The Panel Study of Income Dynamics: overview, recent innovations, and potential for life course research

Katherine A McGonagle  
Institute for Social Research, University of Michigan  
kmcgon@umich.edu

Robert F Schoeni  
Institute for Social Research, Department of Economics and Gerald R. Ford School of Public Policy, University of Michigan

Narayan Sastry  
Institute for Social Research, University of Michigan

Vicki A Freedman  
Institute for Social Research, University of Michigan

Abstract  
Spanning over four decades, the Panel Study of Income Dynamics (PSID) is the world’s longest-running household panel survey. The resulting data archive presents research opportunities for breakthroughs in understanding the connections between economic status, health and well-being across generations and over the life course. The long panel, genealogical design, and broad content of the data represent an unique opportunity for a multi-perspective study of life course evolution and change within families over multiple generations. Based on relational data structures and advanced web-based archiving and delivery tools, the PSID has a publicly available web-based facility for users worldwide to create customized data extracts and codebooks, based on nearly 70,000 variables from over 70,000 individuals over 44 years. This paper provides an overview of the PSID and its supplemental studies, the Disability and Use of Time Supplement, the Child Development Supplement, and the Transition into Adulthood study, and describes features and recent enhancements that have increased the potential of the archive for studying life course development.

Keywords: panel study, life course, data collection, human development, time use, young adulthood, aging

1. Introduction  
The world’s longest-running household panel survey was originally created in the United States of America 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 design and field the Survey of Economic Opportunity (SEO) to provide data for a national assessment of War on Poverty programs. A representative national sample of approximately 22,000 households and an over sample, in census enumeration districts with large non-white populations, of approximately 15,000 households was drawn by the Census Bureau and interviews were completed with 30,000 of these households. Interest in continuing the study with the primary goal of understanding the dynamics of economic well-being led the OEO to approach the Survey Research Center (SRC) at the University of Michigan about continuing to interview a sub-sample of low-income SEO households. Professor James N. Morgan, who eventually became the study’s first director at Michigan, argued successfully for adding a cross-section of households from the SRC national sampling frame so that the study would also include non-poor
households and hence represent the entire population of the U.S. In addition, a fortuitous decision was made to follow family members who moved out of study households, such as children who came of age during the study. This allowed the sample to remain representative of the nation’s families and individuals over time. This study became what is now called the Panel Study of Income Dynamics (PSID; Hill, 1992).

Launched in 1968, PSID will soon mark its 44th anniversary. The data are a cornerstone of social science research in the U.S.: more than 3,200 peer-reviewed publications are based on the PSID, nine U.S. federal agencies utilize the data in fulfilling their missions, annually nearly 25,000 data extracts are downloaded by over 2,500 different users world-wide, and there are nearly 5 million visits to the PSID website each year. As an indicator of its significance, the study’s lead sponsor, the U.S. National Science Foundation (NSF), named PSID as one of the 60 most important innovations that NSF played a key role in creating.

Starting in the late 1990s, several developments increased the potential of the PSID archive for studying life course development. Most notably, content was expanded in the areas of health, wealth, expenditures, philanthropy, child development, the transition to adulthood, and time use. This paper provides a detailed profile of PSID and recent supplemental studies, highlighting the features that enhance the value for longitudinal and life course research. The sections below describe the design and content of the PSID main interview, the Disability and Use of Time Supplement, the Child Development Supplement, and the Transition into Adulthood study, the distribution of data and documentation, funding and administration, the analytic potential for life course research, including research examples designed to stimulate new ideas and discoveries, and what the future holds for PSID.

2. The PSID Main Interview

The primary source of information on PSID sample members has been a survey conducted annually through 1997 and biennially thereafter. This survey, called the "main interview," is described in this section. Subsequent sections describe supplemental data collections on PSID sample members.

2.1. Study design and following rules

The original aim of studying the dynamics of income and poverty led the 1968 sample to be formed from an oversample of 1,872 low income families from the SEO and a nationally representative sample of 2,930 households designed by the SRC. Approximately 18,000 individuals lived in these original families at that time, and are considered to have the "PSID gene" making them eligible to be followed for subsequent interviews. In addition, all individuals born to or adopted by an individual with the PSID gene acquire the gene themselves, becoming PSID "sample persons" who are followed in the study. As members of sample families grow up, move out, and form their own economically independent households, they are interviewed separately, increasing the overall number of interviews conducted each wave.

This unique design of following children of sample members as they themselves become adults, replenishes the sample and helps to maintain its national representation, as well as facilitating the study of outcomes across generations.

