A PANEL STUDY OF INCOME DYNAMICS: PROCEDURES AND CODEBOOKS

(DOCUMENTATION)

GUIDE TO THE 1993 INTERVIEWING YEAR

PROCEDURES

WAVE XXVI
A SUPPLEMENT

Survey Research Center
INSTITUTE FOR SOCIAL RESEARCH
THE UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN

1998

Sandra Hofferth, Frank P. Stafford, Wei-Jun J. Yeung, Greg J. Duncan, Martha S. Hill, James Lepkowski and James N. Morgan. PANEL STUDY OF INCOME DYNAMICS 1968–1993 (Waves I-XXVI) [computer file]. Ann Arbor, MI: Survey Research Center [produced]. 1998. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]. 1998.

Preface

This file documents the twenty-sixth wave of data collected by the Panel Study of Income Dynamics, interviews taken in 1993 about income received in 1992. Volumes I and II of *A Panel Study of Income Dynamics: 1968-1972 Interviewing Years (Waves I-V)* contain codes, indexes, available data, questionnaires and procedures specific to our first five years of data collection (1968-1972). These documents also describe the early history of the study and some of the basic procedures that are common to all twenty-six years of interviewing. Twenty-one supplemental series of documents, including this one, cover procedures, codes and questionnaires for Waves VI-XXVI. Ten volumes of findings entitled *Five Thousand American Families—Patterns of Economic Progress* are available, covering ten years of PSID findings from 1969 through 1978. *Years of Poverty, Years of Plenty* by Greg J. Duncan and colleagues, based on PSID data, is also available. This book is an accessible summary of findings regarding poverty and employment dynamics through the late 1970s. A very helpful guide for data users is Martha S. Hill's *The Panel Study of Income Dynamics: A User's Guide*. This book is the second in Sage Publications' series of guides to major social science databases.

All documentation for the PSID is available from the Inter-university Consortium for Political and Social Research, P.O. Box 1248, Ann Arbor, MI, 48106.

Staff

Greg J. Duncan, Martha S. Hill, and James M. Lepkowski were the study's principal researchers during the data collection and family composition editing phases of Wave XXVI. Charles Brown was in charge of labor market content. William Shay was the project manager and oversaw daily operations, and Mary Wreford was the administrative manager and oversaw financial operations during this period. Margaret Gunnesch and Sarah Olson provided secretarial support. Thomas Gonzales supervised preparation of the computerized interviewing application, and oversaw the data collection effort, with assistance from Ron Amos, Bonnie Bittman, Barbara Browne, Priscilla Hildebrandt, and Anne Sears. Charles Stallman developed new systems for extracting data from the automated interviewing application, and Marita Servais oversaw development of a new automated data editing application. Tecla Loup oversaw family composition editing, with assistance from Ron Amos, Barbara Browne, and Anne Sears.

Sandra Hofferth and Frank Stafford were the study's principal researchers during the data processing and file release phases of Wave XXVI. William Shay was the project manager and oversaw daily operations, and Nancy Ditmar was the administrative manager and oversaw financial operations during this period. Margaret Gunnesch and Kaarin Stahl provided secretarial support. Randy Herbison, Tecla Loup, Marita Servais, William Shay, and Jean Yeung each supervised a part of file processing and release, with assistance from Ron Amos, Robert Bressan, Barbara Browne, Margaret Hoad, Anne Sears, Charles Stallman, and Yun-San Tsai. New systems for data documentation and delivery, including a web site, were developed by Robert Bressan, Randy Herbison, and Marita Servais. Tecla Loup oversaw compilation of the documentation, with the assistance of Barbara Browne, Anita Ernst, Marita Servais, and Nancy Wiegand. Terry Adams and Marita Servais supervised sensitive and other special data file preparation, with support from Barbara Browne and Kaarin Stahl. Jean Yeung and Marita Servais oversaw user services, including a bibliography of publications that use Panel Study data, with assistance from Howell Burnell and Sarah Olson.

Users who wish to communicate with the study staff regarding questions about PSID data content should contact us by e-mail: PSID_staff@umich.edu.

Table of Contents

SECTION I
PROCEDURES FOR THE 1993 INTERVIEWING YEAR

Part 1: The 1993 Questionnaire, Recontact Sample, Interviewing Procedures, Data Processing, Occupation Codes, Data Quality, Independent Part Samples, Weights

The 1993 Questionnaire

The 1993 questionnaire included a large question supplement about the time and money burden caused by poor health, released as a separate file.

We continued to ask about marriages, divorces, adopted and natural children. Updating questions were again asked to account for new children and marital changes for those whose data were collected from 1985 through 1992; new Heads and Wives/"Wives"[1] were asked about all of their children and first and last marriages. Employment event dating questions for 1993 continued with the design instituted in 1988 asking about spells with employers, instead of the position-oriented approach used from 1984 through 1987. Only information about prior-year employers was asked in detail. The food stamp, mortgage detail and other housing questions, including housing-related services for Heads and Wives/"Wives" age 50 or older, were continued. A series of about twenty health questions, begun in 1992, was retained for 1993 for each Head, Wife/"Wife" or other family member age 55 or older, although eliminated for those who had left the family.

The 1993 questionnaire is available on the PSID website in Adobe format.

The 1993 Recontact Sample

Since the PSID began in 1968, following rules stated that those eligible for the next wave of interviewing would include only persons present in the prior year. Thus, individuals who refused or were lost in one wave were never followed in later waves. Only if a nonresponse person subsequently moved into a currently responding family would he or she be followed in the future. While the reappearance of some formerly nonresponse individuals occurred in each wave, it was a relatively rare event; few were reinstated in the study when compared with the overall number of nonrespondents.

The 1993 wave included a large recontact effort of core cases. We identified sample individuals who were last present in a PSID family in 1991 or earlier. From this pool, we selected those who shared an original family identifier (1968 ID Number) with someone who was still responding in 1992. The pool was then subset to 1,419 individuals for 1993 recontact, the remainder to form part of the 1994 recontact sample. In order to make the most of interviewer tracking efforts, the 1993 subset included all of the sample nonresponse individuals within a given original family identifier (1968 Interview Number). The selection of original identifiers was geared toward recovering individuals who had not been followed when they initially became nonresponse because they were under age 18. If more than one individual was last present in the study in the same family, then we counted the group as one recontact family. If the recontact family was successfully interviewed in 1993, anyone who was no longer present was followed as a splitoff. Response rates for the 1993 recontact effort are discussed below.

---

[1] The PSID uses the term "Wife" (in quotes) in referring to long-term female cohabitors.

Nonsample Elderly Group

In 1990, a group of elderly individuals who had become nonresponse between 1985 and 1989 and who were expected to have achieved 65 years of age by 1990 were selected for a special recontact effort. This group included both sample and nonsample individuals. We planned not only to continue to attempt interviews with this special group, but also to follow all nonsample persons in currently-interviewed PSID families as these persons turned age 65 and either survived all of the sample individuals in their family or moved out. This plan was continued for 1993.

Computer Assisted Telephone Interviewing and Interviewing Procedures

For the 1993 wave, the PSID converted the questionnaire to electronic form. That is, respondents were interviewed using a computer application programmed with the 1993 question sequences. This programmed application handled skip logic and permitted the entry of valid codes only, thus eliminating these sorts of recording errors. For further information about interviewer recording error, see Lepkowski et al. (1995).

The software used for data collection was written by Surveycraft Pty. Ltd. of Victoria, Australia. For more information about the company, see their website at http://www.surveycraft.com. Using this software, Survey Research Center programmers wrote and tested the application with feedback from PSID staff.

Interviews were mainly conducted by SRC's Telephone Facility interviewers.

The average length of the interview for all families was 34.5 minutes (Table 1). Each respondent was paid $15.00 for the interview and an additional $5 for returning an address correction postcard in January 1993.

Table 1
AVERAGE LENGTH OF INTERVIEW

Year	Number of Interviews	Average Length In Minutes
1968	4,802	63.1
1969	4,460	61.8
1970	4,645	60.5
1971	4,840	59.1
1972	5,060	66.2
1973	5,285	20.1
1974	5,517	23.1
1975	5,725	26.9
1976	5,862	48.2*
1977	6,007	25.0
1978	6,154	26.9
1979	6,373	28.1
1980	6,533	29.0
1981	6,620	26.5
1982	6,742	20.8
1983	6,852	23.8
1984	6,918	34.7
1985	7,032	49.9*
1986	7,018	34.9
1987	7,061	29.5
1988	7,114	37.0
1989	7,114	35.9
1990	9,371	37.4**
1991	9,363	29.5**

	Year	Number of Interviews	Average Length In Minutes
	1992	9,792	31.1**
	1993	9,977	34.5**

*   Includes both Head's and Wife's interviews.
**   Includes core and Latino, as well as any recontacts.

We expanded our following rules for 1993. In previous waves, we had never followed as splitoffs sample members under 18 years of age if they left the family (usually with a nonsample parent). But in 1993, we began to follow these younger persons and attempt to interview an adult in the new family group. As a corollary to this alteration of a long-established tradition, our family composition rules changed. PSID families had always included a sample member as the Head or Wife/"Wife" of the family unit, but this became impossible in some cases where we followed the underage sample members. Therefore, although all families contain at least one sample member, both the Head and the Wife/"Wife" may be nonsample.

Response Rates

Core Sample Response Rates. Excluding recontacts, interviews were taken with 7,438 heads of families out of 7,929 possible, for an overall response rate of 93.8%. Subtracting from the base 74 respondents who had died since the last interview, had moved into institutions that precluded an interview, were too ill to be interviewed, or had rejoined sample ex-spouses raises the response rate to 94.7%. The interview total includes interviews with 250 splitoffs (out of a total of 368) with a response rate of 67.9%.

We selected 950 eligible recontact families in 1993, of whom 426 were successfully interviewed for an overall response rate of 44.8%. This base included 35 families who were deceased, institutionalized or otherwise unable to co-operate, and 98 more who reunited with another core family. Removing these groups from the recontact base reduces it to 817 and revises the recontact response rate to 52.1%. In addition, nine recontact splitoffs were successfully interviewed from a pool of 19 possible for a response rate of 47.4%.

Latino Sample Response Rates. A total of 2,104 heads of families responded out of 2,510 possible, for an overall response rate of 83.8%. Twenty families had died since the 1992 interview, had moved into institutions that precluded an interview, were too ill to be interviewed, or had joined the families of other sample members. Excluding these from the base raises the response rate to 84.5%. This total includes interviews with 132 splitoffs (of a possible 242) for a splitoff response rate of 54.5%. No Latino sample families were selected for recontact in 1993.

Data Processing

We used an electronic data collection system for the first time in 1993, thus altering our data cleaning procedures significantly. Data cleaning is discussed more fully below in Parts 3 and 4. PSID files are no longer released in OSIRIS format. Our data processing is done using SAS, and the release files are in ASCII format accompanied by SAS and SPSS statements.

Occupation Codes

We continue to use the 1970 Census three-digit occupation and industry codes for the current main jobs of employed Heads and Wives/"Wives." They are also used for the most recent jobs held by Heads and Wives/"Wives" who are not currently working, and in coding employment histories and extra or second job questions. For comparability with past data, one-digit occupation codes are used to code Head's first job and Head's father's occupation, since these data items were collected only for new Heads in 1993.

Data Quality

Tables 2a and 2b show response rates based on original sample individuals, annually and cumulatively. Since it would be impossible to know how many individuals were eligible but did not respond in 1968 for

the core sample or in 1990 for the Latino sample, we used the first-year counts as bases for further calculations. The successfully interviewed Latino cases from 1990 nonresponse were added into the second 1992 column. The tables also include columns that remove the deceased from the base. Known deaths among recontacts are accounted for in the percents. Individuals *born into* the sample are *not* included in these tables, even though some born-in core sample persons are now being interviewed as Heads and Wives/"Wives" of their own families. In all, 7,692 currently response individuals were born into the core sample, raising the total number of sample individuals in responding core sample families in 1993 to 15,925.

Table 2a
INDIVIDUAL ANNUAL AND CUMULATIVE PANEL RESPONSE RATES
FOR CORE SAMPLE

Year	Sample Size	Annual, Deceased Included in Base	Cumulative, Deceased Included in Base	Annual, Deceased Removed from Base	Cumulative, Deceased Removed from Base
1968	18,224	100.0	100.0	100.0	100.0
1969	16,046	88.0	88.0	88.5	88.5
1970	15,476	96.4	84.9	96.9	85.7
1971	15,108	97.6	82.9	98.2	84.1
1972	14,713	97.4	80.7	98.1	82.4
1973	14,295	97.2	78.4	97.9	80.6
1974	13,908	97.3	76.3	97.9	78.8
1975	13,548	97.4	74.3	98.1	77.2
1976	13,096	96.7	71.9	97.2	75.0
1977	12,706	97.0	69.7	97.7	73.1
1978	12,417	97.7	68.1	98.2	71.7
1979	12,056	97.1	66.2	97.7	70.0
1980	11,683	96.9	64.1	97.6	68.1
1981	11,382	97.4	62.5	98.1	66.7
1982	11,125	97.7	61.0	98.5	65.5
1983	10,828	97.3	59.4	98.1	64.1
1984	10,515	97.1	57.7	98.0	62.6
1985	10,183	96.8	55.9	97.7	60.1
1986	9,826	96.5	53.9	97.4	59.1
1987	9,504	96.7	52.2	97.9	57.6
1988	9,225	97.1	50.6	98.0	56.1
1989	8,930	96.8	50.0	97.2	54.5
1990	8,776	98.3	49.1	99.7	54.3
1991	8,518	97.1	47.7	98.2	53.3
1992*	8,467	99.4	46.5	100.7	53.7
1993*	8,233	97.2	45.2	98.7	53.0

*Includes successful recontacts. Deaths of recontacts while they were nonresponse are accounted for in the removal of deceased from the base.

4

Table 2b
INDIVIDUAL ANNUAL AND CUMULATIVE PANEL RESPONSE RATES
FOR LATINO SAMPLE

Year	Sample Size	Annual, Deceased Included in Base	Cumulative, Deceased Included in Base	Annual, Deceased Removed from Base	Cumulative, Deceased Removed from Base
1990	7,122	100.0	100.0	100.0	100.0
1991	6,522	91.6	91.6	93.1	93.1
1992*	6,976	107.0	97.9	107.6	100.2
1993	5,947	85.3	83.5	85.7	85.9

*Includes successful recontacts of 1991 nonresponse and 1,043 sample members picked up in 1992 for the first time through the attempted recontact of 1990 Latino nonrespondents. Deaths of recontacts are accounted for in the removal of deceased from the base.

Table 3
PROPORTION OF INTERVIEWS BY TELEPHONE

Year	Sample Size	Number of Telephone Interviews	Unweighted Percent of Sample
1968	4,802	—	—
1969	4,460	—	—
1970	4,645	67	1.4
1971	4,840	108	2.2
1972	5,060	134	2.6
1973	5,285	4,047	76.6
1974	5,517	4,554	82.5
1975	5,725	4,836	84.5
1976	5,862	5,360	91.4
1977	6,007	5,040	83.9
1978	6,154	5,283	85.8
1979	6,373	5,635	88.4
1980	6,533	5,829	89.2
1981	6,620	6,081	91.9
1982	6,742	6,257	92.8
1983	6,852	6,401	93.4
1984	6,918	6,369	92.1
1985	7,032	6,423	90.6
1986	7,018	6,454	92.0
1987	7,061	6,479	91.8
1988	7,114	6,520	91.5
1989	7,114	6,522	91.7
1990	7,328	6,774	88.7
1991	7,375	6,946	93.9
1992	7,561	7,317	95.9
1993	9,977	9,711	97.3

Table 4
PROPORTION OF FAMILY HEADS INTERVIEWED

Year	Sample Size	Proportion of Interviews by Head
1968	4,802	92.6
1969	4,460	93.1
1970	4,645	93.2
1971	4,840	93.3
1972	5,060	93.5
1973	5,285	91.1
1974	5,517	90.0
1975	5,725	88.3
1976	5,862	92.6
1977	6,007	90.0
1978	6,154	90.2
1979	6,373	88.5
1980	6,533	85.8
1981	6,620	84.3
1982	6,742	83.8
1983	6,852	82.2
1984	6,918	81.0
1985	7,032	87.1
1986	7,018	81.5
1987	7,061	79.0
1988	7,114	76.9
1989	7,114	76.2
1990	9,371	74.1
1991	9,363	72.1
1992	9,829	70.7
1993	9,977	69.5

Sampling Error

We include several variables for use in defining paired sampling error computing units within half-sample strata for repeated replication to compute sampling errors. These variables are present only at the individual level (V31990-V31999). See Chapter 17 of Vol. IX of *Five Thousand American Families* and Section I, Part 5 in this volume for further details.

