some aggregate information about the childbirth or adoption history for the parent.

NOTE: Beginning with the 2005 wave, variable names changed at CAH19 because Hispanicity of the child was added to the questionnaire.

Information specific to the parent:

CAH2 1968 Interview Number of Parent

CAH3 Person Number of Parent

CAH4 Sex of Parent

CAH5 Month Parent Born

CAH6 Year Parent Born

CAH7 Marital Status of Mother When Ind Born (applicable only to childbirth records)

Information specific to the child:

CAH8 Birth Order (childbirth records only)

CAH9 1968 Interview Number of Child

CAH10 Person Number of Child

CAH11 Sex of Child

CAH12 Month Child Born

CAH13 Year Child Born

CAH14 Birth Weight of Child in Ounces (childbirth records only)

CAH15 State in Which Child Born

CAH16 Where Child Was When Last Reported

CAH17 Month Child Moved Out or Died

CAH18 Year Child Moved Out or Died

CAH19 Hispanicity

CAH20 Race of Child, 1st Mention

CAH21 Race of Child, 2nd Mention

CAH22 Race of Child, 3rd Mention

CAH23 Primary Ethnic Group of Child

CAH24 Secondary Ethnic Group of Child, 1st Mention

CAH25 Secondary Ethnic Group of Child, 2nd Mention

CAH27 Year Most Recently Reported This Child

CAH29 Relationship to Adoptive Parent (adoption records only)

Other Information:

CAH1 Record Type (childbirth or adoption)

CAH26 Year Most Recently Reported Number of Children

CAH28 Number of Natural or Adopted Children

CAH30 Number of Birth or Adoption Records

CAH31 Release Number

7.7 Idiosyncrasies, data cleaning, and variable details

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 child

One caution, particularly relevant to event-history analysis, concerns the records for individuals who have never had or adopted any 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 (CAH9 and CAH10); both of these variables are padded with zeroes.

7.7.2 How to identify individuals for whom birth or adoptions were not ascertained

Persons who may or 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 (CAH30) 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 (CAH9) and Child Person Number (CAH10). These two variables contain values of 9s, which do uniquely identify this sort of record.

7.7.3 Treatment of individuals who become nonrespondents

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 up to date only through the last year that nonresponse individuals were living in a responding family.

The childbirth and adoption history is current through 2011 for those in a responding PSID family at the time of the 2011 interview and who are otherwise eligible for the childbirth and/or adoption history questions. For those who were nonresponse in the 2011 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 non-response 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 CAH26.

7.7.4 Location data about the child's place of birth

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 1 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 Childbirth and Adoption History File. The county variable is classified as sensitive. Access to sensitive data must be obtained by a special request and confidential data use contract.

For further information, contact PSID staff by e-mail: psidhelp@umich.edu

7.7.5 Idiosyncrasies, data cleaning, and treatment of incomplete or inconsistent information

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, 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 (CAH16), and death date (CAH17-CAH18) 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, CAH27, 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 gender or birth date, values from the cross-year Individual File were used. Additionally, parental reports of a child's gender 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-one 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 900-990), 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.

The gender and birth date checks resulted in many corrections to both the Childbirth and Adoption History File and the cross-year Individual File. As an added benefit, we were also able to find and correct some spurious child identifiers.

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 1985-2011 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?

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 900-990. Consequently, each parent has been in a PSID family unit and has a record on the 1968-2011 Individual File, but his or her child may or may not.

7.7.7 What cross-year files to use for merging

The 1985-2011 Childbirth and Adoption History File matches the 1968-2011 Public Release Individual File exactly. If you attempt to merge the 1985-2011 Childbirth and Adoption History File with an earlier Individual file, some cases on the Individual File will NOT match and may assign erroneous parent-child relationships! See the next section for details.

7.7.8 Birth order and number of children

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 (CAH8) 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 Head moves into the study. At that time we receive a report about two children living with his ex-wife, but we do not obtain information about their birth dates. The children are assigned values of 98 for birth order (CAH8) because we don't know which one is older. In the next wave, the Head and Wife 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 (CAH28). 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 2011 file has 80 parents who fall into this category.

7.7.9 Adoption dates

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

7.8.1 Using the childbirth and adoption history file in conjunction with the individual file

The Childbirth and Adoption History File is designed to be linked with the Individual File for analysis purposes. The Childbirth and Adoption History File has only a modest amount of information about the parent and his or her child. 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.

When matching the 1985-2011 Childbirth and Adoption History File to individual data, only the 1968-2011 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 1985-2011 Marriage History File, and the 1968-2011 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 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" (CAH2) and "Person Number" (CAH3). The corresponding variables for these identifiers on the cross-year Individual File are V30001 and V30002 (or 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 Section 8 on "Family Composition and Change" in Hill (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 1968-2011 Individual File, but less than the full range (35 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 non-response.

As noted in the previously, all eligible individuals have records on the 1968-2011 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 1968-2011 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 900-990.

Even though all parents have records on the 1968-2011 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 V30001 and V30002 (or ER30001 and ER30002) on the Individual File. On the Childbirth and Adoption File, they are CAH2 and CAH3 for the parent and CAH9 and CAH10 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 the other demographic history files

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" (MH1 and MH2 for the person designated as the individual on the 1985-2009 Marriage History File; CAH2 and CAH3 for the person designated as the parent on the 1985-2011 Childbirth and Adoption History File).

7.9 Childbirth information available on the individual and family files

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 Files. 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.

7.9.1 Individual-level childbirth variables

The following listing shows all of the variables available on current cross-year Individual Files relating to childbirth.

ER32009	1968 ID of Mother
ER32010	Person Number of Mother
ER32011	Year Mother Born
ER32012	Total Number of Children Born To Mother
ER32013	Rank Order of Birth to Mother
ER32014	Birth Weight of This Individual
ER32015	Marital Status of Mother When Individual Born
ER32016	1968 ID of Father
ER32017	Person Number of Father
ER32018	Year Father Born
ER32019	Total Number of Children Born to Father
ER32020	Rank Order of Birth to Father
ER32021	Year Birth Info Most Recently Updated
ER32022	Number of Live Births to This Individual
ER32023 ER32024	Month and Year First/Only Child Born
ER32025 ER32026	Month and Year Youngest Child Born
ER32027 ER32028	Month and Year Second Youngest Child Born
ER32029 ER32030	Month and Year Third Youngest Child Born
ER32031 ER32032	Month and Year Fourth Youngest Child Born

ER32021-ER32032 are compiled from the birth history information collected from 1985 through the year of final release data. ER32009-ER32015 contain childbirth data for the individual's mother, and ER32016-ER32020 contain similar information for his or her father.

7.9.2 Family-level childbirth variables

Four variables are generated for each Public Release family file utilizing childbirth information current through that wave. They are:

Number of Children Born to Head Only in the Prior Calendar Year

Number of Children Born to Wife/"Wife" Only in the Prior Calendar Year

Number of Children Born Jointly to Head and Wife/"Wife" in the Prior Calendar Year

Number of Children Born to Other Family Unit Members in the Prior Calendar Year

7.10 Codebook

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. PUBLIC USE DATA AND DATA PROCESSING: MARRIAGE HISTORY FILE

8.1 Overview

The 1985-2011 Marriage History File is designed to facilitate access to information collected in the 1985 through 2011 waves of the PSID regarding retrospective histories of marriages. This 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 2011. The PSID definition of "eligible" is described in detail below.

Each set of records for a specified individual contains all known cumulative data about the timing and circumstances of his or her marriages up to and including 2011, or those waves during that period when the individual was in a responding family unit. If an individual has never been married, one record indicates that report.

The number of variables are few; they contain relevant identifiers for the individual and his or her spouse, dates of marriages/divorce/widowhood, the order of the specific marriage, total number of marriages and the most recent wave in which data were collected.

The 1985-2011 Marriage History File stores information in an efficient manner that allows a high degree of flexibility in linking with the PSID's Individual File. It is designed to be linked to the 1968-2011 Individual File. Linkages can be done from either the individual's or the spouse's standpoint. These linkages are more limited for spouses than for individuals, since some spouses have never been part of the PSID and hence have no record on the 1968-2011 Individual File.

Data users who want only some of the detail of the marriage data will find marriage history information summarized on Public Release versions of the PSID Individual files in several individual-level summary variables. These variables include number of marriages, beginning and ending dates for the first and the most recent marriages, marital status of the individual at the time of the most recent interview, and the most recent wave when marriage data were collected.

8.2 Individuals for whom the data are available

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 2011; 1985 followed one pattern and 1986 through 2011 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 Head, Wife, or "Wife" of any age;

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

(c) in succeeding waves, details about first and current or most recent marriages were asked for a New Head, New Wife, or New "Wife" 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.

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 Head, Wife, or "Wife", the exit threshold for eligibility is 45.

(c) A PSID individual can change his or her relationship to Head 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 Heads, Wives, and "Wives" but not other family members.

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

(f) While both Heads and Wives/"Wives" were interviewed in 1985, only one person (usually the Head) has been the respondent in each year since then.

In the 1990 wave, a sample of 2,043 Latino households was added to the PSID sample to provide the representative information about Latinos that is now available for blacks and non-Latino whites in the original PSID. The Latino sample size was increased to 2,308 households in 1992. This sample was dropped from the PSID in 1996. The marital history data of the Latino sample are also included in this file.

In the 1992 interview, the PSID undertook a pilot effort to recontact former respondents who had attrited from the study and persuade them to rejoin. Additional information regarding the 1992 recontact samples is discussed in the 1992 PSID Documentation, pp. 1-3; but briefly, four types of recontact samples were identified for inclusion in this round of interviewing. They differed in the detail gathered about their marriage history.

(a) 1992 Reinterview recontact families who were last interviewed during the 1990 wave, but attrited in 1991.

(b) 1992 Splitoff recontacts who were sample members who moved out of a responding 1991 family but who were not interviewed themselves in 1991.

(c) 1992 Sample recontacts who were designated as members of original 1968 Panel families who were last interviewed at some time between 1969 and 1989.

(d) 1992 Latino recontacts who were families in the original Latino National Political Survey sample but either refused or were never contacted by the PSID during the 1990 wave when the Latino sample was added to the study.

Splitoff, Sample, and Latino recontacts (the last three types described above), like all other New Heads/Wives/"Wives", were asked questions detailing their first and most recent marriages. The marriage history of Reinterview recontacts (the first type above) was updated for changes in 1991 and 1992. Since this latter group of people was interviewed in 1990, but not 1991, information was lost for events occurring between the 1990 interview and January of 1991. The unaccounted-for time varies from a month or two up to a maximum of nine to ten months. Reinterview recontacts can be identified on the cross-year Individual Files by data values of 2 or 3 for "1992 Follow Status" (V30799/ER30799).

In 1993, 1994, and from 1997 onward, the PSID also recontacted attriters, but these individuals were asked the same retrospective marriage histories as persons new to the study; that is, reports of their first and current or most recent marriages were sought.

