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

Draft Chapters from the 1990 Revision
of the PSID User Guide

Greg Duncan

Survey Research Center - Institute for Social Research
University of Michigan

1990

THE PANEL STUDY OF INCOME DYNAMICS

Draft Chapters from the

1990 Revision of the User Guide

Survey Research Center
University of Michigan
P.O. Box 1248
Ann Arbor, Michigan 48106

Greg Duncan, Program Director

(313) 763-5186

CONTENTS FOR
CHAPTER A: GENERAL DESCRIPTION OF THE PSID

Chapter A

General Description of the PSID

1. Brief Overview of the PSID

What Is the PSID?

In brief, the Panel Study of Income Dynamics (PSID):

● is a longitudinal study with a national sample of the U.S. population

● began in 1968 and is funded through 1991

● follows individuals (men, women, and children) and their family units

● emphasizes dynamic aspects of economic and demographic behavior, but covers a wide variety of other topics also

● is conducted by the Survey Research Center, University of Michigan.

Who Is In the Study?

The PSID:

● has some information about over 36,000 individuals in its data files, with information collected about 20,487 individuals in 1987

● interviewed 4802 family units in 1968, 7061 family units in 1987

● is representative of the U.S. population, except for in-migration since it began

● has a sizable sub-sample of blacks and of individuals in low-income families in 1968

● has a sizable sub-sample of children followed throughout a large portion of their childhood and on into early adulthood.

What Information Is Collected?

The PSID:

● has adhered to a general design and core content that have remained largely unchanged since its beginning, features which enhance its potential for longitudinal analysis

● includes as core content measures of income sources and amounts, employment, family composition, housing, and residental location

● collects additional supplements in most years, with the topics varying from year to year and including health, wealth, flows of time and money resources between relatives and friends, resources of parents, demographic histories, education histories, and motivation.

How Are the Data Collected?

The PSID:

● conducts annual interviews, which were face-to-face during the first five years of the study and by telephone since then

GENERAL DESCRIPTION

- generally interviews one person per family unit, but collects information about the family unit and all of its members.

What Are the Response Rates and the Quality-Control Measures?

The PSID:

- has maintained high response rates over the years

- hand-edits major variables and does numerous within-wave and across-wave consistency checks.

What Data Files Are Available?

The PSID:

- makes a variety of data files, accompanied by comprehensive documentation, available to the research community, mostly through the data archive known as the Interuniversity Consortium for Political and Social Research (ICPSR)

- currently has disseminated main data files for 20 waves, 1968–1987

- has over 15,000 variables on its main files spanning the study from its beginning up to 1987

- prepares several special files in addition to its main files.

What Have the Data Been Used For?

The PSID:

- has been used for numerous analyses (both cross-sectional and longitudinal) covering a wide range of topics in economics, demography, sociology, psychology, and health

- is a study whose data files have been disseminated widely throughout the United States and to numerous foreign countries, the result being that the PSID is one of the most widely used social science data sets in the world.

2. What Is the PSID?

The Panel Study of Income Dynamics (PSID) is a longitudinal study of a representative sample of U.S. Individuals (men, women, and children) and the family units in which they reside. Its emphasis is on dynamic aspects of economic and demographic behavior, but its content includes a wide range of measures, including sociological and psychological ones. As of 1989, the PSID had collected information about over 36,000 individuals spanning as much as 22 years of their lives. The study has been conducted at the Survey Research Center, University of Michigan since its beginning in 1968. Data collection is on an annual basis, and the data files contain the full span of information collected over the course of the study.

The general design and core content of the study have remained largely unchanged since its beginning, a feature which enhances the PSID's potential for longitudinal analysis. Comprehensive documentation is prepared and made available to

the public along with the data files. The Inter-university Consortium for Political and Social Research (ICPSR), a data archive, handles the public distribution of the data files and documentation. Currently, merged data is available for 20 waves of the study, covering the years 1968–1987. Several special files are available as well. Interviewing on the study continues, with additional waves in process and funding secured through interviewing year 1991. The PSID data files have been disseminated widely throughout the United States and to numerous foreign countries. As a result, the PSID is one of the most widely used social science data sets in the world.

3. Who Is in the Study?

Starting with a representative national sample of U.S. households in 1968, the PSID has traced individuals from those households since that time, whether or not they are living in the same dwelling or with the same people. Adults have been followed as they have grown older, and children have been observed as they advance through childhood and into adulthood, forming family units[1] of their own. Information about the original 1968 sample people and their current co-residents (spouses, cohabitors, children, and anyone else living with them) is collected each year.

Because the original focus of the study was on the dynamics of poverty, the original sample was comprised of a disproportionately large number of households in poverty during the late 1960s.[2] This means, also, that the sample consists of a sizable sub-sample of blacks. Probability-of-selection weights are available to enable analysts to make estimates from the sample that are representative of the U.S. population. In the absence of nonresponse bias, the PSID's rules for tracking individuals and family units over time lead to accurate representation of the U.S. population both cross-sectionally each year from 1968 forward and in terms of demographic change since 1968, with the exception of post-1968 in-migration by family units comprised solely of immigrants. The study's tracking rules also lead to a steady increase in the number of persons and family

---

[1] A PSID "family unit" includes what the U.S. Census Bureau terms "one-person households" as well as what it terms a "family."

[2] The first PSID wave (in 1968) included interviews with 1,872 households with low income in 1966 and 1967 (a sample taken from the U.S. Census Bureau's Survey of Economic Opportunity), plus 2,930 households drawn from a cross-section of dwellings in the coterminous U.S. selected from the Survey Research Center's multi-stage, national master sample at a constant overall sampling rate.

units about whom information is gathered.[3] The number of individuals for whom the PSID has gathered some information has increased from about 18,000 in 1968 to a cumulative total of over 36,000. The number of family units about whom information is gathered in a given year has increased from just under 5,000 at the beginning of the study to about 7,000 currently.

4. What Information Is Collected?

The PSID provides a wide variety of information about both families and their individual members, plus some information about the areas where they live. The central focus of the data is economic and demographic, with substantial detail on income sources and amounts, employment, family composition changes, and residential location. Content of a more sociological or psychological nature is also included in some waves of the study. Information gathered in the survey applies to the circumstances of the family unit as a whole (e.g., type of housing) or to particular persons in the family unit (e.g., age, earnings). While some information is collected about all individuals in the family unit, the greatest level of detail is ascertained for the primary adults heading the family unit. A large core of topics (e.g., income, employment, family composition) has been addressed consistently throughout the study, but many additional topics (e.g., health, wealth, retirement plans, flows of time and money help between relatives and friends, and motivation and efficacy) have been covered from time to time. The amount and variety of data are enormous; over 300 pages are required to list, by topic and year, the variables on the study's 1968–1987 main file.

5. How Are the Data Collected?

Each year the study interviews one primary adult per family unit containing someone from an original 1968 household. Generally this is the "head" of the family

---

[3]The PSID's tracking rules call for following members of the original family units plus their offspring to whatever living arrangements they experience, gathering information about them and their co-residents if they are living in a household (i.e., non-institutional) situation. Attrition in the PSID has been modest, and new families have formed when children have grown up and established separate households or when marriage partners have gone separate ways. This has resulted in growth over time in both the number of family units and the number of people residing with an original sample member at some time during the study.

unit, who in married-couple units is defined as the husband.[4] The greatest detail has been collected each year about the heads of family units. Since 1976, however, the study has sought to collect the same detail for wives as for Heads. In 1976 and 1985, the study conducted separate interviews with all wives of Heads as well as their husbands.

The PSID has also varied its interviewing mode. Each year 1968–1972, the PSID interviews were *face-to-face*, but in 1973, to reduce costs, the study began taking the majority of interviews by *telephone* rather than in person. Face-to-face interviews are now performed only for respondents with no telephone or with special circumstances which make a telephone interview unfeasible. As a further cost-saving measure and because long interviews are especially difficult over the telephone, the length of the interview was reduced with the change in mode. During interviewing years 1968–1972, the annual interview averaged about one hour in length. With the switch to telephone interviewing, the questionnaire was reduced in size to require no more than 20 to 30 minutes of interviewing time, on average.

6. What Are the Response Rates and Quality-Control Measures?

Response rates have remained high since the initial wave of the study, with about *97 percent* of the prior-year sample continuing to participate from one year to the next. Even small nonresponse from wave to wave cumulates over time, however, and the response rate in the first wave of the study was more modest— 76 percent of the households from the selected sample of households was successfully interviewed that year. As a result of all of these factors, approximately *66 percent* of the still-living original sample of individuals was participating in the study in interviewing year 1985. A number of analyses provide reasonably reassuring evidence of the *absence of substantial nonresponse bias.*[5]

The study devotes extensive effort to assuring data quality, including hand-editing of variables of major importance to the study's overall purpose, with assignment of values based on past as well as current information. Numerous within-wave consistency checks are also made in the course of data preparation, and extensive between-wave consistency checks take place each year.

---

[4]The PSID has adhered to the old Census Bureau definition of "head" of household. This means that the husband in a married-couple family unit is automatically designated as the "head" of that unit unless he is severely disabled.

[5]See, for example, Becketti, et al. (1983) and Duncan and Hill (1989).

GENERAL DESCRIPTION

7. What Data Files Are Available?

The PSID staff merges each new wave of data with prior waves to provide comprehensive coverage of information collected for an individual or family unit over the entire course of the study. These multi-wave data files become publicly available as soon as the merging and final checks for data quality are completed. The most recent file currently available covers the years 1968 through 1987. This file is one of the PSID's main files. There are other types of files that the PSID produces in addition to this.

The *main PSID files* have been assembled and publicly disseminated each time a new wave of data has become available. There are now three distinct main PSID files. Two of the main PSID files are *cross-year family-individual files*. Each record on one of these files represents an individual in a PSID family unit, and the record contains all years' observations about the individual and the family units in which the individual has resided while participating in the study. One of the cross-year family-individual files is termed *"response"*, the other is termed *"nonresponse"*: the distinction reflects the type of individuals represented by the file. The *cross-year family-individual response file* contains records for individuals interviewed in the most recent wave of the study. The *cross-year family-individual nonresponse file* contains records for individuals who participated in the study at some time in the past but who are not still participating due to death, refusal, etc.. The nonresponse file was first prepared for the 1968–1984 cross-year family-individual file, and has been updated each wave since 1984. The response and nonresponse files are stored on separate tapes that can be concatenated to form a *merged file of all individuals* for whom any information has ever been collected. Currently, the main files for the PSID data contain as much as two decades of information about the approximately 18,000 people in the original 5,000 family units plus some information about their over 18,000 subsequent co-residents.

A third main PSID file is the *cross-year family file*; each record on this file represents a separate family unit participating in the study in the most recent interviewing year. Current-year information about a family unit is provided along with past years' information about the family units to which the Head of the current-year family unit has belonged.

Special files containing very detailed information about particular topics have also been added in recent years. The fine detail is relegated to separate files because it is too large to efficiently store on the main files. Two public-release files of this type are

updated with each additional wave of data: a file containing detail about *demographic events* such as marriage, childbirth, adoption, and substitute parenting; and a file containing very detailed *work history* information. A few other special files are also available, some requiring special contractual arrangements.

PSID data files, with a few exceptions, are *available through the Inter-University Consortium of Political and Social Research (ICPSR).*[6] Comprehensive documentation is provided with the data files the ICPSR distributes. (See Chapter ?? and Appendix ?? for details about ordering files.) The set of documentation for the PSID's main files, covering 20 waves (1968-1987), is contained in the 16 volumes of A Panel Study of Income Dynamics: Procedures and Tape Codes.

8. What Have the Data Been Used For?

The rich content of the PSID data, its longitudinal design, its span of two decades and multiple generations of families, its high level of data quality, and its comprehensive documentation make it well-suited to a wide range of analysis possibilities, especially longitudinal approaches. Areas of *basic economic research* that have been addressed by the data include: labor supply, consumption, life-cycle earnings, unions, compensating wage differentials, dynamic aspects of income distribution and various methodological studies. PSID *topics of interest to several disciplines*— demographers, sociologists, psychologists, and economists—include poverty and welfare experiences during adulthood or childhood; motivation and economic mobility; changes in family structure (e.g., births, divorce. remarriage); child support; out-of-wedlock births; teenage child-bearing; and the intergenerational transmission of poverty and welfare dependency. This diversity of topics reflects the philosophy of the PSID to ask modest sets of questions about a wide variety of topics rather than extensive questioning about only a few topics, couching the multi-faceted information in the context of substantial detail about income, employment, and family composition.

A *50-page bibliographical listing* of papers and books produced by users of the PSID data is available from the ICPSR or the PSID staff upon request. Table A.1 provides some illustrative references for work done in each of a number of substantive areas. Areas are organized according to the dependent variable.

---

[6]The ICPSR is an archive for a large variety of social science data sets. It is member-based, with a fixed fee for participation, but distributes machine-readable data files to non-member users on a per-item costing basis.