Weights

The addition of the Latino sample to the PSID has meant that more weights are necessary. Besides the core sample weight, we added a Latino sample weight and a combined core-Latino sample weight. See Section I, Parts 5 and 13 in the 1990 (wave XXIII) documentation and Section I, Part 5 in the 1992 (wave XXV) documentation for details regarding the Latino sample and the creation of the core-Latino combined weight. Refer to the *PSID User Guide* for a more general discussion of reweighting theory and techniques.

The major recontact effort in 1993 required a new approach to weighting because of potential bias that might exist for the reappearance of formerly nonresponse individuals. In addition, the method of calculation

6

of family weights was altered, so that not only are they updated for marriages and divorces, but also for ingress and egress of all individuals. See Part 5 of this section for further information.

Part 2: Changes in the 1993 Variable List

Because of the switch in data collection methods with its reduction of individual case handling by study staff, some variables that have been included in the annual data files for many years were eliminated. Equivalents or the necessary components for user generation are available for most of these items, but some were dependent on hand coding and thus were not created for 1993.

Dropped Variables

Because of our conversion of most hand procedures to machine-based cleaning and variable generation, we were forced to eliminate and/or restructure a number of variables for 1993. The number of variables eliminated completely and not generatable by the user is relatively small, approximately three dozen. All other variables present in 1992 but absent in 1993 can be generated by the user, some with more difficulty than others. The following six paragraphs discuss dropped variables that we expect a user can generate; subsequently, we list those 1992 variables for which there is no 1993 equivalent and that are ungeneratable from the data available in 1993.

Current family composition, the number of individuals who moved in or out of the family, and the number of rooms required for the family (1992 V20315, V20311-V20314, and V20399, respectively) can be generated from individual data. The number of movers-out is tricky, however, as those who move to another responding PSID family are somewhat difficult to identify.

The variables for yearly interview numbers (family ID numbers) for 1969 through 1992 (1992 variables V20622-V20644 and V20302) are those of the individual who is Head of the 1993 family and may be gathered from his or her individual data record.

The component variables for income-to-needs ratios, total family income and needs are available on the 1993 family file, and users should find these measures simple to calculate. Similarly, creation of a decile for family money income equivalent to 1992 V21506 should cause no difficulty.

USDA food needs (1992 V21487 and V21492) were generated in most waves through 1992 for the then-current family based on the ages and genders of family members. These can be computed from 1993 individual data using the rules listed in the 1992 codebook descriptions.

Average hourly earnings of Head and Wife/"Wife" were not created because our method of dividing business and farm income into labor and asset components has changed and thus comparability could be an issue for some cases. See Part 5 for the new rules regarding this split. However, annual work hours and all income components are included in the family file; users may generate this variable as they see fit.

The variables indicating numbers of family members in various age and gender categories (1992 V21507-V21515) can be created from individual data using 1993 relationship to Head, age and sex variables. Similarly, numbers of associated persons in institutions (1992 V21516-V21519) are generatable from 1993 individual data using the sequence number and reason for nonresponse variables from the 1993 individual file.

The following five lists of 1992 variables indicate those measures dropped for 1993 that are not generatable:

These two variables were included through 1992 as possible indications of interview quality:
   V20308 WHETHER HEAD REFUSED
   V20322 QUALITY OF MATCH

Some variables describing household composition (not to be confused with family composition!) include the following:

    V20316 CURRENT HOUSEHOLD COMPOSITION
    V20317 1968 ID NUMBER OF HOUSEHOLDER
    V20318 PERSON NUMBER OF HOUSEHOLDER
    V20319 AGE OF HOUSEHOLDER
    V20320 SEX OF HOUSEHOLDER
    V20321 RELATIONSHIP OF HOUSEHOLDER TO THIS FU

These two variables cannot be generated by users because the information necessary to their coding has always been obtained from marginal notes and thumbnail sketches. Creation of these variables entails case-by-case hand coding:

    V20397 # MAJOR ADULTS
    V20455 EXEMPTIONS FOR BLINDNESS OR AGE

Estimations of income tax liability were dropped from the PSID data for the 1992 wave, but some variables greatly aiding their generation were still included in the 1992 family data. These variables had been hand coded by staff for each case. Because that is no longer a possibility, equivalents were omitted for 1993. It is possible that these items could be generated for some cases with little chance of error, but for those families whose membership includes one or more embedded "subfamilies" (e.g., a working daughter with children), the tax scenario becomes quite complex. The tax-related variables dropped for 1993 are:

    V20456 TOTAL EXEMPTIONS FOR FEDERAL INCOME TAX-HEAD & WIFE/"WIFE"
    V20457 FEDERAL INCOME TAX TABLE USED-HEAD & WIFE/"WIFE"
    V20578 # EXEMPTIONS OF FIRST HIGHEST OTHER EARNER
    V20579 FEDERAL INCOME TAX TABLE USED-1ST OTHER
    V20583 # EXEMPTIONS OF SECOND HIGHEST OTHER EARNER
    V20584 FEDERAL INCOME TAX TABLE USED-2ND OTHER
    V20588 # EXEMPTIONS OF THIRD HIGHEST OTHER EARNER
    V20589 FEDERAL INCOME TAX TABLE USED-THIRD HIGHEST OTHER EARNER
    V20593 # EXEMPTIONS OF FOURTH HIGHEST OTHER EARNER
    V20594 FEDERAL INCOME TAX TABLE USED-FOURTH HIGHEST OTHER EARNER
    V20598 # EXEMPTIONS OF FIFTH HIGHEST OTHER EARNER
    V20599 FEDERAL INCOME TAX TABLE USED-FIFTH HIGHEST OTHER EARNER

The following variables are indicators of imputations to income measures with missing data. Because imputation procedures were changed significantly between 1992 and 1993, creation of these variables was eliminated. See Part 3 of this section for a detailed discussion of imputations.

    V21482 PERCENT FAMILY MONEY INCOME WITH MAJOR ASSIGNMENT
    V21483 PERCENT FAMILY MONEY INCOME WITH MINOR ASSIGNMENT
    V21485 PERCENT HEAD'S TOTAL LABOR INCOME WITH MAJOR ASSIGNMENT
    V21486 PERCENT HEAD'S TOTAL LABOR INCOME WITH MINOR ASSIGNMENT
    V21501 ACCURACY OF MONEY INCOME COMPONENTS
    V21502 # MINOR ASSIGNMENTS
    V21503 # MAJOR ASSIGNMENTS

Miscellaneous items omitted for 1993 include the following four variables. V21399 and V21400 are from background information asked about the family unit Head. A substitute variable that asks whether the Head has ever lived in another state now replaces them.

    V20682 REASON FOR FREE RENT
    V21399 # REGIONS LIVED IN-HEAD
    V21400 # STATES LIVED IN-HEAD
    V21520 SPLIT SAMPLE INDICATOR

Added Variables and a New File

The major change to the 1993 dataset is the addition of much more income detail than was formerly available. Each type of income received is collected using three questions: the amount, the time unit during which that amount was received (e.g., per week, per month, per year), and the months in which the income was received. Now we release the actual raw data as a set of 14 variables for each type of income and include an additional two variables in which we present an annualized, imputed amount and an indicator of whether or not an imputation occurred.

Incomes of other Family Unit members are simply summarized on both the family and individual files, as we now also separately release a new file, the 1993 OFUM Income Detail File. This file includes a record for each other FU member with all the detail about that person's income. See Part 7 of this section for more information.

Year Variables

Beginning with the 1993 wave, all year variables were converted from two to four digits representing the actual year (in preparation for the new millennium). Note that ALL year variables on the 1968-1993 individual file underwent this change, so file positions for 1968-1992 and earlier cross-year individual files, as well as all early release individual files through 1996, are different.

Part 3: Editing and Imputation Procedures

The PSID editing process serves three main purposes: (1) accounting for all year-to-year changes in family membership, (2) rectifying discrepancies within the interview before coding, and (3) imputing and calculating numeric data.

Family composition editing is the first step. Next, extensive cleaning and editing of income and work is done. A detailed discussion of past techniques for data editing is located in *A Panel Study of Income Dynamics: Study Design, Procedures, Available Data 1968-1972 Interviewing Years (Waves I-V)*, Volume I, pp. 270-339. Specific changes since that time have been included annually in Section I, Part 3 of the succeeding documentation volumes.

Editing and data processing techniques changed significantly for Wave XXVI because of our conversion from paper questionnaires to computer-assisted telephone interviewing (CATI). The major difference between the two modes of data collection is that most items are coded by the interviewer during the interview. Coding of all raw variables is no longer necessary. However, interviewers can err in several ways, including making typographical errors, but perhaps most significantly in their application of study concepts to actual cases. Below we discuss some of the measures we took to correct the data.

Family Composition Editing

All people in a panel family at the time of the previous year's interview must be accounted for in the current year. They may remain in the family or may have moved out, died, or entered an institution since the prior wave. Sample members who move out and form their own households are followed and interviewed as new panel families (i.e., splitoffs). More detailed relationship to Head and birthdates for individuals have been coded since 1983, and since 1985, we have also coded the type of institution for families in the armed forces, educational or health facilities, etc. The marital and childbirth histories collected since 1985 have placed more demands on the task of family composition editing through the addition of a unique individual identifier for each spouse or child mentioned.

Beginning in 1993, for the first time we followed not only sample members age 18 or older who left the family but also younger sample members who moved out. Many of these younger sample members left with

a nonsample parent; others moved to another relative's home. As mentioned in the discussion of interviewing procedures in Part 1 above, the outcome of following young sample members is that we can no longer insist that within each family unit, either the Head or the Wife/"Wife" must be a sample member.

Because of the change in data collection procedures, we were presented with a data set, albeit a "dirty" one, when interviewing was complete and data were pulled from the CATI system. Many of the items, such as Sequence Number, that were formerly hand coded were now machine generatable, so there was no longer any need to view every case. Our editing procedures altered to focus on those interviews showing change. We reviewed the family listings for every interview in which a family member had moved in or out (minimally, we needed to assign individual identifiers to the former). We also checked every interview in which a change in relationship to Head, an untoward age change, or a sex or name change was indicated. We examined not only the listings but also the interviewers' thumbnail sketches and marginal notes for these cases to be sure that we concurred with the interviewer's family composition decisions.

Our following rules dictate that we not only continue to interview the Head in succeeding waves, but also attempt to interview family members who leave to form their own households. A corollary is that returning family members who have been successfully followed are not reintegrated into the family (with the notable exception of recombined married couples). The result is that several related individuals may share the household but are treated by us as separate families, each with its own family unit Head. Family composition editing also checked for this coresidency to ensure that all families sharing a household were mutually indicated.

Income and Hours (Economic) Editing

Economic edit procedures also altered drastically. Although the questionnaire content remained similar to 1992 and we continued our procedures begun in 1988 for collecting work histories about the prior calendar year, CATI produces a data set, already coded, upon completion of the interview. As with family composition, the interviewer may violate study concepts during the collection phase. Economic editing, therefore, consisted of developing and implementing checks to ensure the integrity of income and hours reports.

Hours. Although much hand editing was done to ensure that weeks reported in the employment sections of the questionnaire for Head and Wife/"Wife" (Sections B, C, D and E) summed to 52, we found that we could clean many cases of erroneous weeks reports by machine. Furthermore, if data for only one of the seven categories of weeks (six types of time off and the work week report) were missing, then this was resolved by subtraction, meaning no imputation was necessary. For example, if the weeks worked were reported as 52, but some time off was also reported, then the work weeks were adjusted downward as 52 minus the time off.

For those who were retired, were not currently working and had underreported time for the prior calendar year, missing weeks were added to the out-of-labor-force category. Also, sick time spent subsequent to the end of the last job in 1992 was converted to weeks out of the labor force using the work history dates for timing of employment spells.

Some retirees continue to work, but only for part of the year. In these and other cases of continuously-employed persons with underreports of total weeks, the excess time was designated as vacation.

Similarly, for anyone who was employed all of the prior calendar year and who reported time out of the labor force, weeks were recategorized as vacation.

If work weeks were underreported and all time-off questions were answered in the negative and the most recent job began in 1992, then all missing time was assumed to be out of the labor force.

If an individual had no employer at all during the prior calendar year, he or she was not allowed to report any time as illness of self or someone else, vacation, or strike. All such weeks were moved to the out-of-labor-force category.

If an individual with no employer in the prior calendar year reported 52 weeks of unemployment, then all other categories were set to zero.

Income. Income editing was less complex and, surprisingly, less amenable to machine cleaning than hours. In general, transfer income sources required the most case-by-case attention, as reports of ADC, Social Security, and child support receipt by underage persons were common. Almost invariably, this income is received by an adult because of a child; seldom is it the child's income. Such amounts were moved to the parental record where indicated. The final question in the income series for any individual in the study asks about receipt of any income not heretofore reported for this person. At least 50% of the time, this miscellaneous income belongs in a category already mentioned. Surprisingly, even labor income is sometimes reported here! These and other erroneously-categorized reports were moved to their correct sets of variables.

Another effort in economic editing was to recode where possible income time unit reports of "other" into code categories more favorable to annualization. The information needed for recoding was contained in marginal notes.

As we always have, we checked cases of persons reporting labor income but no work hours and vice versa. Such inconsistencies are resolvable through information in marginal notes and thumbnail sketches.

Imputation Procedures

Through the 1992 wave, imputations were done case by case during the hand economic edit. Prior-wave questionnaires were often consulted in the hope they could provide the editor with a case-specific report. Failing that, amounts were assigned from subgroup means. The codes for the accuracy variable connected with and following each imputed amount indicated the degree to which we believed our estimates differed from the expected amount. Often, only minor adjustments were necessary.

For 1993, however, imputations were done by machine. Details for each item are listed later in this discussion, but in general hot-deck procedures were used where possible and practical. The scheme for accuracy codes also changed; every dollar amount has an associated accuracy variable. The codes were reduced to two: a value of 0 indicating no imputation was necessary and a value of 1 indicating the amount was imputed.

Data collected for most dollar values consisted of an amount, a time unit variable, and a series of twelve month indicators. These latter are dummy variables indicating whether or not that item was received during the specified month. Amounts were annualized from the components as follows, assuming none of the fourteen component variables were missing.

If the time unit was yearly, then the given amount was used as is.

If the time unit was monthly, then the given amount was multiplied by the number of months in which receipt was indicated.

If the time unit was weekly, then the given amount was multiplied by the number of months in which receipt was indicated times 4.33.

If the time unit was biweekly, then the given amount was multiplied by the number of months in which receipt was indicated times 2.17.

Hourly amounts were given only for labor income; these were multiplied by the number of hours (which may have been imputed).

Annualization of some items is not month-specific, as these items are not asked for the prior calendar year. These are housing costs, most food costs, and housework hours, and were annualized using the associated time unit variables.

Missing information on any one of the fourteen component variables (the amount, the time unit, or one or more of the month indicators) caused an imputation to occur. To aid the analyst in deciding whether to use our imputations or to create his or her own, we include the reported amount, the time unit, and the twelve month variables along with our annualized amount and accuracy code for each reported item.

Housing. Housing imputations were divided into those for homeowners and those for renters. Homeowning families could have missing data on one or more of the following items: house value, property taxes, remaining mortgage principal, and mortgage payments.