8.3 Background for the marriage history files

The 1985-2011 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 are included in the 1985 Ego-Alter File. This file is mostly of interest for the 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. For more details about the 1985 Ego-Alter File, refer to A Panel Study of Income Dynamics: 1985 Ego-Alter File Documentation.

Beginning in 1986, we decided to release the demographic history data 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.

8.4 How to obtain the file and whom to contact about questions

The 1985-2011 Marriage History File is available on the Data page of the PSID website.

8.5 Questionnaire detail

The flow of the PSID questionnaire is complex, and the types of individuals asked history information and the detail gathered about their history has changed over time.

The analyst may find it helpful to actually see the questionnaires. The 1985 sequence is available in the PDF format version of the main 1985 documentation volume on pp. 23-31 and pp. 58-68, located on our website. The 1986-2011 sequences are identical to each other, with the addition, beginning in 1997, of questions about birth location, race and ethnicity. PDF format versions of the 1991-1996 main questionnaires are also available on our website. They contain the questions for Head/Wife/"Wife" marriage history, but do not include the supplementary forms used for details about childbirth and adoptions, or for OFUMs' marriages and childbirths. These supplements are not available on the website as of this writing.

8.6 Structure of the file

The 1985-2011 Marriage History File contains a total of 54,630 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 1985-2011 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.

The 1985-2011 Marriage History File is sorted, in ascending order, by "1968 Interview Number of Individual" (MH1), "Person Number of Individual" (MH2), and "Order of This Marriage" (MH8).

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 the Overview of the 1985-2009 Marriage History File, complete marriage histories were gathered for Heads, Wives and "Wives" in 1985, but information on only first and most recent marriages was initially collected for New Heads, Wives, and "Wives" in 1986 through 2011 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.

The 1985-2011 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.

Information specific to the individual:

MH1 1968 Interview Number of Individual

MH2 Person Number of Individual

MH3 Sex of Individual

MH4 Month Individual Born

MH5 Year Individual Born

Information specific to the spouse:

MH6 1968 Interview Number of Spouse

MH7 Person Number of Spouse

Information specific to the marriage:

MH8 Order of This Marriage

MH9 Month Married

MH10 Year Married

MH11 Status of This Marriage

MH12 Month Widowed or Divorced

MH13 Year Widowed or Divorced

MH14 Month Separated

MH15 Year Separated

Aggregate Information for the marriage history of this individual:

MH16 Year Most Recently Reported Marriage

MH17 Number of Marriages of This Individual

MH18 Last Known Marital Status

MH19 Number of Marriage Records

MH20 Release Number

8.7 Idiosyncrasies, file cleaning, and variable detail

Several aspects of the 1985-2011 Marriage History File merit particular attention. This section discusses what they are and how to handle them.

8.8 How to identify individuals who were never married

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 (MH17) has a value of zero. Codes indicating "Inapplicable" (9s) are padded in the fields for marriage details, with the exception of Spouse ID (MH6) and Spouse Person Number (MH7); both of these variables have values of zero.

8.9 How to identify individuals for whom no marriage data were ascertained

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 (MH17) 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 (MH6) and Spouse Person Number (MH7). These two variables contain values of 9s, which do uniquely identify this sort of record.

8.10 Treatment of individuals who become nonresponse

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 is current through 2011 for those in a responding PSID family at the time of the 2011 interview and who are otherwise eligible for marriage history questions. For those who were nonresponse in the 2011 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 47 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 MH16.

8.11 Idiosyncrasies, data cleaning, and treatment of incomplete or inconsistent information

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 currentness of the spouses' reports differs. The variable indicating the wave in which the marriage history was most recently updated is MH16. As an example, in 1990 a female sample member marries and her new husband, Head, 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 MH16. In each successive wave their marriage is reconfirmed and values for MH16 are updated until 1994, when they separate and divorce. He, the non-sample husband and former Head, leaves the PSID but the sample Wife 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 MH16 receives a value of 1994. But his record is not updated; his status remains married, and MH16 retains a value of 1993.

The preparation of the 1985-2011 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 131 cases, indeed the couples did not separate until after their divorces were finalized.

In 30 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 (MH11).

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 1985-2011 Childbirth and Adoption History File to ensure that no spurious cases of intergenerational incest occur. If both spouses had been in the study, their genders 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.

8.12 Who has cross-year information?

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" (MH7) in the range 900-990. Consequently, each individual has been in a PSID family unit and has a record on the 1968-2011 Individual File but his or her spouse may or may not.

8.13 What cross-year filed to use for merging

The 1985-2011 Marriage History File matches the 1968-2011 Individual File exactly. If you attempt to merge the 1985-2011 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.

8.14 Marriage in order

As described above, in the initial wave of demographic event history collection in 1985, all Heads, Wives and "Wives" were asked to provide details about all of their marriages. But in subsequent waves, the retrospective marriage history questions for new Heads, Wives and "Wives" 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 3844; 2663 of them have reported all their marriages, but 1181 have not.

The order of each marriage for an individual is indicated in MH8; values are assigned to each marriage in chronological order. The 1181 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 (MH17) are greater than values for the number of their records (MH19), 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, 307 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 (299). These persons have a known spouse at MH6-MH7, a known marital status at MH11, but the number of their marriages is not ascertained (MH17=98) and the number of records for them (MH19) 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 five 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 2009 wave, however, no person has more than one marriage with order unknown.

8.15 Linking records

8.15.1 Using the marriage history file in conjunction with the cross-year individual file

The Marriage History File is designed to be linked to 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.

When matching the 1985-2011 Marriage History File to individual data, only the 1968-2011 Individual File should be used. During file merging, a number of unique individual identifiers were corrected. Special care was taken to ensure perfect correspondence in individual identifiers between this file, the 1985-2011 Childbirth and Adoption History File, and Public Release of the 1968-2011 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 unique individual identifiers. Nonmatches 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" (MH1) and "Person Number" (MH2). The corresponding variables for this unique identifier on the cross-year Individual File are V30001 and V30002 (or 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.

Individuals and spouses 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 those files only in the years that the 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 Section 8 on "Family Composition and Change", p. 55, in Hill (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 individual or spouse, has a record on the 1968-2011 Individual File but less than the full range (35 waves) of data in that record, variables in the waves 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.

As noted previously, all eligible individuals have records on the 1968-2011 Individual File because they were present in a PSID family unit during that time period. However, many spouses do not have records on that file, since indeed they have never been in a PSID family at any time during the course of the study. Spouses who have been present in a PSID family unit at some time since the study began have records on the 1968-2011 Individual File. Values for their Person Numbers are in the range 001-399. Spouses who have never been present have Person Number values in the range 900-990.

Even though all individuals have records on the 1968-2011 Individual File, that file does not necessarily contain data for all of them for all years of the study. Some of them first entered the study in, e.g., 1985, and thus only have data since that year. Others have logged more than 30 years of inclusion in

the study. Similarly, there is considerable variation among spouses regarding which years, if any, they were included in the study.

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 V30001 and V30002 (or ER30001 and ER30002) on the Individual File. On the Marriage History File, they are MH1 and MH2, respectively, for the person designated "individual" and MH6 and MH7 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.

8.15.2 Using the marriage history file with the 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" (MH1 and MH2 for the person designated as the individual on the 1985-2011 Marriage History File; CAH2 and CAH3 for the person designated as the parent on the 1985-2011 Childbirth and Adoption History File).

8.16 Marriage information available on an individual files

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.

8.17 Codebook

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. PUBLIC USE DATA AND DATA PROCESSING: PARENT IDENTIFICATION FILE

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 1968-2011 Individual File. Linkages can be done from either the child's or a parent's standpoint.

9.1 Sources of parental identifier information

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 1968-2011 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 Head's and Wife's/"Wife'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 Head and Wife/"Wife" 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 Head or Wife/"Wife". 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.

9.2 How to obtain the file

The 2011 Parent Identification is available on the Data page of the PSID website.

9.3 Structure of the file

The 2011 Parent Identification File contains a total of 88,648 records. Included are all individuals from the 1968-2011 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" (PID1) and "Person Number of Individual" (PID2). These two Variables, taken together, constitute a unique identifier or each person and record.

The file contains 30 variables. Besides the pair of identifier variables for the child, four more sets of parental identifiers are present, one set each for birth and adoptive mothers and fathers. Variables indicating the source of the parental information are also included.

Child-specific information:

PID1 1968 Interview Number of Individual

PID2 Person Number of Individual

Mother-specific information:

PID3 1968 Interview Number of Birth Mother

PID4 Person Number of Birth Mother

PID5 1968 Interview Number of Adoptive Mother

PID6 Person Number of Adoptive Mother

PID7 WTR Birth Mom Coded in 1983/84 IW

PID8 WTR Birth Mom Coded in Fam Comp

PID9 WTR Adoptive Mom Coded in Fam Comp

PID10 WTR Mom Indicated by Interviewer

PID11 WTR Mom Coded as 1997 Support Receiver

PID12 WTR Mom Coded as 1997 Support Payer

PID13 WTR Mom Coded in 1988 Parent Supp

PID14 WTR Mom from 1988 Time/Money Supp

PID15 WTR Birth Mom Record in Birth History

PID16 WTR Adoptive Mom Record in Birth History Father-specific information:

PID17 1968 Interview Number of Birth Father

PID18 Person Number of Birth Father

PID19 1968 Interview Number of Adoptive Father

PID20 Person Number of Adoptive Father

PID21 WTR Birth Dad Coded in Fam Comp

PID22 WTR Adoptive Dad Coded in Fam Comp

PID23 WTR Dad Indicated by Interviewer

PID24 WTR Dad Coded as 1997 Support Receiver

PID25 WTR Dad Coded as 1997 Support Payer

PID26 WTR Dad Coded in 1988 Parent Supp

PID27 WTR Dad from 1988 Time/Money Supp

PID28 WTR Birth Dad Record in Birth History

PID29 WTR Adoptive Dad Record in Birth History Other Variable:

PID30 Release Number

9.4 Idiosyncrasies, data cleaning, and variable detail

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 88,648 individuals who do have 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, V32000 (or ER32000).

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 2011 Parent Identification File matches the 1968-2011 Public Release Individual File exactly. If you attempt to merge the 2011 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

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 2011 Parent Identification File to individual data, only the 1968-2011 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 1985-2011 Childbirth and Adoption History File, and the 1968-2011 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.

Data processing is required to link records between these two files. To accomplish linkages, one must match on the unique individual-specific identifier. This unique identifier is a combination of two variables: "1968 Family Interview Number" and "Person Number" (e.g., PID1 and PID2 for a child). The corresponding variables for this unique identifier on the cross-year Individual File are V30001 and V30002 (or ER30001 and ER30002). In doing this matching, care must be taken with regard to the proper files to use and the choice of individuals on the Parent Identification File for whom matches are attempted.