In 1990, 2,043 Latino households, including families originally from Mexico, Puerto Rico, and Cuba, were added to the study. While this sample represented three major groups of immigrants, it did not include all post-1968 immigrants. Because of this crucial shortcoming, and a lack of sufficient funding, the Latino sample was dropped after 1995.

In 1997, because of the escalation in costs driven by the doubling of the sample size during its 30-year history, PSID was forced to drop some families from the SEO oversample. At the same time, a sample of 511 families who had immigrated to the U.S. after 1968 was added, in order to maintain the national representation of U.S. families in the study. Budget constraints limited the size of the addition, preventing it from being large enough to support the analysis of immigrant subgroups. Because some of these families created split-off families, their overall number increased to 638 by 2009.

2.2. Respondents and sample sizes

The definitions and terminology used to describe the PSID sample were adopted from the Census Bureau in 1968 and, although dated, are maintained for consistency and for their straightforward following
rules. A single primary adult serves as the respondent. In a married-couple family, the Head is defined as the husband—unless he is physically or mentally incapable of being interviewed. The Head can also be a single female. The term Wife is used for a female in a married couple and “Wife” is used for a cohabitating female (henceforth, the former term is used for both married and cohabiting female partners). PSID attempts to interview the same family member in each wave to maximize consistency over time; approximately 95% of families that appear in successive waves have the same respondent for both interviews.

PSID respondents provide information about themselves, their spouse/partner, and all other family members living together, who are referred to as the “family unit.” Individuals who are not Heads or Wives are classified as “other family unit members” if, at the time of interview, they are members of the respondents’ family and either residing in the interviewed family unit or temporarily away in an institution (e.g. college, jail, hospital, or the military). Family members are individuals who are related to the Head/Wife by blood, marriage, or adoption, although unrelated individuals are included if they are permanently living together and share income and expenses.

Table 1 shows that in 2009 the PSID sample was comprised of 8,690 family units consisting of 13,410 Heads/Wives and 23,102 individual family members (including Heads/Wives). There are substantial sample sizes of individuals at all ages: 7,511 at ages 0-18, 7,396 at ages 19-39, 5,678 at ages 40-59, and 2,517 aged 60 and older. In 2009, 13,041 individuals with the PSID gene had been part of the interviewed sample for at least five waves; 5,876 had survived and participated in every wave since 1968, providing large samples of individuals who have been observed over a substantial portion of their life course.

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Individuals</th>
<th></th>
<th>Heads/Wives</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0-18</td>
<td>7511</td>
<td>32.5</td>
<td>28</td>
<td>0.2</td>
</tr>
<tr>
<td>19-39</td>
<td>7396</td>
<td>32.0</td>
<td>5700</td>
<td>42.5</td>
</tr>
<tr>
<td>40-49</td>
<td>2891</td>
<td>12.5</td>
<td>2744</td>
<td>20.4</td>
</tr>
<tr>
<td>50-59</td>
<td>2787</td>
<td>12.1</td>
<td>2621</td>
<td>19.6</td>
</tr>
<tr>
<td>60-69</td>
<td>1397</td>
<td>6.1</td>
<td>1287</td>
<td>9.6</td>
</tr>
<tr>
<td>70+</td>
<td>1120</td>
<td>4.8</td>
<td>1030</td>
<td>7.7</td>
</tr>
<tr>
<td>All ages</td>
<td>23102</td>
<td>100.0</td>
<td>13410</td>
<td>100.0</td>
</tr>
</tbody>
</table>