The initial step, assigning house values, called for division of all homeowners into quartiles based on property taxes (question A21). On inspection of nonmissing property tax values, we found cases of actual-report but outrageously high amounts, defined as 10% or more of the house value where the house value was also an actual report. Before beginning with the division into quartiles, these cases were awarded missing data for property taxes. Next, cases needing house value imputations but also having missing data for property taxes were randomly assigned to one of the four property tax brackets.

House values were assigned using hot-deck procedures within each of the four property tax groups.

For property taxes, the reverse was used: the sample was divided into quartiles based on house values, and property taxes were hot-decked within each of the four groups.

Remaining mortgage principals were assigned from 1992 data if the Head remained the same and hadn't moved (question A42) and the 1992 report of years remaining to pay (question A27) was known. If necessary, the number of years remaining to pay in 1993 was created by subtracting 1 from the 1992 report. The multipliers for the 1992 principal are below:

If 13 years remained, then the 1993 principal was assumed to be 98% of the 1992 principal. If 9 to 12 years remained, the 1993 principal was assigned as 95%. At eight years it was 93%, at seven years, 91%, at six years 90%, for five years 87%, for four years 84%, for three years 79%, for two years 70%, for one year 53%, and for no years remaining (i.e., where the 1992 interview reported one year to go) 25%.

We used a special rule for newly mortgaged homeowners: if the number of years in which the mortgage had been paid thus far (A26) equaled 1, then the principal was assigned as 80% of the house value.

Missing mortgage payments were assigned as the 1992 value if the Head remained the same and the move indicator, A42, indicated no move had occurred. If the Head had changed and only one or two years to pay remained, the payments were assumed to be largely principal and so the outstanding principal was divided accordingly.

Forty-two cases were not assignable by the above rules for remaining principal and payments; these were imputed on a case-by-case basis by hand.

Missing rent was imputed from 1992 data if the Head remained the same and the whether moved variable indicated the housing unit was the same. Otherwise, rent was hot-decked.

For those who neither owned nor rented, missing data were NOT imputed. These cases remain in the final dataset with missing data.

Work Hours of Head and Wife/"Wife". As stated in the discussion of economic editing techniques above, hand and machine cleaning of the various kinds of time off was not generally considered an imputation unless two or more components were missing. Most cases not cleanable by those methods, however, were amenable to machine imputations. Hot-decking of missing weeks is not a practical procedure for this, as

exactly 52 weeks must be accounted for in the prior calendar year; writing a program to match cases would be a nightmare.

If work weeks were underreported with no time-off questions answered in the affirmative, the beginning date for the most recent job was 1992, and whether unemployed and whether out of the labor force were NA, then unreported weeks were split 50-50 between time unemployed and time out of the labor force.

Of the remaining cases with underreported work weeks that were not cleanable as described above, unreported weeks were divided between unemployment and time out of the labor force. These cases needed, but didn't get, a hand edit for 1993.

Overreporting of weeks is not usually an imputation, as most of these cases involve respondent confusion with our definitions of vacation and illness versus time out of the labor force—the former can only occur during an employment spell, whereas the latter can only occur when the person has no employer.

If more than one item among the six kinds of time off from main jobs and the weeks worked on main jobs had missing data, they were imputed as follows:

If unemployment and time out of the labor force or vacation weeks were missing, those weeks were split 50-50. The decision as to whether the time was vacation or weeks out of the labor force was made during cleaning based on employment start and stop dates.

If vacation and work weeks were missing, then vacation was assigned as two weeks and the work weeks were assigned all remaining time.

Approximately three dozen cases were unsolvable by the above rules and were assigned by hand.

Missing main-job work hours were imputed using means from age, gender and education categories. Five age brackets (age less than 31, 31-40, 41-50, 51-65, 66 and over) and four education brackets (11 grades or less, 12 grades, 13-15 years, 16 or more years) were defined. If age or education was missing, then it was hot-decked for the mean calculation.

Missing overtime hours were imputed as 5% of the total main job work hours.

Assignment of missing extra job work weeks used the month indicator variables for each extra job. The procedure counted the number of months in which the values for these variables contained 1 or 9. This number was multiplied by 4.33 to calculate the maximum number of possible weeks worked. The maximum was then divided in half for the weeks imputations. If the months were all missing data, then weeks were assigned as 10% of main job hours for the first extra job or 5% of main job hours for each succeeding extra job.

All missing extra job weekly work hours were assigned using the 10%-5% rule above.

Total annual work hours were calculated and summed from the above three components: main jobs, overtime hours, and extra jobs. The total was checked for maximums above the allowable 5,840 level, and components and totals were adjusted downward if necessary. Three Heads and one Wife/"Wife" needed adjustment.

If Heads and Wives/"Wives" last worked before 1992, they were only asked whether they spent any time looking for work (C7/E7). All remaining weeks in 1992 are assumed to be spent out of the labor force. To impute missing unemployment here, the months looking for work at C8/E8 were used to determine maximum weeks, as in the extra job procedure described above. If months were missing, then weeks were divided evenly between unemployment and time out of the labor force. Note that for these cases, either both or neither is imputed.

Housework of Head and Wife/"Wife". Missing data here were imputed using the subgroup means algorithm for these variables from 1992 editing.   These means are based on employment status, gender, work hours, number in the family, and for Wives/"Wives", the Head's housework.

Food.   Food cost imputations were also based on subgroup means from 1992 processing; these amounts were adjusted for inflation using the Consumer Price Index from the U.S. Bureau of Labor Statistics for 1992 and 1993 (137.9 and 140.9, respectively).

Food cost variables are amount for food eaten at home, amount spent on meals eaten out (not counting meals at work or school), value of food stamps received in the prior month, and annual value of food stamps received for the prior calendar year.

The subgroup means were based on house value, total family income, family size, sex and age of Head, age of Wife/"Wife", number of children and adults in the family, and age of the youngest child.

Income of Head and Wife/"Wife".   In general, missing income items were hot-decked from among all cases receiving the specific type of income.   There are, however, a few exceptions and embellishments, as follows:

Farm income was imputed from 1992 where possible. The few remaining cases with missing information were hot-decked from among all farmers with 1993 reports. Because farmers may break even, the user should be aware that, unlike wage income, it is possible for farm income to equal zero but have an accuracy value of 1.

Missing business income was assigned using 1993 medians for nonmissing cases. The median profit in 1993 was $20,000; the median loss was $2,000.

For wages, age, gender, and education categories were assigned. The five age groups were: aged less than 31, 31-40, 41-50, 51-65, and 66 and over. The education groups were: 11 grades or less, 12 grades, 13-15 years, and 16 or more years. If age or education was missing, then it was hot-decked.

Wage rates were calculated for each Head or Wife/"Wife" using the total annual work hours and the response to question G13 or G52. Note that this wage rate is skewed downward slightly because of those who have other sorts of labor income in addition to wages. Cases with missing wage data were hot decked within each age-sex-education group; this wage rate was then multiplied by the person's work hours.

The remainder of missing income items were hot-decked from among all Heads with that type of income for Heads and from within all Wives/"Wives" for Wives'/"Wives'" imputations, with no further division into age or other subgroups.

Income and Work of Other Family Unit Members (OFUMS). All income and work hours amounts are prorated for each OFUM after imputations for the amount received while the OFUM was living in the family. OFUM income questions were asked in two groups, one set for those age 16 and older and another, smaller set for those under age 16.

The family and individual files contain only summaries of labor, asset and transfer incomes. Because of user demand for more specific information about each OFUM's income receipt, we are releasing this year, for the first time, a file containing all the income detail about each OFUM. See Part 7 of this document for more information.

The labor data for the "older" OFUMs consisted of as many as four reports of jobs. Missing information was imputed for each of these jobs. The procedure for work weeks used the set of twelve months variables; each month worked was counted as 4.33 weeks. If all the months were NA, then 26 weeks was assigned. Missing hours were hot-decked from among those with nonmissing reports for the job mention. Total annual hours were inspected for cases reporting more than 5,840 hours per year; a half dozen were found (most had more than one job with missing information) and adjusted downward by hand. Incomes of all types were hot-decked within income categories if missing.

Procedures for the "young" OFUMs were somewhat different. Unfortunately, data collection did not include month-of-receipt questions. Thus, all amounts not reported as annual received accuracy codes of 1, with the exception of labor income; it could sometimes be annualized using the weeks and hours variables.

Missing weekly work hours were simply hot-decked. Missing work weeks were imputed from the unweighted mean (18.9 weeks) for those young OFUMs who worked in the prior year. Using the same assignment of 26 weeks as for older OFUMs with missing months produced unbelievably high work hours because some PSID children work a lot during the summer, thus increasing the likelihood of higher weekly hours being assigned to those with both weeks and hours missing. Similarly, the median of 52 weeks for young OFUMs wasn't feasible because quite a few youngsters have paper routes—but their reported weekly commitment is 10 hours.

Young OFUMs' missing asset and transfer incomes were hot-decked.

Part 4: Coding Procedures

Coding was not performed in 1993 as it had been in prior waves because the data were collected electronically. However, a few open-ended responses still required skilled coding, most notably the three-digit occupations and industries for Heads' and Wives'/"Wives'" current main jobs, for their prior main jobs in work histories, and for their extra jobs. The two-digit industries for businesses (G7a) and one-digit occupations for Head's first job and Head's father's usual occupation, asked for new Heads in the background section (L5 and L4) were also coded. Other coded background items were the states and counties for the locations where Head and his or her father and mother grew up (L8, L2 and L3 respectively). The remaining item that was judged too difficult for interviewers to handle during the interview was how a job had ended (B55, C15, and C47 for Heads; D55, E15 and E47 for Wives/"Wives").

All of this open-ended coding was performed by the Coding Section at ISR's Survey Research Center.

Part 5: Generated Variables, Additional Data and Hot Topics

Various indices and complex measures of economic status have been constructed each year using variables derived directly from coded interview data. Inter-year changes in the interview schedule have made addition and deletion of indices necessary. In general, if an index could not be built to be exactly comparable to a previous index, it was not constructed.

Federal Income Taxes

Sadly, we no longer compute federal income taxes. The tax codes have increased in complexity through the years, and the burden to the processing staff of generating taxes has increased to a level where it would significantly delay the release of the final data. With this in mind, we have decided that we cannot currently afford these variables.

State and County Codes

Beginning with the release of the 1989 family files, county codes for the current county of residence have been suppressed and those positions are filled with zeroes. The codes are available in separate files to qualified users under special contractual arrangements with us. For information about obtaining the special files, contact Kaarin Stahl by e-mail: PSID_Staff@umich.edu. The affected variables are as follows:

Table 5
SUPPRESSED COUNTY VARIABLES

Year	Current County	Current State and County	FIPS County
1968	94	-	-
1969	-	538	-
1970	1104	1105	-
1971	1804	1805	-
1972	2404	2405	-
1973	3004	3005	-
1974	3404	3405	-
1975	3804	3805	-
1976	4304	4305	-
1977	5204	5205	-
1978	5704	5705	-
1979	6304	6305	-
1980	6904	6905	-
1981	7504	7505	-
1982	8204	8205	-
1983	8804	8805	-
1984	10004	10005	-
1985	11104	11105	12381
1986	12504	12505	13633
1987	13704	13705	14680
1988	14804	14805	16154
1989	16304	16305	17540
1990	17704	17705	18891
1991	19004	19005	20191
1992	20304	20305	21497
1993	21604	-	23329

The county variables for Head's background, including the counties where Head and parents grew up (1993: V23247, V23249, V23255) are not affected. These variables still contain actual values.

Income

Several measures of economic status have been generated for all twenty-six years of the study, including money income variables and measures of income adequacy. Family Money Income, one of the simplest indices, is the total of all members' earnings, transfers, and asset income from the prior calendar year (1993: V23322). This variable and its components are adjusted for movers into and out of the family, in that we only include income of non-Heads and non-Wives/"Wives" if it was earned during the time that these other family members were present in the family unit. See the discussion below entitled "Proration of OFUM Incomes" for a full explanation of proration procedures.

New Income Variables. Most income amounts are collected in three pieces: the amount received, the time unit indicating that amount's receipt (per week, month, etc.), and the months in which the income was received. These data are now included on the family file for each type of Head's or Wife's/"Wife's" income listed in Section G of the questionnaire as sets of fourteen variables (one for the amount, one for the time unit, and twelve dummy variables for the twelve months of the prior calendar year). An additional two variables, one for our annualized version, imputed if necessary, and another for its accuracy code, are also

included. Each type of income has sixteen associated variables. See V21731-V22366 for Head and Wife/"Wife" incomes. Imputation procedures are detailed in Part 3 of this section.

Labor and Asset Income Splits. Income received from farming and businesses was split into labor and asset components.

If the Head's current main job is self-employed farming, then this income is reported in the questionnaire at question G4 (1993 V21731). This amount is split half and half into labor and asset income if there was a profit. If the farm showed a loss, then labor income is set to zero and the amount is reported as negative asset income. If the enterprise broke even, then Head was given zero labor and zero asset income from this source.

A few wives indicated that they were co-owners of and worked on the farm. In that case, the profit was divided evenly four ways, with Head and Wife/"Wife" each receiving a quarter of the amount as labor and a quarter as asset. If the farm showed a loss, then it was divided evenly between the two of them as negative asset income; if it broke even, then labor and asset income from farming was coded zero.

These four variables appear on the 1993 family file: V21733 and V21809 for Head's labor and asset income from farming, V21803 and V21813 for Wife's/"Wife's".

Information about as many as five unincorporated businesses was collected for Heads and Wives/"Wives". Profits and losses for each were divided as follows:

If both owned, the business showed a profit, and both worked in the business, then each received a quarter of the profits as labor and a quarter as asset.

If both owned, the business showed a profit, and only one worked in the business, then the worker's half was split equally into labor and asset, while the nonworker's half was all asset.

If both owned, the business showed a profit, and neither worked in the business, then the profit was split in half, all asset for each.

If both owned and the business showed a loss, then each received half of the loss as negative asset income, regardless of work put in.

If a business was owned solely by the Head or the Wife/"Wife", then only the owner profited from the business; even if the other worked in the business, this is ignored when splitting the income. We assume (and check) that a nonowning worker's earnings were paid elsewhere as wages and salaries. The owner's income was divided as for farm income above: profits were split 50-50 labor and asset if he or she worked in the business, and if he or she did no work, then the income was all asset. A loss shows as negative asset income, regardless of whether work was involved.

The labor and asset amounts for Head and Wife/"Wife" were summed across all businesses (1993 V21734) to create the family-file totals (for 1993, V21738 and V21810 for Head's labor and asset income from businesses; V21806 and V21814 for Wife's/"Wife's").

Heads who received income from farming but for whom farming was not their current main job reported this income in questions G18b-G21b (1993 V21767-V21782). This income was split into 75% for labor and 25% as asset if there was a profit; any losses showed as negative asset income (1993 V21783 and V21811).

If Other Family Unit Members had business profits, this income was split with half designated as labor and half as asset. Only the labor portion is included in the family-level and individual-level variables labeled Labor Income (1993 family variable V22367; 1993 individual variable V30821). The asset portion and all business losses are included in family and individual variables labeled Asset Income (1993 family variable V22731; individual variable V30822).

17

Proration of OFUM Incomes

As described in the discussion in Part 3 about imputations of Other Family Unit Members' income, any income these individuals may have is only counted for the portion of the year that they were members of a PSID family unit because we wish to match prior calendar year income with the family's needs last year.

Proration factors were calculated for each type of income reported, as each could have been received in a different portion of the year. Proration factors compared the months in which an income was received with the months in which an individual lived in the family. A person is considered to have moved on the first day of the month coded on his or her individual data record. Those with missing move dates are assumed to be in the family for half of the year. Those who "appeared" or "disappeared" (i.e., whose moves occurred before the present wave but were not discovered until now; the moved in-out indicator [1993: V30813] equals 2 or 8) were treated as present for all or none of the year, respectively. Individuals in splitoff families are assumed to be present for the entire year.