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 1968-2011 Individual File, that file does not necessarily contain data for them for all years of the study. Some first entered the study in, e.g., 1990, and thus only have data since that year. 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 the individual is 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 Hill (1992).

If a person, either child or parent, has a record on the 1968-2011 Individual File but less than the full 35-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 previously in Idiosyncrasies, Data Cleaning and Variable Detail , 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 900-996.

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 V30001 and 30002 (or ER30001 and ER30002) on the Individual File. On the Parent Identification File, they are PID1 and PID2, respectively, for the child and PID3 and PID4, PID5 and PID6, PID17 and PID18, or PID19 and PID20 for a parent.

9.6 Codebook

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

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, PSID creates longitudinal individual weights, longitudinal family weights, and cross-sectional individual weights. Documents describing the construction of the PSID weights are available on the questionnaire and supporting documents page of the PSID website.

11.SUPPLEMENTAL DATA COLLECTIONS

In addition to the main interview, numerous supplemental data collections have been conducted throughout the years. A supplemental data collection is defined as data collected through interviews not conducted as part of the main interview. In this section we provide an overview of each of the supplemental data collections that have been conducted since 1997 and point users to relevant documentation.

Child Development Supplement and Transition to Adulthood Study

Children have always played an important role in the PSID. However, only limited data were collected about children until the first wave of the PSID-Child Development Supplement (CDS-I) was fielded in 1997. The CDS-I collected information on up to two randomly selected 0-12-year-old children residing in PSID families, resulting in a total sample of 3,563 children in roughly 2,400 families. The second wave of the CDS was fielded in 2002/2003 with a response rate among primary caregivers of 91 percent. The same children originally interviewed in PSID-CDS-I, aged 5-18 in 2002/2003, were re-interviewed. A third wave – CDS-III - was collected in 2007 for those children in this birth cohort who were under 18 or were still in high school.

The CDS interviews of children, their parents, and teachers included extensive assessments in a variety of domains: (1) cognitive, behavioral, and health status; (2) assessment of basic skills in reading, mathematics, and written language; (3) a comprehensive accounting of parental and caregiver parenting styles and time inputs; (4) two 24 hour time diaries for each child – one for a random weekday and another for a random weekend day; (5) other-than-time-use measures of other resources, for example, the learning environment in the home; teacher and administrator reports of school resources and decennial-census-based measurement of neighborhood resources; and (6) child-reported indicators of socio-emotional characteristics.

When children in the CDS cohort are older than 18 and have left high school, information is obtained about their circumstances through a telephone interview completed shortly after the Main Interview. This study, called Transition into Adulthood (TA), has been implemented in 2005, and biennially thereafter. Information includes measures of time use, psychological functioning, marriage, family, responsibilities, employment and income, education and career goals, health, social environment, religiosity, and outlook on life.

All CDS and TA documentation and data are available on the PSID website.

Disability and Use of Time Study

The Disability and Use of Time Study (DUST) collected information from older couples in PSID families about disability, time use, and well being. The study was conducted immediately following the collection of data during the fielding of the main PSID in 2009. PSID Heads and Wives each provided two separate interviews about two randomly selected days—one week day and one weekend day. Information was obtained using time diaries about what respondents did, where they were, who did the activities with them, who else was there, how they felt, and - for household and care-related activities - for whom the activity was carried out. Diaries also include more detailed measures of well-being for 3 randomly selected activities per diary. The first interview also included in-depth questions about the respondent's health, functioning, well-being, and stylized time use/participation measures. The 2009 DUST research files are available as supplemental files on the PSID website.

12. RESTRICTED USE DATA

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 PSID website.

13. DATA DISTRIBUTION

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.

13.1 Internet-based Data Center

The PSID began distributing data through an on-line 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, 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 and named, 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.

13.2 Online cross-year variable index

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.

13.3 Family Identification Mapping System

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 split-off 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 Heads or Wives, 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, in 2005 the PSID created 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 SAS code that uses the map file and a file containing the variables the analysts wishes to examine to create a new data file that includes the variables for the relatives of interest. For example, an analyst could choose "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. The user would then create an individual level data file that contains all of the variables they want for either the grandchild or the grandparent. PSID then provides SAS code that will merge these two files and 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.

13.4 Tutorials

A series of tutorials have been created to help users learn how to examine the data. Each tutorial walks users through the process of creating an analytical extract within the PSID Data Center and analyzing the data to answer a specific scientific question. These tutorials are available on the PSID website.

13.5 Cross National Equivalent File

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, are available here

13.6 Tax information

Prior to 1992 the PSID provided estimates of federal income tax payments for each tax unit in a PSID family based on simulations. From 1992 forward it has been up to users to make these calculations. Most users take advantage of NBER's TAXSIM, which makes it quite easy to calculate this information. Butrica and Burkhauser (1997) describes how best to use the PSID data with TAXSIM.

14.DATA QUALITY

PSID staff members and other researchers in the scientific community regularly assess the quality of the data. On the PSID website is a Data Quality Bibliography containing references to such studies.

PSID staff members have written technical papers that contain cross-sectional comparisons of total family income between the PSID and the March Current Population Survey, and cross-sectional estimates of health status and health behavior between the PSID and the National Health Interview Survey. A detailed description of how these comparisons are conducted is contained in these technical papers, and they provide estimates through the 2007 wave. Beginning with the 2009 wave, the key tables and figures contained in these technical papers are updated and presented within Appendix III of this user manual.

A technical report was published in 2011 that reports rates of item nonresponse for some of the most salient questions in the PSID from 1968 to 2009. The technical paper, which is available on the PSID website, provides a detailed description of the variables and the approach for calculating the item nonresponse rates.

15.GETTING HELP

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.

16.FUNDING AND ADMINISTRATION

The PSID has been funded from a variety of sources through the years. For the past few decades, the National Science Foundation has been the primary funder. During the current funding cycle 2012-2016, support has also been received from the following organizations: The National Institute on Aging, the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, the Economic Research Service of the United States Department of Agriculture, the Indiana University Lilly Family School of Philanthropy, and the Assistant Secretary for Planning and Evaluation of the United States Department of Health and Human Services.

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 directed at the Survey Research Center, University of Michigan by James Morgan. 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-PI. 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-PIs in 2008 and 2010, respectively. During 2010 – 2011, Robert Schoeni and Charles Brown were co-Directors, with McGonagle continuing as Assistant Director, and the co-PI team consisting of: Charles Brown, Vicki Freedman, Katherine McGonagle, Narayan Sastry, Robert Schoeni, and Frank Stafford. Starting in 2012, Charles Brown assumed the role of Director, with Narayan Sastry and Vicki Freedman as Associate Directors, and Katherine McGonagle as Assistant Director. Co-Investigators included Robert Schoeni, Frank Stafford, and Fabian Pfeffer.

17.INDICATORS OF SCIENTIFIC IMPACT

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. Table 13 reports totals for each year.

Table 13. Number of publications using PSID data, registered users, website usage, and data downloads

	Number of peer-reviewed publications based on PSID~	Number of registered users*	Number of total page hits to the website	Number of unique IP addresses visiting PSID website	Number of data downloads	Number of unique users downloading PSID data
1971	8					
1972	11					
1973	6					
1974	23					
1975	21					
1976	31					
1977	32					
1978	39					
1979	41					
1980	56					
1981	71					
1982	48					
1983	75					
1984	59					
1985	78					
1986	57					
1987	74					
1988	75					
1989	62					
1990	64					
1991	91					
1992	57					
1993	69					
1994	104					
1995	74					
1996	88					
1997	119					
1998	105					
1999	96					
2000	91					
2001	108					
2002	99					
2003	101				4,856	1,024
2004	127		1,078,837	145	5,940	839
2005	136		1,920,071	8,913	7,070	1,039
2006	122	4,641	3,166,668	41,259	11,890	1,472
2007	154	6,833	4,684,659	43,426	22,931	2,467
2008	136	9,136	4,948,254	33,498	23,694	2,642
2009	155	11,575	4,846,543	31,307	25,094	2,841
2010	144	13,734	4,931,770	30,674	24,503	2,656
2011	47	14,742	2,203,266	14,075	12,710	1,406
Total	3154		27,780,068		138,688	

~Publications include journal articles, books, book chapters, and dissertations. Numbers in most recent years will increase due to the delay in identifying publications.

*Users were required to register beginning in 2007.

17.1 Peer-reviewed publications using the PSID

As of June 2013, PSID staff has located 3,492 peer-reviewed publications based on PSID data. This total includes 2,321 journal articles, 537 books and book chapters, and 634 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: Journal of Human Resources, American Economic Review, Review of Economics and Statistics, Journal of Labor Economics, Journal of Marriage and the Family, Demography, Journal of Political Economy, Journal of Econometrics, Social Science Research, and American Sociological Review. Articles have appeared in many journals from a variety of scientific disciplines, including economics, sociology, demography, public health, medicine, child development, geography, and psychology.

17.2 Grants awarded by NSF and NIH using the PSID

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 the data is its use in grant supported research. As of June 2013, there have been at least 260 awards made by NSF and NIH to support the collection and secondary analysis of PSID data, with NSF making about one-third and NIH making about two-thirds of all awards. About 88% 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.

17.3 Website activity, data downloads, and numbers of users

Information on general website activity is monitored by PSID with summary statistics for each year reported in Table 12. In calendar year 2012, there were 7 million total page hits to the PSID website made by 85,000 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 over 38,000 data downloads during this period. Across both types of downloading, these datasets were created by nearly 3,000 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. As of June 2013, there are nearly 19,000 registered users, which is an increase of nearly 2,000 registered users (11%) over the past year: 70% have identified their major field as economics, 9% sociology, with the remainder distributed across education, psychology, demography, child development, medicine, geography and "other."

18.REFERENCES

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 TAXSIMModel. 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. Response Rates in National Panel Surveys. *The Annals of the American Academy of Political and Social Science*. 2013. 645, (1): 60-87.

Appendix 1. 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 2009 and 2011 waves of data.