Table A.1

Some Examples of Behavioral Research Using Data from the Panel Study of Income Dynamics

General Area	Subarea	Illustrative Reference	Description
I. Labor Supply	A. Work Hours	Heckman and MaCurdy (1980)	Develops and estimates a dynamic model of female labor supply
		Hausman (1981)	Carefully models labor supply budget constraint
		Cramer (1980)	Investigates causality between fertility and labor supply of married women
		Shishko and Rostker (1976)	Models multiple job holding
	B. Retirement	Boskin (1977)	Estimates effects of Social Security on retirement
		Haveman (1982)	Estimates effects of disability on retirement
		Morgan (1980)	Estimates a general model of the retirement decision
	C. Unemployment	Yoon (1981)	Models unemployment duration
		Corcoran and Hill (1985)	Estimates state dependence influence on unemployment controlling for heterogeneity
	D. Job Change	Freeman (1980)	Estimates effects of unions on job quitting
		Datcher (1983)	Estimates effects of information networks on job quit behavior
		Mincer and Higuchi (1988)	Compares role of wage profiles in explaining differences in turnover rates between U.S. and Japan
	E. On-the-job Training	Rosen (1982)	Estimates a model of OJT with particular attention to tax effects
II. Earnings	A. General	Lillard and Willis (1978)	Estimates a wage model with a sophisticated dynamic error term
	B. Work Experience	Corcoran, Duncan and Ponza (1983)	Estimates effects of work experience on the wages of women
		Abraham and Farber (1987); Altonji and Shakotko (1987)	Does tenure with employer raise wages, or are long-tenure workers just those who found good matches?
	C. Unions	Freeman (1984)	Compares cross-sectional and longitudinal estimates of union effects in the PSID and other data sets
	D. Unemployment	Abowd and Ashenfelter (1981); Ashenfelter and Ham (1979)	Estimates effects on unemployment of wages

Table A.1 (continued)

General Area	Subarea	Illustrative Reference	Description
	E. Social Psychology	Duncan and Liker (1983)	Models reciprocal relationship between efficacy and earnings
	F. Earnings Differentials	Hoffman (1979)	Examines black-white earnings differences over the life cycle??
III. Economic Status	A. General	Duncan and Morgan (1981)	Analyzes effects of various factors on total family income and income/needs
		Duncan and Hoffman (1985)	Tracks economic circumstances of men, women, and children before and after divorce
		Hill, et al. (1985)	Estimates intra- and intergenerational models of motivation and economic status
		Ellwood (1988)	??
	B. Poverty	Bane and Ellwood (1986)	Analyzes length and determinants of poverty spells
		Hill (1981)	Estimates extent of heterogeneity and state dependence in patterns of poverty
	C. Intergenerational	Datcher-Loury (1988)	Estimates effects of mother's child care time on children's outcomes as adults
		McLanahan (1988)	Tests several hypotheses about effects of growing up in a single-parent home on daughters' family formation and economic behaviors in early adulthood
		Solon, et al. (1988)	Examines resemblance across siblings to test effects of family background on children's achievements as young adults
		Hill and Duncan (1987)	Tests several hypotheses about effects of parental income during childhood on children's socio-economic achievements as young adults
		Datcher (1982)	Estimates effects of family background and community on achievement of young adults
		McLanahan (1985)	Estimates effects of coming from a female-headed parental family on educational attainment
IV. Income Transfer Programs	A. AFDC	Bane and Ellwood (1983a)	Analyzes length and determinants of AFDC spells

Table A.1 (continued)

General Area	Subarea	Illustrative Reference	Description
	B. Food Stamps	Moffit (1983)	Estimates model of welfare stigma
		Coe (1979b)	Analyzes participation in the food stamp program
	C. SSI	Coe (1985)	Examines nonparticipation in the SSI Program by the Eligible Elderly
	D. SSDI	DeJong, Haveman and Wolfe (1988)	Effects of Disability Insurance on work by pre-retirement age women
	E. Unemployment Insurance	Katz and Meyer (1988)	Both new job finding and recalls jump when UI benefits expire
V. Living Arrangements	A. General	Richards, White, and Tsui (1987)	Estimates hazard functions for various living arrangements
	B. Single-Parent Family	McLanahan (1988)	??
	C. Departure of Children from the Parental Home	Hill (1977)	Develops and estimates a model of the decision of children to leave the parental home
VI. Fertility and Nuptiality	A. Fertility	Turchi (1975)	Estimates models of fertility??
		Hofferth, et al. (1981)	??
		Moore, et al. (1978)	??
	B. Marriage and Divorce	Ross and Sawhill (1975)	Models marital decisions
		Hutchens (1979)	Models the effects of AFDC program on marital changes
		Hofferth (1985)	Characterizes family structure during childhood
VII. Residential Mobility	A. Migration	DaVanzo and Morrison (1981, in press)	Models migration and return migration
		Gramlich and Laren (1984)	Estimates effect of AFDC benefit levels on migration
	B. Local	Goodman (1976)	Estimates models of local residential mobility
VIII. Consumption	A. Food Expenditures	Hymans and Shapiro (1974)	Estimates an equilibrium model of food expenditures by averaging variables over five years
		Benus, Kmenta and Shapiro (1976)	Dynamic demand analysis of food expenditures
		Hall and Mishkin (1982)	Estimates transitory and permanent income elasticities of food expenditures

Table A.1 (continued)

General Area	Subarea	Illustrative Reference	Description
	B. Housing	Altonji and Siou (1987)	Compares changes in consumption following a change in income to predictions of life cycle model with rational expectations
		Carliner (1976)	Estimates transitory and permanent income elasticities of housing
		Rosen (1979)	Estimates effects of income tax on housing decisions
		Margolis (1982)	Estimates depreciation rates for rental housing
IX. Other	A. Commuting	Cherlow and Morgan (1976)	Describes patterns of commuting time and speed
	B. Social Psychology	Cohn (1978)	Estimates effects of unemployment on self-esteem
		McLanahan (1983)	Relates family structure to stress
		McLanahan and Sorensen (1984)	Estimates effects of life events on psychological well-being
		Hill et al. (1985)	Examines intra- and intergenerational links between efficacy, motivation and economic well-being
	C. Kinship Networks	Hofferth (1984)	Compares kin assistance patterns by race
	D. Household Production	Leuthold (1981)	Estimates home production income and taxation implications
		Graham and Green (1984)	Estimates household production function with joint products of household goods and leisure
	E. Time Use	Wolfe and Haveman (1983)	Examines effects of time allocation on women's health status??
	F. Methodology	Becketti, et al. (1988)	Evaluates the nature of nonresponse in the PSID
		Duncan and Hill (1989)	Compares PSID sample to CPS and aggregate statistics

Bibliography

Abowd, John M.. and Ashenfelter, Orley, 1981. "Anticipated Unemployment. Temporary Layoffs and Compensating Wage Differentials." In Studies in Labor Markets. edited by Sherwin Rosen. Chicago: University of Chicago Press.

Abraham. Katherine, and Farber, Henry. 1987. "Job Duration, Seniority and Earnings." American Economic Review 77: 278–297.

Altonji, J.G.; Martins, Ana; and Siow, Aloysius. February, 1987. "Dynamic Factor Models of Consumption, Hours, and Income." Cambridge, MA: National Bureau of Economic Research.

Ashenfelter, Orley, and Ham, John. 1979. "Education, Unemployment, and Earnings." Journal of Political Economy 87: 99–116.

Bane. Mary Jo, and Ellwood, David. 1983. "The Dynamics of Dependence and the Routes to Self-Sufficiency." Final report to the U.S. Department of Health and Human Services. Cambridge, MA: Harvard University, Kennedy School of Government.

Bane, Mary Jo, and Ellwood, David. 1986. "Slipping Into and Out of Poverty: the Dynamics of Spells." The Journal of Human Resources 21: 1-23.

Becketti, Sean; Gould, William; Lillard, Lee; and Welch, Finis. 1988. "The PSID after Fourteen Years: an Evaluation." Journal of Labor Economics 6: 472–492.

Benus, J., Kmenta, J., and Shapiro, H. 1976. "The Dynamics of Household Budget Allocation to Food Expenditures." The Review of Economics and Statistics.

Boskin, Michael J. January, 1977. "Social Security and Retirement Decisions." Economic Inquiry.

Carliner, Geoffrey. 1976. "Income Elasticity of Housing Demand" (Discussion Paper No. 144–72). Madison: University of Wisconsin, Institute for Research on Poverty.

Cherlow, Jay, and Morgan, James N. 1976. "Commuting Time and Speed." In Five Thousand American Families: Patterns of Economic Progress, Vol. 4, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: University of Michigan, Institute for Social Research.

Coe, Richard D. 1979. "Participation in the Food Stamp Program." PhD dissertation. Ann Arbor: University of Michigan.

_____. April, 1985. "Nonparticipation in the SSI Program by the Eligible Elderly." Southern Economic Journal 51: 891–897.

Cohn, Richard. 1978. "The Effect of Employment Status Change on Self Attitudes." Social Psychology 41: 81–93.

Corcoran, Mary; Duncan, Greg; and Ponza, Michael. 1983. "A Longitudinal Analysis of White Women's Wages." Journal of Human Resources 18: 497–520. Michigan, Institute for Social Research.

Corcoran, Mary, and Hill, Martha S. 1985. "Reoccurrence of Unemployment Among Adult Men." Journal of Human Resources 20: 165–183.

Cramer, James. April, 1980. "Fertility and Female Employment: Problems of Causal Direction." American Sociological Review 45: 167–190.

Datcher, Linda. February, 1982. "Effects of Community and Family Background on Achievement." Review of Economics and Statistics 64: 32–41.

_____. August 1983. "The Impact of Informal Networks on Quit Behavior." Review of Economics and Statistics 65: 491–495.

Datcher-Loury, Linda. August 1988. "Effects of Mother's Home Time on Children's Schooling." Review of Economic Statistics, Tufts University, 70:3, 367–73.

DaVanzo, Julie, and Morrison, Peter A. 1981. "Return and Other Migration Sequences in the U.S." Demography 18: 85–101.

De Jong, Philip; Haveman, Robert; and Wolfe, Barbara. 1988. "Labor and Transfer Incomes of Older Women's Work: Estimates from the U.S." (Working Paper No. 2728). Cambridge, MA: National Bureau of Economic Research. Michigan, Institute for Social Research.

Duncan, Greg J., and Hill, Daniel (1989). "Assessing the Quality of Household Panel Survey Data: The Case of the Panel Study of Income Dynamics." Journal of Business and Economic Statistics 7 (4): 441–452.

Duncan, Greg J., and Hoffman, Saul D. 1985. "A Reconsideration of the Economic Consequences of Marital Dissolution." Demography 22: 485–497.

Duncan, Greg J., and Liker, Jeffrey. 1983. "Disentangling the Efficacy Earnings Relationship Among White Men." In Five Thousand American Families: Patterns of Economic Progress, Vol. 10, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: University of Michigan, Institute for Social Research.

Duncan, Greg J., and Morgan, James N. 1981. "Persistence and Change in Economic Status and the Role of Changing Family Composition." In Five Thousand American Families: Patterns of Economic Progress, Vol. 9, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: University of Michigan, Institute for Social Research.

Ellwood, David. 1988. Poor Support: Poverty and the American Family. New York: Basic Books.

Freeman, Richard B. 1978. June, 1980. "The Exit-Voice Tradeoff in the Labor Market: Unionism, Job Tenure, Quits, and Separations." The Quarterly Journal of Economics 94: 643–674.

GENERAL DESCRIPTION

_____. January, 1984. "Longitudinal Analysis of the Effects of Trade Unions." Journal of Labor Economics 2: 1–26.

Goodman, John L. 1976. "Housing Consumption Disequilibrium and Local Residential Mobility." Environment and Planning 8: 855–874.

Graham, J., and Green, C.A. May, 1984. "Estimating Parameters of a Household Production Function with Joint Production." Review of Economics and Statistics: 277–282.

Gramlich, Edward M., and Laren, Deborah S. Fall, 1984. "Migration and Income Redistribution Responsibilities." Journal of Human Resources 19: 489–511.

Hall, Robert E., and Mishkin, Frederic S. March, 1982. "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." Econometrica 50: 461–481.

Hausman, Jerry A. 1981. "The Effect of Taxes on Labor Supply." In How Taxes Affect Economic Behavior, edited by H. Aaron and J. Pechman. Washington, DC: Brookings Institute.

Haveman, Robert H. April, 1982. "Disability Transfers and Early Retirement: a Causal Relationship?" Journal of Public Economics.

Heckman, James, and MaCurdy, Thomas. 1980. "A Life Cycle Model of Female Labor Supply." Review of Economic Studies 47: 47–74.

Hill, Martha S. 1977. "The Decision by Young Adults to Split-Off from Their Parents' Households." PhD dissertation. Ann Arbor: University of Michigan.

Hill, Martha S. 1981. "Some Dynamic Aspects of Poverty." In Five Thousand American Families: Patterns of Economic Progress, Vol. 9, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: University of Michigan, Institute for Social Research.

Hill, Martha S.; Augustyniak, Sue; Duncan, Greg J.; Gurin, Gerald; Gurin, Patricia; Liker, Jeffrey K.; Morgan, James N.; and Ponza, Michael. 1985. "Motivation and Economic Mobility" (Research Report Series). Ann Arbor: University of Michigan, Institute for Social Research.

Hill, Martha S., and Duncan, Greg J. 1987. "Parental Family Income and the Socioeconomic Attainment of Children." Social Science Research 16: 39–73.

Hofferth, Sandra L. 1984a. "Kin Networks, Race and Family Structure." Journal of Marriage and the Family 46: 791–806.

_____. 1984b. "Long-Term Economic Consequences for Women of Delayed Childbearing and Reduced Family Size." Demography 21: 141–155.

_____. 1985. "Updating Children's Life Course." Journal of Marriage and the Family 47: 93–115. Michigan, Institute for Social Research.

Hofferth, Sandra L.; Moore, Kristin; Wertheimer, R.; Waite, L.; and Caldwell, S. 1981. "Teenage Childbearing: Consequences for Women, Families, and Government Welfare Expenditures." In Teenage Parents and Their Offspring, edited by Scott, Field and Robertson. New York: Grune and Stratton.

Hoffman, Saul D. December, 1979. "black-White Life Cycle Earnings Differences and the Vintage Hypothesis: A Longitudinal Analysis." American Economic Review 69 (5): 855–867.

Hutchens, Robert M. June, 1979. "Welfare, Remarriage, and Marital Search." The American Economic Review 69: 369–379.

Hymans, Saul, and Shapiro, Harold. 1974. "The Allocation of Household Income to Food Consumption." In Five Thousand American Families: Patterns of Economic Progress. Vol. 2, edited by James N. Morgan. Ann Arbor: University of Michigan, Institute for Social Research.

Katz, Lawrence, and Meyer, Bruce. 1988. "The Impact of the Potential Duration of Unemployment Benefits on the Duration of Unemployment" (Working Paper No. 2741). Cambridge, MA: National Bureau of Economic Research.

Leuthold, Jane H. 1981. "Taxation and the Value of Nonmarket Time." Social Science Research 10: 267–281.

Lillard, Lee, and Willis, Robert. September, 1978. "Dynamic Aspects of Earnings Mobility." Econometrica 46: 985–1012.

Margolis, Stephen. February, 1982. "Depreciation of Housing: an Empirical Consideration of Filtering Hypothesis." Review of Economics and Statistics: 90–96.

McLanahan, Sara S. May, 1983. "Family Structure and Stress: a Longitudinal Comparison of Male and Female-Headed Families." Journal of Marriage and the Family: 347–357.

_____. January, 1985. "Family Structure and the Reproduction of Poverty." American Journal of Sociology: 873–901.

_____. February, 1988. "Family Structure and Dependency: Early Transitions to Female Household Headship." Demography 25: 1–16.

McLanahan, Sara S., and Sorensen, A.B.. 1984. "Life Events and Psychological Well-Being: a Reexamination of Theoretical and Methodological Issues." Social Science Research 13: 111–128.

Mincer, Jacob, and Higuchi, Yoshio. 1988. "Wage Structures and Labor Turnover in the U.S. and Japan." Journal of Japanese and International Economics 2: 97–133.

GENERAL DESCRIPTION

Moffitt, Robert A. December, 1983. "An Economic Model of Welfare Stigma." American Economic Review: 1023–1035. Arbor: University of Michigan, Institute for Social Research.

