Published in final edited form as:

Ann Am Acad Pol Soc Sci. 2012 November ; 680(1): 193–212. doi:10.1177/0002716218791751.

USES OF PANEL STUDY OF INCOME DYNAMICS DATA IN RESEARCH ON AGING

Douglas A. Wolf

Maxwell School of Citizenship and Public Affairs and Aging Studies Institute, Syracuse University, dawolf@maxwell.syr.edu, Aging Studies Institute, 314 Lyman Hall, Syracuse University, Syracuse, New York 13244

Abstract

The Panel Study of Income Dynamics (PSID) has, over its 50-year history, proven to be a useful source of data for research on virtually all of the major topics in the area of social gerontology. This usefulness reflects three of the leading features of the PSID: its longitudinality, its content, and its tracking rules, which permit users to develop family-based and generationally-linked measures. This paper summarizes key areas of survey content, including both routinely-collected data and several one-time or occasional supplements to the routine items. The paper also illustrates how these data elements have been used, providing examples of published papers in several areas of social gerontology. Finally, the paper points out some methodological issues associated with the PSID design; these methodological issues arise, in varying degrees, in longitudinal studies other than the PSID, and should be acknowledged by both the producers and consumers of longitudinal-data research.

Keywords

Life course; social gerontology; contextual effects; longitudinal data; selection effects

In 2018, the Panel Study of Income Dynamics (PSID) celebrated its 50th birthday, if counting from its first year of data collection. This was a remarkable and unprecedented achievement. Having reached what for humans would generally be considered “middle age,” it seems fitting to survey the presence of, and prospects for, PSID data in research on aging. “Aging” is, to be sure, a very broad field of study, and comes in numerous narrower flavors such as “demography of aging,” “economics of aging,” “epidemiology of aging,” “psychology of aging,” “biology of aging,” and so on. Some of these narrower flavors are, to varying degrees, multidisciplinary. There are, as well, varying amounts of overlap of substantive foci, if not of disciplinary perspective, across these subareas. To place some bounds on my task, I will focus mainly (although loosely) on what is generally viewed as *social gerontology*, a field “... concerned mainly with the social, as opposed to the physical or biological, aspects of aging ... [and whose topics of interest include] family relationships, health, economics, retirement, widowhood, and care of the frail elderly” (Quadagno 2011, 4). Likewise social gerontology also adopts the “life course” perspective, with its emphasis

NOTE: Previous versions of this article were presented at the 2017 PSID Annual User Conference, Ann Arbor, September 14, 2017, and at the annual meeting of the Population Association of America, Denver, April 28, 2018. Vicki Freedman has provided helpful comments on the earlier versions.

on “linked lives” (Setterstein 1999)—a perspective to which PSID data are particularly well suited.

This article joins a literature devoted to elucidating the usefulness of the PSID in empirical research in general, and for research on aging in particular (for just a few examples, see McGonagle et al. 2012; Schoeni et al. 2015; McGonagle and Sastry 2016). Given the very large body of PSID-based research—more than 5,000 items as of early April 2018, of which nearly 1,400 include the keyword “aging”ⁱ—as well as the existence of several guides to the data and the ways in which they have been used, I offer a selective review of research topics and their associated methodological issues.

The PSID has for decades demonstrated its usefulness in the study of individual and family income, employment, poverty, and the time-paths of each of these dimensions of well-being. Among the hundreds of published aging-related works using PSID data, the subjects of retirement, pensions, savings and wealth, social security and economic well-being are well represented. This is unsurprising, given the PSID’s inception as a source of data focused on the components of family income. The PSID is also known for its data on family and household composition, intergenerational relationships, and health. Accordingly, it is no surprise that the PSID has contributed much to the field of social gerontology.

As an admittedly rough way of characterizing the field, I have obtained a list of the most frequently used keywords attached to the articles published in the *Journal of Gerontology: Social Sciences (JG:SS)*, one of the leading publications in this area, during the period 2011–2015. The top ten keywords used to describe these papers are: cognition; social support; depression; health; disability; caregiving; widowhood; life course; well-being; and gender.ⁱⁱ And, when I perform a “title” search on these keywords, using the PSID online bibliography, I find that all but the first (“cognition”) appear in the database.ⁱⁱⁱ In the remainder of this article, I review selected features of the PSID data, and summarize several important research contributions that have used those features. I go on to mention a few methodological issues that are particularly relevant to the social-gerontology topical areas; those issues have confronted past PSID data users, and will continue to do so. My conclusions include a few suggestions about public policy issues, which the PSID is especially well-positioned to inform.

KEY FEATURES OF THE PSID

For purposes of supporting research in social gerontology, three of the most salient features of the PSID are its *longitudinality*, its *content*, and its *tracking rules*. With respect to *longitudinally*, the fact that data have been collected over a 50-year period makes the PSID unique. According to cohort life tables produced by the Social Security Administration,^{iv} a man born in 1968, the first year of PSID data collection, had a life expectancy at birth of

ⁱAccording to the PSID’s “Bibliography” (<https://psidonline.isr.umich.edu/Publications/Bibliography/Search.aspx>).

ⁱⁱI am grateful to Deborah Carr, the editor of *JG:SS*, for providing me with this list.

ⁱⁱⁱ“Full text” search of abstracts or articles is not available in the online PSID Bibliography, so my search results undoubtedly understate the true picture of PSID-based research on these topics. For instance, while the word “cognition” gets no hits in a PSID Bibliography title search, the word “memory” does.

^{iv}See <https://www.ssa.gov/oact/tr/2012/lr5a4.html>.

76.4 years (for a woman the corresponding figure is 81.3 years). Thus, PSID data have already been collected for what will turn out to be roughly two-thirds of this cohort's average life span (and more years of data collection are, of course, anticipated). This allows for a remarkably thorough depiction of people's lifetime circumstances.