Thus, those who were present throughout 1992 received a proration factor of 1 for each income; those who had moved out in January 1992 received a proration factor of 0.

Since young OFUMs, who were asked a separate set of income questions, had no months variables indicating income receipt, we assumed their incomes were received evenly throughout the year; amounts for movers were prorated simply by the fraction of time in which they lived in the family.

Month-by-month receipt of each type of income for an older OFUM who was present for only part of the year was compared with that person's move date to determine the correct fraction of FU-worthy income. For example, if a son worked from July through December but moved out in June, none of that labor income was counted; whereas if he had worked from January until June, all of his income would have been counted.

If months of income receipt were missing, then the income was assumed to be received evenly across the year.

The dates of move out for deceased persons may or may not be the death dates. These are the dates on which they left the FU; there may have been some time spent living outside the FU before death occurred. In general, deceased persons' incomes are prorated as for other movers, but an exception was made for their pensions and Social Security income; if these were received in the same month as the move out, that month's amount was retained.

Needs

We continue to generate the same annual needs standard (V23325) that we have since the study began in 1968. This version of needs is described in detail in Section I, Part 5 of the 1974 documentation and in the codebook in this wave's documentation. The standard was taken from the June, 1967 issue of *Family Economics Review* and is based on food costs for March 1967 using the USDA low-cost plan, not the "economy" budget used by the Census for CPS. The needs have been adjusted for family composition change since the previous interview to match prior-year income as explained in the preceding subsection. Essentially, the PSID calculates twelve separate needs, one for each month of the prior year. Each of these is based on the number, age and gender of the persons in the family during that month using the individual-level data. The values for all twelve months are summed and averaged to create V23325.

This version of needs is still in terms of 1967 dollars. Table 6 shows CPI deflators for 1967 through 1993:

Table 6
ANNUAL AVERAGE CONSUMER PRICE INDEX, 1982-1984=100

Calendar Year	CPI Deflator
1967	36.3
1968	37.7
1969	39.4
1970	41.3
1971	43.1
1972	44.4
1973	47.2
1974	51.9
1975	56.2
1976	59.4
1977	63.2
1978	67.5
1979	74.0
1980	82.3
1981	90.1
1982	95.6
1983	99.6
1984	103.9
1985	107.6
1986	109.6
1987	113.6
1988	118.3
1989	124.0
1990	130.7
1991	136.2
1992	140.3
1993	144.5

Beginning with the 1990 wave, we generate another needs standard (V23326). This poverty measure, from the U.S. Bureau of the Census' *Current Population Reports*, Series P-60, uses family size, age of the householder, and the number of children under age 18 as threshold determinants. Below we have reproduced the table we used for this wave, taken from Table A on p. vii of the Census Bureau's *Poverty in the United States: 1992*, Series P-60, No. 185. The income levels are in terms of 1992 dollars. PSID values for these variables were calculated for part-year family membership to match our family income measures, as described in the preceding paragraphs.

Table 7
POVERTY THRESHOLDS IN 1992
By Size of Family and Number of Related Children Under 18 Years

Size of family unit	Related children under 18 years								
	None	**One**	**Two**	**Three**	**Four**	**Five**	**Six**	**Seven**	**Eight or more**
One person under 65	$7,299								
One person 65+	6,729								
Two persons, Head under 65	9,395	$9,670							
Two persons, Head 65+	8,480	9,634							
Three persons	10,974	11,293	$11,304						
Four persons	14,471	14,708	14,228	$14,277					
Five persons	17,451	17,705	17,163	16,743	$16,487				
Six persons	20,072	20,152	19,737	19,339	18,747	$18,396			
Seven persons	23,096	23,240	22,743	22,396	21,751	20,998	$20,171		
Eight persons	25,831	26,059	25,590	25,179	24,596	23,855	23,085	$22,889	
Nine persons+	31,073	31,223	30,808	30,459	29,887	29,099	28,387	28,211	$27,124

Labor Market Measures

We collected county labor market information from state agencies through the 1989 wave, but this has been discontinued. Instead we have replaced our variables about availability of unskilled jobs and unemployment rates with average annual unemployment rates for respondents' counties of residence at the time of the 1993 interview from the U.S. Bureau of Labor Statistics (1993: V23335).

Sampling Error Computation Unit (SECU) Variables

The 1993 individual data file includes variables that may be used for computation of variances under the stratified multistage design (V31998-V31999), as well as variables for balanced half-sample replication (V31996-V31997). These variables are available only at the individual level. Please see the 1983 Documentation volume, pp. 89-90, for details about SECUs for the core sample.

The development of the Latino SECUs (V31990-V31995) was based on listing area numbers in the LNPS. The exact specifications cannot be released for reasons of confidentiality.

See the SECU variable descriptions in the 1968-1993 individual file codebook for more information about use of these variables.

Dates and Year Variables

The PSID collects dates of events, such as month and year jobs began and ended. Beginning with the 1993 wave, all year variables were converted to four digits representing the actual year in preparation for the millennium.

New Heads and New Wives/"Wives"

Two variables indicate the year in which the current Head *most recently* became Head (1993: V23339) and the year in which the current Wife/"Wife" *most recently* became Wife/"Wife" (1993: V23340). (It is possible that an individual becomes Head or Wife/"Wife" more than once in the course of the panel due to marital breakups, reconciliations and remarriages.) These variables contain as code values the year in which the background data for Heads or Wives/"Wives" was most recently asked. Most background information was reasked in 1985 for the core sample. Because of this, all core sample Wives/"Wives" answered these data items afresh for the 1985 interview. Therefore, V23339 equals 1985 for most core cases. New Wives/"Wives" since then were asked the entire sequence and thus have values in the range 1986-1993 for this variable. The elderly recontact and Latino sample Wives/"Wives" were asked the series in 1990, and so have values of 1990 unless the family acquired a new Wife/"Wife" in 1991 or 1992. Both core and Latino 1992 reinterview recontacts have values from 1990 for this variable, as background information was not reasked for them (unless, of course, the Wife/"Wife" had changed). All other 1992 and 1993 recontact cases with Wives/"Wives" were asked the background questions, and so have values of 1992 or 1993 for V23340.

New core Heads in 1985 were, as usual, asked the entire sequence (1993: V23246-V23317). For 1985 Heads who were also Heads in 1984, however, only 1985 variables V11924-V11981 were asked. The values for 1985 variables V11907-V11923 were simply transferred from 1984 or earlier years' data, as most of these items should not have changed from year to year. Variable 23339 indicates the year in which these items were most recently collected. Refer also to p. 72 of the Wave XX (1987) documentation for information regarding specific background variables. Core elderly recontact and Latino Heads for 1990 were considered new Heads, so background data are current as of 1990 unless a new Head moved in for 1991, 1992 or 1993. Both core and Latino 1992 reinterview recontacts were not reasked background information unless the Head had changed by 1992. All other 1992 and 1993 recontact cases were asked the background questions, and so have values of 1992 or 1993 for V23339.

Education of Head and Wife/"Wife" at the Individual Level

The income and work sequence that we ask for all of last year's family unit members besides the current Head and Wife/"Wife" includes some questions about completed years of schooling. These questions are reasked and coded each year for such individuals (1993: V30817-V30820.) We have generated equivalent data for Heads and Wives/"Wives" for the completed education variable from this series (1993: V30820) from 1976 through the present wave, although their years of schooling and much more education detail are available at the family level (Head: V23279-V23314, Wife/"Wife": V23215-V23241). Beware, however, that their education and other background items are not reasked each year, as completed education is for other individuals. See each completed education variable (V30197, V30226, V30255, V30296, V30326, V30356, V30384, V30413, V30443, V30478, V30513, V30549, V30584, V30620, V30657, V30703, V30748, V30820) in the 1968-1993 individual codebook for details. The variable descriptions for each year's completed years of education document the variables we used to generate values for Heads and Wives/"Wives" from the family-level data.

Bracketed Education of Head and Wife/"Wife" at the Family Level

Since the study began, Head's and Wife's/"Wife's" education was available as a single-digit bracketed variable. Beginning in 1992, however, we include in its place the two-digit completed education variable described immediately above. This variable is *not comparable* with the past years' bracket (1993: V23333 for Head, V23334 for Wife/"Wife") because years spent in nonacademic training are *not* included in the two-digit total. To create a comparable one-digit bracket, the user must also take into account information from the background variables about whether or not nonacademic training was received (1993: V23228 for Wives/"Wives" and V23298 for Heads).

Identifying Long-Term Cohabitors at the Family Level

Couple Status of Head (1993: V23338) allows users to easily identify cases with female Heads and husbands at the family level. Additionally, through the coding of relationship to Head in more detail, long-term female cohabitors ("Wives") are distinguishable from legally married couples at the family level.

1993 PSID Analysis Weights

The final data release for the 1993 Panel Study of Income Dynamics (PSID) includes both individual and family-level weight variables for cross-sectional or longitudinal analyses that are based on the 1993 PSID data.

The 1993 PSID weight variables are designed to enable unbiased estimation of descriptive statistics for the U.S. individuals and families that are eligible for the 1993 PSID survey population. The 1993 PSID core survey population includes all living individuals who were resident in the contiguous-U.S. household population in 1968 and their direct descendants. The survey population for the 1993 Latino supplement includes Hispanic individuals of Mexican, Cuban and Puerto Rican descent who were resident in households in selected U.S. counties in 1989. Direct "born in" descendants of individuals who were eligible for the core and Latino baseline samples are also included in the respective 1993 PSID survey populations. The 1993 PSID family reference populations for the core and Latino supplement panels are defined as the sets of U.S. family units that include at least one individual member who is eligible for the corresponding 1993 PSID survey population.

1993 PSID core and Latino analysis weights do not adjust estimates of descriptive statistics for noncoverage of individuals and families who are not eligible for the corresponding 1993 PSID survey populations. The uncorrected noncoverage for the core panel includes primarily those individuals and families that immigrated to the United States after 1968. Noncoverage in the Latino supplement panel includes persons and families of Mexican, Cuban and Puerto Rican descent whose 1989 residence was outside the geographic boundaries of the survey population or who immigrated to the U.S. after the 1989 baseline sample of Latinos was selected.

The remainder of this discussion of weights and weighting for 1993 is divided into four parts. Subsection 1 below describes issues and concepts in PSID weighting. Subsection 2 includes a detailed description of the calculation of the 1993 core longitudinal analysis weights. Separate paragraphs of subsection 2 describe the sample selection factors, nonresponse adjustment factors and mortality adjustments required for the computation of the individual weights for core sample persons. The 1993 Latino longitudinal analysis weights for individuals and families are described in subsection 3. Subsection 4 provides documentation for the 1993 combined (core plus Latino) longitudinal analysis weights.

1.A Weighted Analysis of the PSID Data

The 1993 PSID analysis weights may include three factors: (1) a design-based adjustment for differences in the *sample selection* probabilities for sample individuals and families; (2) a model-based adjustment for the *nonresponse attrition and death* of sample members over the life of the panels; and (3) a ratio adjustment to *poststratify* weighted sample distributions for important demographic and household variables to population

values measured in the decennial Census or a larger, more precise sample such as the U.S. Current Population Survey (CPS).

Sample selection factor: Weighting for differential sample selection probabilities in the estimation of descriptive statistics is a fundamental principle of the theory of unbiased, sample-based estimation for survey data (Horvitz and Thompson, 1952; Skinner et al., 1989). To understand weighting in lay terms, think of the sample selection weight factors as the number of finite population elements that are represented by the sample element to which the weight is assigned. A sample family that is selected with a probability of 1/1000 can be viewed as representing the characteristics of 1000 families in the survey population. If that family is low-income, think of it as representing 1000 low income families in the population. Sample selection probabilities for individual population elements are generally fixed by the features of the sample design (i.e., population definition, stratification and sample allocation) and sample selection procedures (i.e., probability-based rules for choosing respondents within households). The calculation of sample selection weighting factors is therefore a deterministic procedure that rarely involves stochastic modeling of probabilities and relationships.

Weighting for nonresponse attrition: The 1993 PSID core and Latino weights also include nonresponse adjustment factors. Adjustment for nonresponse attrition through "weighting cell" factors or "propensity scores" is inherently a model-based procedure that builds in explicit assumptions concerning the underlying mechanism or process that leads to nonresponse (Little and Rubin, 1987). For example, adjustments computed using simple weighting cell methods assume that observations are missing at random (MAR) within cells. That is, the process that resulted in nonresponse is equivalent to a further simple random sampling within cells defined by selected characteristics of individuals (e.g., age, gender, race, education, etc.). If the chosen model for the adjustment is correct and the attrition process is MAR, applying the nonresponse adjustment factor to the sample selection weight factor will enable the researcher to restore the unbiased representation of weighted sample estimates.

Nonresponse attrition in cross-sectional and panel surveys is a complex outcome that involves many population as well as individual factors. It is unrealistic to expect that we as statisticians and researchers can define a single model that will be correct for all population subclasses and all statistics of interest. Recognizing this, the need to develop general purpose weights for PSID analysis requires the definition of simple models of the nonresponse process that enable us to adjust for the key predictors of nonresponse attrition. Researchers who are interested in developing their own models of attrition or in dynamic modeling that includes a specification for the attrition process may wish to reference Little and Rubin (1987).

Poststratification of Analysis Weights to External Data on Population Distributions: Analysis weights for design-based estimation of population statistics from sample survey data may also include post-stratification factors. Post-stratification factors adjust the combined weights for sample selection and nonresponse so that weighted distributions of sample cases match population distributions that are known from external data sources such as a decennial Census or a large sample-based data set such as the Current Population Survey (CPS) that has smaller levels of sampling variability. The purpose of poststratifying weights to known population distributions is two-fold: (1) to reduce the variability of estimates by adjusting the weighted distribution of the sample across population subclasses; and (2) to correct the analysis weights for sample noncoverage that is not reflected in the sample selection weight or nonresponse adjustment factors.

Poststratification adjustments have not been included in the longitudinal analysis weights for PSID core data sets for 1968 to 1993. The historical reason for not poststratifying PSID core analysis weights for individuals and families to Census or other data sets has been the lack of detailed data on the current segment of the U.S. population that is represented by the original 1968 PSID sample families and their descendants. Individuals and families who immigrated to the United States after 1968 are not represented by the PSID core sample. Unfortunately, there is no comprehensive federal or other data source that enables us to clearly delineate household and population counts for the segment of the U.S. population that is represented by the PSID core in year t=1969,...,1993.

Poststratification factors are introduced to the 1990, 1991, 1992 and the 1993 PSID analysis weights for the Latino supplement sample of individuals and families.

1.B Weights for 1993 PSID Longitudinal and Cross-section Analysis

PSID public use data files for the 1968 to 1992 annual releases contain individual and family weights intended for use in all cross-sectional and longitudinal analysis of the core and Latino panel data. A set of weight variables that extends the historical series of annual weights for sample persons in the core and Latino panels is again included in the 1993 PSID data sets. The descriptive label for these weight variables has been modified to *"longitudinal analysis weights"* to distinguish this set of weights from new sets of 1993 *"cross-sectional analysis weights"* that will soon be available to PSID analysts in a supplemental file. As in previous years of PSID data, the series of longitudinal weights may be used to conduct either longitudinal or cross-sectional analysis of the 1993 PSID core and Latino data. The new series of cross-sectional weights will be developed only for the PSID core and, as the label implies, are intended solely for cross-sectional analysis of the 1993 PSID core data. The advantage in using the new weight over the longitudinal weight for cross-sectional analysis is that all members of sample families—both *sample and nonsample persons*—will receive a nonzero weight. Analysts who choose this weight for cross-sectional analysis of the 1993 PSID core data will have significantly larger case counts in the combined set of observations of PSID core sample and nonsample persons. The practical and theoretical basis for the decision to develop both longitudinal and cross-sectional analysis weights for the 1993 PSID is described in more detail in the following paragraphs.