Moore, Kristin A.; Waite, Linda J.; Caldwell, Steven B.; and Hofferth, Sandra L. August, 1978. "The Consequences of Age at First Childbirth: Educational Attainment" (Working Paper No. 1146–01). Washington, D.C.: The Urban Institute.

Morgan, James N. 1980. "Retirement in Prospect and Retrospect." In Five Thousand American Families: Patterns of Economic Progress, Vol. 8, edited by Greg J. Duncan and James N. Morgan. Ann Arbor: University of Michigan, Institute for Social Research.

Richards, Toni; White, Michael; and Tsui, Amy Ong. 1987. "Changing Living Arrangements: a Hazard Model of Transitions Among Household Types." Demography 24: 77–97.

Rosen, Harvey S. February, 1979. "Housing Decisions and the U.S. Income Tax: an Econometric Analysis." Journal of Public Economics 11: 1–24.

Rosen, S.R. August, 1982. "Taxation and On-the-Job Training Decisions." Review of Economics and Statistics: 442–449.

Ross, Heather, L., and Sawhill, Isabel V. 1975. Time of Transition: the Growth of Families Headed by Women. Washington, D.C.: The Urban Institute.

Solon, Gary; Corcoran, Mary; Gordon, Roger; and Laren, Deborah. Summer, 1988. "Sibling and Intergenerational Correlations in Welfare Program Participation." Journal of Human Resources 23 (3): 388–396.

Shishko, Robert, and Rostker, Bernard. June, 1976. "The Economics of Multiple Job Holding." American Economic Review 66: 298–308.

Turchi, Boone A. 1975. The Demand for Children: the Economics of Fertility in the United States. Cambridge, MA: Ballinger Publishing.

Yoon, Bong Joon, 1981. "A Model of Unemployment Duration with Variable Search Intensity." The Review of Economics and Statistics 63: 599–609.

CONTENTS FOR
CHAPTER B: INFORMATION COLLECTED IN THE PSID

Chapter B

Information Collected in the PSID

1. Coverage of Topics

The PSID contains a large number of variables that have been asked year after year in much the same, if not the identical, manner. They constitute what we term *"core" PSID content*, which is available on the main PSID data files. A wide variety of other topics have been covered intermittently in the study. Many of the variables based on these *intermittent topics* are also available on the main PSID data files, although some of the fine details have been relegated to special data files. A comprehensive (300-page) listing, arranged alphabetically by topic area, of the variables available on the main files is provided at the end of this Guide. That listing provides both variable numbers and tape locations for all variables from 1968 through the most current interviewing year for which data is available. Any problems of comparability in variables over the course of the study are noted there as well. Updates to that listing are scheduled for distribution when new waves of data are released. Below we try to provide a much more abbreviated summary of the entire content of the study.

1.1 *Core Content*

The "core" content of the PSID is listed in some detail in Table B.1. Most of this information comes directly from PSID questionnaires and coversheets. The greatest level of detail on these topics is available for the head of the family unit and, in cases where a male head is married or cohabiting for the same woman for one year or more, the wife or long-term cohabitor (termed "wive") of the head. Information about the wife/ "wife" is more limited in the early years of the study than the later ones. Throughout the study a modest set of information has been collected for individuals who were neither head, wife, nor "wife."

INFORMATION COLLECTED

Table B.1
Core Topics in the PSID*

A. Income Sources and Amounts:
Earnings of family members
Business/farm income
Income from professional practice or trade
Income from farming or gardening
Income from roomers or boarders
Income from rent
Dividends, interest, trust fund, royalties
AFDC/ADC
SSI
Other welfare
Social security
VA pension, service disability or GI bill
Retirement pay, pensions, or annuities
Unemployment compensation
Alimony
Child support
Help from relatives/non-relatives
Other income

B. Poverty Status:
Family poverty thresholds**

C. Public Assistance In the Form of Food or Housing:
Use of food stamps
Public assistance with housing:
. If in public housing project
If rent is publicly subsidized
Government assistance with heating bills

D. Other Financial Matters:
Estimate of federal taxes paid**
Financial assistance to people living elsewhere

E. Family Structure and Demographic Measures:
Marital events and status
Fertility events
Adoptions of children
Number of siblings (total and number still living)
Ethnic group
Race

F. Employment Information:
Annual and monthly information on:
Weeks worked
Weeks unemployed
Weeks out of labor force
Work missed because sick
Work missed because family members was sick
Weeks of vacation

Table B.1 (continued)

Weeks on strike
For each main job and second job:
  Occupation and industry
  Whether government worker
  Rate of pay on job
  Hours per week working
For each main job:
  Whether union worker
  If self-employed, whether business is incorporated
Work experience:
  Total
  Employer-specific
Employment status:
  Employment status at time of interview
  Whether have been looking for work and if so how
Event-history dating employment changes during past year:
  Movements between employers
  Title changes with the same employer
  Occupation and industry
  Pay and work load at start and end with each employer
  Reason for changing employers

G. Housework Time

H. Housing:
Size and type of housing structure
Whether own home, pay rent, or what
House value
Remaining mortgage

I. Geographic Mobility:
Moves during last year—when and why
Plans about moving in future-how certain and why
State and county of residence
Where head grew up—rural vs. urban, state and county
All states head has lived in
Whether head ever moved to take a job

J. Socio-Economic Background:*
Education history
Parents' completed education
Number of siblings
Race and ethnicity
Father's occupation
Parents' poverty status

K. Health, Religion, Military Service:
General health and disability of family members
Religious preference
Ever in military service

L. County-level Data:

INFORMATION COLLECTED

Table B.1 (continued)

Unemployment rates
Wage rates for unskilled workers
Labor market demand conditions

NOTES:
*The amount of detail for these topics is most extensive for the head and wife of the family unit, but some information is often provided for other family members as well.

**Estimates are generated for this information from indirect indicators collected in the annual interviews.

***Questions regarding an individual's socio-economic background are asked the first year the individual appears as a head, wife, or "wife" in an interviewed family unit. This information is *not updated* on a regular basis, although pertinent information may have been gathered subsequently that allows some updating. If the individual switches from a head one year to a wife the next, or vice versa, all of the socio-economic background questions are re-asked. In addition, in a few waves of the study socio-economic background information has been asked of all heads, wives, and "wives," regardless of whether they are new to that role that year.

INFORMATION COLLECTED

Most of the variables in Table B.1 are obtained directly from measures collected in the PSID's annual interviews. Inclusion of county and state of residence on the data files allows analysts to add environmental data for the region, state. or county of residence if they wish. There are some further possibilities for linking environmental data to the PSID's data files as well.[1] The PSID also supplements the interview information annually with some county-level detail obtained directly from state employment security officials. This county-level detail includes information about unemployment rates, unskilled wage rates, and labor market demand conditions.

Some additional variables are generated each year by the study staff working from indicators, but not direct measures, collected in the interview. These generated variables include annual estimates of family poverty thresholds and federal taxes paid. To construct estimates of family poverty thresholds, family needs variables are estimated from information about size and composition of the family unit in accordance with the USDA budget guidelines underlying the official set of U.S. poverty thresholds. The PSID's needs variable can be used to construct a ratio of family income to family income needs, which is useful both in defining whether a family unit is officially poor and for a general adjustment of income for family size. Estimates of the amount of federal taxes paid are based largely on reported income and a question about whether itemized deductions were used. These estimates are calculated each year for each tax unit persumed to be present in the household.

Timing of many events is recorded as part of the PSID's core information. This allows construction of a wide variety of event histories. (See Table B.2.) Some events

---

[1]For some years of the study (1968–1978) information is available regarding the name of the city closest to the family unit's dwelling. The codes for the variables reflecting this information were not provided in the documentation volumes for those years because of concern about confidentiality. This concern also led to suppressing the information on the original versions of the data files for those years. The potential research value of this information has grown in recent years, with heightened interest in effects of neighborhood factors. In response to this change, the PSID staff has re-inserted the information on recent cross-year files, starting with the 1968–1981 cross-year files. These codes are also available upon written request to:

    ICPSR, attn.: Janet Vavra
    University of Michigan
     P.O. Box 1248
    Ann Arbor, Mi. 48106.

A special file identifying the Census tract or enumeration districts in which respondents have lived during the course of the study is also available under special contractual arrangements with the Survey Research Center.

INFORMATION COLLECTED

can be dated to the month, whereas others can only be dated to the year. The precision of the dating for some events varies over the course of the study, as Table B.2 indicates. Monthly dating of employment events and income sources began with the 1984 interviewing year, and retrospective marital and childbirth histories collected in the 1985 interviewing year and updated since that time facilitated monthly dating of those events.

1.2 *Intermittent Content*

A variety of topics have been addressed by the PSID on a less regular basis than those noted above. These intermittent topics are listed in Table B.3. During the first five years of the study an hour-long personal interview allowed collection of an extensive set of information not as well-suited to collection with the study's switch to the telephone as its main mode of interviewing. Between 1973 and 1983 the intermittent topics varied more from one year to the next than they had in the early years of the study, but many of them were included in several waves of interviewing. With the advent of NSF funding in the 1984 interviewing year, a systematic review of PSID content was made by the PSID's NSF Board to define "core" topics. After that time, any other topics that were candidates for inclusion in the PSID needed a separate source of funding. This resulted in special supplements, each focused specifically on a given topic and generally included in only one interviewing year.

2. Types of Variables in the PSID

PSID variables vary in terms of who they apply to, where they come from, and how much hands-on processing has been done to them by the PSID study staff. The variables can be categorized as in Table B.4, which also provides information about the number of variables of various types.

2.1 *Family-Level Variables*

Most of the information from any year's data collection is categorized as *family-level variables*. The family-level variables include not only information that applies to the family unit as a whole (such as total family income or number of children), but also almost all information about the head of the family unit and, if present, the wife or "wife" of the head, plus a small set of information about the current county of residence.

Table B.2

Event Histories Recorded in the PSID

Event	Available Information About the Event
Events with Timing Recorded to the Month:	
Births	Complete birth histories for most PSID individuals, collected 1985 on [a]
Marital Status Change	Complete marital histories for 1985 heads, wives, and "wives" plus abridged marital histories for most other PSID individuals, collected 1985 on [a]
Formation and Dissolution of Cohabitating Couples	Gathered for heads and "wives" as part of every annual interview, 1976 on
Out-of-Wedlock Births	Can be derived from birth and marital histories for most all PSID individuals, collected 1985 on
Raising of Children By Non-Parents	Complete histories for heads, wives, and "wives," collected in 1985 [b]
Change in Living Arrangements, Including Children Leaving Home	For all PSID individuals, collected all years of panel for most living arrangements [c]
Change in Education	Complete educational histories for heads, wives, and "wives" (when last attended school; timing of high school diploma, GED, college degree, other degrees or certificates such as vocational or apprenticeship), collected in 1985
Residential Moves	For all PSID families, 1975 on (except 1982) [d]
Change in Receipt of Transfer Income	Separately by type of transfer income (including ADC/AFDC, other welfare, head's unemployment compensation, wife's unemployment compensation, head's workers' compensation), 1983 on
Transitions between Employment/ Unemployment/Out of Labor Force	For head's unemployment spells, 1981 on; for wife's unemployment spells and for all labor force transitions for heads and wives, 1983 on
Change in Job/Position	For head and wife, 1983-1987
Change in Employer	For head and wife, 1983 on
Change in Second Jobs	For head and wife, 1983 on
Change in Occupation	For head and wife, 1983-1987

Table B.2 (continued)

Event	Available Information About the Event
Events with Timing Recorded to the Year:	
Change in Receipt of Public Assistance	Based on report of calendar-year assistance given separately by type of assistance (including ADC/AFDC, other welfare, SSI and food stamps), for all years of panel (except 1973 for food stamps)
Change in Poverty	Based on report of calendar-year income and needs, for all years of panel
Transitions between Employed/ Unemployed/Out of Labor Force	Based on calendar-year reports of work hours and unemployment hours, for all years of panel
Retirement	Based on report of age at retirement, first reported in 1981 for heads and 1983 for wives. Information on annual work hours is gathered every year, providing an alternative way to define retirement
Change in Job/Position	Captures one change per year, for all years of panel
Change in Disability of Head	Based on disability status reported at time of interview, for all years of panel except 1973, 1974, and 1976

Notes:
[a] In 1985 retrospective histories were collected for 1985 head, wives, "wives," and other family members aged 12-44. In each subsequent years retrospective histories were gathered for a more limited set of individuals--heads, wives, and "wives" new to the study in that capacity and family members aged 12-44 who were neither head, wife, nor "wife." Each year beginning in 1986, the histories for prior-year heads, wives, and "wives" have been updated annually for events since the prior year.

[b] Dates for the raising of children by a non-parent includes only first and most recent spell for any given child/non-parent pair.

[c] Relationships between people living together can be more finely distinguished from 1983 on than during 1968-1982.

[d] Data only capture one move per year but indicate any changes in county or state.

Table B.3
Schedule of Intermittent Topics in the PSID

1968–1972, 1977–1987: Housing Utilities

1969–1986: Commuting to Work

1968–1972: Housing and Neighborhood Characteristics
        Attitudes and Behavior Patterns
        Do-It-Yourself Activities
        Saving (Crude measure)
        Disability of Family Members
        Fertility and Family Planning
        Child Care
        Time Use

1972 Only: Achievement Motivation
        Cognitive Ability (sentence completion test)

1973–1974: Child Care

1975: Neighborhood and Housing Problems
    Satisfaction
    Attitudes
    Disability of the Head

1976: Wives' Interview
    Employment History**
    Fertility and Family Planning*
    Characteristics of Job (including training required)**
    Attachment to Labor Force**
    Child Care*
    Attitudes**

1977: Child Care
    Disability of the Head

1978: Job training
    How got jobs
    Retirement plans and experiences
    Disability of Family Members

1979: Do-It-Yourself Activities
    Child Care
    Impact of Inflation
    Savings (Crude measure)
    Retirement Plans
    Disability of the Head

1980: Time and Money Help with Emergencies
    Food Stamp/SSI Eligibility
    Impact of Inflation
    Child Care
    Disability of the Head

INFORMATION COLLECTED

Table B.3 (continued)

Extended Family
Savings (Crude Measure)

1981–1983: Retirement plans & expectations (most detail in 1983)
Spells of Unemployment/Out of Labor Force
Hospitalization Over the Year
Disability and Illness of Family Members

. 1984: Wealth (level of assets of various types)
Fringe Benefits
Pension plans and rights
Retirement plans
Inheritances
Savings (Crude Measure)
Job Training
Spells of Unemployment/Out of the Labor Force
Disability and Illness of Head and Wife

1985: Wives' Interview
Retrospective Childbirth History**
Retrospective History of Adoptions**
Retrospective History of Substitute Parenting**
Retrospective Marital History**
Retrospective Education History**
Child Care**
Housework**
Family Planning**
Disability and Illness of Head and Wife**
Job Training**

1986: General Health of All Family Members
Activities of Daily Living**
Hospitalization Over the Year**
Height and Weight**
Smoking and Exercising Behavior**

1988: Kinship Ties
Financial Situation of Parents
Health of Parents
Time and Money Help of Most Kinds

1989: Wealth (level of assets of various types)
Saving Behavior 1984–1989

1990: Health and Health Care of the Elderly

NOTES:
   * Questions asked of Wife
 ** Questions asked of both Head and Wife

Table B.4
Types of Variables in the PSID

Type of Variable	Number of Variables in 1987 Wave
I. Family-Level Variables	1,037
A. Edited variables	387
B. Coded variables	567
C. Generated Variables	80
D. County Variables	5
II. Individual-Level Variables	84
A. Year-Specific Variables	36
B. Summary Variables	48

The main PSID files contain about 1,000 family-level variables for the 1987 interviewing year.