The *content* of the PSID is limited by several practical considerations—the ability to elicit useful information in a closed-end telephone-administered survey, the need to limit respondent burden, a desire to maintain longitudinal consistency once question-wording patterns have been established, and so on—as well as its scientific ambitions. In addition to questions about employment and income, which have been asked in every interview, the following “routine” data elements are of particular relevance for research in aging:

- Whether a household head's income includes “help from relatives,” a rough measure of intergenerational financial support;
- Beginning in 1969, indicators of monetary contributions given to others—including, but not limited to, family members—outside the household, another rough measure of intergenerational financial support;^v
- Coresidence with family members (parents, children, siblings), another measure of intergenerational “solidarity” and resource flows (or, possibly, joint production or economies of scale);
- A work-disability item (“[d]o you have any physical or nervous condition that limits the kind of work you can do?”); and
- Geographic codes that can be used to locate individuals in a particular environmental, social, or public-policy and programmatic context, or to determine the distance traveled in a residential move, or—of particular importance—the distance between a PSID respondent and his or her parent or child. Available codes include the state, county, ZIP code, Census tract, and Census block-group identifiers.

Several additional content areas were added at some point after 1968, and have subsequently become routinely included. Important among these are several indicators of health, disease, and disability. Questions on self-rated health were introduced in 1984; on Activities of Daily Living (ADL)-type disabilities^{vi} in 1992; and on selected chronic diseases and conditions in 1999 (Insolera and Freedman 2017). The longitudinal coverage of even these “late starters”—nearing, and in some cases surpassing, 20 years—compares favorably to many other data sources.

The PSID also periodically includes additional, more detailed, questions or modules devoted to a topic of particular interest for research on social-gerontological topics. Among the examples of these questionnaire modules are:

^vin early years the question took the form “Do you help to support anyone who doesn't live here with you?” In later years, greater specificity was introduced, e.g., “... did you give any money toward the support of anyone who was not living with you at the time, including child support, alimony, money given to parents, and things like that?”

^{vi}Specifically: bathing or showering; dressing; transfer from bed or chair; mobility (walking); getting outside; and using a toilet.

- **Intergenerational resource flows.** In 1988, and again in 2013, the PSID interview included an extra module of questions on several types of intrafamily support. The 1988 questions addressed time help given to and received from, and monetary assistance given to and received from, respondents' parents and others during the preceding calendar year. Some of those questions were restricted to resource flows involving nonhousehold residents only. The questions asked in 2013 repeated, and for some types of resource flows expanded on, these items. The 2013 module included questions distinguishing resource flows between respondents and their parents from flows between respondents and their children, and also added questions about transfers over a long time horizon (since age 18) as well as the short-term (1-year) items used in 1988 (Schoeni et al. 2015).
- **Time use and well-being.** Special "supplement on disability and use of time" (DUST) modules were fielded in 2009 and 2013. The 2009 version was completed both by couples in which both spouses were at least 50, and one was at least 60; the 2013 version was extended to include single individuals age 60 and older, and included spouses of all ages (Freedman and Cornman 2012; 2015). These supplements include items on physical, mental, and sensory impairments; adaptations and behavior changes that can be associated with the impairments; well-being; care activities; and a 24-hour time diary.
- Retrospective data on childhood circumstances (2014). In this supplement, individuals who were 19 or older and a household head, or the spouse of a household head, in the 2013 PSID interview, completed a 2014 survey about a range of childhood circumstances—physical and mental health conditions, socioeconomic status, neighborhood quality, friendships, school experiences, exposure to the criminal justice system, and the respondents' parents (McGonagle and Freedman 2015).
- Another supplement on **well-being and daily life**, fielded in 2016, collected—through self-administered questionnaires—a number of items relating to life satisfaction, positive and negative emotions, daily activities and experienced well-being, and personality types (Freedman 2017).
- A final example of data particularly suited to research on aging consists of items derived from the 1940 U.S. Census, which are to be linked to PSID records as part of a recently begun NIH-funded project. Given the "72-year rule"—a legislatively established prohibition on the release of identifiable information about an individual until 72 years after it was collected for the decennial census^{vii}—data of this type are, by definition, relevant to aging-related research.

Given the broad expanses of time and subject matter that they cover, the PSID data can support research on numerous topics, and can be applied using numerous methodologies. In addition to their value as a source of descriptive information, PSID data elements can be used to support a variety of multivariate statistical models. Greatly simplified, three basic forms of models include:

^{vii}See https://www.census.gov/history/www/genealogy/decennial_census_records/the_72_year_rule_1.html (accessed 8/31/2017).

$$y_t = f(x_t), \quad (1)$$

i.e., a simple cross-section;

$$y_t = f(x_1, \dots, x_t), \quad (2)$$

a type of “cumulative exposure” model; and

$$\{y_t, y_{t+1}, \dots, y_T\} = f(x_t), \quad (3)$$

a “trajectory” model. In (1) – (3), t stands for an arbitrary year somewhere in the time period spanned by PSID, “1” is a baseline year (which could be, but need not be, as early as 1968), while T is some terminal year (which could be, but need not be, as late as 2015, or—eventually—some later year).

The special modules mentioned before (e.g., intrafamily transfers) include data elements that are more detailed for the years in which they are fielded, and may have no counterparts in other years. This situation tends to make the cross-sectional model the most obvious choice, although studies of this type often incorporate additional data elements from other years to enrich the measures used. The “cumulative exposure” and the “trajectory” approaches, examples of each of which will be mentioned in the following section, make fuller use of the longitudinal aspect of the PSID.

The preceding model types are implicitly formulated to represent an individual (or possibly a household) at, or over, time. However, each can—thanks to the PSID’s tracking rules, which allow users to link records along lines of descent—be extended to encompass intergenerational linkages. For example, an individual’s health or disability trajectory [represented by (3)] could influence her capacity to provide, or her claims on receiving, support late in life to or from members of her offspring generation, and those intergenerational flows of support at a point in time might be conditioned on other contemporaneous factors in either generation’s circumstances [represented in a variant of (1)].

EXAMPLES OF THE PSID’S USES IN AGING RESEARCH

In this section I mention several topical areas in which PSID data have been, but also could be more fully, applied to research that falls under the generic social gerontology heading. Some of the examples given could be placed into more than one of the topical areas listed.