1 Includes individuals currently residing in family unit

2.3. Response rates

The PSID has consistently achieved response rates equal to or higher than other panel surveys worldwide (Schoeni et al forthcoming). Response rates are calculated for each of the "sample types" within PSID. The sample is defined across two different strata: first, whether it is considered "core" versus "immigrant refresher", and second, whether in the previous wave, the sample type is "re-interview" versus "split-off" versus "re-contact." In regards to the first strata, the "core" sample consists of all families except those added in the 1997 immigrant refresher sample. In regards to the second strata, the "re-interview sample" includes families who were successfully interviewed in the previous wave. “Split-off” families consist of individuals who left a PSID family unit and established their own economically independent unit. Finally, the “re-contact” sample consists of families who did not respond in the previous wave, but were respondents in the wave before the previous wave. PSID attempts to re-contact and interview these families in subsequent waves, as a way to minimize attrition and maintain the representativeness of the sample.
<table>
<thead>
<tr>
<th>Year</th>
<th>Total</th>
<th>Re-interview</th>
<th>Split-off</th>
<th>Re-contact</th>
<th>Total</th>
<th>Re-interview</th>
<th>Split-off</th>
<th>Re-contact</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1968</td>
<td>76.0</td>
<td>89.0</td>
<td>60.4</td>
<td>81.4</td>
<td>1969</td>
<td>81.4</td>
<td>84.0</td>
<td>95.7</td>
<td>1970</td>
</tr>
<tr>
<td>1972</td>
<td>97.8</td>
<td>98.4</td>
<td>88.6</td>
<td>97.8</td>
<td>1973</td>
<td>97.6</td>
<td>98.0</td>
<td>92.5</td>
<td>97.6</td>
</tr>
<tr>
<td>1976</td>
<td>97.6</td>
<td>98.2</td>
<td>86.5</td>
<td>97.5</td>
<td>1977</td>
<td>97.5</td>
<td>98.0</td>
<td>90.3</td>
<td>97.6</td>
</tr>
<tr>
<td>1980</td>
<td>97.7</td>
<td>98.0</td>
<td>86.0</td>
<td>98.0</td>
<td>1981</td>
<td>98.0</td>
<td>98.3</td>
<td>88.3</td>
<td>98.0</td>
</tr>
<tr>
<td>1984</td>
<td>97.3</td>
<td>97.7</td>
<td>92.0</td>
<td>97.3</td>
<td>1985</td>
<td>97.1</td>
<td>97.4</td>
<td>89.5</td>
<td>97.1</td>
</tr>
<tr>
<td>1988</td>
<td>97.4</td>
<td>97.9</td>
<td>83.3</td>
<td>97.4</td>
<td>1989</td>
<td>91.7</td>
<td>98.3</td>
<td>89.2</td>
<td>98.0</td>
</tr>
<tr>
<td>1992</td>
<td>92.2</td>
<td>95.5</td>
<td>52.1</td>
<td>94.7</td>
<td>1993</td>
<td>95.5</td>
<td>47.4</td>
<td>87.7</td>
<td>94.7</td>
</tr>
<tr>
<td>1996</td>
<td>97.6</td>
<td>97.6</td>
<td>na</td>
<td>na</td>
<td>1997</td>
<td>95.7</td>
<td>95.7</td>
<td>na</td>
<td>na</td>
</tr>
<tr>
<td>2000</td>
<td>91.7</td>
<td>96.7</td>
<td>52.0</td>
<td>93.0</td>
<td>2001</td>
<td>92.7</td>
<td>96.6</td>
<td>57.6</td>
<td>93.4</td>
</tr>
<tr>
<td>2004</td>
<td>94.3</td>
<td>97.0</td>
<td>53.5</td>
<td>94.7</td>
<td>2005</td>
<td>94.2</td>
<td>97.0</td>
<td>53.5</td>
<td>94.7</td>
</tr>
<tr>
<td>2008</td>
<td>94.3</td>
<td>97.0</td>
<td>53.5</td>
<td>94.7</td>
<td>2009</td>
<td>94.3</td>
<td>97.0</td>
<td>53.5</td>
<td>94.7</td>
</tr>
</tbody>
</table>

Notes. na=not available
Table 2 presents response rates for each wave from 1968 to 2009. The core re-interview sample accounts for roughly 85% of the entire sample in recent waves and has a wave-to-wave response rate in almost all waves of 95-98%. The core split-off sample experiences response rates typically in the mid-80% range. The core re-contact sample has response rates in the 50% range in recent waves, and the core re-contact split-off sample is small with variable response rates (e.g. 71% in 2007, 54% in 2009). The response rates for the 1997 immigrant refresher sample are slightly lower than for the core sample; in 2009 the response rates were 96% for the immigrant re-interview sample, 85% for the immigrant split-off sample, and 44% for the immigrant re-contact sample.

2.4. Data collection and panel maintenance

The Survey Research Operations group at the Institute for Social Research has fielded every wave of PSID. Interviews were conducted annually between 1968 and 1997 and biennially thereafter through to the most recent wave in 2011. Since 1973 the majority of interviews in each wave have been conducted via telephone and a computer-assisted telephone collection technology has been used since 1993 (see Couper and Nicholls 1998).

Although most cross-sectional surveys have experienced major declines in response rates over the past several years (de Leeuw and de Heer 2002; Curtin et al 2005; Card et al 2010), response rates in the most recently completed wave were as high as any in the entire history of the PSID. These high rates are attributable to a number of factors, including: incentive payments that translate into roughly $1 per average minute of the interview, that are mailed to respondents within a few days of the interview; an additional small incentive for completion of the interview on a cell phone to compensate for use of paid minutes; maintaining experienced interviewers who are repeatedly matched with the same families; tailored refusal conversion techniques; a respondent newsletter highlighting the importance of the study; inter-wave communications with respondents that include a small incentive for providing updated contact information (McGonagle et al 2011). PSID regularly runs methodological experiments to further improve response rates. Results from a recent experiment indicate that sending respondents a small incentive for providing updated contact information between waves reduces field effort and costs (McGonagle et al 2011a, b; McGonagle et al 2009).