Table A1.1: Components of Head and Wife Taxable Income

Head Wage and Salary Income	Wife Wage and Salary Income	Head Income From Assets	Wife Income From Assets	Net Profit from Farm or Business
Head Wages ER46811 (2009) ER52219 (2011)	Wife Wages ER46841 (2009) ER52249 (2011)	Head Interest Income ER46834 (2009) ER52242 (2011)	Wife Interest Income ER46847 (2009) ER52255 (2011)	Head Net Business Income ER46809,ER46809 (2009) ER52216,ER52217 (2011)
Head Bonus ER46813 (2009) ER52221 (2011)		Head Dividend Income ER46832 (2009) ER52240 (2011)	Wife Dividend Income ER46845 (2009) ER52253 (2011)	Wife Net Business Income ER46838,ER46839 (2009) ER52246,ER52247 (2011)
Head Overtime ER46815 (2009) ER52223 (2011)		Head Rental Income ER46830 (2009) ER52238 (2011)	Wife Rental Income ER46843 (2009) ER52251 (2011)	Net Income from Farm ER46806 (2009) ER52214 (2011)
Head Tips ER46817 (2009) ER52225 (2011)		Head Trust Funds ER46836 (2009) ER52244 (2011)	Wife Trust Funds ER46849 (2009) ER52257 (2011)	
Head Commissions ER46819 (2009) ER52227 (2011)				
Head Market Gardening ER46823 (2009) ER52231 (2011)				
Head Professional Practice ER46821 (2009) ER52229 (2011)				

Table A1.2a: Head of Household Wage and Salary Income Imputation Process in the 2009 PSID

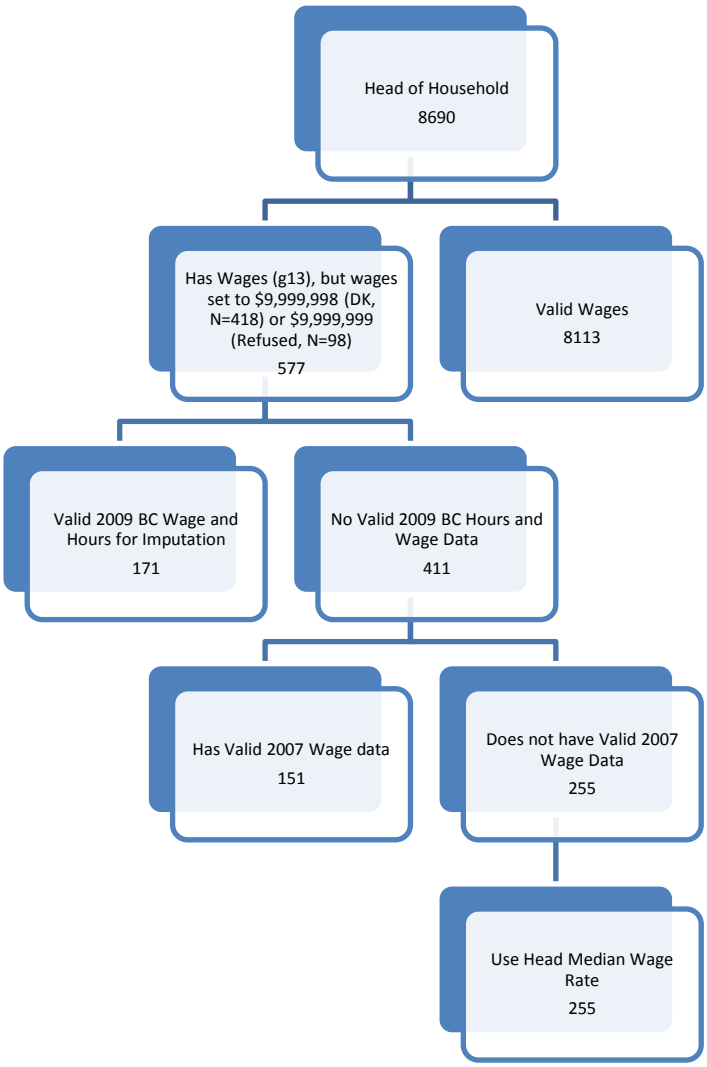


Table A1.2b: Head of Household Wage and Salary Income Imputation Process in the 2011 PSID

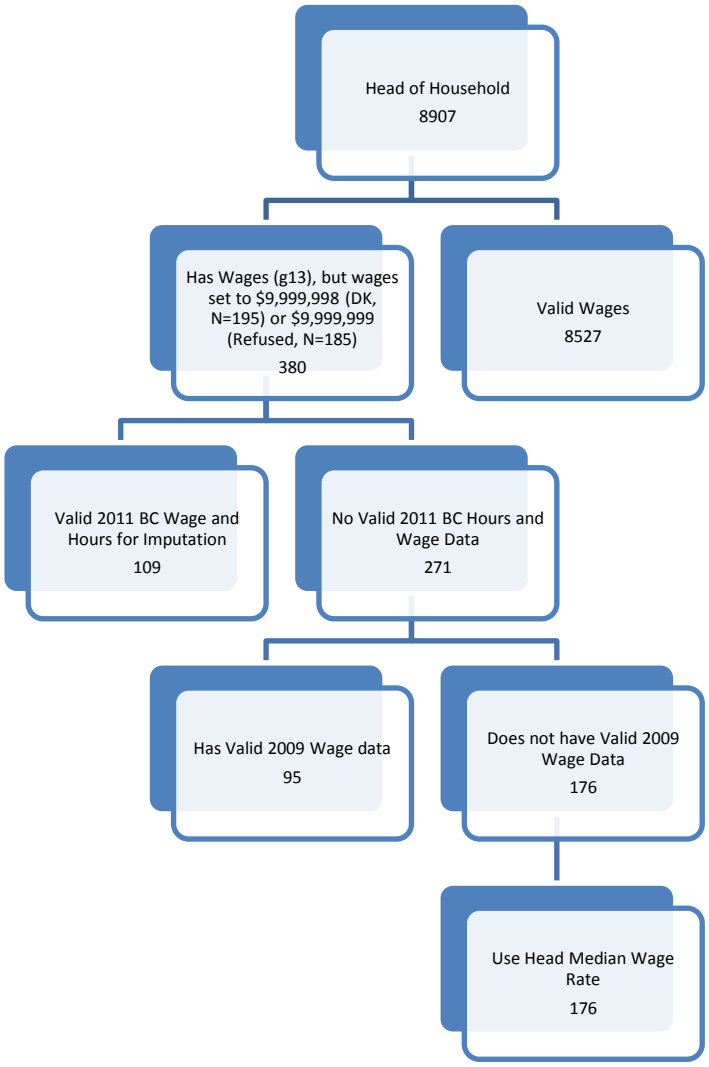


Table A1.3a: Wife Wage and Salary Income Imputation Process in the 2009 PSID

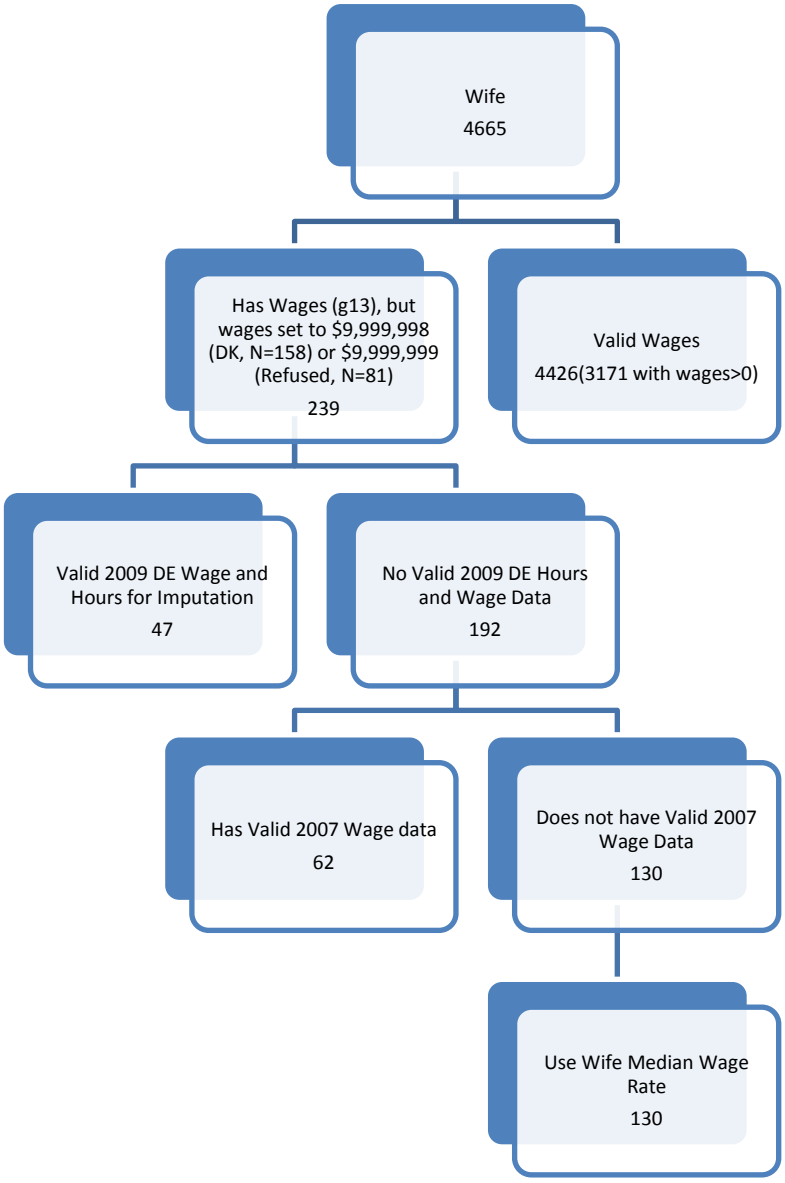


Table A1.3b: Wife Wage and Salary Income Imputation Process in the 2011 PSID

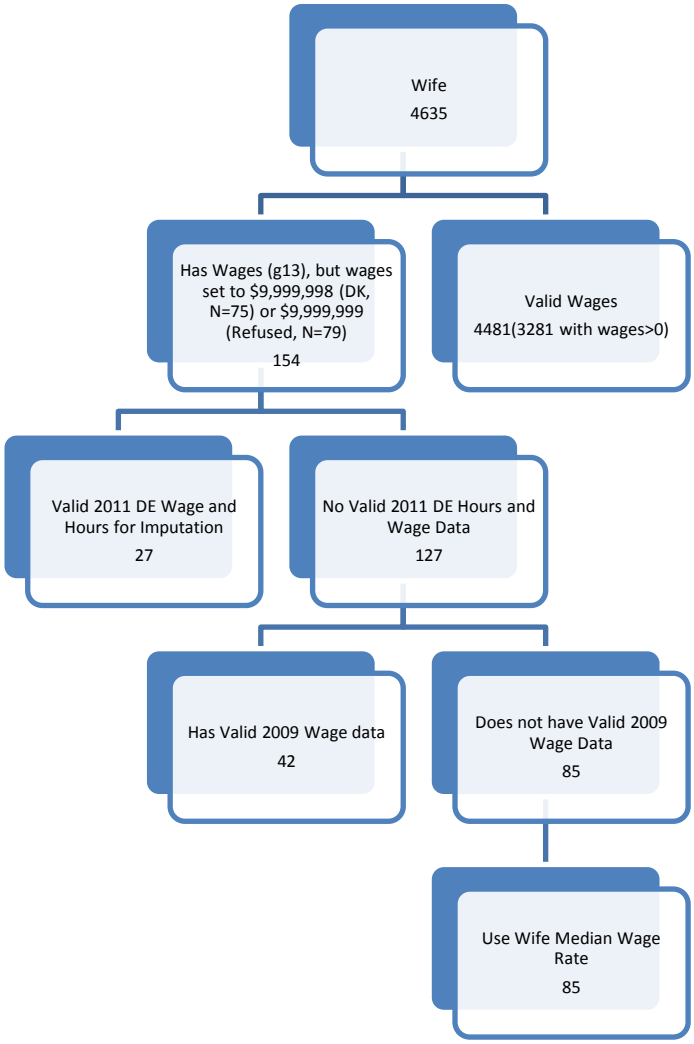


Table A1.4a: 2009 Overtime, Tips and Commission Imputation

Income Source	# Heads 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	306	42	37	16
Tips	75	0	10	2
Commission	72	1	9	1