Late-life consequences of early-life exposures

There is a growing recognition that conditions and contexts of early life shape later health, well-being, and mortality outcomes. This is partly due to accumulating evidence produced

using data from the PSID and other longitudinal data sources. One of the markers of items in this literature is an allusion to the “long arm of childhood” (e.g., Hayward and Gorman 2004; Latham 2015). The contexts considered in this literature include the household, the neighborhood, and the place of residence (i.e., city, county, or even state); they can also be traced to social institutions (e.g., the military) and to interaction between individual-level behaviors and contextual factors. While much contextual information can be collected in, or inferred from, responses to survey items such as those used in the PSID, the existence of geographic identifiers (such as census tract or county-of-residence codes) in the PSID opens up a vast potential for linkage of individual longitudinal records to area-specific information.

Several PSID-based studies adopt some form of this long-time-horizon research design. For example, Karraker, Schoeni, and Cornman (2015) select a sample of male PSID respondents aged 20 to 40 in 1972 (their baseline year), tracking their survivorship and mortality over a 35-year follow-up period. Although their principal focus is on the association between several psychological traits and longevity, they do control for one early-life household-contextual factor, namely living in a household that contains smokers. Having lived in a household containing a smoker significantly increases the chances of early death, whether in a baseline-only model or a model that includes measures of marital status and income over the life cycle; however, if the respondent’s own later smoking behavior is added to the model, the effect of the earlier-life exposure to smoking loses significance.

The own-household context depicted in the Karraker, Schoeni, and Cornman (2015) study is located near the “micro” level end of a continuum for which one’s country of residence might represent the most “macro” of levels. Lillard et al. (2015) adopt the latter focus, analyzing the effects on self-reported health status over the adult life cycle of exposure to income inequality prior to adulthood. Their measure of income inequality is constructed for the United States as a whole. Using 25 years of PSID data, they find a consistent and robust health-reducing effect of early life exposure to income inequality, even after controlling for additional early-life exposures (such as childhood health and poverty status), parents’ education, and one’s own lifetime income.

To introduce neighborhood-, city-, or county-level contextual factors into a PSID data analysis, it is necessary to link external information using geo-coded identifiers. An example of this approach is the study by Spring (2017), who uses PSID sample members age 45 or older in 1999, following them through 2013, at which time the surviving individuals are 59–104 years old. Spring focuses on the built environment, using indicators of the per-capita density of supportive services (seven types of establishments) and of commercial decline (using three categories of establishments). Counts of both categories of establishments come from the economic census. Quartiles of the two density measures are then determined, and each neighborhood (zip code) is categorized, using judgmental coding, into one of five built-environment types (e.g., “low density,” “service dense,” “commercially declined,” and so on). Spring examines the association between short-term (one-year) exposure and long term (a cumulative measure, averaging over all survey waves) exposure to these environmental types, and finds significant differences in self-rated health across selected categories of exposure.

Another example of this category of research design is found in Johnson and Schoeni (2011), who use data on nearly 4,400 children who were in the initial (1968) PSID sample and followed up to 2007. Johnson and Schoeni investigate the distinct effects of weight at birth, childhood poverty and other family-background characteristics, along with local-area factors measured at a very fine level of geographic detail—census tracts—on the incidence of chronic diseases and conditions during adulthood. They find significant effects of childhood poverty on the onset of hypertension, diabetes, and cardiovascular diseases, all of which are well-established factors in late-life disability (Freedman and Martin 2000).

While Spring's (2017) results reveal the power of local-area contextual factors to explain individual-level health outcomes, they also serve to illustrate some of the challenges posed by the linkage of longitudinal survey data to area-specific contextual measures. For example, there are multiple levels of relevant context, and these distinct contexts may not fall into neatly nested hierarchies. Whereas multiple zip codes—widely used in applied research to represent “neighborhood”—may align neatly with county boundaries, there are many examples of local labor markets that encompass more than one county, and may even cross state lines (for example, W. Connecticut–New York City–Northern New Jersey; or Philadelphia–Camden and adjacent areas). Moreover, not only may features of differing spatial aggregates within a hierarchy exert influences on individual or household outcomes, there may be additional influences that can be traced to adjacent contexts as well (e.g., housing prices in San Francisco influence housing and travel behavior many miles away). State-level policies or programs may influence individual-level outcomes throughout a state, but they may, as well, be modified at the county level within the state. Finally, data-reduction techniques—such as those employed in Spring's study, which maps the population densities of ten types of establishments into a single five-level categorical variable—are an understandable response to a need to conserve degrees of freedom when there may be a large number of contextual dimensions. However, the resulting summary measures might obscure which of the several underlying observable attributes are the most influential. These difficulties are further amplified when we consider multi-year sequences of contextual exposures, whether experienced over a child's or an adult's lifetime: adverse exposures may matter more at some ages than others; they may matter only if sustained over some well-defined set of years; their consequences may fade over time, or may be overcome by a favorable change in environment, and so on. All such questions remain as additional research opportunities, presenting potential users of the PSID with essentially limitless possibilities.

As a final point concerning the usefulness of area-specific measures, it is worth noting that linking individual-level survey data to local-area contextual measures can serve the purpose of addressing omitted variable bias in evaluation or outcomes research. One example (although not one based on PSID data) is the use of geographic distance to the nearest specialty hospital as an instrumental variable, in a study of the effectiveness, with respect to later survival, of catheterization and revascularization in the treatment of acute myocardial infarction (Newhouse and McClellan 1998). Many possible variations on this approach can be envisioned.

Patterns of functioning and disability

Health, functioning, and disability status are among the leading topics of interest in social gerontology. The PSID lends itself to a wide variety of study designs in this area. One such design is the use of longitudinal data on health or disability transitions to calculate active life expectancy (ALE); this can be viewed as a type of “trajectory” model [see (3) above]. Even though the PSID did not begin to routinely ask about ADL-type disability until 1992, well into its existence, the temporal span of the resulting sequence of disability measures compares favorably with other leading data sources used for this type of research (in particular, the Health and Retirement Study or HRS).