2.5. Sample representativeness

Despite consistently high response rates, there is evidence that lower income families have higher cumulative attrition (e.g. Fitzgerald et al 1998; Fitzgerald 2011). However, parameter estimates of interest have not been found to be biased. In a recent analysis of the effects of cumulative attrition in PSID up to 2007, Fitzgerald (2011) finds little-to-no evidence of biased estimates of sibling correlations, or of parameters, in inter-generational models of health outcomes. The close alignment of weighted estimates from PSID with those from other U.S. benchmark studies – the March Current Population Survey for income (Gouskova et al 2010), the Survey of Consumer Finances for wealth (Bosworth and Anders 2008), the National Health Interview Survey for health status and health behaviors (Andreski et al 2009), the Consumer Expenditures Survey for expenditures (Li et al 2010), and the American Time Use Survey for time use behaviors (Cormcan et al 2011) – support the claim that PSID remains representative of the U.S. population. In other words, cumulative effects of modest wave-to-wave attrition do not appear to have biased the PSID’s cross-sectional representation of key economic or health factors. Although families comprised entirely of post-1997 immigrants are not part of the sampling frame, this group is a small segment of the U.S. population and over time joins the PSID sample through intermarriage. Nonetheless, the addition of a post-1997 immigrant refresher sample remains a high priority for the project’s strategic plan.

Item non-response is also low, with very few questions missing responses for more than 3-4% of cases (Killewald et al 2011). Sample weights are provided for each wave to account for differential probabilities of selection due to the original sample design and subsequent attrition, including longitudinal individual weights, longitudinal family weights, and cross-sectional individual weights. Information about the construction of these weights and their analytic use is provided in companion documents on the PSID website.

2.6. Content domains

A central aim of the PSID since its inception has been to collect and distribute detailed information on income and family demographics in order to support research on the dynamics of economic well-being. With input from its Board of Overseers and the broader scientific community, the Study has been expanded over time to collect and distribute data in many domains, to allow the study of topics of emerging scientific and policy interest. Data have been collected on a wide array of social, demographic, economic, and health topics, which have in turn supported unique multidisciplinary research. In 2011, the 89 minute interview collected data on: employment, earnings, income from all sources, expenditures covering 95-100% of total household spending (Charles et al 2004), housing, foreclosure and mortgage distress, geospatial data, health status, health behaviors, adult and childhood health conditions, health insurance, marriage and fertility, program participation, vehicle ownership, wealth and pensions, debt, and philanthropy. As Table 3 shows, many of these areas have been included in the instrument since 1968. Thousands of additional variables in other domains have been collected throughout the history of the PSID.

Table 3. Content of the PSID 2011 main interview

<table>
<thead>
<tr>
<th>Economic content</th>
<th>Year first collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>Housing characteristics (foreclosure and mortgage distress), utilities</td>
<td>1968 (2009)</td>
</tr>
<tr>
<td>Employment</td>
<td>1968</td>
</tr>
<tr>
<td>Income</td>
<td>1968</td>
</tr>
<tr>
<td>Government transfer program participation</td>
<td>1968</td>
</tr>
<tr>
<td>Consumption expenditures</td>
<td>1984</td>
</tr>
<tr>
<td>Wealth and active savings</td>
<td>1968</td>
</tr>
<tr>
<td>Pensions</td>
<td>1976</td>
</tr>
<tr>
<td>Philanthropic giving and volunteering</td>
<td>2001</td>
</tr>
<tr>
<td>Housework and child care</td>
<td>1968</td>
</tr>
<tr>
<td>Health content</td>
<td></td>
</tr>
<tr>
<td>Work and activity limitations</td>
<td>1968</td>
</tr>
<tr>
<td>Health insurance</td>
<td>1968</td>
</tr>
<tr>
<td>Smoking and alcohol</td>
<td>1968</td>
</tr>
<tr>
<td>Workers’ compensation</td>
<td>1977</td>
</tr>
<tr>
<td>Hospital and nursing care</td>
<td>1981</td>
</tr>
<tr>
<td>Social security disability insurance</td>
<td>1984</td>
</tr>
<tr>
<td>General health status</td>
<td>1984</td>
</tr>
<tr>
<td>Exercise</td>
<td>1986</td>
</tr>
<tr>
<td>Height and weight</td>
<td>1986</td>
</tr>
<tr>
<td>Activities of daily living / Instrumental activities of daily living</td>
<td>1992</td>
</tr>
<tr>
<td>Health conditions</td>
<td>1999</td>
</tr>
<tr>
<td>Emotional distress</td>
<td>2001</td>
</tr>
<tr>
<td>Overall life satisfaction</td>
<td>2009</td>
</tr>
<tr>
<td>Health expenditures</td>
<td>1999</td>
</tr>
<tr>
<td>General health status in childhood</td>
<td>1999</td>
</tr>
<tr>
<td>Health conditions in childhood</td>
<td>2007</td>
</tr>
<tr>
<td>Medication use</td>
<td>2011</td>
</tr>
</tbody>
</table>
Two modules were recently added that present special opportunities for life course research. Added in 2007, the childhood health calendar was designed to provide data on the reasons for the strong and persistent relationship between health status and socio-economic status and its evolution over the life course. The module includes a series of retrospective questions on whether and when Heads and Wives experienced a set of health conditions during their own childhood, including: asthma, diabetes, respiratory disorders, speech impairment, depression, drug or alcohol problems, and other emotional problems. The collection of these data enhances the power of the PSID to examine the effects of health on life course development, by providing markers very early in the life span. Analyses of these data have demonstrated their validity (Smith 2009).