About 400 of the 1,000 family-level variables are *edited variables* in the 1987 interviewing year. These variables include income, work hours, remaining mortgage, and family needs. They are variables which are considered crucial enough to the overall purposes of the study to merit special treatment in terms of assigning missing data (assignments are always made for these variables when there is missing data for them) and performing data quality checks (such as year-to-year consistency checks). These variables are constructed from a detailed set of instructions. Most of the edited variables have an associated variable that indicates the extent of editing done for any given case; this associated variable codes whether major, minor, or no assignments for missing data were made when processing the data for that case.

The set of family-level variables termed *coded variables* is the largest of the various types of variables. These variables involve information taken directly from the interviews, with no special editing. A small set of family-level variables is calculated by computer from the edited and coded variables, and, in some cases, from previous years' variables. These are known as *generated variables*. A few additional family-level variables take the form of *county-level variables*. These are based on county-specific information collected from state officials; they apply to the family unit's county of residence at the time of the interview.

INFORMATION COLLECTED

2.2 *Individual-Level Variables*

A small set of *individual-level variables* is available for each individual in a family unit interviewed by the study. This set is comprised of both year-specific variables and summary variables that may span many years. About 40 *year-specific individual variables* are provided for the 1987 interviewing year, with valid data for each individual in a family unit interviewed in 1987. These variables cover basic demographic and economic data about an individual, collected in the 1987 wave. If the individual was head of a family unit or the wife or "wife" of a head, much of the information in these variables is also available among the family-level variables, often in substantially greater detail. The *summary variables* at the individual level first appeared with the 1985 interviewing wave, but they can cover a wide time frame since retrospective historical information forms the basis of many of these variables. Either time-invariant information (such as birth date, identity of parents, and status as an original sample member), cumulative counts of rare events (such as number of marriages or number of childbirths), or the timing of rare events (such as month and year of various marriages or childbirths) are represented by these variables.

CONTENTS FOR
CHAPTER C: PSID DATA FILES

Chapter C

PSID Data Files

1. Types of Data Files

The PSID routinely prepares a number of data files. Most are updated with each new wave of data collection and then made available, along with comprehensive documentation, through the Inter-university Consortium for Political and Social Research (ICPSR). The ICPSR is a member-based social science data archive that maintains machine-readable data files for distribution to interested users.

Most PSID files contain information dating back to the study's first wave (1968) and include records for either family units, individuals, or pairs of individuals. The types of files and longest interval of data they contain are listed in Table C.1. The files fall into three major categories—main files, special public files, and special restricted files.

1.1 *Main Data Files*

The PSID's main data files contain information gathered since the beginning of the study. The files vary with regard to whether there is one record per family unit or one record per individual. Files with one record per individual contain either individuals in a family unit interviewed in the most recent interviewing year (*response*) or individuals who participated in the study at some time in the past but not in the last interviewing year (*nonresponse*). Whether the record represents individuals or family units, the data date back to the start of the study (1968). The **cross-year family file** contains information about family units, which includes all the detail collected about the Head and Wife/"Wife" in a given family unit as well as about the family unit more generally. The **cross-year family-individual** files have one record per individual and contain a small set of information about that individual, gathered each year the individual has participated in the study, and a large set of information about both the family units in which the individual has resided over the years and about the primary adults heading those units when the individual lived there.

The distinction between the *response* and *nonresponse* version of the *cross-year family-individual* files began with the 1968–1984 cross-year file and has been made for all subsequent cross-year family-individual files since then. The **cross-year family-**

Table C.1
Types of PSID Data Files Currently Available

Type of File	Widest Range of Interviewing Years Currently Covered by the File	Is File Updated With Each Subsequent Wave of Data?	Is File Available Through ICPSR?
Main Files			
Cross-Year Family File	1968-1987	Yes	Yes
Cross-Year Family-Individual Response File	1968-1987	Yes	Yes
Cross-Year Family-Individual Nonresponse File	1968-1987	Yes	Yes
Special Public Files			
Ego-Alter File	1985*	Yes	Yes
Work History Supplemental File	1984-1986	Yes	Yes
Relationship File	1968-1985	No	Yes
Special Restricted Files			
Census Tract/Enumeration District File	1968-1986	No	No**
Death Certificate File	1968-1984	No	No**

NOTES:
*The information in this file can date as far back as the early 1900s, since it contains marital and child birth histories for all PSID adults in family units interviewed in 1985, some of whom were quite elderly by then.

**See Chapter 22 on Special Files for details about access to these files.

individual response file contains records for all individuals in family units interviewed in the most recent wave of data collection. Prior to the 1984 interviewing year, this was the only available form of the cross-year family-individual files. The **cross-year family-individual nonresponse file** has been prepared each wave since the 1984 interviewing year. It contains the same type of information as the response file, but for a different set of individuals—individuals who were part of a family unit interviewed prior to, but not in, the most recent wave. These "nonresponse" persons have left the study for reasons such as death, refusal, or failure to locate them after a move. The non-response file is structured in exactly the same way as the response file (except, of course, it contains no family or individual information after the point of nonresponse). The *cross-year family-individual response* and *nonresponse* files can easily be concatinated. The combined file is very useful, since it enables an analyst to go back to a prior year and gather data from all of the individuals present in that prior year.

1.2 Special Files

Several special PSID files contain detailed information that would be cumbersome to store on the study's main files. Hence, the details have been relegated to special files and the information presented in a more summarized form on the main files. Analysts wanting the complete details on the special topics must turn to these special files. The special files may have some stand-alone uses and contain some of the same information as the main files, but they are of greatest value if merged with the main PSID data files. Most of the special files are publicly released through the ICPSR. One public-release special file, called the **Ego-Alter file**, contains records for pairs of individuals who are related by marriage, childbirth, adoption, or substitute parenting. The information on this file pertains to the timing and circumstances of the demographic event relating the individuals—parenting or marriage. Another public-release file, the **Work History Supplemental file**, contains complete information, for Heads and Wives/"Wives", about all of the spells of employment, unemployment, second jobs, etc., reported each wave, beginning with in the 1984 wave. A third public-release file, known as the **Relationship file**, shows the blood, marital, or cohabitational relationship between pairs of individuals up to the 1985 interviewing year. Relationships among all individuals who were members of family units that have descended from a common, original 1968 family unit have been assembled on this file.

PSID DATA FILES

Two additional data files are prepared by the PSID staff but are available only under special contractual arrangements with the Survey Research Center. One contains Census tract or enumeration district and Zip Code identifiers for all PSID family units for all years of the study. The second shows cause of death for PSID individuals who died between 1968 and 1985 and for whom official death certificate information is available. Since both of these files contain information that would make it much easier to identify participants in the PSID, they are not part of the public releases and are available only under special contractual arrangement.

2. Documentation of the Files

Detailed documentation is provided, through the ICPSR, for all PSID data files released to the public. Variables on the main PSID files—the cross-year family file and the cross-year family-individual files (both response and non-response)—are detailed in the 16 volume set of documentation entitled A Panel Study of Income Dynamics— Procedures and Tape Codes. This multi-volume set covers interviewing years 1968–1987. The first five waves of data are documented in the first two volumes of the set. For each subsequent wave a separate volume of documentation has been prepared and made available. Beginning with the 1985 interviewing year, the documentation is available in machine-readable as well as printed form.[1]

The documentation volumes for the main PSID files contain the questionnaire, a description of the processes used for edited variables such as income and work hours, and a complete listing of the code categories for all variables, including missing data codes. Each volume of documentation also contains an alphabetical index of all variables to date listed by topic area, plus a concordance of all variables available for the given wave of data, listed in the variable number order for that wave. These listings of variables give the variable numbers and tape locations for comparable variables for all waves of the study up to and through the given one. Questionnaires are presented in the front of the documentation volumes. These documentation volumes are quite comprehensive, and there is no real substitute for them for the analyst who wants to discover the details of the data.

---

[1]The ICPSR routinely distributes the printed form of documentation. The machine-readable documentation is available upon special request.

Special documentation volumes are also available for the PSID special files that are publicly released—the ego-alter file, the work-history supplemental file, and the relationship file.

3. How to Order Data Files and Documentation

Most of the data files from the Panel Study of Income Dynamics are available through the ICPSR.[2] More than 270 academic institutions are currently members of that organization. Member institutions pay a fixed annual fee, which provides access to all data tapes in the archive. Requests for data at member institutions are coordinated through an Official Representative (OR) at that institution. Data are available at a per-item charge to users at non-member institutions. Individuals at non-member institutions, persons who are uncertain if they are at member institutions, and persons not knowing the name of their Official Representative should contact the ICPSR directly. The ICPSR writes the data onto magnetic tapes at technical specifications and in data formats that are compatible with the user's installation.

When ordering PSID files that are updated on a regular basis, it is almost always advantageous to order the *most recent* version the file. Information from prior waves are also contained on the most recent files. With the advent of the cross-year family-individual nonresponse file, prior waves can be accurately represented with the main PSID files by concatinating the most recent response and non-response cross-year family-individual files. The PSID staff corrects any errors discovered in prior waves of data when it merges on subsequent waves. This produces some change in values of prior variables, including an occasional change in the identifiers for individuals. The ICPSR does not contain any version of PSID files other than the most recent one; hence, earlier versions have to be ordered directly from the PSID staff.

---

[2]The Inter-University Consortium for Political and Social Research (ICPSR) has archived thousands of data files on such topics as social indicators; social institutions and behavior; mass political behavior and attitudes; organizational behavior; education; census enumerations; community and urban studies; economic behavior and attitudes; health care and health facilities; environment and natural resources; and instructional packages. A complete listing of all data collections in the holdings and a brief description of each can be found in the Guide to Resources and Services which is published annually by ICPSR.

Persons interested in obtaining more information about the ICPSR, any data in its holdings or a copy of the Guide to Resources and Services should address their inquiries to Janet Vavra, ICPSR, P.O. Box 1248, Ann Arbor, Michigan 48106. Phone (313) 763-5010.

CONTENTS FOR
CHAPTER D: CRUCIAL POINTS ABOUT USING THE PSID

1. CHECKLIST OF CRUCIAL POINTS

This checklist is intended as a brief listing of important points to consider when working with the PSID. Many of the points touched on in this chapter are discussed in greater detail in subsequent chapters.

1.1 *For Any Analysis of the PSID*

1) **Reference Period for Variables:** Are you aware that some of the information gathered in the year t interview refers to calendar year t-1, but other information refers to the situation of the family at the time of the year t interview? It may be necessary to "line up" information from two different interviewing waves. A check of question wordings or headings for variables in the yearly documentation's tape codes provides information on whether the present or the past is the frame of reference.

2) **Weighting:** If you are estimating simple statistics such as means, variances and correlations, are you weighting the data? If you are estimating a more elaborate model, are you confident enough that it is properly specified so that weights are unnecessary? If your more elaborate model has a dependent variable related to family income (e.g., Head's wage rate or annual labor income), then any unweighted analysis should exclude the low-income subsample. (Chapter E provides details.)

3) **Proxy Respondents:** Do you realize that only one person per family unit provides an interview in a given year and that this person is generally, but not always, the Head of the family unit? The respondent is usually the Head of the family unit, but sometimes the Wife/"Wife" of the Head agrees to grant an interview when the Head does not, and, in a few cases, someone in the family unit other than the Head or Wife/"Wife" grants an interview when the Head has no Wife/"Wife" and does not want to be interviewed. Do any of your key variables (e.g., attitudes) require that the Head be the respondent? If so, you need to subset your file on the basis of *Who Was Respondent?* to eliminate cases with someone other than the Head as the respondent.

4) **Missing Data:** Have you properly handled missing data values on all of your variables? If your key variables are ones that were "edited" by the study staff, have you considered eliminating instances where major or minor assignments were made? (Chapter F provides details.)

5) **Extreme Cases:** Have you checked your data for extreme cases? This is not a problem that is specific to the PSID, but rather something to be considered in analysis of any dataset. (Chapter I provides details.)

CRUCIAL POINTS

6) **Background Measures:** If you are using variables from the "Background" (new Head or new Wife/ Wife") section of the questionnaire, do you realize that the values for these variables are not updated annually? Questions in the Background section on retrospective work history, asked of Heads after 1985, for example, have been asked only of new Heads each year. Beginning in 1983, a family-level variable has been coded each year indicating when the Head of the family most recently became a new Head and thus answered the background questions. (Chapters D and F provide details.)

7) **Measures Requiring Special Understanding:** Does your analysis require an official definition of poverty or call for measures of disability, religious preference, taxes, education, transportation, background information or aggregate information matched on the basis of state or county? Such measures in the PSID have many idiosyncratic aspects to them. (Chapter D provides details.)

8) **Multiple Family Units within the Same Household:** Are you aware that some family units that we treat as separate units may be living in the same household? Some individuals who set up independent households during the panel period subsequently return to their original households (as when a child leaves the parental home for a new home of his or her own, and then returns to the parental home). Splitoffs such as these usually continue to be interviewed as separate family units, even after they return to their original households. (Chapter G provides details.)

9) **City Size and Urban/Rural Measures:** Does your analysis require measures of city size, or urban/rural measures that are comparable to Census definitions? If so, differences between the PSID and the Census Bureau regarding these measures are important. (Chapter D provides details.)

10) **Changes Over Time In Primary Adults Heading Family Units:** If your data come from several interviewing waves, have you considered that the Head of the family in year t may be different from the Head of that "same" family in years before or after t? Do not get caught, for example, trying to relate the hourly earnings reported by a young male Head in 1987 to the educational attainment that his father reported in 1976, when the father was the Head of the family in which the son resided in 1976. *Changes in the family "head" have caused serious errors in analyses with PSID data. It is important that analysts understand the implications of these changes for the structure of the data.* (Chapter G provides details.)