Moreover, compared to other data sources used in ALE research, the PSID allows for an “origin” point much earlier than age 65 or 70 that is often used. Laditka and Laditka (2016) use a sample of nearly 9,000 PSID respondents age 40 and older in 1999 and followed until 2011. They emphasize differences in ALE by race (white and black, only), gender, and level of education, finding (as have others working in this area) that less-educated people live shorter lives, on average, than more-educated people; they also spend a higher percentage of their lives after age 40 living with a disability than do more-educated people.

One question that is gaining increasing attention among researchers, practitioners, and advocates is the differences between people “aging with” disability—that is, spending most or all their adult years with a disability—versus “aging into” disability—that is, making what is generally assumed to be a first transition into having a disability late in life. In particular, one issue on which almost no information can be found is how people’s late-life circumstances differ (with respect to material resources, health status, living and care arrangements, use of services, and so on) according to whether their period of late-life disability began late in life or was preceded by a possibly lengthy period of adult life spent with disability. The PSID, with its long span of coverage, would seem to be an obvious source of data with which to investigate the “with versus into” question. However, I have not located any such study. Clarke and Latham (2014) cover one of the two periods—the “aging with” period, but not the other. Their sample consists of PSID respondents age 20–34 in 1979, and followed until 2009; thus, they obtain good coverage of the adult part of the life cycle. They classify people as “living with” disability if they report having a physical or nervous condition that limits the kind of work they can do—the work-disability question mentioned before—on four or more occasions during the 1981–1994 period. Persons so classified are found to report poorer health, to work fewer hours, to have lower probabilities of employment, and to have lower household income, than those not aging with disability.

Shuey and Willson (2017), who use a slightly longer sequence of PSID years (1981–2013) and a very different approach to classifying disability patterns compared to Clarke and Latham (2014), address patterns of work disability. Shuey and Willson apply a latent class estimator to their 33-year sequences of work-disability indicators, concluding that a five-class solution provides the best fit; interestingly, they find that about two-thirds of their sample falls into a “never disabled” group, while 4 percent have a high probability of living with a work disability throughout adulthood (ages 22–32 through 54–64). A surprisingly high percentage of the sample (more than 24 percent) arrives at the threshold of “old age” (age 65) with a high chance of having a work disability. And, using probabilistically

assigned class-membership indicators as predictors, they show that both the early and midlife-onset (of disability) classes are less likely to own a home, more likely to be in poverty, more likely to have fewer assets, and less likely to have pension coverage, in 2013 (when they are 54–64), compared to the never-disabled group.

Given the findings from both the preceding studies, it is reasonable to anticipate that the late-life circumstances of those with disabilities during their adult lives will be less favorable than those who have comparable levels of disability that came late in life. This is an interesting and important question, and has implications for policy and practice, yet one that remains largely in the not-yet-addressed realm. The PSID also presents opportunities for modeling trajectories in other domains, such as body weight change and cognition, both of which have been shown to exhibit distinctive patterns of downturn in the years immediately preceding death (see Alley et al. 2010 and Sliwinski et al. 2006, respectively). With its long time span and detailed coverage of adult-life circumstances, the PSID seems like a promising basis for detecting early markers of these end-of-life phenomena.

Finally, the topic of late-life disability can be linked to the contextual-factors issue discussed above. Montez, Hayward, and Wolf (2017), for example, demonstrate an association between disability and selected features of one's state-level policy context (cigarette tax rate and supplementation of the Earned Income Tax Credit). However, these findings are based on cross-sectional data from the American Community Survey, and reflect one's current state of residence only. Using the PSID, it would be possible to account for a much longer history of contextual influences on disability in various policy domains.

Intrafamily and intergenerational financial support and caregiving

An extensive body of research has been directed toward determining whether resource flows within families reflect *altruistic* motives—the idea that person A's generosity toward person B increases A's well-being—or are a form of *exchange* between self-interested parties—the creation or discharge of a debt for which repayment, in some form and at some time, is expected. This distinction matters, because altruistic resource flows can be seen as a manifestation of “family solidarity” (Bengston et al. 2002); however, altruistically motivated transfers may be “crowded out” by public transfer programs. Moreover, different motivations for intergenerational support may lead to different patterns of inequality within families in sharing the “burden” of parent care. However, despite the many theoretical models that have been developed to explain observed resource flows, there remains a “... dominant feeling of disillusion in the [economics] profession about the explanatory power of these models” (Arrondel and Masson 2006, 978).

Apart from the lack of discriminatory power among these various theoretical models, there is an empirical problem: it is difficult to establish whether an observed intrafamily flow of resources is part of a self-interested exchange transaction when the data cover only a brief period in the family members' lives. A parent may transfer resources to a young-adult child (helping, for example, with a home purchase) in anticipation of repayment in the form of coresidence or personal-care services decades later. Much of the available data on intrafamily support, however, are temporally limited. The PSID's 1988 intergenerational transfers module, for example, provides measures of personal-care support and financial

support between parents and children during only one year—1987—of the lives lived in either of those generations. Those data have been used in many studies (Altonji, Kayashi, and Kotlikoff 2000; Couch, Daly, and Wolf 1999; Furstenburg, Hoffman, and Shrestha 1995; Schoeni 1996), but they cannot provide much resolution to the altruism-exchange issue. The more recent 2013 version of PSID intergenerational transfers data includes measures of support over much longer time periods, suggesting that a more powerful test of the altruism-versus-exchange issue could be undertaken.

In addition to the newer 2013 transfer data, the PSID presents even greater opportunities for research on lifetime patterns of intrafamily support. Researchers in this area often refer to the three “currencies” of familial support—shared housing space (i.e., coresidence), money, and time devoted to personal-care service (Soldo and Hill 1993). Regarding the time horizon for these various forms of support, a recent paper showed that among a surprisingly large percentage of mother-child pairs, and during the part of the life cycle beginning when the mother is 58 years old, the child has never left the parental home (Wiemers et al. 2017). If this shared living arrangement were interpreted as one governed by the terms of an implicit exchange-motivated “contract” then the time period over which the exchange is conducted is evidently quite long. And, as noted earlier, in addition to the 50-year coresidence histories produced in the PSID, there are equally long sequences of annual information on income received from, and financial help provided to, family members outside the household. Despite the lack of specificity about kin relationships in those latter measures, they clearly represent a form of intrafamily resource flows, and would seem to provide a resource with which to investigate possible exchange-motivated behaviors.