Table A1.4b: 2011 Overtime, Tips and Commission Imputation

Income Source	# Heads 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	362	17	28	6
Tips	68	0	8	0
Commission	57	0	5	1

Table A1.5a: 2009 Head Bonus Imputation

Heads with non-zero Bonus income	599
Heads requiring Bonus income imputation	33
Use average bonus % by occupation (Step 1)	26
Use overall median bonus amount (Step 2)	7

Table A1.5b: 2011 Head Bonus Imputation

Heads with non-zero Bonus income	627
Heads requiring Bonus income imputation	31
Use average bonus % by occupation (Step 1)	29
Use overall median bonus amount (Step 2)	2

Table A1.6a: 2009 Professional Practice and Market Gardening Imputation

	Imputation Condition	#Heads with non-zero income in 2007	Step 1: Use Prior Wave Income for Head	Step 2: Use Median hourly rate.
Professional Practice (ER46821)	Income outside the range of $0 and $9,999,996, or Don't Know (9,999,998) /Refused (9,999,999)	97	0	7
Market Gardening (ER46823)	Income outside the range of -$50,000 and $9,999,996, or Don't Know (9,999,998; -999,998) /Refused (9,999,999;-999,999)	35	1	2

Table A1.6b: 2011 Professional Practice and Market Gardening Imputation

	Imputation Condition	#Heads with non-zero income in 2007	Step 1: Use Prior Wave Income for Head	Step 2: Use Median hourly rate. If hours not available, use 500 (done for 1 observation)
Professional Practice (ER52229)	Income outside the range of $0 and $9,999,996, or Don't Know (9,999,998) /Refused (9,999,999)	51	0	3
Market Gardening (ER52231)	Income outside the range of -$50,000 and $9,999,996, or Don't Know (9,999,998; -999,998) /Refused (9,999,999;-999,999)	18	0	1

Table A1.7a: 2009 Asset Income Imputation

	# Head/Wifes with Income>0	# Heads/Wifes with Imputed Values
Rent Head	488	21
Dividends Head	1009	146
Interest Head	2073	267
Trust Head	69	7
Rent Wife	186	10

Dividends Wife	545	77
Interest Wife	1138	146
Trust Wife	19	3

Table A1.7b: 2011 Asset Income Imputation

	# Head/Wifes with Income>0	# Heads/Wifes with Imputed Values
Rent Head	500	19
Dividends Head	984	96
Interest Head	1974	174
Trust Head	71	8
Rent Wife	200	10
Dividends Wife	483	45
Interest Wife	1071	84
Trust Wife	15	0

Table A1.8a: 2009 Net Business Income Imputation Prevalence by Family Business Ownership Type

Ownership	**# Businesses with Non-missing Net Income**	**# Businesses with Imputation Required**
Head	412	84
Wife	142	16
Head & Wife	89	12
OFUM Only	14	6
Head & OFUM	4	1
Mystery - All N/A	6	4

Table A1.8b: 2011 Net Business Income Imputation Prevalence by Family Business Ownership Type

Ownership	# Businesses with Non-missing Net Income	# Businesses with Imputation Required
Head	469	60
Wife	121	17
Head & Wife	82	10
OFUM Only	17	6
Head & OFUM	3	1
Head, Wife & OFUM	1	0
Mystery - All N/A	13	2

Table A1**.9a: 2009 Net Business Income Imputation Methodology**

Methdology	Number of Businesses Imputed
Using Self Employment data from Jobs section (Step 1)	10
Use Prior Wave's Net Business Income (Step 2)	4
Hot Deck Methodology (Step 3)	109

Table A1**.9b: 2011 Net Business Income Imputation Methodology**

Methdology	Number of Businesses Imputed
Using Self Employment data from Jobs section (Step 1)	11
Use Prior Wave's Net Business Income (Step 2)	0
Hot Deck Methodology (Step 3)	85

Table A1.10a: 2009 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	5
Don't Know Gain (9,999,998)	Assign Random Positive Income	83
N/A, Refused Gain (9,999,999)	Assign Random Positive Income	21

Table A1.10b: 2011 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	5
Don't Know Gain (9,999,998)	Assign Random Positive Income	51
N/A, Refused Gain (9,999,999)	Assign Random Positive Income	29

Table A1.11a: 2009 Transfer Income Imputation

Transfer Income Source	Who	Number where income amount>0	Number Imputed
Alimony (ER46872)	Head	39	2
Annuity (ER46862)	Head	121	14
Child Support (ER46870)	Head	426	14
Help Non-Relatives (ER46876)	Head	493	51
Help Relatives (ER46874)	Head	1017	74
Other Pension (ER46864)	Head	6	0
Other Transfer Income (ER46878)	Head	229	10
Retirement (ER46860)	Head	798	53
SSI (ER46854)	Head	255	16
TANF (ER46852)	Head	144	7
Unemployment (ER46866)	Head	519	24
VA Pension (ER46858)	Head	228	14
Welfare (ER46856)	Head	39	3
Workers Comp (ER46868)	Head	67	4
Child Support (ER46892)	Wife	159	6
Help Non Relatives (ER46896)	Wife	17	2
Help Relatives (ER46894)	Wife	170	16
Other Transfer Income (ER46898)	Wife	99	2
Pension (ER46886)	Wife	228	16
SSI (ER46882)	Wife	43	1
TANF (ER46880)	Wife	26	3
Unemployment (ER46888)	Wife	142	11
Welfare (ER46884)	Wife	8	0
Workers Comp (ER46890)	Wife	22	0

Table A1.11b: 2011 Transfer Income Imputation

Transfer Income Source	Who	Number where income amount>0	Number Imputed
Alimony (ER52280)	Head	37	4
Annuity (ER52270)	Head	156	12
Child Support (ER52278)	Head	473	16
Help Non-Relatives (ER52284)	Head	473	48
Help Relatives (ER52282)	Head	1086	65
Other Pension (ER52272)	Head	1	0
Other Transfer Income (ER52286)	Head	226	8
Retirement (ER52268)	Head	814	41
SSI (ER52262)	Head	269	5
TANF (ER52260)	Head	145	1
Unemployment (ER52274)	Head	699	23
VA Pension (ER52266)	Head	264	10
Welfare (ER52264)	Head	65	2
Workers Comp (ER52276)	Head	71	5
Child Support (ER52300)	Wife	153	4
Help Non Relatives (ER52304)	Wife	11	0
Help Relatives (ER52302)	Wife	160	4
Other Transfer Income (ER52306)	Wife	87	4
Pension (ER52294)	Wife	247	12
SSI (ER46882)	Wife	38	0
TANF (ER52288)	Wife	24	1
Unemployment (ER52296)	Wife	223	6
Welfare (ER52292)	Wife	24	0
Workers Comp (ER52298)	Wife	8	1

Table A1.12a: 2009 Labor Income Imputation for Other Family Members

	# Jobs with income>0	# Imputations
Job 1	2388	765
Job 2	412	139
Job 3	88	40
Job 4	18	10

Table A1.12b: 2011 Labor Income Imputation for Other Family Members

	# Jobs with income>0	# Imputations
Job 1	2102	506
Job 2	329	80
Job 3	51	16
Job 4	3	1

Table A1.13a: 2009 Transfer Income Imputation for Other Family Members

	# OFUMS with Income Source>0	# Imputations
ADC (ER46908)	13	3
SSI (ER46910)	197	9
Welfare (ER46912)	36	12
VA Pension (ER46914)	12	3
Pension (ER46916)	49	17
Unemployment (ER46918)	20	5
Workers Comp (ER46920)	1	0
Child Support (ER46922)	35	4
Support from Relatives (ER46924)	41	10
Other (ER46926)	134	16

Table A1.13b: 2011 Transfer Income Imputation for Other Family Members

	# OFUMS with Income Source>0	# Imputations
ADC (ER52316)	13	3
SSI (ER52318)	197	9
Welfare (ER52320)	36	12
VA Pension (ER52322)	12	3
Pension (ER52324)	49	17
Unemployment (ER52326)	20	5
Workers Comp (ER52328)	1	0
Child Support (ER52330)	35	4
Support from Relatives (ER52332)	41	10
Other (ER52334)	134	16

Table A1.14a: 2009 Social Security

	Number with Social Security Income> $0	# Imputations
Heads (ER46929)	1436	106
Wives (ER46931)	507	25
OFUM (ER46933)	397	53

Table A1.14b: 2011 Social Security

	Number with Social Security Income> $0	# Imputations
Heads (ER52337)	1518	93
Wives (ER52339)	554	32
OFUM (ER52341)	452	44

Table A1.15: Weeks and Hours Variables

Variable Name		Variable Description	Variable Group
2009	2011		
ER46761	ER52169	HEAD WORK WEEKS	Weeks
ER46763	ER52171	HEAD WEEKLY WORK HOURS	Hours Worked
ER46765	ER52173	HD OVERTIME WORK HOURS	Hours Worked
ER46767	ER52175	HEAD TOTAL HOURS OF WORK	Hours Worked
ER46768	ER52176	HD WEEKS MISSED FOR ILLNESS OF OTRS	Weeks
ER46770	ER52178	HD WEEKS MISSED FOR OWN ILLNESS	Weeks
ER46772	ER52180	HEAD WEEKS OFF FOR VACATION	Weeks
ER46774	ER52182	HEAD STRIKE WEEKS	Weeks
ER46776	ER52184	HEAD WEEKS LAID OFF	Weeks
ER46778	ER52186	HEAD UNEMPLOYMENT WEEKS	Weeks
ER46780	ER52188	HEAD WEEKS OUT OF LABOR FORCE	Weeks
ER46782	ER52190	WIFE WORK WEEKS	Weeks
ER46784	ER52192	WIFE WEEKLY WORK HOURS	Hours Worked
ER46786	ER52194	WF OVERTIME WORK HOURS	Hours Worked
ER46788	ER52196	WIFE TOTAL HOURS OF WORK	Hours Worked
ER46789	ER52197	WF WEEKS MISSED FOR ILLNESS OF OTRS	Weeks
ER46791	ER52199	WF WEEKS MISSED FOR OWN ILLNESS	Weeks
ER46793	ER52201	WIFE WEEKS OFF FOR VACATION	Weeks
ER46795	ER52203	WIFE STRIKE WEEKS	Weeks
ER46797	ER52205	WIFE WEEKS LAID OFF	Weeks
ER46799	ER52207	WIFE UNEMPLOYMENT WEEKS	Weeks
ER46801	ER52209	WIFE WEEKS OUT OF LABOR FORCE	Weeks