11) **Possible Across-time Inconsistencies in the Data:** If you use data from several interviewing waves, have you checked for possible inconsistencies in key variables over time? (Chapter F provides details.) Have you checked for possible differences in the way that variables have been coded over time? (Appendix 2 provides details.)

12) **Representative of U.S. and major regions, but not states or cities:** Are you aware that the PSID sample is drawn to be representative of the U.S. as a whole and within major Census regions, but it is *not* designed to be representative within states, cities or similar geographic divisions? (Chapter H? provides details.)

1.2 *When Individual is Unit of Analysis*

13) **Seemingly Valid Data for Persons Not in Interviewed Family Units:** Do you realize that seemingly valid data may appear on the family-level and individual-level portions of an individual's data record in a given year even if that individual was not residing with an interviewed family in that year? To eliminate instances where individuals were in institutions (e.g., college, the armed forces. jail) or were temporarily or permanently nonresponse and are therefore "carried along" by the study, you should select individuals whose "*Sequence Number*" (typically the second individual-level variable coded each year) is in the 01–20 range in the year in which you want to require membership in an interviewed family. (For 1968, use a combination of the variables *Person Number* with code values of 001–019 and *Relationship to Head* with code values of 01–09). (Chapter G provides details.)

Chapter D

Crucial Points about using the PSID

2. Introduction

The PSID is a multi-purpose data set for use in both cross-sectional and longitudinal analyses, and for studying both individuals and families. Its design is complex because of its multipurpose nature and because there is enormous diversity in the experiences of PSID members (if it can happen, it will in our study!). All of this means that adaptation of the data files to particular purposes can be quite complicated. The goal of this chapter is to provide sufficient detail about the PSID to enable an analyst to understand the basics of, and possible pitfalls in, using the data. Finer levels of detail about the PSID are relegated to subsequent chapters of this Guide.

This chapter is oriented toward the types of analyses a researcher might want to do with the PSID data. Analysis possibilities with such a multi-year, longitudinal study are so diverse that it is impossible to detail the full flavor of them in a short amount of space. Chapters H and I provide details about the rich variety of analysis possibilities, whereas this chapter provides a broad overview. Analysis possibilities include cross-sectional analyses based on a single year of data, cross-sectional analyses based on multiple years of data, and a variety of longitudinal approaches. Examples include the following:

- describing or modelling individual change in measures of interest;

- averaging a measure over several years to reduce the effects of random errors of measurement or of transitory fluctuations;

- taking different measures from different years (since not all questions were asked in every year);

- using the long series of year-to-year reports of rare events to construct "event histories" of various demographic and economic behaviors;

- relating, for large numbers of individuals who were children in the first year of the study, own reports of attainments in adulthood to the characteristics of the family in which they grew up, reported by their parents during those years;

- pooling several years of data to perform "pooled cross-section time series" analyses;

- pooling pairs or triplets of years, perhaps surrounding an event of interest (such as divorce, death, or childbirth) to look at antecedents or consequences of the event;

- matching ex-husbands and ex-wives to examine the comparative effects of divorce and the potential for larger child support transfers;
- matching siblings to estimate "sibling" models that are designed to determine the effects of parental background on attainment.

Data analysis is easiest when using only the most recent year's information and treating it as a single year of cross-sectional data. Care must be taken to distinguish between family-level and individual-level variables. and a decision must be made whether or not to weight the data, but the data complications are relatively minimal. *Still, to do any analysis of the PSID—even the simplest—the analyst must understand the structure of the data files and possible complications introduced by family composition changes.*

The first section of this chapter reviews key points about the PSID noted in prior chapters. The second section discusses key issues to consider in any analysis of the PSID. The third section describes the structure of the PSID data in greater detail than prior chapters, and the section following that indicates which file is best suited to what type of analysis. The remainder of the chapter discusses the PSID in the context of particular analysis approaches. The approaches vary in terms of use of single versus multiple waves of data, use of the most recent wave versus a past wave of data. focus on individuals versus families as the units of analysis, and cross-sectional versus longitudinal uses of multiple waves of data.

3. Review of Key Points about the PSID

Since 1968, the PSID has:

- changed the content of the questionnaire somewhat from year to year, but kept a large core of questions constant,
- interviewed one person, usually the "Head," in each of the family units in its sample.

The PSID is a family-oriented study, but it collects differing types of information about different types of family members. Because of this, it is important to keep in mind some rules regarding the relationships of the family members and how the type of information collected varies with the relationships:

CRUCIAL POINTS

• If there is only one adult in a family, that adult is designated as "Head" of the family unit.[1] In family units with a husband and wife, the husband is almost always designated as Head.[2]

• A great deal of the information collected in the interview applies to the entire family (e.g., total family income, food expenditures, and number of children under the age of 12).

• Extensive information has also been gathered on the individual designated as Head (e.g., work hours, labor market earnings, unemployment experience, disability status, and social-psychological attitudes).

• Since the 1979 wave, much of the detail gathered for the Head has also been gathered for the Wife/"Wife,"[3] but it is usually reported by the Head.

• More limited information is collected for members of the family unit other than the Head or Wife/"Wife."

The information gathered each year in the PSID is assembled into variables that are considered either *family-level* or *individual-level*, and data files are produced that contain both current-year and past years' information. These merged cross-year files can represent either families or individuals.

• Family-level variables deal with the family as a whole or the family unit Head or Wife/"Wife", if present.

• Individual-level variables are a small set of demographic and economic measures for each individual in a family unit.

• *Cross-year family files* contain only family-level variables and represent family units.

• *Cross-year family-individual files* represent individuals, but each record contains both a given person's individual-level information and the family-level information for the family unit with which he or she is associated.

This means that details for the Head and Wife/"Wife" are contained in the record of each and every family unit member represented on the cross-year family-individual files. It also means that some information for a Head or Wife/"Wife" is repeated on the cross-year family-individual files, first as family-level variables and then as an individual-level variables.

---

[1]Unlike the Census Bureau, the PSID makes no distinction between "families" that contain only one person living alone and "families" that contain more than one individual. The Census Bureau terms the former group "unrelated individuals" and often analyzes them separately from the rest. See Chapter G for a discussion of how the Census Bureau's concept of "household" and "family" compare with the PSID's concept of "family unit."

[2]For exceptions and further explanation, see Chapter G.

[3]The term Wife refers to a *legal* spouse; the term "Wife" (in quotes) is of PSID coinage and denotes a long-term female cohabitor.

Certain PSID data files can represent the U.S. population as of the most recent wave. These are:

- the *cross-year family file* (containing one record for each family interviewed in the most recent wave) and.

- the *cross-year family-individual response file* (which contains one record for each individual in a family unit interviewed in the most recent wave).

Another file is useful for helping represent the population of families or individuals at some time in the past:

- the *cross-year family-individual nonresponse file* contains information for all individuals who had been members of families interviewed in the past, but not in the most recent wave).

The response and nonresponse versions of the cross-year family-individual file follow the same structure and can be concatenated.

- The *concatenated cross-year family-individual response + nonresponse file* contains the information for all individuals ever part of the PSID study.

4. Key Analysis Issues

Five questions are important to address before doing an analysis of the PSID. These are: analysis using the PSID:

1. **Can family composition change influence the variables or the sample, and, if so, what should the analyst do about this?**

2. **What variables are available and how are they constructed?**

3. **Should the data be weighted?**

4. **Should the analyst accept the estimates of sampling errors and standard errors provided in most statistical programs?**

5. **Should the analyst use all cases from the start, or save part of them for a final test of the model?**

Each of these questions is explored in more detail below.

4.1 *Family Composition Change*

Even cross-sectional analyses can be affected by family composition change, because not all family measures taken at the time of the interview hold for the same time period. Annual family income and annual family needs, for example, both refer to the calendar year preceding the time of interview, but interim changes in family composition are certainly possible. Family composition change can play an even greater role in longitudinal analyses, and a vital point to keep in mind regarding the PSID is

CRUCIAL POINTS

that it tracks both individuals and families. This is often a distinct advantage, but it also makes it very important to maintain clear distinctions between family units and the individuals within them when planning and implementing PSID analyses.

The sizable amount of family composition change from one year to the next initially came as quite a surprise to the study staff. This was something about which little was known when the study started, and, as the study has progressed, the cumulative level of family composition change has become quite striking. The study began, in 1968, with the naive notion that families were like T.V.'s "Cleaver family," (Ward, June, Wally, and the Beaver), comprised of a husband, his wife, and their children, who remained together through time. Not only was Ward head of the family year after year, but June was always his wife. Such families make things simple for data structure and for the analysis of a family-based data set. Looking at the Cleaver "family unit" at two points in time:

> one can analyze changes in Ward's wage rate by simply comparing the wage rate of the Head of the family unit at those two points in time;

> one can analyze changes in the Cleaver family income by simply comparing family income in different calendar years; and

> one can study June's labor force participation decisions by analyzing the patterns of the Wife's/"Wife's" work hours.

However, only about *one in twenty* of the Panel Study families has remained completely intact for the first 18 years of the study. Couples divorce, spouses die, children grow up and leave home, and new children enter families at birth. These life-cycle changes are routine. More complicated changes include: children leaving home in "false starts" (later returning to their parental household); married couples separating for a few years and then reuniting (or even divorcing and remarrying one another later); grandchildren, aunts, nephews, or other relatives moving into or out of the family; and children being born to one of several unmarried daughters or sons still living in their parental home.

These changes can have profound implications for analysis of the PSID, especially if the analysis involves two or more waves of data. Some of these implications are simply annoying complications to the structure of the data, but *many of them raise substantive issues that analysts must address*. For example:

1. A number of variables available in any one wave of the data have been adapted to account for family composition change. Such measures include annual family income of all types and annual family needs. These measures are likely to be included in cross-sectional as well as longitudinal analyses, and it is

important for the analyst to understand how they are constructed to best match them with measures of family structure.

2. The very concept of the family loses its traditional meaning when placed in a dynamic context. For example. if there is a divorce, then a single family becomes two families—a "family" of the ex-husband and a "family" of the ex-wife. If one wants to describe changes in the economic status of families. then which family is to be thought of as the "same" family? Any rule that tries to use the family as the unit of analysis can be confronted with a case that makes it look absurd. As explained below, this problem can be solved if the unit of analysis is the individual rather than the "family."

3. Relationships defined by family status, e.g.. "wife." "child" or "head." are unique at one time but not across time. Thus, if one wants to analyze the labor force patterns of married women. one must realize that to restrict the analysis to *continuously* married women is to leave out a sizable and potentially interesting group of women who changed marital status (and may well have changed their labor supply patterns at the same time).

4. Longitudinal analyses over several years are most naturally conducted on individuals who were alive during all of those years. But one might want to include individuals who died during that period. This would avoid selection biases that might result if these individuals are ignored. By the same token, analysts may want to include children born into the Panel during the period under investigation.

Even if these substantive issues are not relevant, technical problems caused by family composition changes can be substantial. For example:

1. The Head (or the Wife/"Wife") of a family may not be the same in year t as in year t + 1 or year t-1. If Wally left home between 1975 and 1976 to form his own family unit. then his family history contains family-level information reported by Ward (during the years through 1975 when Wally was a "child") and then family-level information with Wally as Head (from 1976 on). *Changes in the family Head have caused serious errors in analysis with PSID data. It is important that analysts understand the implications of these changes for the structure and interpretation of the data.*

2. Not all individuals living in a family in year t share the same family history. If a couple has reunited after a separation of several years, the family-level variables (such as annual family income) for those years of separation will differ for the two individuals.

4.2 *Construction of Variables*

When choosing or adapting variables in the PSID, there is no substitute for studying the specific volume of documentation (or, for recent years, machine-readable documentation file) that accompanies each year of data. Each volume contains the questionnaire, a description of the process by which crucial income and work hours variables are edited, frequency distributions for each variable, and a complete listing of

CRUCIAL POINTS

the code categories for all variables, including missing data codes. Each documentation also contains an alphabetical index (since 1984 the index has been published separately as Volume II of each year's documentation) and concordance of current and past variables. In this Guide there is a summary of content areas in Chapter B. In addition, Appendix 2? of this Guide provides the most current version of the alphabetic index of variables, showing for what years comparable variables are available.

4.3 *Weighting the Data*

Decisions about weighting the data can be difficult. There are four reasons why unweighted estimates made from PSID data might not correspond to U.S. population totals.[4] First, the initial sample consisted of about 3,000 families who were chosen from a Survey Research Center self-weighting probability sample, but they were combined with about 2,000 low-income families that had previously been interviewed as a part of another study.[5] (See Chapter E for details.) Second, the dynamics of family composition change produce a larger proportion of younger family units and individuals than appears in the population as a whole. (See Chapter E for an explanation.) Thus, even the SRC cross-section portion of the sample has become "overloaded" with the young and will not produce unbiased estimates of simple population parameters unless weighted. Third, there has been some differential nonresponse over the years.[6] Fourth, immigrants have joined the population of the United States since 1968 but have not been added to the PSID sample.

Although the PSID cannot be adjusted in a way that makes its sample representative with respect to recent immigration, it can be adjusted in ways that help overcome the other three problems. Weight variables (one at the family level and one at the individual level) have been constructed each year to account for the effects of initial

---

[4] Not included among these reasons is the notion that the PSID sample has "aged" over time and no longer represents young families and individuals. This notion is false because the PSID does have a mechanism for adding new families and individuals (births) to the sample just as new families and individuals are added to the U.S. population. See Chapter G.

[5] These two groups can be distinguished through their 1968 Interview Number. The self-weighting part of the total sample has *1968 Interview Numbers* in the 0001–2930 range. The low income subsample has *1968 Interview Numbers* in the 5001–6872 range.

[6] This differential nonresponse also contributes to the nonrepresentativeness of the unweighted SRC portion of the sample.

oversampling of some subgroups, expansion over time in the proportion of younger families in the study, and differential nonresponse.

When should the data be weighted? Clearly, weights should be used whenever the analyst uses cases from both parts of the sample to estimate simple population parameters such as means, variances or simple correlations between variables. If such estimates are not based on the weights, then they describe only what is true for the PSID sample and not for any subgroup within the population or for the population as a whole.

If the analyst wishes to use the data to estimate a properly-specified multivariate model, the case for weighting is less compelling, since the model presumably controls for the effects of the factors that lead to the need for the weights in the first place. (An exception to this is when the dependent variable of the model is income or earnings. In this case, there is no justification for including the low-income subsample if weights are not used.) These issues are detailed in Chapter E.