Migration and family geography

Proximity to one’s grown children during late life has been characterized as a resource for family support and long-term care service use (Wolf 1994). Proximity to family members is, of course, a consequence of (and is altered by) any move one makes, as well as any moves made by other family members, leading to a complex set of interacting move- and proximity-variables sequences.

The PSID has throughout its history included self-reported indicators of whether a move has happened (since the last interview) as well as whether the respondent has nearby relatives. Much more detailed information can, however, be developed through the linkage of survey data to geographic locations (i.e., named spatial entities such as tracts, zip codes, or counties or polar coordinates). Sequences of individual locational data can be used to compute the (approximate)^{viii} distance traversed in a residential relocation. Richer still is the more detailed kin-proximity data that can be derived by calculating the (approximate) distances between two different PSID respondents who are linked by known familial relationships. Migration and location choice are heavily influenced by contextual factors (besides the

^{viii}The approximation results from the fact that we do not know the polar coordinates of individual respondents’ addresses, but rather those of the population centroid of their spatial unit. Accordingly, distance moved or distance from a particular relative will entail greater amounts of measurement error the larger the spatial units being analyzed. Some counties are extremely large; however, we can take some comfort in the fact that the errors in (say) parent-child proximity measures are likely to have a mean close to zero.

presence of kin in a given location, itself a contextual factor), so the sorts of linkages to external data discussed earlier are again relevant in migration studies.

One example of a PSID-based study that entails most of these features is a new paper by Spring et al. (2017). Their central interest is on the effect of proximity to nonresident kin on whether and to where people move, although they limit their sample to individuals who remain within the same metropolitan area throughout the panel period studied (1980–2013). Using linked records for individuals that trace back to the same 1968 households, they are able to compute the distance from each “ego’s” location to the locations of other PSID individuals to whom they are related. The analysis is conducted at the census tract level, which allows for a very detailed depiction of geography. Moreover, using the tract identifiers they are able to incorporate area characteristics—the number and average value of housing units, and racial composition—into their model of destination choice. Among their findings are that movers tend to move toward aging parents, among whites and higher-income individuals there is a tendency to move closer to parents and children, and among lower-income individuals a tendency to move toward extended-family kin. This is a very richly formulated analysis, and one that makes use of many of the distinctive features of the PSID. Yet (as is often the case) despite this richness it leaves open many questions, in particular the long-distance moves, and the role of proximity to kin in such moves, that were removed, by design, from the Spring et al. analysis.

SOME METHODOLOGICAL CHALLENGES

The PSID presents researchers with unparalleled potential to address questions associated with longitudinal patterns of behavior and of exposure to contextual factors, and to investigate intergenerational aspects of these longitudinal patterns within families. However, the inherent complexity of the data needed for such research, and the research-design issues implied by that complexity, can seem daunting. Many of the modeling and estimation issues that arise when analyzing PSID data are entirely standard for anyone using longitudinal data—dealing with right-censored event histories, selective attrition, sample mortality that is correlated with the outcome of interest, and so on. Others reflect more specialized concerns associated with the types of data elements and topical interests reviewed above. In this section I address a few such issues.

Tradeoffs in breadth and temporal scope of analysis

Carriers of the PSID “gene” include members of the original 1968 sample families and their descendants. A sample member’s age in 1968 is, as well, an indicator of how many years of presample experience is unrecorded in the data.^{ix} And, even though 50 years is a very long time, it is less than the entire lifetime of most people, placing bounds on the portion of sample members’ lifetimes that can be studied. These design features create various

^{ix}There are a few exceptions to this statement about “unrecorded experience”: questions about marital and fertility history, or retrospectively-reported childhood experiences, provide a limited view of pre-sample information. The majority of routinely-reported current/recent experience—income, employment, health, living conditions, and so on—are, however, available only for the survey period.

tradeoffs regarding study design that all users of the PSID must confront. These issues are quite self-evident; nevertheless, it is useful to review them.

Figure 1 shows “lifelines” for individuals from a handful of birth cohorts whose lives overlap with the PSID study period through 2018; the horizontal axis represents years and the vertical axis represents age (0 through 90, although many PSID sample members are observed to even older ages). Each cohort is characterized by its year of birth and by the ages covered by the PSID study period. For example, the earliest cohort shown consists of people born in 1923, and observed at ages 45 (in 1968) to 95 (in 2018). Successively later cohorts—e.g., those born in 1938 or 1953—are observed throughout most of their adulthood—ages 30–80 and 15–65, respectively. As fewer of the years lived after turning 65 are observed, more of those lived prior to adulthood are observed. There are, however, limits on the extent to which childhood, the work life, and postretirement experience can all be directly observed. People born in 1968, the first year of PSID data collection, can be traced throughout their childhood, but their adult years are censored at age 50 (for now; as the PSID continues to age, this cohort’s lifetimes will become increasingly visible).

Coverage by Age for Selected Birth Cohorts in PSID Sample

The tradeoffs confronting researchers are evident: (1) people observed late in life (after, say, age 65) have adult lives that are incompletely observed, and childhoods that are completely unobserved; (2) for a few cohorts (e.g., those born in 1953), their entire adult lives are observed, but little or none of either childhood or old age is observed; and (3) those whose childhoods are fully observed (e.g., born in or after 1968) have a range (from very little to the majority) of their adult lives directly recorded. Assuming that the PSID continues to operate and follow these more recent cohorts’ lives, or course, these limitations will gradually be overcome. The limitations associated with the earlier cohorts cannot, however, be eliminated.

As a consequence of these unavoidable design features, comparisons over time in the circumstances of those over age 65, or alternatively those under 18, are limited to those in a well-defined and somewhat narrow set of birth cohorts. And studies that rely on joint observation of parents and children face even more stringent limits. It is important to recall, of course, that however restrictive any such limitations might be, they are much less restrictive for PSID users than for users of any other longitudinal data source (for the United States). Moreover, with each additional round of data collection these constraints are incrementally relaxed.