Table A1.16: Weeks Worked Imputation

	2009	2011
Number of Heads/Wives with Jobs	10549	10337
Number of Heads/Wives with Jobs with Weeks Worked Edits	107	98

Table A1.17a: 2009 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
ER46772	Head Weeks not working due to vacation	75	8	1 Week
ER46770	Head Weeks not working due to illness- self	8	5	0.4 Weeks
ER46768	Head Weeks not working due to illness – other	1	8	0.4 Weeks
ER46774	Head Weeks not working due to strike	0	0	3 Weeks
ER46776	Head Weeks laid off	36	1	2.5 Weeks
ER46793	Wife Weeks not working due to vacation	41	5	1 Week
ER46791	Wife Weeks not working due to illness- self	5	1	0.4 Weeks
ER46789	Wife Weeks not working due to illness – other	0	4	0.4 Weeks
ER46795	Wife Weeks not working due to strike	0	0	3 Weeks
ER46797	Wife Weeks laid off	14	0	2.5 Weeks

Table A1.17b: 2011 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
ER52180	Head Weeks not working due to vacation	124	14	1 Week
ER52178	Head Weeks not working due to illness- self	5	15	0.4 Weeks
ER52176	Head Weeks not working due to illness – other	6	7	0.4 Weeks
ER52182	Head Weeks not working due to strike	0	0	3 Weeks
ER52184	Head Weeks laid off	13	4	2.5 Weeks
ER52201	Wife Weeks not working due to vacation	71	8	1 Week
ER52199	Wife Weeks not working due to illness- self	1	5	0.4 Weeks
ER52197	Wife Weeks not working due to illness – other	2	5	0.4 Weeks
ER52203	Wife Weeks not working due to strike	0	0	3 Weeks
ER52205	Wife Weeks laid off	3	0	2.6 Weeks

Table A1.18a: 2009 Time Not Working Imputations

	# Heads/Wives with Weeks Manually Edited	# Heads/Wives with Weeks Adjusted	# Head/Wife Non-zero Weeks
Unemployment	54	0	1291
OOLF	37	14	3584

Table A1.18b: 2011 Time Not Working Imputations

	# Heads/Wives with Weeks Manually Edited	# Heads/Wives with Weeks Adjusted	# Head/Wife Non-zero Weeks
Unemployment	52	0	1696
OOLF	43	21	3803

Table A1.19a: 2009 Work Hours Imputation Summary

	2009
Number of Head/Wife Jobs	13355
Number of Head/Wife Jobs for which Hours Worked has had Pre-imputation Manual Adjustments Applied	135
Number of Head/Wife Jobs for which we Impute Using a Value of 40 Hours per Week	104

Table A1.19b: 2011 Work Hours Imputation Summary

	2011
Number of Head's/Wife's with Jobs	10300
Number of Head's/Wife's with jobs for which Hours Worked has had Pre-imputation Manual Adjustments Applied	148
Number of Head's/Wife's with jobs for which we Impute Using a Value of 40 Hours per Week	107

Table A1.20a: 2009 Overtime Hours Imputation Summary

	Manual Edits	Imputed	Number of Observations with Non-Zero Overtime
Head Overtime (ER46765)	1	537	3334
Wife Overtime (ER46786)	1	140	1138

Table A1.20b: 2011 Overtime Hours Imputation Summary

	Manual Edits	Imputed	Number of Observations with Non-Zero Overtime
Head Overtime (ER52173)	0	101	1497
Wife Overtime (ER52194)	0	16	482

Table A1.21a: 2009 Number of Families by Number Income Sources Imputed

Number of Income Sources Imputed	Number of 2009 PSID Families	% of Families
0	6645	76.47
1	1292	14.87
2	510	5.87
3	144	1.66
4	53	0.61
5	19	0.22
6	17	0.20
7	4	0.05
8	4	0.05
9	1	0.01
10	1	0.01

Table A1.21b: 2011 Number of Families by Number Income Sources Imputed

Number of Income Sources Imputed	Number of 2011 PSID Families	% of Families
0	7293	81.89
1	1111	12.47
2	325	3.65
3	103	1.16
4	48	0.54
5	18	0.20
6	7	0.08
7	2	0.02

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 A1.23a: 2009 Income Imputation Summary Table

Source	# Observations with non-zero income (including imputed cases)	# Observations requiring imputations	Methodology	Accuracy Variable
Head Wage and Salary Income	6849	577	Step 1, Use PSID Employment Section BC/DE Wages/Hours/Weeks Worked to impute (171), else Step 2, Use prior year income (151), else overall median wage rate (255)	ER46812
Wife Wage and Salary Income	3411	239	Step 1, Use BC/DE Wages/Hours/Weeks Worked to impute (47), else Step 2, Use prior year income (62), else overall median wage rate (130)	ER46842
Head Bonus	599	66	Step 1: Use average bonus percent by OCC code, apply to wages (26), else Step 2, Use overall median bonus percent (7)	ER46814
Head OT	306	95	Step 1, Use BC jobs info (42), else Step 2, avg OT as pct of wages by occ code (37), else median OT amount (16)	ER46816
Head Tips	75	12	Step 1, Use BC jobs info (0), else Step 2, avg tips as pct of wages by occ code (10), else median tips amount (2)	ER46818
Head Commission	72	11	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (9), else median tips amount (1)	ER46820
Head Professional Practice	96	7	Step 1, Use Prior year (0), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (7)	ER46822
Head Market Gardening	35	3	Step 1, Use Prior year (1), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (2)	ER46824
Head Farm Income	31	6	Step 1, Farm Receipts - Farm Expenses (0), else Step 2, Prior Year Income (5) else Step 3, overall median farm income (1)	ER46807
Head Business	412	84	Step 1, Use BC/DE Self Employment Income (8), else Step 2, Prior year's income if same industry (3), else Step 3, hot deck within industry (73)	ER46804

Wife Business	142	16	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 (16)	ER46804
Head & Wife Business	89	12	Step 1, Use BC/DE Self Employment Income (2), else Step 2, Prior year's income if same industry (1), else Step 3, hot deck within industry (9)	ER46804
OFUM Only Business	14	6	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 (6)	ER46804
Head & OFUM Business	4	1	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 (1)	ER46804
N/A Business	6	4	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 (4)	ER46804
Rent Head	488	25	Overall Median	ER46831
Dividend Head	1009	148	Overall Median	ER46833
Interest Head	2073	277	Overall Median	ER46835
Trust Head	69	7	Overall Median	ER46837
Rent Wife	186	10	Overall Median	ER46844
Dividends Wife	545	77	Overall Median	ER46846
Interest Wife	1138	146	Overall Median	ER46848
Trust Wife	19	3	Overall Median	ER46850
OFUM Job 1	2388	765	Overall Median (within Job #)	ER46904
OFUM Job 2	412	139	Overall Median (within Job #)	ER46904
OFUM Job 3	88	40	Overall Median (within Job #)	ER46904
OFUM Job 4	18	10	Overall Median (within Job #)	ER46904
OFUM Interest	61	16	Overall Median	ER46906
OFUM ADC	11	3	Overall Median	ER46909
OFUM SSI	171	11	Overall Median	ER46911
OFUM Welfare	24	6	Overall Median	ER46913
OFUM Veterans	12	4	Overall Median	ER46915
OFUM Pension	48	18	Overall Median	ER46917
OFUM Unemployment	20	5	Overall Median	ER46919
OFUM Workers Comp	1	0	Overall Median	ER46921

OFUM Child Support	33	4	Overall Median	ER46923
OFUM Relatives	38	9	Overall Median	ER46925
OFUM Other	111	17	Overall Median	ER46927
Social Security	2340	191	Overall Median	ER46930, ER46932, ER46934
Head Alimony	39	2	Overall Median	ER46873
Head Annuity	121	14	Overall Median	ER46863
Head Child Support	426	14	Overall Median	ER46871
Head Help Non Rel	493	53	Overall Median	ER46877
Head Help Rel	1017	76	Overall Median	ER46875
Head Other	229	11	Overall Median	ER46879
Head Retirement	798	54	Overall Median	ER46861
Head SSI	255	16	Overall Median	ER46855
Head TANF	144	8	Overall Median	ER46852
Head Unemp	519	25	Overall Median	ER46867
Head VA Pension	228	14	Overall Median	ER46859
Head Welfare	39	3	Overall Median	ER46857
Head Other Retirement	6	0	Overall Median	ER46865
Head Workers Comp	67	4	Overall Median	ER46869
Wife ADC	26	3	Overall Median	ER46881
Wife Child Support	159	7	Overall Median	ER46893
Wife Help Non Rel	17	2	Overall Median	ER46897
Wife Help Rel	170	16	Overall Median	ER46895
Wife Other	99	2	Overall Median	ER46899
Wife Pension	228	16	Overall Median	ER46887
Wife SSI	43	1	Overall Median	ER46883
Wife Unemp	142	11	Overall Median	ER46889
Wife Welfare	8	6	Overall Median	ER46885
Wife Workers Comp	22	0	Overall Median	ER46891

Table A1.23b: 2011 Income Imputation Summary Table

Source	# Observations with non-zero income (including imputed cases)	# Observations requiring imputations	Methodology	Accuracy Variable
Head Labor Income	6673	518	Step 1, Use PSID Employment Section BC/DE Wages/Hours/Weeks Worked to impute (171), else Step 2, Use prior year income (151), else overall median wage rate (255)	ER52220
Wife Labor Income	3281	195	Step 1, Use BC/DE Wages/Hours/Weeks Worked to impute (47), else Step 2, Use prior year income (62), else overall median wage rate (130)	ER52250
Head Bonus	627	31	Step 1: Use average bonus percent by OCC code, apply to wages (26), else Step 2, Use overall median bonus percent (7)	ER52222
Head OT	362	52	Step 1, Use BC jobs info (42), else Step 2, avg OT as pct of wages by occ code (37), else median OT amount (16)	ER52224
Head Tips	68	9	Step 1, Use BC jobs info (0), else Step 2, avg tips as pct of wages by occ code (10), else median tips amount (2)	ER52226
Head Commission	57	6	Step 1, Use BC jobs info (1), else Step 2, avg tips as pct of wages by occ code (9), else median tips amount (1)	ER52228
Head Professional Practice	51	6	Step 1, Use Prior year (0), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (7)	ER52230
Head Market Gardening	18	1	Step 1, Use Prior year (1), else Step 2, mean hourly rate * hours (use 500 hours if hours n/a) (2)	ER52232
Head Farm Income	28	5	Step 1, Farm Receipts - Farm Expenses (0), else Step 2, Prior Year Income (5) else Step 3, overall median farm income (1)	ER52215
Head Business	469	66	Step 1, Use BC/DE Self Employment Income (8), else Step 2, Prior year's income if same industry (3), else Step 3, hot deck within industry (73)	ER52212