The term c*ross-sectional analysis* is defined to include estimation and inference that applies to a single data collection year, say year t=1993. The most common cross-sectional analysis of PSID individual and family data involves estimation of simple descriptive statistics and models for the time t data. Examples of cross-sectional analyses of the 1993 PSID data would be age- and sex- specific estimates of the mean 1992 personal income for individuals or estimates of median 1992 family income for subclasses of 1993 PSID families. By definition, cross-sectional analysis requires non-missing data for year t. To achieve unbiased estimates of population statistics, each non-missing observation at time t should have a non-zero weight that reflects its proportionate representation in the population.

Longitudinal analysis of the PSID panel data focuses on the estimation of models of changes, transitions or events over a time interval (t-k, t). Most longitudinal analyses of PSID data are *"end point" analyses* that model change, transitions or events that terminate or are censored to further observation at time t. For a case to be included in such analyses, it must have non-missing data at time t. Furthermore, observed (or imputed) data must also be available at time t-k and at other points in the (t-k, t) interval as required by the model that is being estimated. Take as an example the problem of estimating the mean value of simple change in individuals' annual income between the time periods covered by the 1990 and 1993 PSID interviews. Ideally, the longitudinal analysis weight for this estimation problem should be based only on those cases with observed or imputed data for both 1990 and 1993. Assigning positive analysis weights to cases that have missing observations for 1990 and 1993 may lead to bias in the weighted estimate of change, since most basic analysis programs will drop cases with missing data on any required variable (casewise deletion).

1.C Special Issue for 1993 Core Longitudinal Weights

A non-zero positive weight value was assigned to each cooperating PSID core sample person in years t=1968 to 1992 (1990 to 1992 for cooperating sample persons in the Latino supplement). Each PSID family that included one or more cooperating sample persons received a positive, non-zero family weight. This historical approach to PSID analysis weights allowed analysts to use all of the respondent data for PSID sample persons and families in weighted, descriptive analysis for of the current year's PSID data. Likewise, since nonzero weights were given to all responding sample persons in year t the same weight could be used for "end point" longitudinal analysis of the data for the interval [t-k, t]. The only difficulty analysts faced in using these weights for longitudinal analysis is that the criteria for assigning a positive weight value to an individual case depended only on the sample person's response status in year t. Therefore, the assigned

weights for sample persons addressed the problem of nonresponse at time *t* but did not take into account that the same individual may have been a nonrespondent in a previous wave.

Prior to 1993, the problem of incomplete time series of annual observations was primarily limited to sample persons who had been nonrespondents in previous waves but through ties to cooperating families reappeared as a respondent in a current wave's data collection. The annual totals for these "reappear" cases were relatively small for the period 1968-1992, averaging 100-200 cases each year. PSID analysis weights for these years ignored the very small potential bias that might result from ignoring the fact that these cases had reentered the sample after a period of nonresponse (PSID, 1992). The weighted estimation bias in question arises whenever a sample person reappears and their preceding years of nonresponse include a year in which the nonresponse adjustment was performed (see subsection 2.C, paragraphs on attrition adjustment factors and 1993 nonresponse/attrition adjustment)—nonresponse adjustments included in the weights of full-time respondents have corrected for their nonresponse in previous waves.

An additional bias for weighted longitudinal analysis arises from assuming that each sample person assigned a positive weight at time t will have the requisite data to conduct the desired analysis for the multi-year interval (t-k, t). Depending on the time interval of interest, the reappear cases may not have the required data for the previous year(s) and will be dropped from the analysis. Researchers who are conducting longitudinal analysis of the PSID core and Latino data are encouraged to check how many cases are being excluded from analysis due to nonresponse for an entire wave or item missing data for key variables.

The small problem that 1969-1992 reappear cases pose for longitudinal data analysis is amplified in the 1993 PSID data collections. Specifically, in 1993 a large number of previous years' nonrespondents reappear as a result of special recontact efforts designed to trace and resume interviews with PSID nonrespondent families and individuals. The consequence of these special recontact efforts is a dramatic rise in the numbers of 1993 responding sample persons with interrupted series of annual observations. The counts of reappear and recontact cases for 1993 reach proportions of the total samples which would introduce nonnegligible bias to weighted longitudinal analysis if the historical rules for handling weights for these cases were applied. A review of response patterns for the period 1988 through 1993 indicates that the 1993 data include approximately 1,800 sample persons who were recontacted or reinterviewed in 1993 and had not participated in the PSID since at least 1987.

To minimize the problem of incomplete data series for the large sample of recontact cases, a decision was made to assign non-zero values of the 1993 core analysis weight to all 1993 sample person respondents, except individuals who were nonrespondents during the 1988 and 1989 data collection years. Thus, all core sample persons who receive a non-zero 1993 longitudinal analysis weight can be presumed to have a near-continuous longitudinal sequence of PSID data beginning back at least as far as the 1988 data collection year (the obvious exception being newborns). Any individuals who reappeared in the PSID sample as a result of the standard processes or the special recontact effort will not be assigned a positive panel attrition weight unless they have data present for the 1988 and 1989 data collection years. 1989 was the last year in which a formal five-year non-response adjustment for 1984 to 1989 attrition was performed. It is the base year for the performance of the current four-year attrition adjustment 1989-1993 (see 2.C, paragraph on 1993 nonresponse/attrition adjustment).

1.D Choosing the Correct Weight Variable for Descriptive Analysis of the 1993 PSID Data

The weight variables provided with the 1993 PSID public use data sets are designed for general use by PSID data analysts. PSID analysts should choose the weight variable that is appropriate for the level of their analysis (sample individuals, family units) and the PSID panel component (core panel, Latino supplement, combined core and Latino). Table 8 identifies the correct choice of weight variable for individual and family-unit analyses of the 1993 PSID core sample, Latino supplement and combined core/Latino data sets.

Table 8
1993 PSID ANALYSIS WEIGHT VARIABLES
Weight Variables for Longitudinal and Cross-Sectional Analysis

Sample	Individual-level Analysis	Family-level Analysis
Core sample only	V30864	V23361
Latino supplement only	V30865	V23362
Combined Core/Latino	V30866	V23363

2. The 1993 Core Longitudinal Analysis Weights

The following sections provide a detailed description of the procedures used to develop the 1993 PSID core longitudinal analysis weights for individuals and families. Experienced PSID analysts will recognize that the 1993 longitudinal weight development procedures closely follow a general methodology of adjustment for initial sample selection probabilities, panel attrition and mortality that has been employed by the PSID since 1968. There are a number of small but important changes to the past procedures that have been introduced in the development of the 1993 weights. Specific changes to note are:

(1) The 1993 PSID individual longitudinal weights include an adjustment for panel nonresponse attrition and mortality that occurred during the period covered by the 1990-1993 PSID data collections.

(2) The adjustment for nonresponse attrition now uses a "propensity" model methodology based on logistic regression of a dichotomous indicator of attrition on selected covariates that are known for each sample individual. Previous nonresponse adjustments used a weighting cell adjustment methodology in which SEARCH classification algorithms were used to define weighting cells.

(3) 1993 PSID family longitudinal weights are now computed as the average weight of all family members. Previously, family weights were computed as the average of the individual weights of the head and spouse (where present).

(4) The 1993 PSID individual longitudinal analysis weight for sample persons is set to zero if the individual did not participate in the 1988 and 1989 PSID interview. This action aims to eliminate a small bias due to respondents who "reappear" after being counted as lost in the standard attrition adjustment (see 1.C).

2.A Overview of the Weight Computation

The 1993 individual longitudinal weight values for PSID core sample persons are the product of three distinct sets of factors: (1) a single factor that represents the reciprocal of the probability by which the sample person was "selected" to the PSID panel; (2) a compound product of attrition adjustment factors developed in 1969, 1974, 1979, 1984, 1989 and 1993; (3) mortality adjustment factors also developed and applied in 1969, 1974, 1979, 1984, 1989, and 1993. A general formula that reflects the composite nature of the 1993 individual weight is therefore:

$$W_{i,1993} = W_{i,sel} \times \prod_{j=1969}^{1993} \left[W_{i,NR(j)} \times W_{i,M(j)} \right]$$

$where: W_{i,sel} =$ selection weight factor – the reciprocal of probability that individual i is selected to the PSID panel by membership in a 1968 PSID sample family or by birth to a PSID sample parent (see below).

$W_{i,NR(j)} =$ the attrition adjustment factor applied to the ith individual weight at time period $j \ \varepsilon$ (1969, 1974, 1979, 1984, 1989, 1993).

$W_{i,M(j)} =$ the age, sex and race-specific mortality adjustment applied to the ith individual weight at time period $j \ \varepsilon$ (1969, 1974, 1979, 1984, 1989, 1993).

Longitudinal analysis weights for PSID sample persons who were alive in 1968 are the product of all 13 factors in this basic equation. Individual weights for children of sample persons "born into" the PSID after the 1968 baseline year include a sample selection factor and attrition/mortality adjustment factors only for the adjustments periods following the year of their birth. Nonsample members of PSID family units receive an individual longitudinal weight of zero.

The 1968 PSID family unit weight includes only a sample selection factor to reflect the probability that the family unit was selected for the baseline interview. The 1993 PSID family weight (and all other family weight variables after 1968) is derived as the average of individual weights for sample and nonsample persons in the family unit. Therefore, the 1993 family unit weight is completely determined by the weight computations for the individuals who comprise the 1993 PSID family.

2.B Sample Selection Weight Factor

The primary building block in the construction of PSID core individual weights is the design-based correction for differential probabilities of selection (sample inclusion) for PSID sample persons. The sample selection weight factor equals the reciprocal of the sample inclusion probability for the individual. The sample selection factor for each sample person is constant over time and is uniquely determined by the relationship of that person to a 1968 PSID family unit.

1968 PSID Sample Design: Unequal probabilities of selection were introduced at the beginning of the PSID when the original Office of Economic Opportunity (OEO) sample of poor families was combined with a new equal probability national sample of households selected from the Survey Research Center 1960 National Sample (Hess, 1985). Compensatory weights were developed in 1968 to account for the different sampling rates used to select the OEO and SRC components of the PSID.

Sample selection in 1968 consisted of sampling OEO families or, in the case of the SRC cross-sectional sample, housing units, with known nonzero probabilities of selection. At each sample housing unit, all individuals related by blood or marriage were listed to create the family unit for that year. All individuals within the family unit were included in the 1968 baseline study and followed in subsequent years. Thus, the probability of selection for the family unit, which is the probability of selecting the OEO family or SRC housing unit, applies to all individuals within the original sample of families.

1969-1993 PSID Following Rules and Panel Definitions: The probability sample of individuals defined by the original 1968 sample of PSID families was then followed in subsequent years. A distinction between original sample individuals, including their offspring if born into a responding panel family during the course of the study (i.e., both those born to or adopted by a sample individual), and nonsample individuals was also made. Only original sample persons and their offspring have been followed. These individuals are referred to as *sample persons* and assigned person numbers in a unique range. If other individuals resided with the sample individuals, either in original family units or in newly created family units, data were collected about them as heads, spouses/long term cohabitors, or other family unit members in order to obtain a complete picture of the economic unit represented by the family. However, these nonsample individuals were not followed if they left a PSID family.

Sample Selection Factors for Sample Persons: Sample persons either stayed within the same family from one year to the next, or they moved out to form new family units. Adult sample persons were followed year to year, and as they continued in the same family, or created new families, interviews were attempted with

these sample members or their spouses/long term cohabitors. Data were collected about families, as well as about sample and nonsample individuals, in each subsequent year. Analysts are able to examine either family characteristics for sample individuals, family characteristics for both sample and nonsample individuals, or could investigate sample individuals as individual units. Sets of weights were needed for cross-sectional and longitudinal analysis of both sample persons and families.

Sample persons who are living members of a 1968 PSID family have a sample selection factor equal to the reciprocal of the selection probability for their 1968 PSID family unit.

The computation of the sample selection weight factor for sample persons who are "born into" a PSID family after 1968 uses a formula that is conditional on the "sample status" of their parents. The probability of the born-in-sample individual being in the sample is approximately equal to the sum of the PSID sample selection probabilities of the parents. Therefore the selection weight for born-in-sample individuals is proportionate to the inverse of this sum of parent selection probabilities. If both parents are sample individuals, the weight for the born-in-sample individual is equal to the average of their parents' full analysis weights at the year of their birth. On the other hand, if one parent is a sample individual and the other a nonsample individual, it is assumed that the nonsample individual had a 1968 probability of being selected that is equal to that of the sample parent. Thus, the born-in-sample individual is assigned a sample selection factor equal to one-half the weight of the sample parent (i.e., the average of the known weight for the sample parent and the assumed or imputed weight for the nonsample parent). Once the weight was assigned to the born-in-sample individual, she or he is handled with respect to weighting just as every other sample individual.

Sample Selection Factors for Nonsample Persons: The PSID is designed to be a longitudinal study of individuals and families. Each year, individual data are collected for nonsample persons who enter a PSID family through marriage or residency. Data for nonsample persons present a problem for longitudinal analysis since the time series for these individuals is left censored at the date at which they entered the PSID family. Furthermore, it is not likely that this left censoring is random with respect to the types of variables that might be considered in longitudinal analysis.

Because of the left censoring of their data series, nonsample persons in PSID families have historically been assigned a zero value selection weight factor and a zero-value for the PSID longitudinal analysis weight. The 1993 core longitudinal weight variables included in the 1993 Final Release data file continue this practice. All nonsample persons interviewed in 1993 have a zero value for the individual longitudinal weight.

Over time, there have been several proposals to develop a new weight that would enable PSID analysts to incorporate data for nonsample persons in cross-sectional and short term longitudinal analyses (Kalton, 1987; Little, 1989). Beginning with the 1993 wave, PSID is providing users with a file that includes special weights that will enable analysts to include all 1993 sample and nonsample person respondents in cross-sectional analysis of the 1993 PSID data set. This new series of 1993 core cross-sectional weight values is documented separately and is provided to users as a supplemental data release.

2.C Attrition Adjustment Factors

The second set of factors in the 1993 core longitudinal weights are adjustments for panel attrition due to nonresponse and mortality. Prior to 1993, attrition adjustments were performed in 1969 and every five years thereafter (1974, 1979, 1984 and 1989). The last attrition adjustments to the core longitudinal weights were performed in 1989 and these adjustments were carried forward in the individual and family weights for the 1990, 1991, and 1992 public release files. The 1993 core Longitudinal weights include updated attrition adjustments for the time period 1989-1993.

The process of correcting for attrition involves two steps. The first step requires the building of a model of the joint probability of losing an individual from the panel either due to nonresponse or death. The second

step of the process is to transform the model estimates of the joint probability of nonresponse and survival to a conditional probability of response given that the sample person is still alive. Subsection 2.D explains why this two-step procedure is needed and how these conditional probabilities of nonresponse are derived.

The model of attrition is developed for a specific time period and is always conditional on the surviving set of PSID respondents for the base year of the adjustment period. For example, the attrition adjustment introduced in 1993 is designed to compensate for nonresponse and mortality losses that occurred between the base year (1989) and the close of the 1993 data collection. That attrition process was modeled using only PSID sample persons who were alive and responded in 1989. Since each attrition factor is computed conditionally for each adjustment interval (1968-1969, 1969-1974, 1974-1979, 1979-1984, 1984-1989, 1989-1993), the product of the attrition factors for sequential adjustment periods reflects the total probability of attrition for the combined period 1968 to 1993.

Attrition Adjustment Factors for 1969, 1974,1979, 1984, 1989: In 1989, attrition adjustments for the periods 1968-1969, 1970-1974, 1975-1979, 1980-1984 were recalculated taking into account corrections and updates of the nonrespondent data bases first developed in 1984 as well regular updates of the information in the standard PSID data sets for individuals and families. A new attrition adjustment factor was also computed to compensate for panel attrition during the period 1985-1989.

The attrition adjustments to individual weights for each of the multi-year intervals was a relatively straightforward weighting cell process. An analysis was conducted at each adjustment year to identify groupings of cases into "adjustment cells". The definition of the weighting cells was intended to maximize the variation in response rates among the defined cells. Sample individuals within cells who responded in the final year of the period had their weights increased to compensate for nonresponding individuals within the same cells.