Completeness of family-network representation

The PSID “gene pool” is limited to members of households included in the original 1968 sample and their descendants. For the many study designs that rely on linking the records of parents and their children, this limitation has consequences for the completeness of family networks studied: any children of original PSID respondents who had already “split off” from their parent’s household before 1968 are excluded. Given the strong role of age in the split (or “emancipation”) process, there are additional study-design limitations that must be superimposed on the cohort-representativeness diagram shown in Figure 1.

We can get a sense of the family-representation problem using public-use microdata from the 1970 U.S. Census (Ruggles et al. 2017), which is reasonably close to 1968 for purposes of illustrating the overlap of families and households.^x I use the census data to address two questions. First, what is the age profile of coresidence with one's parents? Second, what is the age profile of not living with all of one's children (or siblings)? These two questions are just two different ways of addressing the problem of family members who are excluded from the PSID gene pool by design.

An indicator of coresidence with parents is directly provided in the census records. For the analysis of "missing" children (or siblings), I use just two data elements, the number of children ever born (to women), and the number of own children in one's household, both of which are available in the 1970 Census. I classify a woman's household as one with "missing children" if the number of children ever born exceeds the number present in the household.^{xi}

Figure 2 shows the proportions of individuals (not distinguished by gender) living with and without parents present. Having neither parent present is rare up to about age 16, and rises sharply at age 18 and thereafter. This suggests, in turn, that among the original 1968 PSID households nearly all children younger than 18 are present. However, Figure 3 shows that the proportion of young children whose household fails to contain one or more of their (presumably older) siblings rises steadily, from about 0.1 even among the very youngest children to over 50 percent by age 16. Figure 3 also shows that for very young mothers (e.g., 18 or younger), a substantial proportion have had children who are not current coresidents; these missing children may have died (we cannot tell, using these data), or might be living with a grandparent or adoptive parent. Even among mothers in their twenties, 10 percent or more are in households for which at least one child is "missing" (as defined here).

Together, these graphs illustrate a problem facing users of linked parent-child PSID records: the earlier the range of birth cohorts included among the children's generation, the more likely it is that some family members will be missed. Even if a study limits its inclusion of the offspring generation to those born after 1968, there is at least a small chance that some older siblings are irrevocably lost from the sample. Fortunately, the PSID contains many data elements with which the existence, if not the detailed circumstances, of others in the sample members' nuclear-family kin networks can be determined. Using these items, researchers could assess potential selection biases or external-validity limitations of their analytic samples.

The issues reviewed above should be viewed as factors that limit, but do not preclude, broad-coverage life-cycle studies of individual trajectories and of their association with the lifecycle circumstances of family members to which they can be linked. Care must be taken not only to select the ages and years for which both background factors and outcomes are to

^xWe could use the PSID itself to address these questions; however, the PSID did not obtain the crucial information on children ever born until 1985, by which time there had been a good deal of attrition from the original sample.

^{xi}The indicator of "missing children" is erroneous for at least two reasons: first, one or more of the children classified as "missing" may be dead, and therefore are no longer part of the family network; second, "own children" in a household includes step-as well as adoptive children, whereas children "ever born" are limited to biological children. These errors, however, are likely concentrated in a small percentage of the population.

be measured, but to recognize, and convey to readers, the built-in limits on the generalizability of findings from these designs.

In addition to the issue of completeness of family-network coverage just described, it bears mentioning that step-relationships are not fully captured over time: given that one can acquire the PSID “gene” only through birth or adoption, children born to the spouse of a PSID gene carrier prior to that marriage, or after the dissolution of the marriage (and consequent split-off of the ex-spouse’s family), are not considered to be PSID sample members. Therefore, some step- and half-siblings of PSID sample members are excluded from the sample by design.

Survival bias in retrospective reports of early circumstances

The 2014 module of questions on childhood circumstances, the planned linkage to 1940 Census records, and other retrospectively reported characteristics (such as military service) are all examples of data whose distributions can be affected by a type of survival bias: if an earlier-life circumstance (for example, living in poverty during childhood) is correlated with survival to a later age, then those who remain alive to tell us about that earlier-life circumstance are not representative of the population that experienced that earlier-life circumstance. This problem is not always fully recognized, and even when it is recognized it presents challenges.

To illustrate just one of the relevant variants of this problem in the PSID context, we can use a 1940 birth cohort life table.^{xii} This birth cohort will be the youngest, among all PSID individuals, for whom a 1940 Census data linkage is possible. They became 72 (and therefore met the condition of the “age 72 rule”) in 2012. They will be 77 in 2017 and 79 in 2019, which are likely times at which they will be asked to provide any data items that would be used in the linkage effort.

The cohort life table (for all races and both sexes) tells us that already by 1968, the first year of PSID data collection, modest selectivity was present: about 8 percent of the birth cohort had died by 1968. Among survivors to 1968, only about 61 percent will survive to age 77, while 56 percent will survive to age 79. This degree of sample loss through mortality could pose problems for the users of the retrospectively reported information.

There are a number of potential solutions to this selective-retrospective reporting problem, such as the use of external data sources to model the selection process, joint modeling of the reporting and survivorship processes, and possibly others; in general, these approaches are technically complicated, narrow in their applicability, and not widely used. The availability of rich data sources such as the PSID, coupled with a growing recognition of the methodological issues suggested by these data structures, can be expected to help spur further development and applications of appropriate models and methods.

^{xii}https://www.ssa.gov/oact/NOTES/as120/LifeTables_Tbl_7_1940.html.

CONCLUSION

The PSID was evidently not initially designed to support research on aging, but as it has aged it has become a major resource for such research. It has been used in a steadily growing body of scholarly research for more than 40 years. And, notwithstanding the several inherent limitations and analytic challenges posed by the data, the data offer what is essentially limitless potential to support more such research in the future.