Wife Business	121	18	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 (16)	ER52212
Head & Wife Business	82	12	Step 1, Use BC/DE Self Employment Income (2), else Step 2, Prior year's income if same industry (1), else Step 3, hot deck within industry (9)	ER52212
OFUM Only Business	17	7	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 (6)	ER52212
Head & OFUM Business	3	1	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 (1)	ER52212
Head, Wife & 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 (1)	ER52212
N/A Business	13	2	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 (4)	ER52212
Rent Head	500	32	Overall Median	ER52239
Dividend Head	984	100	Overall Median	ER52241
Interest Head	1974	177	Overall Median	ER52243
Trust Head	71	9	Overall Median	ER52245
Rent Wife	201	10	Overall Median	ER52252
Dividends Wife	483	45	Overall Median	ER52254
Interest Wife	1071	88	Overall Median	ER52256
Trust Wife	15	0	Overall Median	ER52258
OFUM Labor Income	2697	469	Overall Median (within Job #)	ER46904
OFUM Interest	45	12	Overall Median	ER52314
OFUM ADC	20	4	Overall Median	ER52317
OFUM SSI	153	12	Overall Median	ER52319
OFUM Welfare	41	6	Overall Median	ER52321
OFUM Veterans	13	3	Overall Median	ER52323
OFUM Pension	56	13	Overall Median	ER52325
OFUM Unemployment	36	4	Overall Median	ER52327

OFUM Workers Comp	1	0	Overall Median	ER52329
OFUM Child Support	32	5	Overall Median	ER52331
OFUM Relatives	16	4	Overall Median	ER52333
OFUM Other	131	10	Overall Median	ER52335
Social Security	2653	194	Overall Median	ER52337, ER52339, ER52341
Head Alimony	37	4	Overall Median	ER52281
Head Annuity	156	14	Overall Median	ER52271
Head Child Support	432	16	Overall Median	ER52279
Head Help Non Rel	473	49	Overall Median	ER52285
Head Help Rel	1086	68	Overall Median	ER52283
Head Other	226	9	Overall Median	ER52287
Head Retirement	814	46	Overall Median	ER52269
Head SSI	269	16	Overall Median	ER52263
Head TANF	145	2	Overall Median	ER52261
Head Unemp	699	31	Overall Median	ER52275
Head VA Pension	264	10	Overall Median	ER52267
Head Welfare	65	2	Overall Median	ER52265
Head Other Retirement	1	0	Overall Median	ER52273
Head Workers Comp	71	6	Overall Median	ER52277
Wife ADC	24	0	Overall Median	ER52289
Wife Child Support	153	4	Overall Median	ER52301
Wife Help Non Rel	11	0	Overall Median	ER52305
Wife Help Rel	160	4	Overall Median	ER52303
Wife Other	87	6	Overall Median	ER52307
Wife Pension	247	12	Overall Median	ER52295
Wife SSI	38	0	Overall Median	ER52291
Wife Unemp	223	8	Overall Median	ER52297
Wife Welfare	8	1	Overall Median	ER52293
Wife Workers Comp	24	1	Overall Median	ER52299

Appendix 2: Tables and figures comparing PSID-based estimates of income and health with external sources

PSID staff members have written technical papers comparing cross-sectional estimates of income and health based on the PSID to external sources through the 2007 wave of data. Details describing these estimates can be found in the respective technical paper on the PSID website. These comparisons have been updated through the 2011 wave, with results presented in the tables and figures below.

Figure A2.1. 10th-40th percentiles of PSID aggregated family income and CPS household income,
1967 – 2010. [Top (blue/square) line=PSID; Bottom (purple/triangle) line=CPS]

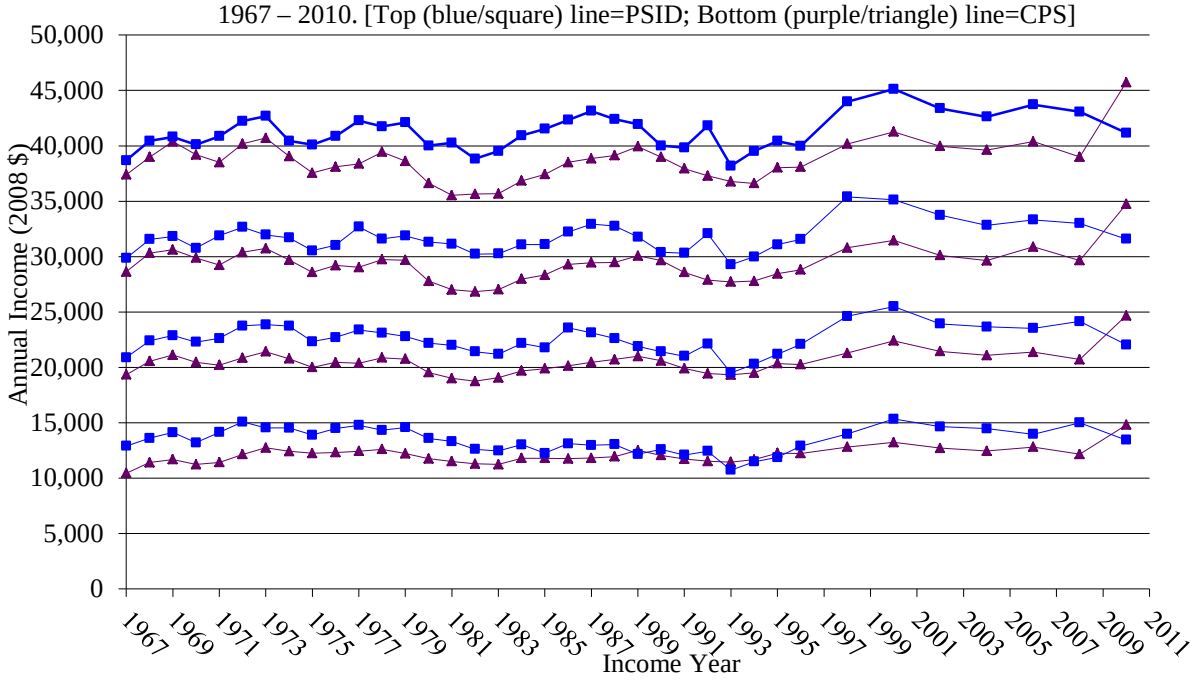


Figure A2.2. 50th-80th percentiles of PSID aggregated family income and CPS household income,
1967 – 2010. [Top (blue/square) line=PSID; Bottom (purple/triangle) line=CPS]

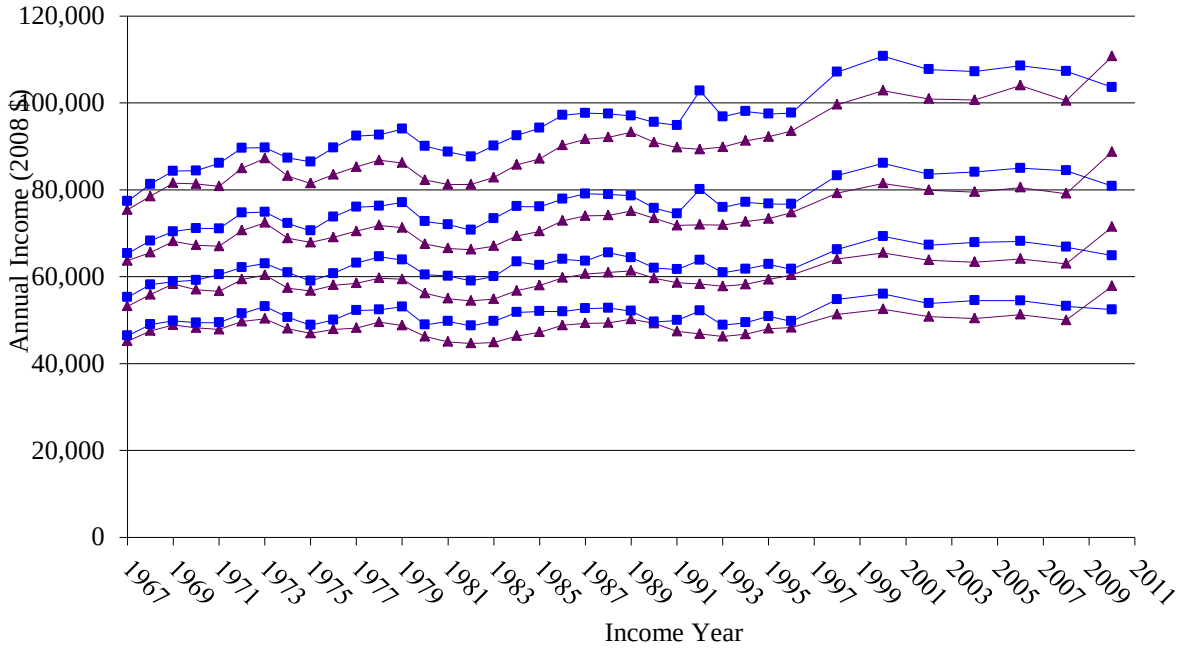


Figure A2.3. 1st-5th percentiles of PSID aggregated family income and CPS household income, 1967 – 2010.
[Top (blue/square) line=PSID; Bottom (purple/triangle) line=CPS]

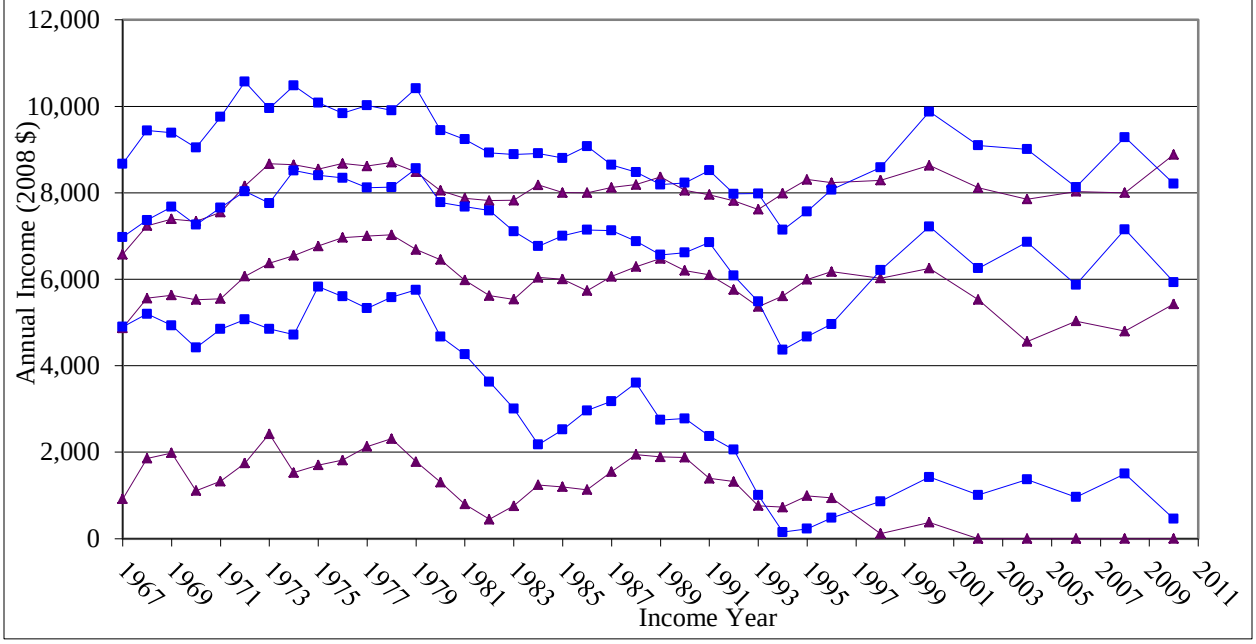


Figure A2.4. 90th-99th percentiles of PSID aggregated family income and CPS household income,
1967 – 2010. [Top (blue/square) line=PSID; Bottom (purple/triangle) line=CPS]