Starting in 2009, new questions were added about mortgage distress, including foreclosure activity, falling behind in payments, mortgage modifications, and expectations about payment difficulties in the coming year. These data can be used to examine the medium-term effects of the recent economic recession on a range of outcomes, including family formation and dissolution, childbearing, and educational attainment. Researchers can examine, for instance, whether higher unemployment rates and declines in income and wealth have led young adults to postpone or avoid higher education, marriage and cohabitation, and childbearing. The estimation of causal effects is made possible by the repeated measures on respondents, the presence in the data of siblings, cousins, and other related individuals, regional variation in the severity of the recession, and the long and rich historical record on individuals and families.

To maintain respondent confidentiality, some PSID data are available only through restricted use contract to users who follow an appropriate data safeguarding plan, including geospatial identifiers and a variety of linked administrative records. Geospatial identifiers, including Census tract, block-group, and block, are available for the current residence at each wave for all families since 1968. County-level data on where PSID individuals and their parents grew up are also available. Links to external administrative data sources have been expanded in recent years, with information now available under restricted use contract on: date and cause of death from the National Center for Health Statistics National Death Index for decedents since 1979; health care claims for 1991-2008 from the Centers for Medicare and Medicaid Services; whether the family has received housing subsidies from the Department of Housing and Urban Development; and school characteristics from the National Center for Education Statistics for primary and secondary public and private schools attended by sample members.
3. Disability and Use of Time Supplement (DUST)

Fielded in 2009, the scientific aim of DUST was to produce a rich and nationally focused data archive to support innovative research on disability, time use, and well-being for older married couples (Freedman and Cornman, 2011). A total of 755 PSID Heads and Wives were each interviewed about two randomly selected days - one weekday 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 performed. Detailed information was obtained about health, functioning, well-being, and stylized time use/participation measures. The inclusion of data from the main PSID allows for analyses about how time use varies, as a consequence of pre-existing health, disability, and economic status of the family, from both care giver’s and care recipients’ perspectives. Linkages to subsequent waves of PSID will allow analysis of the implications of time use for a variety of future health outcomes. Plans are underway to collect a second wave of DUST on all PSID individuals aged 60 and older in 2013.

4. The Child Development Supplement (CDS) and the Transition into Adulthood (TA) study

Beginning in 1997, the CDS collected information on up to two randomly-selected 0-12 year old children (N=3,563) and their caregivers in PSID families. The scientific aim was to provide researchers with a comprehensive, nationally representative (with child-based weights), prospective database of young children and their families, for studying the dynamic process of child development. The same children and their caregivers were re-interviewed in 2002/2003 and again in 2007/2008 with a child-based response rate exceeding 90% in the most recent wave. Topics included: health, skills assessments, parenting styles, time use, school resources and the learning environment in the home, and socio-emotional characteristics of children and their parents.

In 2005, in recognition that the years from 18-24 are critical for life span development, the PSID began a new study designed to follow the children from CDS who had turned age 18 and had completed or left high school and had families still active in PSID, called the Transition into Adulthood (TA) study. The primary scientific aim of TA is to understand the causes and consequences of social, economic, and health transitions of young adults. Information is collected about educational pursuits, employment, occupational choices, education and career expectations, family responsibilities, skills and abilities, intimate relationships, and more. Along with data collected during the CDS, detailed information is available about development from early and middle childhood through adolescence and into adulthood; additional information will be collected on this cohort over its life course as these youth transition to economic independence and become PSID Heads and Wives. TA data have been collected biennially for 2005-2011 and will be collected through 2015 at which time all children from CDS will have been observed at least once in the study. The response rate for TA was 92% in the most recent wave. The CDS-TA-PSID archive is unique in the scientific research opportunities it presents for inter-generational and life course analysis.