The ability of PSID data to address important public policy concerns, and to play a role in the design, implementation, and evaluation of programs, seems similarly vast. One especially salient policy area is the use of long-term care resources. In general, elderly individuals unable to address their personal care needs must rely on some mix of “informal care”—mainly, care provided by family members—and “formal care” provided in institutional settings or through what are mainly publicly funded community-based programs. The supply of informal care services is heavily influenced by the size, composition, and spatial distribution of family members, phenomena for which the PSID is a particularly strong data resource. Indeed, several recent publications use the PSID to address these issues of family composition as they relate to living and care arrangements (Wiemers and Bianchi 2015; Friedman, Park, and Wiemers 2017). However, access to formal care services, and the generosity of public support for such services, is known to vary widely both between and within states (Gornick, Howes, and Braslow 2012). The PSID, with its ability to address individual- and family-level variation at the same time as geographic variation in the service-delivery and programmatic environments, is well-positioned to support research on the consequences of long-term care policy choices.

A second, and closely related, area of policy in which the PSID can play a major role is that of place-based health and mortality disparities. Some of the papers reviewed above relate to the role of contextual, including environmental, factors as they contribute to health and well-being outcomes. Among the “topics and objectives” in the U.S. Office of Disease Prevention and Health Promotion’s *Healthy People 2020*^{xiii} effort is environmental health, components of which include air and water quality, exposure to toxic substances and hazardous waste, and the home and community environment. PSID data on health outcomes and their individual-level precursors, linked to place-specific measures of these components of environmental health, could contribute materially to state- and local-level policies in such areas as zoning, planning, tax policy, remediation, the location of service providers, and other investments in infrastructure.

The PSID can also contribute to the literature evaluating the impacts of policy interventions on public health. The several papers mentioned earlier that address later-life consequences of early-life circumstances—both individual and contextual—suggest an intriguing follow-up question: To what extent can interventions timed during the adult years overcome the harmful consequences of early-life adversity? And what are the relative costs and benefits of such interventions at different points in the life course? Another issue of growing concern relates to the health impacts of preemptive legislation, which can be enacted at the federal

^{xiii}<https://www.healthypeople.gov>.

level but is most often enacted at the state level (Montez 2017; Pomeranz and Pertschuk 2017). The general form of such laws is to restrict or prohibit lower-level governments' (e.g., counties and cities) ability to impose requirements in areas such as minimum wage levels, nutritional labeling, paid sick leave, smoking in various venues, and so on. Although these efforts began in the 1980s and 1990s they have more recently grown in scope and coverage. The PSID, with its national population coverage, small-area geographic detail, and long temporal scope, seems well poised to support estimates of the health impacts of such constraints on public action, while controlling for pre-policy change conditions at both the individual and local level. All the research topics mentioned here are well-suited to the adoption of a life course perspective, and therefore have the potential to make important contributions in social gerontology, in addition to program evaluation and epidemiology.

It seems clear that the PSID has demonstrated a high level of value—of return on investment—even if we consider only one of its many domains of scientific contribution: aging and the life course. It is also evident that it can serve as a resource for a great deal of additional research in this area. Whether one participates as a producer, or as a consumer, or merely as an observer of this future, it promises to be an interesting one.

Bio:

Douglas Wolf is a demographer and gerontologist who conducts research on patterns of late-life disability and on the provision of, and the consequences of providing, informal care to older adults.

References

- Alley DE, Metter EJ, Griswold ME, Harris TB, Simonsick EM, Longo DL, and Ferrucci L. 2010 Changes in weight at the end of life: Characterizing weight loss by time to death in a cohort study of older men. *American Journal of Epidemiology* 172:558–65. [PubMed: 20682520]
- Altonji Joseph G., Kayashi Fumio, and Kotlikoff Laurence. 2000 The effects of income and wealth on time and money transfers between parents and children In *Sharing the wealth: Demographic change and economic transfers between generations*, eds. Mason A and Tapinos G, 306–57. New York, NY: Oxford University Press.
- Arrondel Luc, and Masson André. 2006 Altruism, exchange or indirect reciprocity: What do the data on family transfers show? In *Handbook of the economics of giving, altruism and reciprocity, volume 2*, eds. Kolm Serge-Christophe and Mercier Jean, 971–1053. Amsterdam: Elsevier B.V.
- Bengston Vern, Giarusso Roseann, Mabry J. Beth, and Merrill Silverstein. 2002 Solidarity, conflict, and ambivalence: Complementary or competing perspectives on intergenerational relationships? *Journal of Marriage and Family* 64:568–76.
- Clarke Phillipa, and Latham Kenzie. 2014 Life course health and socioeconomic profiles of Americans aging with disability. *Disability and Health Journal* 7:S15–S23. [PubMed: 24456680]
- Couch Kenneth A., Daly Mary C., and Wolf Douglas A.. 1999 Time? Money? Both? The allocation of resources to older parents. *Demography* 36:219–32. [PubMed: 10332613]
- Freedman Vicki A. 2017 The Panel Study of Income Dynamics' Well Being and Daily Life Supplement (PSID-WB); User Guide, Final Release 1. University of Michigan, Institute for Social Research Available from <https://psidonline.isr.umich.edu/WB/WBUserGuide.pdf>.
- Freedman Vicki A., and Cornman Jennifer C.. 2012 The Panel Study of Income Dynamics Supplement on Disability and Use of Time (DUST) User Guide [Release 2009.1]. Available from <https://psidonline.isr.umich.edu/Guide/documents.aspx>.