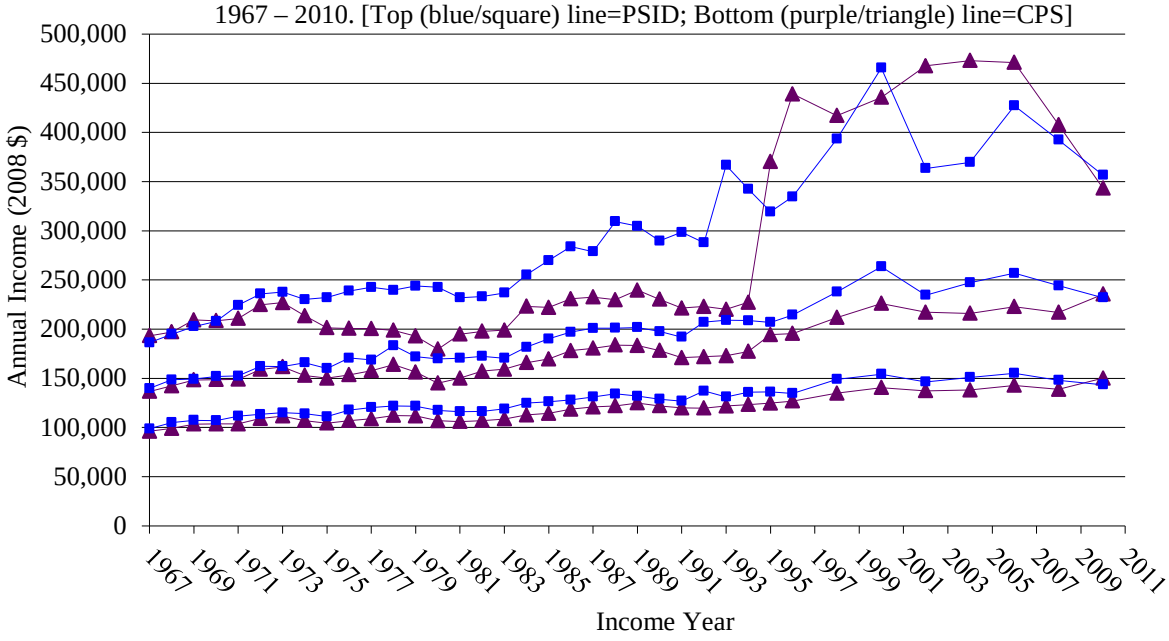


Table A2.1. Health Status, Health Behaviors, and Health Insurance Coverage (%) in the PSID and NHIS: 1999-2011

	PSID 1999	NHIS 1999	PSID 2001	NHIS 2001	PSID 2003	NHIS 2003	PSID 2005	NHIS 2005	PSID 2007	NHIS 2007	PSID 2009	NHIS 2009	PSID 2011	NHIS 2011
Smoking														
Currently smoke	20	23	21	23	20	21	19	21	19	19	19	21	17	19
Never smoked	52	53	53	55	53	56	53	57	53	58	54	58	54	59
Has Health Insurance	87	87	87	85	87	84	85	83	86	84	85	82	83	82
Obese (BMI > 30)	22	22	21	22	22	23	23	25	25	26	26	28	27	28
30-Day Distress (K6*)	--	--	3.2	3.1	3.9	3.1	--	--	3.3	2.7	3.5	3.2	3.5	3.1
Health doesn't limit amount/kind of work	82	88	81	89	82	89	82	89	81	89	81	88	81	88
Health Conditions														
Asthma	2.8	2.1	7.6	11.8	8.6	9.7	9.6	10	10.5	10.9	10.7	13.1	11.5	12.6
Cancer	4.6	6.5	5	6.9	5.3	6.6	6.9	7.4	7.2	7.3	7.8	8.2	8.1	8.2
Diabetes	7.1	5.4	7.7	6.5	8.3	6.6	9.1	7.5	9.3	7.7	10.2	9	11.2	8.9
Hypertension	21.8	22.6	23	23.5	25.4	25.2	27.5	25.8	29.6	27.1	30.8	28.9	32.3	29.7
Myocardial Infarction	3.7	2.9	3.7	3.4	3.5	3.2	4.3	3.3	4.2	3.2	4	3.4	4.7	3.2
Stroke	7.4	8.5	3.1	2.4	3	2.5	3.7	2.4	3.6	2.4	3.8	2.6	3.8	2.7
Self-Rated General Health														
Excellent	23	32	22	31	22	29	19	29	20	29	17	29	16	28
Very Good	33	32	35	32	35	32	33	32	34	31	35	32	36	32
Good	29	25	28	25	28	26	30	27	29	27	31	27	31	27
Fair	11	8	11	9	11	9	13	9	11	9	12	9	12	10
Poor	4	3	4	3	4	3	5	3	5	3	5	3	5	3

Note: The PSID sample consists of all heads and wives, and the NHIS sample consists of people 18 years or older. Weights are used for estimates from both surveys.

* The K-6 Non-Specific Psychological Distress Scale includes six items. The scores of the six items are summed; a score of 13 or higher indicates sensitivity around the threshold for the clinically significant range of the distribution of non-specific distress. For additional information see Kessler, et al 2002.

Table A3.2. Mean height and weight by age and sex:
PSID and NHIS, 2007, 2009 and 2011

	Height						Weight					
	PSID 2007	NHIS 2007	PSID 2009	NHIS 2009	PSID 2011	NHIS 2011	PSID 2007	NHIS 2007	PSID 2009	NHIS 2009	PSID 2011	NHIS 2011
Males												
Age 20-24	71.1	70.2	70.8	70.1	70.4	70.0	190	182	186	180	181	180
Age 25-29	71.2	70.0	71.0	70.3	71.0	70.0	198	189	199	190	199	188
Age 30-34	71.1	70.0	71.0	70.0	70.5	70.1	203	192	203	193	204	194
Age 35-39	70.5	70.0	70.6	69.9	69.9	69.8	201	197	204	196	204	195
Age 40-44	70.1	70.0	70.4	69.8	70.4	69.9	199	196	204	197	205	197
Age 45-49	70.5	69.9	70.1	69.8	70.0	69.8	197	196	196	196	199	197
Age 50-54	70.3	69.9	70.5	69.8	70.0	69.9	197	196	200	198	199	199
Age 55-59	70.1	69.8	69.9	69.7	70.0	70.0	194	196	195	196	200	196
Age 60-64	70.1	69.8	70.1	69.8	69.5	69.9	201	193	198	196	198	196
Age 65-69	70.1	69.8	70.2	69.7	70.1	69.7	195	193	203	196	200	195
Age 70-74	69.5	69.8	69.8	69.4	69.4	69.6	192	190	193	193	192	196
Age 75-79	69.8	69.1	69.3	69.2	69.6	69.4	184	183	185	187	183	191
Age 80+	68.9	69.2	69.2	69.0	68.4	68.9	172	174	176	175	176	178
Females												
Age 20-24	64.8	64.4	64.6	64.4	64.8	64.3	150	144	148	149	151	149
Age 25-29	65.1	64.4	64.9	64.5	64.4	64.3	158	151	161	154	157	154
Age 30-34	65.1	64.6	65.1	64.4	64.9	64.3	164	158	164	157	165	157
Age 35-39	64.8	64.5	64.9	64.3	65.0	64.5	163	159	168	162	169	161
Age 40-44	64.5	64.4	64.5	64.4	64.6	64.5	163	158	160	160	164	161
Age 45-49	64.6	64.4	64.5	64.6	64.3	64.2	163	158	165	160	165	163
Age 50-54	64.5	64.3	64.5	64.2	64.5	64.2	166	161	166	164	167	162
Age 55-59	64.6	64.0	64.5	64.2	64.1	64.1	164	163	167	164	166	163
Age 60-64	64.2	64.2	64.3	64.0	64.4	64.0	160	163	163	167	165	162
Age 65-69	64.0	63.9	64.0	63.8	63.9	63.9	163	160	162	160	161	162
Age 70-74	63.7	63.7	63.7	63.8	63.7	63.8	153	157	159	158	159	161
Age 75-79	63.8	63.7	63.4	63.5	63.2	63.6	154	152	150	154	152	152
Age 80+	63.3	63.3	63.4	63.1	63.4	63.1	142	146	145	143	146	145

Note: The PSID sample consists of all heads and wives, and the NHIS sample consists of people 18 years or older.

Table A2.3. Logistic Regression of Being in Poor/Fair Health (Odds Ratio): 1999 – 2011

	PSID 1999	NHIS 1999	PSID 2001	NHIS 2001	PSID 2003	NHIS 2003	PSID 2005	NHIS 2005	PSID 2007	NHIS 2007	PSID 2009	NHIS 2009	PSID 2011	NHIS 2011
Age	1.044*	1.043*	1.044*	1.044*	1.030*	1.041*	1.044*	1.042*	1.031*	1.042*	1.034*	1.039*	1.037*	1.038*
Race														
White (Reference Group)														
Black	2.14*	1.79*	2.021*	1.838*	1.702*	1.674*	1.786*	1.605*	1.517*	1.568*	1.517*	1.573*	1.393	1.657*
Other	1.44*	1.38*	1.337	1.347	1.128	0.965	1.667	0.919	1.38	0.952*	1.163	1.026*	1.741*	0.954
Female	1.16*	1.04*	1.149	1.073*	1.188*	1.093*	1.176*	1.082*	1.264*	1.108*	1.196*	1.055*	1.151*	1.048
Not currently married	1.28*	1.30*	1.181*	1.306*	1.520*	1.319*	1.438*	1.340*	1.398*	1.375*	1.525*	1.473*	1.632*	1.430*
High school graduate	0.41*	0.36*	.385*	.421*	.331*	.428*	0.350*	0.400*	0.354*	0.400*	0.345*	0.396*	0.337*	0.391*

Note: *Indicates statistical significance at the .01 level.
The PSID sample consists of all heads and wives, and the NHIS sample consists of people 18 years or older.