4.4 *Sampling Errors*

Analysis programs typically assume simple random sampling. The PSID sample (and the samples of virtually all other national studies of this kind) is a clustered, multi-stage stratified sample, often producing higher sampling errors and standard errors than estimates produced under the assumption of simple random sampling. See Chapter E for details.

4.5 *Hypothesis Testing*

Standard statistical tests are not valid if the analyst has "searched" through the data before arriving at the final specification of the model. For each year there is a *Split Sample Filter* variable (a family-level variable) that can be used to divide the sample into independent quarters.[7] As explained in Chapter I, the analyst can search through one-quarter, one-half or three-quarters of the data and then test the final model on the remaining portion of the sample.

---

[7]The complex nature of the sample makes it impossible to do this using a random number table or similar device.

CRUCIAL POINTS

5. Basic Structure of the Main PSID Files

Diagrams of flat-file versions of the PSID cross-year files may assist in better picturing how the data are assembled and which main data file is best to use for what purpose. Here we present such diagrams and a brief explanation of them. (Greater detail about the structure of the files is provided in Chapter ?.)

5.1 *Cross-Year Family File*

To illustrate the structure of the *family* (as opposed to *family-individual*) file, we focus on one year—1987. In 1987 the PSID interviewed 7.061 family units and coded values for 1037 family-level variables for each of them. If the PSID had produced and released a 1987 single-year family file (which it did not), the structure of that file would look like that of Figure D.1. Each record represents one family unit, and the file strings together the 1037 family-level variables for the first family (as the first record), followed by the 1037 family-level variables for the next family (as a second record), and so on, up to the 7,061st family (as the 7,061st record).

Such a file would be sufficient for an analyst interested in only a single year of family data. However, many analysts want more than just one wave of data, and to serve the broad needs of researchers, the PSID documents and releases the single-year family information as part of a much larger file combining the latest year of family-level data with all of the previous years of family-level data for each family's "root" family back through 1968. This file is called the *cross-year family file*.

The structure of the cross-year family file (See Figure D.2.) is a simple extension of the illustrative single-year family file, with family-level data for each year 1968–1987 in each record and in order from the earliest year (1968) to the most recent year (1987). In effect, this file merges a 1968–1986 family-level history with a 1987 family unit's 1987 data. Different members of the same 1987 family unit may have been in different families at some time during the 1968–1986 period, which would mean they would not have experienced the same 1968–1986 family history. The PSID's rule regarding this is that the family history is that of the Head of the family unit in the most recent year. So, each of the 7,061 families interviewed in 1987 has a history of yearly family-level data dating from the beginning of the study (1968) up to the preceding year (1986). This family history is that of the 1987 family unit Head. Added to that family history is the family-level information collected in 1987 for the 1987 family unit. The family history portion of a record contains a total of 13,687 family-level variables, and the

1987 family-level data consists of 1,037 variables, bringing the total to 14,724 variables in each record on the 1968–1987 family file. The file strings together 14,724 variables for one family, followed by 14,724 variables for another family, on up to the 14,724 variables for the 7,061st family.[8] The variable numbers and tape locations for the family-level variables are listed in the appropriate documentation volumes or machine-readable documentation files, and also in the OSIRIS dictionary file.[9]

The 1968–1986 cross-year family file produces a cross-sectional file representative of 1987 data on 1987 families. It cannot be used to represent cross-sections of families in any prior year. This is because some prior-year families will have become nonresponse by 1987 and others will have had several members "splitoff" to form multiple 1987 family units. Thus, the sample of family units interviewed in 1987 is representative of 1987 families but not of 1986 families, or 1968 families, or families for any other year.

5.2 *Cross-Year Family-Individual File*

Variables available in the cross-year family files apply to the family as a whole or to the family Head and Wife/"Wife".[10] But, each year, information is also gathered on a small set of variables about *all* individuals living in the family unit that year. In addition, a number of individual-level summary variables with broad time frames are now assembled and updated on the rare occasions when they change. The files that combine, within a single record, family-level variables and individual-level variables for an individual are called the family-individual files. These are the most versatile of the PSID's main files. *A thorough understanding of the cross-year family-individual files is often the key to success in analyzing the PSID.*

---

[8]For simplification of documentation and sanity retention, we start numbering each year's family variables at a round number, leaving a small gap between the last family variable from year t and the first family variable in year t + 1. Thus, although the last 1986 family variable is numbered 13,687, there are several dozen fewer actual family variables coded between 1968 and 1986.

[9]The first five years of the study (1968–1972) are documented in A Panel Study of Income Dynamics: 1968–1972 Interviewing Years (Waves I-V), a two-volume set. Additional volumes have been published for each wave since then. From 1985 (Wave XVIII) onward, we have also made the material in the documentation volumes available on machine-readable files. See Appendix 4 for more information about these volumes and files.

[10]Individual-level information is used, however, in the construction of certain family-level variables, such as total family income, housework hours, etc.

CRUCIAL POINTS

For illustrative purposes, Figure D.3 depicts the structure of a single-year family-individual file that ignores the largely time-invariant "individual summary variables." This file is much larger than a comparable single-year *family* file. Although there are just a few more additional variables in each record (the 1987 individual-level variables comprise a small set of variables), the number of data records increases dramatically. There are 20,487 family-individual records on the 1987 family-individual file as compared with 7,061 family records on the 1987 family file. (To illustrate why such a large increase occurs, suppose a 1987 family unit contained three members. There would be three records on the family-individual file for the one family, one for each of its members.) The structure of this family-individual file is fairly simple to understand, although inefficient, since exactly the same family-level data are repeated for every individual in a given family.[11]

Recently we have added one more wrinkle to the individual data: a set of *time-invariant* individual-level variables — at the end of each record — that summarize such things as an individual's nonresponse, fertility and marital histories and, if known, the PSID identification numbers of the individual's father and mother.

The huge file that merges the most recent year of family-level and individual-level variables for an individual with all past years' data for that same person is called a cross-year family-individual file. This file orders each individual's data record so that all of the cross-year family-level variables for that person appear first, followed by all of his or her cross-year individual-level variables. (See the top part of Figure D.4.)[12]

Response and Nonresponse. An unfortunate feature of PSID data files released prior to ones containing the 1984 interviewing year — both the cross-year

---

[11]For data processing and in-house use we work with data structures that are more efficient than the rectangular ones described here. (See Chapter H.) However, most data users strongly prefer the conceptual simplicity of rectangular files to hierarchies or networks. For that reason, the PSID data are distributed in rectangular format.

[12]Prior to 1984, we began numbering the individual-level variables where the family-level variables ended. This resulted in changes each year in all individual-level variable numbers, a fact clearly evident if one examines any pair of pre-1984 documentation volumes. From 1984–1988 we numbered individual-level variables beginning with V30001. Although we have tried to keep the individual-level variables consistent since then, we haven't always succeeded. *To obtain the correct individual-level variable numbers, always consult the documentation volume corresponding to the most recent year of data on your data file.* (Family-level variables *are* numbered consistently from one year to the next so that variable numbers appearing in older documentation volumes *are* exactly the same as those appearing on recent data files).

family and cross-year family-individual files — was that information from families and individuals who had become nonresponse prior to the most recent year was omitted. Analysts interested in data from only the most recent year were not affected by this exclusion, but this was a potential problem for analysts wanting information for individuals or families who had become nonrespondents. To facilitate analysis involving persons and families who have become nonrespondents during the course of the study, the PSID now distributes two cross-year family-individual files (unless the user specifies otherwise) to all users acquiring the PSID cross-year family-individual data. The first, the **cross-year family-individual response file**, contains all of the individuals living in or currently associated with all of the families interviewed in the most recent year. This includes individuals who were part of a PSID family unit in 1968, were born to such a person since 1968, or joined a PSID family unit since 1968 other than through birth (e.g., marriage) and are in family units interviewed in the most recent year. All of these individuals have actual data in their family-level variables for all years back to 1968.[13] The individual-level variables in the records of the born-in or joiners contain a mixture of valid and zeroed-out data for the years before they entered the study. The **cross-year family-individual nonresponse file**, on the other hand, contains records for individuals who have been part of the study but are not part of a family unit interviewed in the most recent year. Concatenation of these two files produces a complete history of all individuals (and therefore all families) that have ever existed in the PSID.

Suppose, for example, that one wants to describe economic conditions of widows in the years following the death of a husband. There are between 40 and 50 widowhood events each year in the PSID, which, if pooled over all years of the study, provide a reasonably good-sized sample for such an analysis. The cross-year family-individual response file contains information on surviving widows only if the widows themselves survived in the PSID until the most recent interviewing year. Clearly it makes sense to use cases where the widow survived her husband's death for a year or two, even if the

---

[13]Even though a child born during the panel period or a nonsample member who joined a sample family during the panel period was not present in all years of the study, historical information about the family they entered has many uses. For example, one may wish to describe the poverty history of a child's mother prior to the child's birth and relate that history to the child's birth weight. Our treatment of family and individual-level information for these "joiners" prior to their point of birth or moving in is described in Chapter G.

CRUCIAL POINTS

widow did not survive until the most recent interviewing wave. Indeed, failing to include such cases may well impart a serious bias to the analysis sample.

The response and nonresponse files are structured identically, as indicated in Figure D.4. Whether or not a particular record contains actual data in any given year and whether or not the data pertains specifically to the given individual depends on a number of factors. The important factors are: whether the individual was classified as response in the given year; whether the individual is an original sample member, a born-in sample member, or a nonsample person who has joined the study since its start; and when the person has entered or left the study. Figure D.4 provides examples of how these factors influence the content of the records on the cross-year family-individual files. Valid data values appear on the family-level and individual-level portions of the nonresponse data record until the point of nonresponse, and zeroes are inserted in the family-level portion and in most of the individual-level portion of the data record after the point of nonresponse. An exception is that the time-invariant individual-level variables may have valid data values. Of particular interest among these variables is a series of summary nonresponse variables, documenting key aspects of the individual's history concerning nonresponse in this study.

The parallel structures of the response and nonresponse files make it easy to concatenate the two files into a single one. We strongly encourage analysts to do this, and indeed the frequency distributions printed in our published documentation since the 1984 wave reflect values obtained from a merged response-nonresponse file. Using a merged response-nonresponse file is proving to be a relatively easy task (although expensive because of its size). Selecting individuals who have valid data in a given year requires simply following the rules laid out later in this chapter.

5.3 *Crucial Variables*

There are several important variables to consider in using the files that combine family-level and individual-level variables:

1. Two variables, used in conjunction, constitute the unique identifiers for individuals in the study. These two variables are *1968 Family ID* (the identification number for the 1968 family unit that the individual lived in or has since become associated with) and *Person Number* (a variable that uniquely identifies the different people associated with a given 1968 family unit.

2. A variable that might easily win the title of "Most Useful PSID Variable" is cleverly disguised with the rather nondescript label of "Sequence Number". The **Sequence Number** variable is usually the second variable in the individual-level portion of the tape code each year. (For 1968, a combination of Person Number and 1968 Relationship to Head serves the same purpose). As each year's completed questionnaires arrive to be coded, one of the first processing tasks is to assign a **Sequence Number** to each individual who moved into, or out of, or continued to reside in an interviewed family. The "sequence" part of this variable's label refers to the order of a given family unit's members, ranking them on the basis of their relationship to Head in that family unit in that year. Ranges of code values for **Sequence Number** are meaningful, as are the distinct values for the variable. These ranges are as follows:

    01-20. Individuals in a given responding family unit at the time of the given year's interview (In the study's first year, 1968, the corresponding range was 001-019)

    51-59. Individuals in institutions (i.e.. college, the military, jail. or a hospital) at the time of the given year's interview

    71-80. Individuals who had moved out of the family unit or out of institutions but were not included in another responding family unit in the given year

    81-89. Individuals who were living in an interviewed family unit the prior year but died by the time of the given year's interview

    00. Individuals who had become nonresponse between the prior year and the given year

    Analysts wanting to analyze *only those individuals who were actually present in the family at the time of the interview* should subset their data file to include only those individuals with **Sequence Number** in the 01-20 range for that year (*Person Number*=001-019 and *Relation to Head*=01-09 for 1968). IS THIS

    One of the important functions of **Sequence Number** is in assisting in creating the current cross-year family file by subsetting the cross-year family-individual response file. To do so, include only individuals who, in the most recent year, had **Sequence Number** equal to 01 (Head).

3. Another very useful variable is *Relationship to Head*, an individual-level variable that describes the individual's relationship to the Head of his or her family unit in a given year. Since 1983, the code values for the *Relationship to Head* variable have spanned 34 categories.[14] Most analysts will never need to use the details built into this variable and will be content with the major categories of Head, Wife (presumed to be the legal spouse of the head), "Wife"

---

[14]Prior to 1983 the code for this variable consisted of only 10 categories. Code values also differ in 1968, the study's first year.

CRUCIAL POINTS

(i.e., a spouse-like female partner who has been co-habiting for over one year)[15], children[16], and perhaps grandchildren.

Values on the **Sequence Number** and *Relationship to Head* variables are year-specific, and these variables can be used to select the proper subsample for analysis.

6. Which Files for What Analysis

Table D.1 outlines the major types of analyses possible with PSID data files and notes which of the main PSID data files are most suitable for which types of analysis. As noted in Table D.1, while the cross-year family-individual response + nonresponse concatenated files can be used for all of these analyses, frequently a smaller file is adequate.

7. Using a Single Wave of Data

Even with a single wave of PSID data, a wide variety of analyses are possible. The analyses are all cross-sectional, but they could be based on either the most recent year or an earlier one and could use either individuals or families as the unit of analysis.

7.1 *Families in the Most Recent Year*

Analyses involving a single wave with the family unit as the unit of analysis can be performed using the PSID's cross-year family file. Examples of such cross-sectional analyses include:

1. **Description of families as of the most recent year.** If one wants to describe what fraction of *families* owned their homes in 1987, had 1986 family incomes above $50,000 (the interview gathers income information for the *prior* calendar

---

[15]The code categories for these relationships are:

10. Head in the given year, or the prior year's Head who moved out of this family unit but became nonresponse

20. Wife (presumed legal spouse) in the given year, or the prior year's Wife who moved out of this family unit but became nonresponse

22. "Wife"—female cohabitor who was living with Head in the prior year as well as the given year, or the prior year's "Wife" who moved out of this family unit but became nonresponse.

[16]The code values for children distinguish a number of different types of children. These code values consist of codes 30, 33, 35, 37, 38, and 83. See Chapter ? for details.