5. Data distribution and documentation

The majority of PSID data and documentation have been freely and publicly available on the Internet since 1996 through the PSID Data Center (www.PSID.org). Information is currently available on nearly 70,000 variables, on nearly 70,000 individuals, and for all waves of main PSID data and supplements. Users can create customized data extracts from any set of waves by searching or browsing for variables, obtain customized codebooks specific to their data extract, and can archive data extracts for shared and future use. Users “load” their data carts with variables by wave. Variable descriptions, including univariate statistics and names of the same variable in other waves, are viewed by clicking an “open-book” icon next to each variable. Users can edit their cart by removing or adding variables through a return to the “data aisle.” Data carts may be saved, allowing specific extracts to be shared with colleagues,
reviewers, and students. A range of file formats is available when the user is ready to “check out,” including SAS, STATA, SPSS, dBase, Excel, and ASCII. PSID developed an online cross-year variable index that facilitates searching and browsing all variables across the full archive from 1968 to the most recent wave, and for all waves of CDS and TA. Organized by content domains, the index is integrated with the Data Center so that users can view the codebook and add variables directly to their data cart from the index.

PSID has developed user tutorials covering a variety of topics. The most recent tutorial provides instruction on creating parent-adult child pairs to examine inter-generational transmission of wealth, health, and other outcomes. These tutorials have been positively received by the research community and by teachers (Stafford and Chiteji 2004).

Finally, an application is available called the Family Identification Mapping System (FIMS) that automatically creates a customized file containing identifiers for PSID respondents’ relatives (including biological and adoptive children, parents, grandparents, great-grandparents, and siblings). FIMS greatly simplifies the process of creating genealogical samples by automatically generating a data extract based on user input.

6. Ethical approval, funding, and administration

The majority of PSID data and documentation are freely and publicly available on the Internet. Before downloading public data, users must register with the PSID and agree to conditions of use, which include pledging to make no attempt to identify study participants, and to report any identification of study participants or data errors immediately to PSID. Individuals wishing to obtain restricted data such as geospatial identifiers must provide a data security plan that meets PSID data safeguarding requirements, and have approval from their institution’s human subjects review and/or privacy board (described here: http://simba.isr.umich.edu/restricted/RestrictedUse.aspx). The University of Michigan Health Sciences and Behavioral Sciences Institutional Review Board conducts annual reviews of the PSID data collection and distribution protocols and survey instruments to ensure the rights and welfare of its research participants are protected.

The National Science Foundation has been the lead funder of PSID over the past several decades. Recent support has also been received from the following organizations: National Institute on Aging, Eunice Kennedy Shriver National Institute on Child Health and Human Development, Economic Research Service of the U.S. Department of Agriculture, the Center on Philanthropy at Indiana University, the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services, and the Department of Housing and Urban Development. The study has been housed at the Institute for Social Research, University of Michigan, from its beginning in 1968 to the present and is directed by a team of social scientists. Since 1982, a Board of Overseers has fostered input from the national community of scholars, researchers, and policymakers.

7. Research potential of the PSID for life course research

The extended time series and consistently high response rates of the PSID provide substantial analytic power to study antecedents and consequences of a range of social, health, and economic factors. For older sample members in particular, the history available is extraordinarily rich. Since individuals are followed over the entire life course, the number of waves that an individual appears in the sample is related to their age. However, individuals who were born into PSID families also have substantial information on their childhood circumstances from their parents’ and grandparents’ reports. This combination of information on childhood circumstances and later adult behavior and outcomes represents a major analytical strength that supports a variety of different analyses. Moreover, the analytical samples are large. For instance, approximately 3,000 individuals in the original cohort under age 18 in 1968 were ages 50 or older by 2007 and are therefore represented in the data for a full 40 years of their lives. This cohort, born 1950-1968, overlaps substantially with the 1946-1964 Baby Boom generation and is now reaching the ages where life events such as acute health shocks, chronic disease, and labor force transitions occur. For
children in the CDS and TA, a wealth of information exists on their behavior, health, and development as they passed through early, middle, and late childhood and into adolescence and young adulthood, and, increasingly, as members of the main PSID. The resulting data can be used to better understand how early circumstances, such as childhood socio-economic status, health, neighborhood and school characteristics, and educational choices, shape health and well-being over the life course.