A detailed description of the procedures used to develop the attrition adjustment factors for 1969, 1974,1979, 1984, and 1989 is provided in Section I, Part 5 of the 1992 documentation.

1993 Nonresponse/Attrition Adjustment: The 1993 core longitudinal weights for sample persons include a model-based adjustment factor for panel attrition that occurred in the 1990-1993 annual PSID data collections. It is important to note that this adjustment includes attrition from all known sources: nonresponse of living panel members; known deaths to panel members; and loss to follow-up where the death of the sample person is not known. A mortality correction factor based on age, sex and race-specific standardized rates is applied to the individual weights to correct for the obvious bias due inclusion of both living and dead sample persons in the development of the attrition weight. (See 2.D for a discussion of the mortality correction to the attrition adjustment.)

Prior attrition adjustments (1969, 1974, 1979, 1984 and 1989) to PSID core individual weights were developed using a weighting class methodology (Little and Rubin, 1987). In 1993, a new methodology for estimating the attrition model and corresponding adjustment factors was applied. The probability that a sample person remained in the core panel at the 1993 interview was estimated as a linear logistic function of a set of covariates that include individual and family characteristics as measured in 1989. The 1989 individual and family data are used as the model covariates, since 1989 is the last year when complete data are available for some sample persons who were lost from the PSID panel in the period 1990-1993.

The dependent variable in the logistic "response propensity model" is a dichotomous indicator variable coded 0 for sample persons who responded in 1989 but did not respond in 1993 and coded 1 for sample persons who responded in both 1989 and 1993. 1989 PSID core sample persons who are known to have died in the period 1990-1993 are also coded 0—they are treated as part of the panel attrition in the calculation of the attrition adjustment factors. Nonsample persons and sample persons who were lost to follow-up prior to 1989 are excluded from the estimation of the model. Likewise, sample persons who were born into the PSID core panel after 1989 are not included in the estimation of the response propensity model.

The set of 1989 individual and family variables used as covariates in the logistic regression includes: whether the individual moved in/out in 1989; relationship to head or spouse (whether the relationship is lineal vs. collateral or none), sex-race-age group (race was divided into white and other groups), whether SRC or SEO sample, occupation of 1989 family Head (white-collar or not), marital status of Head, and whether Head was unemployed more than 36 weeks in 1988. Model selection was based on the Schwarz criterion, which takes the number of covariates into account. The final logistic model specification with coefficients is shown in Table 1.

The following equation is then used to transform the estimated logit functions to the uniform probability scale.

$$p(response) = \frac{e^{x\hat{\beta}}}{1 + e^{x\hat{\beta}}}$$

$where : X$ is the vector of covariates; and
    ~

$\hat{\beta}$ is the vector of estimated logistic regression coefficients.
~

The response probabilities for all sample persons who responded in 1993 are then ranked into deciles. In order to prevent improbable and meaningless outliers (i.e., with extremely high or low predicted probabilities of response), cases with predicted probabilities in the lowest and highest decile receive the median value from the respective decile, while cases in the eight central deciles retain the predicted response probabilities. The 1989-1993 attrition adjustment factor applied to each 1993 sample person respondent is then computed as the reciprocal of his/her estimated probability of response.

The attrition adjusted individual weight then becomes the product of the 1992 core weight or reference weight (if nonresponse in 1992 but not in 1989) and the attrition adjustment factor.

Table 9
1993 PSID CORE LONGITUDINAL WEIGHT
Logistic Model of Attrition 1989-1993

Covariate	Parameter Estimate
Intercept	-2.7423
Individual is "mover-in" or "mover-out"	0.7194
Relationship to head/spouse	0.5183
Individual is male and head is white	0.4333
Individual is male and head is black	0.1586
Individual is male and head is Hispanic	0.2485
Individual is less than 25 years	0.4171
Individual is 25-44 years old	0.3652
Individual is 45-64 years old	0.3525
SEO sample indicator	0.3427
Head's occupation (white collar vs. others)	-0.3718
Head is married	-0.3062
Head has never been married	-0.0838
Head is widowed	-0.1434
Head was unemployed/on strike for 36 weeks or more in 1988	-0.6437

2.D Mortality Correction to the Attrition Adjustment Factors

Explanation of the Mortality Correction: In theory, the death of sample members of longitudinal panels should require no correction. Deaths of sample persons can be viewed as representative of the mortality in the PSID survey population. Mortality reduces the size of the panel sample over time, but it should not in theory diminish the representativeness of the survey population at a given point in time. Unfortunately, the PSID and similar longitudinal surveys that follow sample members over time share a problem in that the determination of nonresponse and mortality status is confounded in the data on sample dispositions for panel members. A major source of PSID nonresponse is inability to relocate/recontact the sample person for the annual interview. Sometimes spouses, relatives or neighbors will inform the interviewer that the respondent has died, and the death is recoded in the annual disposition code for the case. However, in other cases when death occurs, there is no report of the death, and the case can only be assumed to be a living nonrespondent.

The confounding of attrition due to nonresponse and mortality becomes even more serious as future waves of longitudinal data are considered. Barring special recontact efforts such as those fielded in the 1993 PSID, little or no information is obtained on deaths that occur to nonrespondents after they have left the PSID core panel. Uncorrected attrition adjustments applied to the individual weights for responding sample persons in 1993 assume that the nonrespondents are alive in 1993 and will continue to live indefinitely. An attrition adjustment that ignores current and future mortality among nonrespondent panel members will in the long run lead to increasing bias in the weighted representation of the living sample respondents. Since non-respondents are no less mortal than respondents, the attrition adjustment factors must be corrected for the mortality that is occurring in the pool of nonresponding sample persons. The PSID individual weights use a correction that is based on standardized age, sex and race specific mortality tables for the United States population.

Statistical Explanation of the Mortality Correction Factors: The mortality correction to the attrition adjustment factors can be formalized as follows. Let $R_t = 1$ indicate that a sample person responds at year t of the PSID data collection and $R_t = 0$ denote nonresponse in year t. Similarly, let $S_t = 1$ denote that a sample person is alive at year t and $S_t = 0$ denote that the sample person has died prior to year t data collection. The joint probability of responding and surviving to year t thus becomes:

$$\Pr\{R_t = 1 \text{ and } S_t = 1\} = \Pr\{S_t = 1\} * \Pr\{ R_t = 1|S_t = 1\}.$$

That is, the joint probability of a sample individual surviving a period and responding in year *t* is the product of the probability that he/she survived and the conditional probability that he/she responded, given that he/she survived. The reciprocal of the conditional probability, $[\Pr\{R_t = 1| S_t = 1\}]^{-1}$, is the adjustment factor that corrects for pure nonresponse among living sample persons. Since mortality and nonresponse in longitudinal surveys are confounded over time—we do not know which nonrespondents are alive at time *t*—the conditional probability cannot be directly estimated from PSID sample control data. However, it can be estimated as the ratio of two known or estimable probabilities:

$$\Pr\{R_t = 1|S_t = 1\}=\Pr\{R_t = 1 \text{ and } S_t = 1\}/\Pr\{S_t = 1\}.$$

The numerator of this ratio is the joint response and survival probability that is estimated in the combined attrition model described in subsection 2.B. The denominator, $\Pr\{S_t = 1\}$, is estimated independently using standardized age, sex and race specific mortality rates derived from United States Vital Statistics data published by the National Center for Health Statistics.

1969, 1974, 1979, 1984 and 1989 Mortality Corrections: The 1969 mortality correction factor was computed directly using 1969 single-year-of-age mortality rates prepared by the U.S. National Center for Health Statistics (NCHS).

The 1974, 1979, 1984 and 1989 attrition adjustments and mortality correction factors apply to five years of PSID data collection. For example, five-year mortality rates for persons ages 10-14 at the middle of the adjustment period were applied to individuals who were 8-12 at the beginning of the period.

One further adjustment was made to the mortality rates before using them in the 1969 to 1989 PSID mortality adjustment process. Rates were not available for the oldest age group (85 years or older in 1969, 83 years or older in later years). Five-year mortality data for those older ages were pooled across four age-race groups: white males, white females, nonwhite males, nonwhite females. A linear model was then fitted for each race-sex group regressing the logarithm of the rates on the mid-point of five-year intervals for persons in five-year age intervals starting with the interval 40-45 years and extending through the interval 80-85 years. The model coefficients were then used to extrapolate beyond the available data to provide a predicted mortality rate for the oldest age group, 85 years and older. The race-sex specific predicted rates from the pooled data were then used in each mortality table for the oldest age group.

1993 Mortality Correction: The 1993 mortality correction applies to the four year period 1990-1993. The correction is computed as the compound product of age, sex and race specific survival rates that correspond to the age of the sample person in 1990, 1991, 1992, and 1993. The single-year-of-age mortality rates for each race by sex subclass were obtained from the 1992 Vital Statistics data published by the National Center for Health Statistics. The actual rates used in the calculations are provided in Table 10. Mortality rates for single years above 85 were estimated using the 1992 pooled rate for the 85+ category and historical data on the relative size of single-year-mortality rates from 85-100 (Shryock and Siegel, 1971) .

2.E Final Core Longitudinal wieghts for Sample Persons

The final 1993 core longitudinal analysis weight for sample persons is computed as the product of the sample selection factor (2.B), the attrition adjustment factor (2.C) and the mortality correction (2.D). A nonzero weight value is assigned to each PSID sample person except 1993 recontact cases that are missing a substantial number of waves of data (see 2.H).

2.F Computation of Newborn Weights

Weights for 1993 core born-in-sample (newborn) cases are computed as the average of the parents' weights if there are two sample parents, and one-half the parent's weight if there is one sample parent in the family.

2.G Longitudinal Weights for Recontacts/Reappear Cases

Beginning in 1992, efforts were made to recontact and reinterview previous years' nonrespondents. In 1992, the number of successful recontacts was about 300 while in 1993 there were over 1000 recontacts/reappears. The total for 1993 recontacts/reappears is no longer a trivial number. Therefore, recontacts can no longer be assigned a weight based on their last appearance in PSID. Instead, for the 1993 core longitudinal weights, these sample persons were assigned a zero weight if they had not been interviewed in 1988 or 1989 in order to prevent serious bias in analysis. From a user's point of view, these sample persons could not well be used in longitudinal analysis anyway, because they have a long stretch of missing data. (See 1.C.)

As mentioned above, reappears are handled differently from previous years. Beginning a major recontact effort caused the number of reappears to rise dramatically. As a consequence, a lot of people for whose loss we had already compensated by increasing similar respondents' weights are coming back into the interview cycle. But we cannot justify including them all with their reference weights, since this would throw off the balance created by taking their attrition into account. Hence, for longitudinal analysis weights we had to exclude them from receiving a positive weight if they had no interview in 1988 or 1989. People who were only short-term nonrespondents were allowed back in, as had been the practice before.

2.H Final 1993 Core Longitudinal Weight for Families

Family weights are constructed as the average of the weights of all members of the family. Non-sample family members carry a zero weight and hence tend to decrease the family weight. Children already have their parents' average weight, so they tend not to affect the family weight unless other people are present.

Table 10
1993 PSID CORE LONGITUDINAL WEIGHT
Mortality Correction to Attrition
Mortality Rates by Race and Gender for Single Year of Age

Age	White Male	White Female	Other Male	Other Female
0-1	0.00768	0.0061	0.01571	0.01309
1-2	0.000624798	0.000462823	0.000985482	0.000790346
2-3	0.000453766	0.000352311	0.000701704	0.000638861
3-4	0.000343001	0.000281948	0.000529192	0.000507357
4-5	0.000272477	0.000221593	0.000448015	0.000406091
5-6	0.000242268	0.000191418	0.000407469	0.000325005
6-7	0.000232229	0.000161225	0.000387253	0.000264152
7-8	0.000222184	0.000151173	0.000356819	0.000223573
8-9	0.00020203	0.000131036	0.000305954	0.000193129
9-10	0.000171761	0.000120973	0.000234636	0.000172833
10-11	0.000141474	0.000110905	0.000173468	0.000172863
11-12	0.000151601	0.000121001	0.000173498	0.000183063
12-13	0.000212273	0.000151269	0.000275603	0.00020344
13-14	0.000353864	0.000201723	0.000520727	0.000244178
14-15	0.000536041	0.000262292	0.000878546	0.00029512
15-16	0.000758956	0.000343088	0.001278079	0.000346105
16-17	0.000951947	0.000413866	0.001658511	0.000417507
17-18	0.001115042	0.000464529	0.001968908	0.000468618
18-19	0.001197471	0.000474848	0.002188566	0.000519798
19-20	0.001249708	0.000464966	0.002316936	0.000571056
20-21	0.00128179	0.000444957	0.002456495	0.000622398
21-22	0.00132418	0.000445155	0.002597053	0.000673834
22-23	0.001356534	0.000435231	0.002697179	0.000735587
23-24	0.001389017	0.000445547	0.002756483	0.000797473
24-25	0.001421631	0.000455876	0.002795394	0.000869734
25-26	0.001433897	0.000476355	0.00281369	0.000931938
26-27	0.001466727	0.000496862	0.002842608	0.001004562
27-28	0.001509969	0.000517399	0.002903307	0.001087659
28-29	0.001573977	0.000537967	0.00301726	0.001160748
29-30	0.001669191	0.000568724	0.003174536	0.001244369
30-31	0.00176487	0.000599533	0.003333263	0.001328295
31-32	0.001871381	0.000640563	0.003504175	0.001422856
32-33	0.001968117	0.00068167	0.003687513	0.001528136
33-34	0.002055029	0.00072286	0.003894265	0.001644227
34-35	0.002152864	0.000774325	0.004103348	0.001771231
35-36	0.002251313	0.000836103	0.004336542	0.001898892
36-37	0.002360856	0.000908237	0.00459438	0.002037655
37-38	0.002471152	0.000970348	0.004844729	0.002177242
38-39	0.002582243	0.001032635	0.005109537	0.002317715
39-40	0.002683646	0.001105346	0.005367219	0.002459136
40-41	0.002817496	0.001188537	0.005651032	0.002612062

Age	White Male	White Female	Other Male	Other Female
41-42	0.002952443	0.001282275	0.005961733	0.002787185
42-43	0.00312039	0.001386635	0.006288956	0.002963729
43-44	0.003289859	0.001522274	0.00663335	0.003152372
44-45	0.003482311	0.001668813	0.007006984	0.003363966
45-46	0.003708865	0.00182638	0.007433839	0.003577627
46-47	0.003970133	0.002015796	0.007892796	0.003836249
47-48	0.004266811	0.002227033	0.008362078	0.004129926
48-49	0.004610545	0.002470776	0.00884251	0.0044702
49-50	0.005013351	0.002747481	0.009346788	0.004847331
50-51	0.005454844	0.003057689	0.00986438	0.005251484
51-52	0.005958347	0.003391569	0.010468618	0.005694549
52-53	0.006525882	0.003749724	0.01124895	0.006221831
53-54	0.007182129	0.004143384	0.012238817	0.006858104
54-55	0.007908069	0.004552336	0.013425078	0.007573733
55-56	0.008718396	0.005009199	0.014719632	0.008372239
56-57	0.009583205	0.005504724	0.016042369	0.009189887
57-58	0.010563949	0.006061844	0.017398671	0.009983323
58-59	0.011667483	0.006682743	0.018714769	0.010741647
59-60	0.012890215	0.007359105	0.020031358	0.011488086
60-61	0.014193548	0.008104581	0.021406306	0.012246921
61-62	0.015597731	0.008900535	0.022903453	0.013091569
62-63	0.017112958	0.009739064	0.024493329	0.014051947
63-64	0.0187637	0.010623437	0.026202552	0.015160629
64-65	0.020526699	0.011568723	0.028031281	0.016391195
65-66	0.02242959	0.012568111	0.029933291	0.017716161
66-67	0.024464057	0.01367306	0.031952072	0.019093571
67-68	0.026594509	0.01491531	0.034221165	0.020556863
68-69	0.0288093	0.016305849	0.03682905	0.022129717
69-70	0.031152873	0.017857797	0.039721715	0.023784021
70-71	0.033631752	0.019549601	0.042929246	0.02560135
71-72	0.036315351	0.021384477	0.046294855	0.027529955
72-73	0.039318102	0.023341364	0.04958273	0.02945237
73-74	0.04269513	0.025437666	0.052577342	0.031309748
74-75	0.046463415	0.027693857	0.055381826	0.033174383
75-76	0.050591051	0.030190211	0.058244974	0.035178259
76-77	0.055095836	0.03296291	0.061414234	0.037493441
77-78	0.059937679	0.036086321	0.065161343	0.040255355
78-79	0.065189133	0.039635846	0.069616211	0.043538013
79-80	0.070917055	0.043657703	0.074887668	0.047473993
80-81	0.077278056	0.048229939	0.081050998	0.052112874
81-82	0.084398908	0.053507855	0.088126262	0.057587846
82-83	0.09256252	0.059583068	0.096160039	0.064011348
83-84	0.102036832	0.066681259	0.105232688	0.071588367
84-85	0.113341546	0.075048504	0.115269759	0.080632206
85-86	0.12195	0.08302	0.12195	0.09208
86-87	0.13067	0.09208	0.13067	0.10219
87-88	0.14380	0.10219	0.14380	0.11244
88-89	0.15816	0.11244	0.15816	0.12195
89-90	0.17355	0.12195	0.17355	0.13067
90-91	0.19032	0.13067	0.19032	0.14380
91-92	0.20835	0.14380	0.20835	0.15816