Table D.1

Types of Analyses and Appropriate PSID Data Files*

Unit of Analysis	Appropriate Main PSID Data File		
	Cross-Year Family	Cross-Year Family-Individual Response	Concatenated Cross-Year Family-Individual Response+Nonresponse
Cross-Sectional Analysis, Using Single Wave			
Families in Most Recent Year	Recommended	Recommended	Possible but Inefficient
Individuals in Most Recent Year	For Heads and Wives Only	Recommended	Possible but Inefficient
Families in Past Year	*	*	Recommended
Individuals in Past Year	*	*	Recommended
Cross-Sectional Analysis, Using Multiple Waves			
Families	**	*	Recommended
Individuals	**	*	Recommended
Longitudinal			
Individuals Living in All Years Being Analyzed, Including Most Recent Year	**	Recommended	Possible but Inefficient
Individuals Living in All Years Being Analyzed, Not Including Most Recent Year	**	**	Recommended
Individuals Who Were Either Head or Wife in Most Recent Year and Some Past Year(s)	**	Recommended	Possible but Inefficient
Histories of Still-Living Individuals in Most Recent Year	**	Recommended	Possible but Inefficient
Person-Years at Risk Years, Including Most Recent Year***	**	*	Recommended

NOTES:

*Not recommended because some cases would be lost.

**Not recommended because it would present a variety of problems.

***Event-history analysis with monthly dating of some types of events requires merging the main file with a separate, special file. For monthly dating of marriages and births, the Ego-Alter File must be merged with the Concatenated Cross-Year Family-Individual Response+Nonresponse files. Event-history analysis with monthly dating of employer/job changes over the year for Heads and/or Wives/"Wives" also requires use of a special file; that file is the Work History Supplemental File.

year), or had working husbands or wives, then the family unit is the appropriate unit of analysis and the cross-year family file is the shortest relevant file. (However, the family file is *not* appropriate for estimating what fraction of the *population of individuals* lived in owned homes or lived in families in which family incomes were above $50.000. These estimates require using the *individual*, not the *family*, as the unit of analysis.)[17]

2. **Analysis using the Head of the family unit as the unit of analysis.**
Examples here would include a regression analysis relating the Head's current wage rate to his or her education and labor force experience or a cross-sectional look at the economic conditions of women who head their own families. Note that both of these analyses require further subsetting of the data: the wage analysis must be restricted to family units in which the Head was in the labor force, and the latter analysis must be restricted to family units in which the Head was a woman. Further subsetting—e.g.. by age or race—is also possible. *Any subset of a representative probability sample such as the PSID's is a representative sample of that subset of the population.*

Selection bias problems may arise due to important differences between the type of people who become Heads and those who do not. Thus, performing the wage rate analysis on the group of male *Heads* who were between the ages of 18 and 25 is not likely to yield the same results as a parallel analysis on all male *individuals* in that age range. because most men have not left home to head their own families by age 18. (In fact, the median age at which men leave home is 23; for women it is 21.) The analyst who does not want to restrict the analysis to Heads cannot use the cross-year family file. Measures for the variables of interest may be included in the individual-level variables, and these are available only on the cross-year family-individual response file. Other problems can arise because of the source of the information. PSID information is reported by the family unit Head. and is likely to be much less reliable for persons who are not Heads than if it were self-reported. The analyst concentrating on school and labor market activities of young people may be better off with a different data set, such as the National Longitudinal Surveys of Young Men and Women.

3. **Analysis using currently-married women as the unit of analysis.** An obvious example of this would be a study of the labor force participation decisions of married women. Other than the need to restrict the sample of family units to those in which a Wife is present and to select variables from the file that refer specifically to the Wife, this kind of analysis is quite

---

[17]There are two ways to do this. The first is to use the cross-year family-individual response file, which contains one data record for each of the study's *individuals in the most recent wave.* Alternatively, one could use the cross-year family file but weight each family by the number of individuals in it. (See the description of weighting in Chapter E for details.)

straightforward using the cross-year family file and is representative of all married women.[18]

4. **Analysis of women who were either Heads of their own family units or Wives in the most recent year.** Almost all women over the age of 25 fall into one of these two groups. This analysis can also be performed with the cross-year family file or the cross-year family-individual response file. but using either is tricky because one must select either the Head's variables or the Wife's variables depending on the status of the woman—is she a female Head or a Wife? Since we have not always gathered the same information for the Head and Wife. this kind of analysis must be restricted to variables that are identical for both. From 1979 onward, the type of information available for Wives has been much more similar to that for Heads than was the case in the early years of the study. However. information about Heads is usually self-reported. while information about Wives is generally reported by their spouses.

The cross-year family-individual response file can also be used for a single-year cross-sectional analysis of families in the most recent year. This is done by including only cases where the individual's **Sequence Number** in the most recent year equals 01 and analyzing data from the family portion of the data record.

7.2 *Individuals in the Most Recent Year*

The shortest file suited to analysis of the most recent year's sample of individuals is the cross-year family-individual response file. This is. however, a large file. Fortunately, it is a fairly simple matter to select a subset of family-level and individual-level variables for the most recent interviewing year from that file. With 1987 as the most recent year, this subsetting would select the family-level variables V13701-V14737 and the individual-level variables V30554-V30589. An analyst interested in cross-sectional data for the most recent year but wishing to address a question requiring the family-individual file (e.g., about the distribution of family income for all *individuals* rather than for all *families*) would need no other information from the PSID data than that contained in this subset of variables. For an analysis of a cross-section of *individuals living in family units in the most recent year*, select cases with **Sequence Number** in the most recent year in the 01–20 range.

---

[18]The assumption that the analyst is content with data from a single year greatly simplifies the tasks in this case. A look at the work patterns of Wives over more than a single year runs into problems of family composition change—women who are Wives in a given year may be Heads in the next in the event of a husband's death, a divorce, or a separation with no immediate remarriage. These complications are discussed later in this chapter.

Analysts of the most recent year of family- and individual-level information face the same questions about the construction of variables, sampling errors and split sample strategies as those faced by analysts of the most recent year of family information. *The decision about whether or not to use weights is a bit trickier, though, than when the family unit is the unit of analysis.* Should the data be weighted if the individual is the unit of analysis? Many of the arguments for and against weighting are the same whether the analysis is family-level or individual-level. However, for reasons explained in Chapter E, some of the individuals who are members of PSID families do not belong to the sample because they have "married into" the study. Their information is valuable because it is part of the family situation of the individuals who *do* belong to the sample, and they can be used in *family-level* analysis for any year of the study. As *individuals*, however, they do not belong to the sample and therefore receive an individual weight value of zero. An analyst who uses the individual weight will automatically eliminate these individuals. An analyst who chooses not to use weights, in the belief that he or she has a properly specified model, will probably wish to include these individuals in the analysis. Indeed, methodological work suggests that these zero-weight individuals are in many respects indistinguishable from sample individuals.

7.3 *Families or Individuals in a Past Year*

The cross-year family-individual response + nonresponse concatenated files can be used to obtain cross-sectional information from any year prior to the most recent one. (This cannot be done with the cross-year family files.) Suppose you want cross-sectional family information for 1975. Since there is only one Head in each 1975 family, you could concatenate the response and nonresponse family-individual files, then subset this concatenated file by selecting all of the 1975 family-level variables for individuals who had a **Sequence Number** equal to 01.[19] (The reason that this cannot be done with the cross-year family file will become apparent in the next section.) If you wish to weight these data, use the 1975 family weight. (Details of weighting are explained in Chapter E.)

---

[19]Concatenation allows inclusion of both 1975 family unit Heads who are 1987 Heads (from the response file) plus those 1975 Heads who died or became nonresponse for other reasons after 1975 (from the nonresponse file). The 1968–1975 cross-year file includes all Heads interviewed in 1975, regardless of their future status with the study. Use of only one of the files would result in the exclusion of legitimate 1975 Heads (and family-level data) from the subset.

For a cross-sectional analysis of *all individuals in family units in a prior year.* concatenate the cross-year family-individual response and nonresponse files, then select cases where the individual's **Sequence Number** in the prior year is in the 1–20 range.

8. Using Multiple Waves of Data

Because the cross-year family-individual files allow more direct means of addressing the complications of family composition change than do the cross-year family files, *we urge the use of the cross-year family-individual files rather than the cross-year family file when undertaking analysis of multiple waves.* The cross-year family-individual file may be too large for some computer installations, rendering this impossible. However, where the cross-year family-individual files can be used, the higher initial processing costs are more than offset by the ease of understanding and properly managing the data.

8.1 *Cross-Sectional Analyses Using More than One Wave*

The key to conducting analysis that requires more than one wave of the PSID data is the use of the **Sequence Number** variable and the—more descriptively named—*Relationship to Head* variable, described in brief earlier in this chapter and in more detail later on. As an example, for a cross-sectional analysis of the 1978 average hourly earnings of male family unit Heads (as reported in 1979) that uses information on labor market experience reported retrospectively in the 1976 interview: concatenate the response and nonresponse files and select individuals who had **Sequence Number** 01 (current-year Head) in *both* 1979 and 1976. (Failure to make the 1976 restriction will result in the inclusion of cases where, for example, an individual was a son in 1976 and Head in 1979, with the 1976 facts being reported by and about his father.)[20]

8.2 *Longitudinal Analysis*

Concatenation of the cross-year family-individual response and nonresponse files offers the widest range of possibilities for longitudinal analysis using PSID data.

---

[20]This procedure will eliminate young adults who split from their parental homes between 1977 and 1979. These young adults could be included in the analysis if the analyst picked up their retrospective reports of labor experience as reported in the year the young adults set up their own households (see last section of Chapter F for details).

CRUCIAL POINTS

However, the resulting file is extremely large, and in many cases a smaller file can be used. In some cases, the smaller file is the one best suited to the analyses, in other cases the smaller file must be handled very carefully to ensure that it would produce the same results that the larger, better suited PSID file would. We briefly outline related cautions below, and encourage the reading of subsequent chapters for the finer details about the necessary procedures. **Sequence Number** and *Relationship to Head* play key roles in longitudinal analysis.

Analyses With the Family-Individual Files. The following examples give an idea of the types of longitudinal analyses that can be done with the cross-year family-individual files:

1. Analysis of change in the annual earnings of male family unit Heads between, say, 1984 and 1986: Using only the cross-year family-individual response file, select cases where the individual is a male and the individual's **Sequence Number** in *both* 1985 and 1987 equals 01, and use variables from the family portion of the 1985 and 1987 data records that refer to the Head (e.g., V14671 = Annual Earnings of the Head in Calendar Year 1986, and V11397 = Annual Earnings of the Head in Calendar Year 1984).[21]

2. Longitudinal analysis of the 1976–1984 labor supply patterns of adult women who were between the ages of 25 and 50 in 1976: Select all individuals who were female and Heads (*Relationship to Head* code value 1 or 10, depending on the year) or Wives/"Wives" (code value 2, 20 or 22) in each of the interviewing years between 1976 and 1985 and who have a **Sequence Number** in the 01–20 range in all of those years.[22] Form a series of conditional statements selecting the Head's variables from the family-level variables for the years in which the woman headed her own family unit and selecting the Wife's/"Wife's"

---

[21]The interview conducted in year t gathers information at the time of the interview and for the calendar year t-1. In this case, one would have to use the 1987 and 1985 data records to obtain information on calendar years 1986 and 1984. Don't use data from 1986 (1985 income year) unless you check that 1986 **Sequence Number** also equals 01. Note that the response file alone is used because the end year of the analysis is the most recent year for which data are released. See example 2 for an illustration of earlier end years.

[22]The 1985 record is needed because it contains information for calendar year 1984. The 1976 record contains a variable for age in 1976. Note that this procedure will exclude the small number of women in this age range who were neither Head nor Wife/"Wife" in at least one of these years. It is a simple matter to select the women who would be excluded and compare them with the women who were included. If necessary, one could model this "selection" process as well. Since work hour information is collected on *all* individuals in the family, these "other" women could be included in the analysis if it does not require much additional information about them.

variables from the family-level variable for the years in which the woman was a Wife or "Wife". If you weight your data, use the 1985 individual weight.

3. Analysis of the persistence of poverty for the population during the 1970s: Concatenate response and nonresponse, then select all individuals with a **Sequence Number** between 01 and 20 for each of the years 1971 through 1980.[23] Use the family income variable in the family-level variables and compare it with the "income needs" variable, which is also a family-level variable. (Note that the needs variable should be adjusted for inflation each year and further adjusted to make it comparable with the "official" poverty standard. See Chapters D and F, and the third example in Chapter J.)

4. Analysis of the effect of parental income on the occupational attainment of sons heading households: Use the cross-year family-individual response file only (because a response in 1987 is desired), select individuals who had a *Relationship to Head* code equal to three (son or daughter of Head) in 1968 and equal to 10 (family unit Head) in 1987, a 1987 **Sequence Number** in the range 01–20, and a *Sex of Individual* code in 1968 (or in 1987) equal to 01 (male). An age restriction (e.g., between 25 and 30 years in 1987) should probably also be imposed on this group to avoid selection bias problems associated with becoming a Head. The 1968 family income variable for this group is reported by their parents and refers to the parental family's income for calendar year 1967. The 1987 occupational information is reported by and about the son who is Head of his own family in 1987. If weighting, use the 1987 individual-level weight variable.

Other subsets (such as pooling data over several years or matching ex-husbands and ex-wives) are more complicated, but follow the same general principles and include the same weighting procedures. Several of them are detailed in Chapter H.

Analyses with the Family File. Although the cross-year family-individual files are more flexible and better suited to longitudinal analysis, the smaller cross-year family file offers the advantage of considerably fewer and somewhat shorter records. This file allows longitudinal analysis in which the unit of interest is a family that maintains the same head or includes a female who remains a primary adult (Head or Wife/"Wife") over the period being analyzed. Anything that can be done with this file

---

[23]While this selection procedure assures that available data for all of the years will be analyzed, it does have the disadvantage that children born during the observation period are excluded from the analysis. Similarly, individuals who left the sample or died during this period (and hence became nonresponse) are also excluded.

CRUCIAL POINTS

can also be done with the cross-year family-individual response file,[24] probably with less drastic data management procedures. The advantage is the smaller size of this file.

Imposing proper restrictions on the cross-year family file records for a longitudinal analysis can be complicated. These complications are fairly minor for, say, an analysis of change in the annual earnings of the family Head between 1984 and 1986 (Example 1 in the previous section). To ensure that the same person headed the family between 1985 and 1987, the sample must be restricted to families in which the Head of the family did not change between 1985 and 1987. This can be done with a family-level variable in the family data record entitled "Family Composition Change." Code values of 00, 01, and 02 indicate that the Head did not change from the prior year. Thus, the sample must be restricted to cases where the 1985, 1986 and 1987 family composition change variables were in the range from 00 to 02.[25]

It is more difficult to obtain cross-sectional information about families from a year prior to the most recent one. Several 1987 family units may have come from the "same" 1986 (prior-year) family. If, for example, a son left home between 1986 and 1987, then the son's 1986 family-level variables will be identical to his parents' 1986 family-level variables. Similar duplication of records results from divorces when both of the ex-spouses are interviewed subsequent to their divorce. *Unless the sample is restricted to eliminate duplication of records reflecting the same prior-year family, the analysis will yield nonsensical results.*[26] In some cases this means restricting the included cases to families in which the Head remained unchanged between the prior year of interest and the most recent year. In cases where families are of interest if the same woman

---

[24]To analyze families with the same family unit Head over a given period using the cross-year family-individual files, for example, one merely selects individuals whose **Sequence Number** equals 01 over the given period.