- Freedman Vicki A., and Cornman Jennifer C.. 2015 The Panel Study of Income Dynamics Supplement on Disability and Use of Time (DUST 2013 Files) User Guide [Release 2015.3]. Available from <https://psidonline.isr.umich.edu/Guide/documents.aspx>.
- Freedman Vicki, and Martin Linda G.. 2000 Contribution of chronic conditions to aggregate changes in old-age functioning. *American Journal of Public Health* 90:1755–1760. [PubMed: 11076245]
- Friedman Esther M, Sung S. Park, and Weimers Emily E.. 2017 New estimates of the sandwich generation in the 2013 Panel Study of Income Dynamics. *The Gerontologist* 57:191–96. [PubMed: 26672020]
- Furstenburg Frank, Hoffman Saul, and Shrestha Laura. 1995 The effect of divorce on intergenerational transfers: New evidence. *Demography* 32:319–33. [PubMed: 8829969]
- Gornick Janet, Howes Candace, and Braslow Laura. 2012 The disparate impacts of care policy In *For love and money: Care provision in the United States*, ed. Folbre Nancy, 140–82. New York, NY: Russell Sage Foundation.
- Hayward Mark D., and Gorman Bridget K.. 2004 The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography* 41:87–107. [PubMed: 15074126]
- Insolera Noura E., and Freedman Vicki A.. 2017 Comparing health estimates in the PSID and NHIS, 2001–2015. PSID Technical Series Paper #17-01. Ann Arbor, MI: University of Michigan, Institute of Survey Research.
- Johnson Rucker C., and Schoeni Robert F.. 2011 Early-life origins of adult disease: National longitudinal population-based study of the United States. *American Journal of Public Health* 101:2317–2324. [PubMed: 22021306]
- Karraker Amelia, Schoeni Robert F., and Cornman Jennifer C.. 2015 Psychological and cognitive determinants of mortality: Evidence from a nationally representative sample followed over thirty-five years. *Social Science & Medicine* 144:69–78. [PubMed: 26397865]
- Laditka James N., and Laditka Sarah B.. 2016 Associations of educational attainment with disability and life expectancy by race and gender in the United States: A longitudinal analysis of the Panel Study of Income Dynamics. *Journal of Aging and Health* 28:1403–1425. [PubMed: 26690254]
- Latham Kenzie. 2015 The “long arm” of childhood health: Linking childhood disability to late midlife mental health. *Research on Aging* 37:82–102. [PubMed: 25651552]
- Lillard DR, Burkhauser RV, Hahn MH, and Wilkins R. 2015 Does early-life income inequality predict self-reported health in later life? Evidence from the United States. *Social Science and Medicine* 128:347–55. [PubMed: 25577308]
- McGonagle Katherine, and Freedman Vicki A.. 2015 The Panel Study of Income Dynamics' Childhood Retrospective Circumstances Study (PSID-CRCS) User Guide: Final Release 1. Available from <https://psidonline.isr.umich.edu/Guide/documents.aspx>.
- McGonagle Katherine, and Sastry Narayan. 2016 Using the Panel Study of Income Dynamics to analyze housing decisions, dynamics, and effects. *Cityscape* 18:185–99. [PubMed: 27110321]
- McGonagle K, Schoeni R, Sastry N, and Freedman V. 2012 The Panel Study of Income Dynamics: Overview, recent innovations, and potential for life course research. *Longitudinal and Life Course Studies* 3:268–84.
- Montez Jennifer Karas. 2017 Deregulation, devolution, and state preemption laws' impact on U.S. mortality trends. *American Journal of Public Health* 107:1749–1750. [PubMed: 29019768]
- Montez Jennifer Karas, Hayward Mark D., and Wolf Douglas A.. 2017 Do U.S. states' socioeconomic and policy contexts shape adult disability? *Social Science & Medicine* 178:115–26. [PubMed: 28219027]
- Newhouse Joseph P., and Mark McClellan. 1998 Econometrics in outcomes research: The use of instrumental variables. *Annual Review of Public Health* 19:17–34.
- Pomeranz Jennifer L, and Mark Pertschuk. 2017 State preemption: A significant and quiet threat to public health in the United States. *American Journal of Public Health* 107:900–02. [PubMed: 28426304]
- Quadagno Jill. 2011 *Aging and the life course: An introduction to social gerontology*. 5th ed New York, NY: McGraw Hill.

- Ruggles Steven, Genadek Katie, Goeken Ronald, Grover Josiah, and Sobek Matthew. 2017 Integrated Public Use Microdata Series: Version 7.0 [dataset]. Minneapolis, MN: University of Minnesota Available from 10.18128/D010.V7.0.
- Schoeni Robert F. 1996 Does Aid to Families with Dependent Children displace familial assistance? RAND Working Paper Series 96-12, Santa Monica, CA.
- Schoeni Robert F., Bianchi Suzanne M., Hotz V. Joseph, Seltzer Judith A., and Wiemers Emily E.. 2015 Intergenerational transfers and rosters of the extended family: A new substudy of the Panel Study of Income Dynamics. *Longitudinal and Life Course Studies* 5:319–30.
- Settersten Richard A. Jr. 1999 *Lives in time and place: The problems and promises of developmental science*. Amityville, NY: Baywood.
- Shuey Kim M., and Willson Andrea E.. 2017 Trajectories of work disability and economic insecurity approaching retirement. *Journal of Gerontology B: Psychological Sciences and Social Sciences*. DOI: 10.1093/geronb/gpx096.
- Sliwinski MJ, Stawski RS, Hall CB, Katz M, Verghese J, and Lipton R. 2006 Distinguishing preterminal and terminal cognitive decline. *European Psychologist* 11:172–81.
- Soldo Beth J., and Hill Martha S.. 1993 Intergenerational transfers: Economic, demographic, and social perspectives In *Annual review of gerontology and geriatrics, volume 13*, eds. Maddox George L. and Lawton M. Powell, 187–216. New York, NY: Springer.
- Spring Amy. 2017 Short- and long-term impacts of neighborhood built environments on self-rated health of older adults. *The Gerontologist*. DOI:10.1093/geront/gnx119.
- Spring Amy, Ackert Elizabeth, Crowder Kyle, and South Scott J.. 2017 Influence of proximity of kin on residential mobility and destination choice: Examining local movers in metropolitan areas. *Demography*. DOI: 10.1007/s13524-027-0587-x.
- Wiemers Emily E., and Bianchi Suzanne M.. 2015 Competing demands from aging parents and adult children in two cohorts of American women. *Population and Development Review* 41:127–46. [PubMed: 26594071]
- Wiemers Emily E., Slanchev Vladislav, Kathleen McGarry, and Joseph Hotz V. 2017 Living arrangements of mothers and their adult children over the life course. *Research on Aging* 39:111–24. [PubMed: 28181865]

Figure 1:
Coverage by age for selected birth cohorts in PSID sample

Figure 2:
Proportion of children coresiding with parent[s], by age, 1970 Census

Figure 3:
Proportion of mothers (children) with 1 or more children (siblings) not in household, 1970 Census