Age	White Male	White Female	Other Male	Other Female
92-93	0.22709	0.15816	0.22709	0.17355
93-94	0.24598	0.17355	0.24598	0.19032
94-95	0.26477	0.19032	0.26477	0.20835
95-96	0.28284	0.20835	0.28284	0.22709
96-97	0.29952	0.22709	0.29952	0.24598
97-98	0.31416	0.24598	0.31416	0.26477
98-99	0.32915	0.26477	0.32915	0.28284
99-100	0.34450	0.28284	0.34450	0.29952
100-101	0.36018	0.29952	0.36018	0.31416
101-102	0.37616	0.31416	0.37616	0.32915
102-103	0.39242	0.32915	0.39242	0.34450
103-104	0.40891	0.34450	0.40891	0.36018
104-105	0.42562	0.36018	0.42562	0.37616
105-106	0.44250	0.37616	0.44250	0.39242
106-107	0.45951	0.39242	0.45951	0.40891
107-108	0.47662	0.40891	0.47662	0.42562
108-109	0.49378	0.42562	0.49378	0.44250
109-110	0.51095	0.44250	0.51095	0.45951
110-111	0.52810	0.45951	0.52810	0.47662
111-112	0.54519	0.47662	0.54519	0.49378
112-113	0.54519	0.49378	0.54519	0.51095
113-114	0.54519	0.51095	0.54519	0.52810
114-115	0.54519	0.52810	0.54519	0.54519
>115	0.54519	0.54519	0.54519	0.54519

3. 1993 Latino Longitudinal Analysis Weights

This section describes the development of the 1993 Latino longitudinal analysis weights for sample persons and families.

3.A Overview of the Weight Calculation

The methodology used to develop the 1993 Latino longitudinal analysis weights follows the same basic steps used to compute the 1993 core longitudinal weights (see 2.B-2.D).

The 1993 Latino individual longitudinal weight values for sample persons are also the product of three distinct sets of factors: (1) a single factor that represents the reciprocal of the probability by which the sample person was selected to the PSID Latino supplement; (2) a compound product of longitudinal attrition adjustment factors developed in 1991,1992 and 1993; and (3) mortality adjustment factors also developed and applied in 1991,1992, and 1993. A general formula that reflects the composite nature of the 1993 Latino individual longitudinal weight is therefore:

$$W_{i,1993} = W_{i,sel} \times \prod_{j=1990}^{1993} \left[W_{i,NR(j)} \times W_{i,M(j)} \right]$$

$where: W_{i,sel} =$ selection weight factor – the reciprocal of probability that individual i is selected to the PSID panel by membership in a 1990 Latino sample household or by birth to a Latino sample parent.

$W_{i,NR(j)} =$ the attrition adjustment factor applied to the ith individual weight at time period $j \ \varepsilon$ (1990, 1992, 1993).

$W_{i,M(j)} =$ the age, sex and race-specific mortality adjustment applied to the ith individual weight at time period $j \; \varepsilon$ (1990, 1992, 1993).

Longitudinal analysis weights for PSID Latino sample persons who were alive in 1990 are the product of all seven factors in this basic equation. Individual weights for children of Latino sample persons "born into" the PSID after the 1990 baseline year include a sample selection factor and nonresponse/mortality adjustment factors only for the adjustments periods following the year of their birth. Not shown in this general formula is the fact that in 1992, a poststratification factor was also applied to the individual weights of Latino sample persons to align the weighted sample proportions to 1989 Current Population Survey (CPS) percentages for the three major ethnic groups. Nonsample members of Latino supplement families receive an individual longitudinal weight of zero.

3.B Sample Selection Factor for the 1990 PSID Latino Sample

The 1990 PSID Latino supplement is derived from an area probability sample of family unit members associated with all respondents to the Latino National Political Survey (LNPS). The LNPS was conducted in 1989 by Temple University for the University of Texas at Austin. For more detailed information about the LNPS sample, the reader is referred to Section I, Part 13 of the 1990 documentation.

In the spring of 1990, Survey Research Center (SRC) interviewers attempted to contact each 1989 LNPS respondent. (No attempt was made by the PSID to follow LNPS nonrespondents.) The 1990 PSID Latino supplement family unit was defined to include all related persons living with the 1989 LNPS respondent at the time he or she was recontacted and interviewed by the PSID. These 1990 family members were designated as "original" sample persons in the PSID Latino supplement. Persons who were living with the LNPS respondent in 1989 but not at the time of the 1990 PSID interview were listed as "movers out" on the PSID coversheet and were not considered sample persons in the Latino supplement. If the 1989 LNPS respondent was in jail or in prison at the time of the 1990 PSID interview, the PSID family consisted of all related persons in the household with which the 1989 LNPS respondent was last associated.

Sampling statisticians at the Temple University Institute for Social Research provided the sample selection probabilities for 1989 LNPS cases to the PSID staff. The sample selection factor for the Latino supplement individual weights were computed from the 1989 LNPS respondent selection probabilities. The 1990 PSID documentation provides a detailed description of the derivation of the original sample selection factors for 1990 Latino supplement individuals and families.

3.C Attrition Adjustment Factors

1989-1990 and 1990-1992 Latino Supplement Attrition Adjustments: The 1992 Latino weights for individuals account for the differential selection probabilities among the three Latino groups: Mexican-Americans, Puerto Ricans, and Cuban-Americans, and for the panel attrition in the periods 1989-1990 and 1990-1992.[2] In addition, the 1992 weights included a poststratification adjustment to 1989 Current Population Survey (CPS) distributions by age and gender within Latino group. Procedures used to develop the 1989-1990 attrition adjustment are described in the 1990 documentation. The 1990-1992 attrition adjustment and poststratification of the 1992 Latino weights is documented in the 1992 documentation.

1992-1993 Latino Supplement Attrition Adjustments: The procedure used to develop the 1992-1993 Latino supplement attrition adjustment is identical to that described in subsection 2.C for the 1992-1993 core sample attrition adjustment.

Repeating the procedure used for the 1992-1993 attrition adjustment for the core sample, a logistic regression procedure was used to model 1993 response probabilities for the 7,363 Latino sample persons

---

[2] The 1991 Latino weight did not include an adjustment for nonresponse between 1990 and 1991.

who responded in 1992. The dependent variable is a response indicator which is 1 if the 1992 sample person responded in 1993 (n=6317) and 0 if the 1992 sample person was a 1993 nonrespondent (n=1046).

The final logistic model of response propensity included the following independent variables: (1) whether a person moved in or out in 1992; (2) the relationship to the head or spouse (whether lineal vs. collateral or none); (3) sex of 1992 respondent; (4) age group in 1992: <25, 25-44, 45-64, 65 and over; (5) whether or not the head of the family is in a professional, technical, or managerial occupation; (6) head's marital status; (7) whether or not the head had 36 or more weeks unemployed or on strike in 1992; (8) whether or not the 1992 interview was conducted in English; and (9) Hispanic group (Mexican-American, Puerto Rican or Cuban). The first seven variables were also used in the PSID core attrition adjustment for 1992-1993. The last three were specifically chosen for the Latino supplement attrition adjustment. The last variable, Hispanic group, has a high correlation with region of the country—a variable which was used in the core model but not used in the Latino model. Table 11 shows the independent variables in the logistic regression model and the parameter estimates for each.

Table 11
1993 PSID LATINO LONGITUDINAL WEIGHT
Logistic Model of Attrition 1992-1993

Variable	Parameter Estimate
Intercept	-2.0412
Individual is "mover-in" or "mover-out"	1.1403
Relationship to head/spouse	0.6535
Individual is male and head is white	0.0375
Individual is less than 25 years old	0.5136
Individual is 25 - 44 years old	0.5444
Individual is 45 - 64 years old	0.3144
Head's occupation	-0.4872
Head is married	-0.4682
Head has never been married	0.2198
Head is widowed	0.0600
Head was unemployed/on strike 36 weeks or more in 1991	0.6286
Interview in English	-0.3601
1992 family is of Mexican descent	-0.0729
1992 family is of Puerto Rican descent	0.2545

3.D Mortality Adjustment

The procedure for applying the mortality adjustment to Latino panel attrition weight factors is identical to that described in subsection 2.D for the 1993 core longitudinal weight. Since the nonresponse and mortality corrections for the Latino longitudinal weights were updated for 1992, the 1993 attrition and mortality adjustments are required only for the transition from 1992 to 1993. Single-year mortality tables which give the death rates by age, sex, and race were obtained from the Monthly Vital Statistics Report published by the Centers for Disease Control and Prevention, National Center for Health Statistics. Although it was possible to obtain death rates for five-year age groups for Hispanics, these statistics were not used because they appear to seriously underestimate the death rates for Hispanics. Death certificates often do not identify a person as Hispanic; therefore the numerator of the Hispanic death rate is too low. The mortality tables for whites were used for the 1993 Latino mortality adjustment (see Table 10).

3.E Computation of Weights for Newborns and Reappears

Individual weights for the 1993 Latino sample newborns are computed as the average of the weights of the head and spouse (parents of the newborn). There were 23 respondents in 1993 who were not interviewed in 1992. These "reappears" did not have a 1992 Latino weight. The 1993 weight for these cases is either the individual reference weight if available or otherwise the family reference weight.

3.F Constructing Family Weight and Updating Reference Weight

Family weights for the 1993 Latino supplement are constructed as the average of the weights of all members of the family. Nonsample family members carry a zero weight and hence tend to decrease the family weight. Children already have their parents' average weight, so they tend not to affect the family weight unless other people are present.

4. 1993 Combined Longitudinal Analysis Weight

4.A Overview

If an analyst wishes to create a subset that includes members from both the core and the Latino samples, neither the core nor the Latino weight is appropriate for a pooled analysis of the data.

The calculation of the separate core and Latino longitudinal analysis weights do not take into account the fact that sample persons in the core sample could have been selected as part of the Latino sample as well, nor do the calculations take into account the possibility that many sample members in the Latino sample could have been selected as part of the original PSID in 1968. The joint probability of selection of these "overlapping" core and Latino sample members is not accounted for in the core or Latino sample weights for either families or individuals. Therefore, a separate "combined weight" has been constructed for analyses which pool cases from these two components of the PSID sample.

The calculation of the 1993 combined longitudinal weights also addresses a small problem of the relative scaling of the values of the core and Latino weights. Scaling of weights (dividing or multiplying weights by a constant) has no effect on the values of weighted estimates of survey statistics. Some survey analysts scale weights so that the sum of the weight values equals the nominal sample size. Weights scaled to total population values may be found in other survey data sets. Either scaling is appropriate, but when weighted data sets are pooled in analysis the combined weights must scaled to a common base that ensures the correct population representation of each weighted case. Prior to 1989, the PSID core sample weights had been constrained to be integer values from 1 to 99 to save file space on the very long merged data records. The sum of these relative weights was arbitrary, bearing no direct relationship to the total U.S. population, the PSID unweighted sample size, or any other particular feature of the data. The original 1990 Latino weights, on the other hand, were obtained from the LNPS base weights. The LNPS weights were scaled to a different reference base. The 1993 combined core-Latino longitudinal analysis weights have been scaled to guarantee that weighted sample totals for pooled analyses will conform to the correct proportionate representation of Latinos and non-Latinos in the U.S. population.

4.B Calculation of the 1993 Combined Core-Latino Longitudinal Analysis Weights

The first step in the calculation of the 1993 combined longitudinal weights is to determine which cases could have been selected from both samples.

Sample persons in the 1993 core sample persons are divided into the following four groups:

1. Non-Latino or resident in a non-LNPS county in 1989.

2. Latino and residing in one of the 382 LNPS counties in 1989, in the following areas:

  a. Florida (the "Cuban" sample);

b. the Northeast states of Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, or Vermont (the "Puerto Rican" sample);

c. any one of the remaining states of the coterminous U.S. (the "Mexican" sample).

The first group of core sample persons was not eligible for the LNPS sample. The remaining three groups contain individuals who could have been selected in the LNPS sample. The second, third and fourth groups include a number of 1993 core sample persons individuals who were not in the study in 1990 because they had joined a panel family in 1991 or 1992 or 1993 by moving in[3] or by being born in. These individuals were assigned to the area designation of the family into which they were born or had moved.

The 1993 core sample weights of those individuals in the first group can be used directly in analyses based on combined samples, after adjusting for the differences in scaling of the core and Latino weight variables (see above). However, sample persons in the other three groups were jointly eligible for the PSID core and the LNPS. Joint selection probabilities for the LNPS sample persons depend on the 1990 geographic sample area in which they resided—those in the Cuban sample geographic area having the highest joint selection probabilities and those in the Mexican geographic sample area the lowest. The weight for core sample persons in these latter three groups must reflect their joint probability of selection for the core and Latino samples. The joint probability of selection of a sample person in these three groups is the sum of the probability of selection for the core sample and the probability of selection for the Latino sample group for which they were eligible, minus the product of those two probabilities. The calculation of these joint probabilities of selection for the core and all of the Latino sample groups is detailed in the Wave XXIII (1990) documentation.

Sample persons in the Latino supplement are assigned to six groups:

1. Eligible for core sample in 1968 and residing in 1989 in

    a. Florida (the "Cuban" sample);

    b. the Northeast states of Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, or Vermont (the "Puerto Rican" sample);

    c. any one of the remaining states of the coterminous U.S. (the "Mexican" sample).

2. Known ineligible for core sample and residing in 1989 in

    a. Florida (the "Cuban" sample);

    b. the Northeast states of Connecticut, Delaware, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, or Vermont (the "Puerto Rican" sample);

    c. any one of the remaining states of the coterminous U.S. (the "Mexican" sample).

The difficult task of determining 1968 core sample eligibility for all 1990 Latino sample persons was initially completed in 1990. Core eligibility for some Latino sample persons who first appeared after 1990 was imputed. For sample members who were born in or "appeared"[4] for 1993 and who were born after 1968, eligibility was determined from 1990 parental information, if possible, according to the following rules. If one parent was known to have been in the U.S. in 1968 or if both parents were known not to have been in the U.S. in 1968, then their eligibility was assigned to the new person. If, however, both parents' eligibility was unknown or one was unknown and the other was known not to have been in the U.S., then eligibility

---

[3] These movers-in had been in the study previously, were nonresponse in 1990, but were selected for recontact in 1992, or subsequently rejoined a response family in 1991 or 1992.