[25]It is not necessary to restrict family composition change in 1984 in this case because it applies to changes between 1983 and 1984. The annual earnings information reported by the 1985 Head applies to calendar year 1984 and is not affected by the possibility that the 1985 Head may not have been Head of a family in part of 1984.

[26]Furthermore, this restricted group will not really be a complete cross-section for that prior year. Suppose that a family in the crucial prior year was headed by a woman. If she married between that time and the current year, then the person designated as Head of the family has changed and that family will not be part of a cross-sectional analysis performed on a sample restricted to having the same Head each year. The cross-year family-individual files solve this problem neatly because the woman was Head of the family in the prior year and that relationship-to-Head status is the one used to select the prior cross-section.

remained as one of the primary adults, it means restricting the cases to family units with either the Head and the Wife/"Wife" (if initially present) remaining the same, the prior-year Head becoming a Wife/"Wife," or the prior-year Wife/"Wife" becoming a Head. In each of these cases, it is clear from the family-level variables alone when individuals change from Head to Wife/"Wife" or the reverse.

Still more complicated are analyses such as Example 2 above in which the analysis topic is the labor supply of women who were Heads or Wives/"Wives" between 1976 and 1985. With the cross-year family-individual response and nonresponse files, it is a simple matter to select all individuals who were either Wives/"Wives" or female Heads in each of those years. With the cross-year family file it is almost impossible to follow women who change from Head to Wife/"Wife" and vice versa. In addition, some women who are no longer in the study would be missed. (They are nonresponse, so they would not be included in the cross-year family file.)

We hope that this summary of crucial points has not been so ponderous as to discourage further reading. The remaining chapters in this Guide need not be read in sequence; we have tried to organize them more or less as steps in the collection and processing of the data. They provide crucial, if oftentimes very tedious, details that help ensure proper use of the data set. We do, however, encourage a complete reading of the rest of the Guide, especially before you decide to call us with a question!

FIGURE D.1

Structure of a Single-Year Family File
(1987 Family Data)

Case
Numbers

1st
Case
(1st Family
Interviewed
in 1987)

> 1037 Family-level Variables
> Gathered in 1987
> For the First Family
> Interviewed in 1987

2nd
Case
(2nd Family
Interviewed
in 1987)

> 1037 Family-level Variables
> Gathered in 1987
> For the Second Family
> Interviewed in 1987

.
.
.

.
.
.

7061st
Case
(7061st Family
Interviewed
in 1987)

> 1037 Family-level Variables
> Gathered in 1987
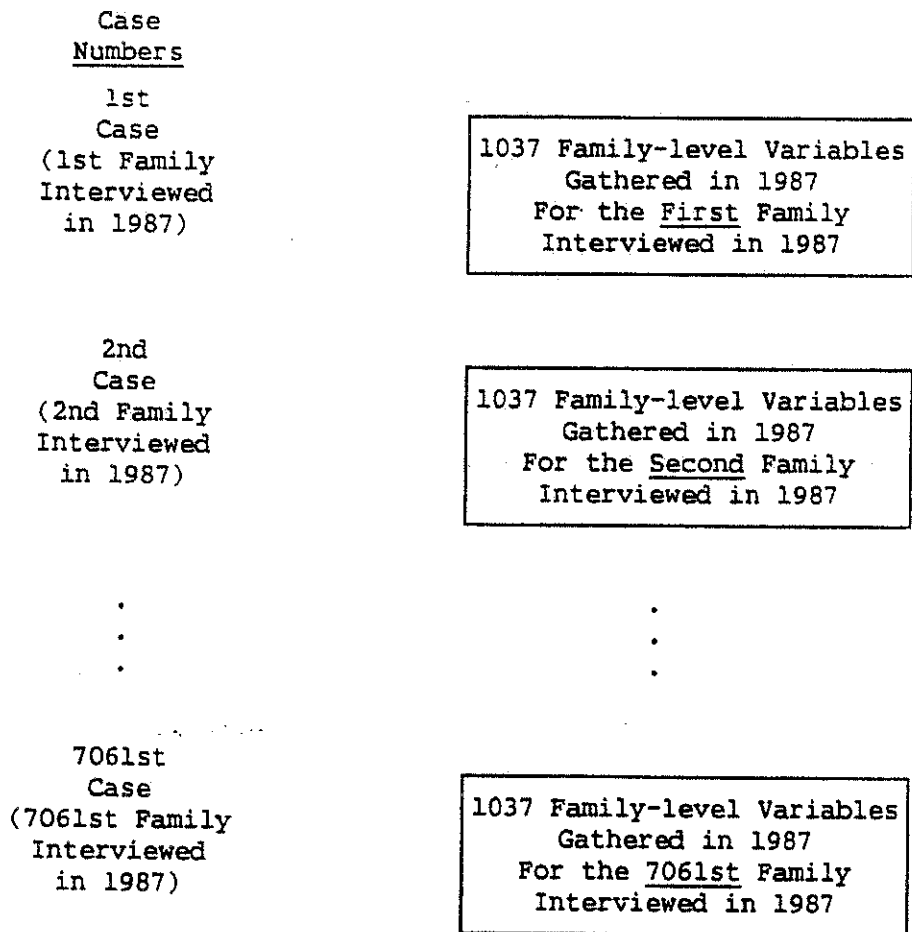> For the 7061st Family
> Interviewed in 1987

FIGURE D.2
Structure of 1968-1987 Cross-Year Family File*

Ordering of Data in Each Record	1968 Family-level Variables	1969 Family-level Variables	1970-1986 Family-level Variables	1987 Family-level Variables

1st Case (First Family Interviewed in 1987)	Data for 1968 Family Unit of 1987 Head	Data for 1969 Family Unit of 1987 Head	Data for 1970-1986 Family Units of 1987 Head	Data for 1987 Family Unit

2nd Case (Second Family Interviewed in 1987)	Data for 1968 Family Unit of 1987 Head	Data for 1969 Family Unit of 1987 Head	Data for 1970-1986 Family Units of 1987 Head	Data for 1987 Family Unit

.
.
.

7061st Case (7061st Family Interviewed in 1987)	Data for 1968 Family Unit of 1987 Head	Data for 1969 Family Unit of 1987 Head	Data for 1970-1986 Family Units of 1987 Head	Data for 1987 Family Unit

*Note: The size parameters for this file are N=7,061 LRECL=26,023 #Megabytes=184.

CRUCIAL POINTS

FIGURE D.3
Structure of a Single-Year Family-Individual File
(1987 Family-Individual Data)

Ordering of Data in Each Record	1987 Family-level Data	1987 Individual-level Data

1st Case (First Individual)	1987 Family-level Data for 1987 Family Unit of First Individual	1987 Individual-level Data for First Individual

2nd Case (Second Individual)	1987 Family-level Data for 1987 Family Unit of Second Individual	1987 Individual-level Data for Second Individual

.
.
.

20,487th Case (20,487th Individual)	1987 Family-level Data for 1987 Family Unit of 20,487th Individual	1987 Individual-level Data for 20,487th Individual

Note: This figure omits the time-invariant individual-level summary variables that appear at the end of the individual data record.

FIGURE 4

Structure and Examples of Content of Records on Cross-Year Family-Individual Files
(1968-1987 Cross-Year Family-Individual Response Data or Nonresponse Data)+

Structure of Each Record

Ordering of Data in Each Record	Family-level Data				Individual-level Data				Summary Variables
	1968 Data	1969 Data	1970-1986 Data	1987 Data	1968 Data	1969 Data	1970-1986 Data	1987 Data	
Examples of Content of Records on Response File									
Response Sample Member in PSID since 1968	Own Data	Own Data	Own Data	Own Data	Own Data	Own Data	Own Data	Own Data	Own Data
Response Sample Member entering PSID in 1969	1968 Data of Head of 1969 Family Unit	Own Data	Own Data	Own Data	Data Filled with Zeroes	Own Data	Own Data	Own Data	Own Data
Response Non-sample Member entering PSID in 1969	1968 Data of Head of 1969 Family Unit	Own Data	Own Data	Own Data	Data Filled with Zeroes	Own Data	Own Data	Own Data	Own Data
Examples of Content of Records on Nonresponse File									
Nonresponse Sample Member leaving PSID in 1969	Own Data	Data Filled with Zeroes	Data Filled with Zeroes	Data Filled with Zeroes	Own Data	Data Filled with Zeroes	Data Filled with Zeroes	Data Filled with Zeroes	Data Up to Date as of 1969
Nonresponse Non-sample Member entering PSID in 1969 and leaving in 1987	1968 Data of Head of 1969 Family Unit	Own Data	Own Data	Data Filled with Zeroes	Data Filled with Zeroes	Own Data	Own Data	Data Filled with Zeroes	Data Up to Date as of 1986

*Note: Cases involving response sample or non-sample members appear on the Response File, whereas cases involving nonresponse sample or non-sample members are on the Nonresponse File. The size parameters for the Response File are N=20,487 LRECL=27,356 #Megabytes=561. The size parameters for the Nonresponse File are N=16,093 LRECL=27,356 #Megabytes=440.

CONTENTS FOR
APPENDIX 1: A BRIEF HISTORY OF THE PSID

April 26, 1990

Appendix 1

A Brief History of the PSID

1. Origins of the Study

As part of Lyndon Johnson's War on Poverty, the Office of Economic Opportunity (OEO) directed the Census Bureau to mount a nationwide assment of the extent to which the War on Poverty had affected people's economic well-being. This Census Bureau study, called the Survey of Economic Opportunity, took interviews at 30,000 households, first in 1966, and again in 1967.

There was interest in continuing this national study of economic well-being, but an annual interview with 30,000 households was too costly to maintain over time. The OEO asked researchers at the Survey Research Center (SRC), University of Michigan, if they could take up the study and continue to interview a subsample of about 2,000 of the Survey of Economic Opportunity families for several more years. James Morgan, who became the new study's director at The University of Michigan, suggested adding a fresh cross-section of households from the SRC national sampling frame so that the new study would be representative of the population of the United States, and include non-poor as well as poor households. It was also decided to follow, and keep as part of the sample. members of the families who moved away from their original households to set up new households. such as children who came of age during the study. In this way, the sample could remain representative of the nation's families and individuals over time. The new study came to be known as the Panel Study of Income Dynamics. It began interviewing in 1968, with an initial sample of 1,872 households from the Survey of Economic Opportunity and 2,930 households drawn from the SRC sampling frame.

The first annual PSID wave in 1968 included interviews with a total of 4,802 families or households across 40 states. As of 1989, the study collected its 22nd annual interview, with about 7,100 family units from 50 states and some areas outside the continental U.S. Ten volumes of analysis and innumerable articles and papers on the study have been published to date, as well as 16 volumes which document the 1968–1987 data.

2. Procedures

The study has followed the 1968 original panel families, and also all "split-offs" or members of the 1968 families who left home to establish separate households. Each year. one primary adult is interviewed—usually the man adultle Head. if there is one—in all families containing a member of one of the 1968 families. The person being interviewed provides information about him/herself and about all other family unit members. This procedure produces an unbiased sample of families each year. and thus the panel continues to be representative with respect to its basic sampling design. In 1976 and 1985, the study conducted interviews with all Wives/"Wives" in the sample as well as the Heads. Since 1976, an effort has been made to collect the same detail of information for Wives/"Wives" as for Heads regarding income, employment history, education. etc. The interviews conducted in 1985 also gathered extensive marital. fertility. and educational information about both Heads and Wives/"Wives."

In 1973, to reduce costs, the study began taking the majority of interviews by telephone rather than in person. In-person interviews are now performed only for respondents who do not have telephones, or who have special circumstances which make a telephone interview unfeasible. The interview averaged about one hour in length when it was conducted in the households; the telephone interviews range from an average of 20 to 30 minutes in length.

3. Funding

In the course of its 22-year history at The University of Michigan, the study has been funded principally by a collection of federal agencies, including the Office of Economic Opportunity, the Departments of Health, Education and Welfare (now Health and Human Services), Labor, and Agriculture; the National Institute of Child Health and Human Development (NICHD); and the National Institute on Aging (NIA). The Sloan, Rockefeller and Ford foundations provided important supplementary grants to the PSID. Since 1984, the National Science Foundation (NSF) has been the principal sponsor of the study, with substantial continuing support provided by the Office of the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services. Funding from NSF is secured through the study's 24th wave (1991).

4. Board of Overseers

Since 1982, development and refinement of the study have been guided by a Board of Overseers. The board was created by the National Science Foundation to insure input into the study from the national community of scholars, researchers and policy makers. This board oversees the design, content and future course of the study. Through autumn 1989, Board members included:

Current Members

Lee Lillard (Chair), The Rand Corporation
Ralph Folsom, Research Triangle Institute
Michael Grossman, CUNY
Arne Kalleberg, University of North Carolina
Rod Little, University of California
Daniel McFadden, Massachusetts Institute of Technology
Sara McLanahan, University of Wisconsin
Joseph Newhouse, The Rand Corporation
Harriet Presser, University of Maryland
Aage Sorensen, Harvard University
Nancy Tuma, Stanford University
David Wise, Harvard University

Previous Members

Orley Ashenfelter, Princeton University
Mary Jo Bane, Harvard University
Burt Barnow, ICF Corporation
Larry Bumpass, University of Wisconsin
William Butz, Bureau of the Census
Gary Chamberlain, University of Wisconsin
Andrew Cherlin, Johns Hopkins University
Glen Elder, University of North Carolina
David Ellwood, Harvard University
David Featherman, SSRC
Zvi Griliches, Harvard University
James Heckman, Yale University
Robert Michael, National Opinion Research Center
Guy Orcutt, Yale University
Kenneth Prewitt, Social Science Research Council
John Quigley, University of California
Lee Rainwater, Harvard University
Isabel Sawhill, The Urban Institute
Carol Weiss, Harvard University
Finis Welch, Unicon Research Corporation
Charles Westoff, Princeton University
Kenneth Wolpin, University of Minnesota

BRIEF HISTORY

The current project directors of the PSID at the Institute for Social Research are James Morgan. Greg Duncan, and Martha Hill.