The following of sample members who are descendants of the original 1968 sample yields inter- and intra-generational samples, which are powerful additions to the study of life course development. The inter-generational sample is comprised of adult children who split off from their sample parents to form their own households and who are recruited into the study. Intra-generational analyses have focused on outcome comparisons among siblings, as a means to control for, or examine, unmeasured family effects (e.g. Conley and Glauber 2007). Relative-pairs that can be examined in the PSID include dyads formed from parents and children, from siblings, and from grandparents and grandchildren. There are hundreds, and in many cases thousands, of these pairs within the PSID; they are typically observed for many years, leading to a large number of relative-pair-year observations. In 2009 alone, PSID had 4,626 Heads/Wives who had a parent in the study, and 4,927 had a sibling in the study who was also a Head or Wife in that year (yielding about 2500 sibling pairs). Among Heads or Wives aged 50 and older in 2009, 2,858 had an adult child participate in the main interview (many more had children interviewed in CDS-TA), 1,285 had at least one sibling interviewed, and 722 had a parent interviewed. For many PSID families, self-reported information is currently available on up to four generations within the same family, at various points in their life course.

7.1. Illustrative research examples
This section provides a small sampling of ways the data have been used for life course research, that may generate new ideas and future scientific discovery. The PSID website provides a complete bibliography of all studies using the data.

Effects of the Great Recession on young adults
PSID and TA data provide an unprecedented opportunity to examine how recent economic adversity starting at the end of 2007 in the U.S., in combination with secular changes in federal financial and mortgage policies, have shaped schooling and employment choices of young adults. Recent work has shown that changes in a household’s housing wealth in the four years prior to a child being of college-age affects the likelihood that the child attends college (Lovenheim 2010). Data from the recent PSID foreclosure and mortgage distress module can be used to study how exposure to financial adversity defines early labor force participation and its link to later employment opportunities and educational trajectories over the life course. The choices and experiences in these cohorts can be compared with those of young adults who reached age 18 or 19 before the financial crisis.

Marital transitions
Studies using the PSID have examined trends in marriage, cohabitation, divorce, and separation, and the causes and consequences of marital transitions. Studies have found that factors influencing marital transitions include risk preferences (Schmidt 2008), education (Ono 2009), household bargaining (Cooke 2006), child support (Chiappori and Weiss 2007), and job loss (Charles and Stephens 2004). Studies have advanced scientific understanding of the impact of marriage, cohabitation, divorce, and separation on numerous outcomes, including labor supply (Aughinbaugh 2010), expenditure patterns (DeLeire and Kalil 2005), child development (Page and Stevens 2005; Bjorklund et al 2007), assets and debt among women (Fisher 2005), and bankruptcy filings (Fisher 2006). The data have been used to document the weakening over the past three decades of marriage and the nuclear family in the U.S. (Clark and Withers 2009; Seltzer et al 2005).

Fertility and birth outcomes
PSID is an important data source for understanding the causes and consequences of teen childbearing, with recent work (e.g. Sullivan et al 2010; Wolfe et al 2007; Lopoo 2005) building on numerous studies from the 1990s and early 2000s (Geronimus and Korenman 1992, 1993; Haveman et
al 1997, 2001; Corcoran and Kunz 1997; Hoffman et al 1993; Foster et al 1998a, 1998b). Studies of the determinants of fertility among women of all ages have examined the effects of child support (Plotnick et al 2007), tax policy (Huang 2008), state-level welfare programs (Ryan et al 2006), and neighborhood poverty (South and Crowder 2010). The data have been used to examine the effects of birth outcomes and health in early childhood on health and well-being in later life (Johnson and Schoeni 2009; Haas 2007; Smith 2009).

Effects of neighborhoods over time

The availability across the life course of PSID individuals’ census tract of residence (and census block beginning in 2009), through a restricted use contract, permits the study of neighborhood effects on development. A number of studies have used the data to examine neighborhood effects on child, adolescent, and young adult development (Dearing et al 2009; Jackson and Mare 2007; Timberlake 2009a, 2009b; Wimer et al 2008), health (Do and Finch 2008; Halliday 2007; Halliday and Kimmitt 2008; Johnson and Schoeni 2009), education (Brooks-Gunn et al 1993; Crowder and South 2003; Galster et al 2007; Harding 2003), income and earnings (Sharkey 2008), the inter-generational transmission of neighborhood context (Dawkins 2005a; Sharkey 2008; Solon et al 2000), and fertility behavior (Clark 2009; South and Crowder 2010; South 1999, 2001a, 2001b). Because of its oversample of African American families, the data have been used to examine levels and trends in racial equality in neighborhood economic status and migration (Sharkey 2008; Dawkins 2005b, 2006; Freeman 2005a, 2005b, 2008; South and Crowder, 2005; Timberlake 2007; White et al 2005; Vartanian et al 2007; Crowder and South 2005).

Socio-economic outcomes across and within generations of the same family

Because PSID follows adult children as they form their own households, it is uniquely positioned to investigate intergenerational effects. The data have been used to examine correlations across generations in health (Davis et al 2008; Valerio et al 2009), socio-economic status (Charles and Hurst 2003; Eberharter 2008; Gouskova et al 2010; Lee and Solon 2009; Mayer and Lopoo 2005; Solon 1992; Vartanian et al 2007), and philanthropic behavior (Wilhelm 2008). In fact, because of the length of the panel, PSID is now being used to estimate changes over time in the intergenerational transmission of economic status (Lee and Solon 2009; Mayer and Lopoo 2008). Researchers are also beginning to use the health data in conjunction with the socio-economic data to examine the extent to which transmission of health across generations within the same family accounts for the transmission of socio-economic status, and vice versa (Johnson and Schoeni 2009).

Life course consumption, savings, and wealth accumulation

The economic position of individuals nearing retirement is heavily influenced by saving and consumption decisions made during their working years. The experience of the large Baby Boom generation – the leading edge of which turns age 65 in 2011 -- is particularly well-represented in the PSID. Boomers were ages 4 to 22 years old when the PSID began in 1968, and most were still living in their parents’ home. Substantial data collected on savings and wealth over several decades can be used to provide insights into how the financial behaviors and decision-making over the lifetimes of this generation has shaped their economic well-being as they reach retirement age.

Comparative research over time and across countries

The design of the PSID offers both flexibility to study emerging trends, and harmonization of key content to facilitate cross-national comparisons across all age groups. Comparable panels which facilitate study of the interplay of social and economic policies, health trajectories, and economic well-being over the life course (Banks et al 2003, 2009; Bjorkland et al 2007; Burkhauser et al 2005; Eberharter 2008) include the UK Understanding Society study, the German Socio-Economic Panel Study, the Household, Income and Labour Dynamics in Australia Survey, and many others. The Cross-National Equivalence File Project at Cornell University has further facilitated comparative analysis, by aligning variables between PSID and its companion surveys.
8. The Future of PSID

As of this writing, data collection of the 2011 main interview has concluded and the data are being processed for release in 2012. Plans are underway to collect new information in 2013 that will respond to emerging research questions and policy issues. For example, in recognition of the role of the extended family as an increasingly important source of assistance over the life course (Bianchi et al. 2008), PSID will provide a more complete enumeration of members of extended families and assess short-term transfers of time and money and long-term, life-cycle transfers for education and housing. Moreover, to enhance information available describing the earliest stage of life, additional data will be collected about between-wave births, and a protocol for obtaining links to birth records will be tested. An education update will be collected from all family members, yielding timely information about investments in education over the life course. In 2014, PSID plans to collect new data on psychological well-being, providing an unique opportunity to study influences of social and economic factors on well-being across all age groups. Finally, two high priorities in the next few years are the addition of a post-1997 immigrant sample refresher, and the implementation of a new PSID Child Development Supplement, in order to collect detailed information on the health, development, and well-being of all (nearly 7,000) children under age 18, including biomeasures such as DNA (Sastry et al 2009).

Acknowledgements

Support for this work was provided by the National Science Foundation (SES 0518943), the National Institute on Aging (P01-AG029409; R01-AG19802), and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (R01HD052646; R01HD33474).

References


Freeman L. (2008) Is class becoming a more important determinant of neighborhood attainment for African-Americans? *Urban Affairs Review*, 44, 3-26. [http://uar.sagepub.com/content/44/1.toc](http://uar.sagepub.com/content/44/1.toc)


---

**Endnotes**


2 See Table 2 of the PSID User Manual for the number of original sample members who have been interviewed in each wave and Table 3 which lists the number of sample members in each wave who have left the study, died, or were dropped from the sample: [http://psidonline.isr.umich.edu/data/Documentation/UserGuide2009.pdf](http://psidonline.isr.umich.edu/data/Documentation/UserGuide2009.pdf).

3 See the PSID Technical Series Paper here for details on the immigrant sample addition [http://psidonline.isr.umich.edu/Publications/Papers/tsp/2000-04_Imm_Sample_Addition.pdf](http://psidonline.isr.umich.edu/Publications/Papers/tsp/2000-04_Imm_Sample_Addition.pdf).

4 Starting with the 2007 wave, three levels of PSID geospatial data are available, in order of descending precision: Census Tract, Block Group, and Census Block. Census Block is the smallest level of data made available by PSID. Several Blocks make up Block Groups, which in turn make up Census Tracts.

5 For more information see the CDS and TA User Guides here: [http://psidonline.isr.umich.edu/Guide/documents.aspx](http://psidonline.isr.umich.edu/Guide/documents.aspx)