[4] An "appeared" individual should have been included in the 1990 family but was not listed until 1992 because the respondent failed to mention him or her during the 1990 interview. Such an individual is considered original sample.

was imputed as per the 1990 procedure in such cases. If an individual was born by 1968, then eligibility was always imputed. Imputation of 1968 core sample eligibility for cases with missing information in 1990 is detailed in the 1990 (Wave XXIII) documentation.

The probability of selection of individuals for the last three groups who were ineligible for the 1968 core was derived directly from the inverse of the current Latino sample weight.

The joint probability of selection for individuals in the first three groups was calculated as the sum of the core and Latino sample selection probabilities, minus the product of those probabilities. Calculation of the core probability for these groups is detailed in the 1990 (Wave XXIII) documentation. To state that briefly, a single overall selection probability was derived for these individuals. The selection probability for core sample persons is 15,755/238,451,375, or 1 in 15,135. For the Latino sample groups, the calculation yielded the selection probabilities shown in Table 12.

Table 12
SELECTION PROBABILITIES FOR LATINO SAMPLE GROUPS

Group	1990 Census	1990 Latino Sample	Probability	Inverse
Mexican	13,495,938	4,304	0.00031891	3135.67
Puerto Rican	1,043,932	1,215	0.00116387	859.20
Cuban	2,727,754	980	0.00035927	2783.42

The joint probability of selection for each sample person in the first three core-eligible Latino groups as well as for the Latino-eligible sample persons in the core sample was calculated as for the 1990, 1991 and 1992 data. Details of this calculation for all ten groups, four core and six Latino, are available from PSID staff upon request.

A poststratification adjustment was then applied to bring the sum of combined weights into agreement with 1990 Census population distributions. Prior to this adjustment, the 1993 combined weight produced the following weighted sample proportions for the three Latino ethnic groups and all other persons: 5.20% (Mexican-American), 0.42% (Cuban-American), 1.01% (Puerto Rican) and 93.37% (Others). The 1990 Census percent of population for these same four groups is: 5.43% (Mexican-American), 0.42% (Cuban-American), 1.10% (Puerto Rican) and 93.06% (others). Adjusting corresponding weights for each of the 10 groups yielded a poststratified weight that matched the overall Census distribution.

As a final step, the values of the individual combined core-Latino longitudinal weights were rescaled to sum to the same total as the 1993 core Longitudinal weight (i.e., 342,286).

Table 13 provides a summary of the distribution of the final combined weight values for each of the ten groupings of core and Latino cases.

The 1993 combined core/Latino weight for families is derived as the average of the individual weight for all persons in the family.

We recommend that anyone conducting analyses that combine cases across the core and Latino samples use the combined weights. For 1993, the family-level combined weight variable is V23363; at the individual level, V30866. For comparisons of core and Latino sample groups or analysis of only the core or only the Latino samples, the core (family: V23361; individual: V30864) or Latino (family: V23362; individual: V30865) weights are sufficient to account for unequal selection probabilities and nonresponse.

40

Table 13
1993 PSID LONGITUDINAL WEIGHTS
COMBINED WEIGHT MEANS

Group	Mean	Standard Deviation	Minimum	Maximum
Core:				
Non-Latino*	21.34	16.58	0.27	138.89
Mexican	3.75	2.50	0.30	10.54
Cuban	1.12	0.60	0.17	2.55
Puerto Rican	3.55	2.36	0.51	9.16
Latino, core eligible:				
Mexican	4.51	3.38	0.48	16.26
Cuban	1.19	0.60	0.31	4.38
Puerto Rican	4.05	2.69	0.62	17.65
Latino, core ineligible:				
Mexican	3.75	2.75	0.36	15.92
Cuban	1.12	0.66	0.29	4.29
Puerto Rican	3.55	2.81	0.43	16.26

* Non-Latino means ineligible for the LNPS sample. The LNPS sample was restricted to areas deemed likeliest to have Latino populations and included only those of Mexican, Puerto Rican or Cuban descent.

Linking Data: Splitoffs

From the 1981 wave onward, data have been provided to assist the user in linking splitoff records with those of their main families. The family-level data for each reinterview or recontact main family in 1993 (V21606=1 or 3) contain values for V23342 representing the actual number of successfully interviewed 1993 splitoff families generated from this family. Thus, splitoff nonresponse cases are not included. On each splitoff data record (V21606=2 or 4), the family portion of the record contains the current year's interview number (V21602) of the associated main family at V23343. The *individual-level* record of each member of a splitoff family also contains this interview number (V30858), as well as month and year the splitoff family was formed (V30856 and V30857).

The month and year in which the splitoff family was formed are derived from actual move-out dates of splitoff individuals as reported on the main family coversheets. Thus, in the relatively rare event that two or more individuals move at different times from the main family to form one splitoff family, each individual receives his or her actual date of move as code values for V30856 and V30857. Any other splitoff individuals who did not move out of a main family but simply appeared for the first time in the splitoff family, such as nonsample spouses, friends, miscellaneous relatives and newborn children, receive the move-out date given for the splitoff mover-out. When more than one splitoff mover-out date exists (a rare occurrence), these new persons receive the earlier date.

For those individuals who move from institutions to form their own splitoff families, code values of 98 are inserted into the file positions for both month and year. The other miscellaneous splitoff family members appearing in the study for the first time receive missing data code values of 99 for these month and year variables.

Linking Data: Families Sharing Households

It is not uncommon for two or more family units to share living quarters. Panel families involved in such a situation may live with persons or families who are not sample members and who are not included by the study as family members because the arrangement is supposedly temporary. The situation resembles that of roommates, where expenses are split between the individuals involved. However, the members of one panel

family may also move in with the members of another panel family. Financial disasters such as divorce and unemployment contribute heavily to such patterns of behavior. Most frequently, a former splitoff child, already separately interviewed, returns home to live with panel parents for a period of time until resuming life on his or her own. Occasionally, siblings who are each being interviewed move in together to share an apartment, or aging panel parents go to live with their panel children. We continue to interview each of these smaller groups separately, as if they were living apart.

We provide information to identify situations of multiple family units sharing the same household and to facilitate the linking of PSID family units in the same household. The precise set of variables used for these purposes has varied over the course of the study, and some were added retroactively. Section I, Part 5 of the 1992 documentation details the history.

As explained in Part 2, because of the change in data collection procedures and our greater reliance on machine generation of variables, we are no longer able to produce the information specifically identifying the householder, that individual among all the family members sharing the household who would be considered Head were it not for our separate interviews. The five householder variables, consisting of the householder's identifiers, age, sex and relationship to family unit Head, are no longer coded. We also no longer code the household composition variable, which indicated primacy of families within the household.

However, we have retained the set of variables describing and identifying each PSID family unit sharing the same household. As many as five PSID family units were living in the same dwelling unit in 1993, and so four sets of these variables are included for 1993 data. For each other PSID family unit in the household, the Family Interview (ID) Number is given as a separate variable labeled ID for First Other Family Unit, ID for Second Other Family Unit, etc. (1993: V23344, V23347, V23350, V23353). The ID number variable is filled with zeroes if no other PSID family unit of the specified rank shares the household (e.g., if a family unit shares the household with only one other PSID family unit, then the ID number for the second sharing family unit contains zeroes). A measure of the kinship ties with the other PSID family unit in the same household is included (1993: V23345, V23348, V23351, V23354), as is a measure of family size (1993: V23346, V23349, V23352, V23355) for that other family. This information helps analysts select which family units within a given household they will link.

The Household ID Number variable (1993: V23356) takes on a common value for all PSID family units sharing the same household in a given wave. This variable can be used to link all PSID family units in the same dwelling. If multiple PSID family units live in the same household, then Household ID equals the lowest value for the current- year family ID number (1993: V21602) of any of the PSID family units in that household. If a family unit does not share the household with another interviewed PSID family unit, then Household ID Number simply takes on the same value as that family unit's ID Number.

Institutionalization

To facilitate analysis of family units living in institutions, a variable (1993: V21609) indicating the type of institution in which the family resides has been included in the data each year since 1985. Thus, the analyst need not employ oblique methods to isolate these families, as was formerly the case.

FIPS State and County Codes and Beale's Urbanicity Code

Population density of the area in which the family lives is a very important item. The urbanicity code (1993: V23330) devised by Calvin Beale and Peggy Ross of the USDA has been added to each wave beginning in 1985. Also, the FIPS system of coding state and county (1993: V23328-V23329) as used by Beale to assign urbanicity was added to the data. We retain our usual state and county codes with 1993 (V21603-V21604). Appendix 1, pages 701-721 of *A Panel Study of Income Dynamics, Procedures and Tape Codes, 1985 Interviewing Year*, Vol. I, lists the FIPS codes and the ways in which they differ from the PSID's codes.

Please note that county codes for current county of residence, including both the PSID code (1993: V21604) and the FIPS code (1993: V23329) have been filled with zeroes for the publicly released files. See "State and County Codes" at the beginning of this part (Part 5) for details.

Marriage and Birth Histories—Family-Level Variables

No marriage history variables are included at the family level for 1993, but a few birth history variables are available. We have simply counted the number of children born during calendar year 1992 to Head only (V23357), to Wife/"Wife"/husband of Head/first-year cohabitor only (V23358), to Head and Wife/"Wife"/ husband of Head/first-year cohabitor jointly (V23359), and to other family unit members (V23360). These totals are based exclusively on 1993 reports. Much more detail about births and marriages is available at the individual level (see below) and through the Demographic Event History files (see Part 6 of this section).

Marriage and Birth Histories—Individual-Level Summary Variables

The individual-level marriage and birth data available on the 1968-1993 cross-year file (V32009-V32049) contain information from the initial retrospective data collection effort in 1985 through the current wave. Because an individual can age out of updating questions (in the case of births) or become nonresponse, these data are not up to date through the current wave for everyone. Variables are provided to indicate the recency of the birth or marriage information.

The data record for each individual whose marriage and birth histories were collected—a Head, Wife, "Wife", or other FU member age 12-44 at any time during the 1985 through the 1993 waves—contains birth dates of the oldest and the four youngest children (V32023-V32032), as well as the total number of births (V32022). Births to this individual are current as of the wave indicated in V32021. Marriage data include the total number of marriages (V32034), month and year dates, and separation, widowhood and divorce events of the first and last marriages (V32035-V32048). The last known marital status of the individual (V32049) is included, as is a variable for the wave in which the marriage data were most recently gathered or updated (V32033).

Data are also provided about the individual's parent if birth and marriage histories were collected for the parent at any time from 1985 through the current wave and the individual is reported as a birth in the parent's birth history. The parental variables include identifiers for the parent (mother: V32009 and V32010; father: V32016 and V32017). (Mother identifiers are present in the data records of some individuals who were not reported as births in any female's birth history collected from 1985 onward. These identifiers were coded during 1983 and 1984 data processing from uncoded information for individuals who were present in 1983 or 1984 families. No information about the mother other than her identifying information is provided in such cases.) The parental variables also include parent's year of birth, total number of children, and where the given individual ranks in the birth order of the parent's children (mother: V32011-V32013; father: V32018-V32020). Birth weight of the individual (V32014) is also derived from a parent's birth history data. If the mother's birth history was collected, then birth weight is taken from her birth history report, but if just the father's birth history was collected, then the values are taken from his birth history information. The variables indicating each parent's total number of children and rank order of this individual in that total are current through the most recent wave in which birth history was collected for that parent. Detail about all children is available on the 1985-1993 Childbirth and Adoption History file. Comprehensive data on marriages of the given individual and the parent are available on the 1985-1993 Marriage History file.

Part 6: The Demographic History Files

Several special public-release files contain detailed information collected by the PSID 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 summarized form on the main files. Analysts wanting complete details on these topics must turn to the special public-release files. These 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. The special files are obtainable through ICPSR or our website; see Part 9 of this section for more information.

Each year since 1985, the interview has contained questions about a number of demographic events asked of PSID family members eligible for such events. The events include childbirth, adoption, marriage, separation and divorce. Retrospective histories of substitute-parenting activities were also collected in one wave— 1985. Since the full detail on the various demographic events is desired by only a relatively small subset of potential data users, but a sizable number of data users may want some of the detail, we disseminate two types of data products. One is the addition of individual-level summary variables to the main PSID data file, discussed under Marriage and Birth Histories—Individual-Level Summary Variables in Part 5 of this section. The other data products are special publicly-released, fully documented files containing all present-year and past-year detail of collected demographic history information.

One of these files, the 1985 Ego-Alter file, contains all of the demographic history detail collected in the 1985 wave. A record on that file represents a pair of individuals related by marriage, childbirth, adoption, or substitute parenting (one of the variables indicates the type of record—marriage record, childbirth record, adoption record, or substitute-parenting record). The demographic history detail from the 1985 wave is based on comprehensive retrospective histories collected at that time. It includes detail about the timing and circumstances of the demographic event relating the pair of individuals—parenting or marriage—up to and including 1985.

Since then, updates to this information, as well as retrospective histories for those new to the study, have been collected. These files cover marital events or childbirth and adoption events and are built from the 1985 Ego-Alter file, adding events reported since 1985. The files are known as the Marriage History file and the Childbirth and Adoption History file. Like the 1985 Ego-Alter file, they follow a one-event-per-record format (each record represents a pair of individuals related by the event specified in the file's title— marriage, childbirth or adoption). They differ from the 1985 Ego-Alter file in that (1) separate files are created for the different types of demographic events; (2) individuals reporting zero events of the specified type are included on the files (they were not included on the 1985 Ego-Alter file); (3) reports of post-1985 events are recorded, as are events from retrospective histories reported for individuals entering the PSID since 1985; and (4) they do not include substitute parenting events. The current set of these demographic history files covers 1985-1993 information and is now publicly available.

Part 7: The 1993 OFUM Income Detail File

The change in our data collection procedures to computer-assisted interviewing caused much more information about other FU members' incomes to be available in a coded format. Laborious hand coding was a thing of the past. Spurred by expressed user interest in details about these persons' incomes, particularly person-specific receipt of transfers, we saw that, rather than releasing summarized totals for transfers, we could release all the collected detail about each and every type of transfer. Since the set of variables containing the detail amounted to some 300 items and since we plan to continue releasing this information on an annual basis, we were reluctant to add the data to the main individual file. Thus the 1993 OFUM Income Detail File was conceived.

This file consists of one record for each OFUM associated with the family. Each type of income has its own set of detailed variables for amounts, time units and months of receipt, as well as annualized imputed amounts and accuracy indicators, just as we now provide for Head's and Wife's/"Wife's" incomes. The sets of variables for labor income additionally include amounts for work weeks and hours; the information is

available for as many as four jobs. A proration factor is added for each income report. This factor indicates the fraction of that income earned while the OFUM was living in the family and was used as a multiplier for annualization.

The 1993 Family (ID) Number and the 1993 Sequence Number are included for matching each individual to the cross-year individual file.

Part 8: 1993 Health Care Burden (HCB) Supplement

The 1993 Health Care Burden (HCB) file contains detailed information collected in the 1993 PSID interview concerning health events of the elderly and their (primarily financial) burden on immediate and extended families. The purpose of the HCB supplements was to provide a better understanding about the impact of these events on the families of the elderly. The questionnaire covered major health-related events of the elderly including some ADL and IADL items; home-based, nursing home and other forms of health care and expenditures; and financial and time help given to and received by the elderly.

Data were collected for persons aged 55 or more, including Heads and Wives/"Wives" of PSID families; the parents of the Head or Wife/"Wife", whether in the family or not; and other family unit members.

A more comprehensive set of information was collected about parents, with fewer questions asked for Heads, Wives/"Wives" and non-parental OFUMs. More data were obtained for persons aged 65 and older than for those aged 55-64.

The file contains one record for each eligible individual, for a total of 8,720 records—2,013 records for Heads, 721 for Wives/"Wives", 5,848 for parents and 138 for non-parent OFUMs. Most of the approximately 140 HCB variables are individual-level, although a few pertain to the family as a whole.

Part 9: Data and Documentation Files Available for 1993

Data files

Before 1990, PSID main files for each interviewing wave consisted of a cross-year family-individual response file, a cross-year family-individual nonresponse file, and a cross-year family file. The Cross-Year Family-Individual Response and Nonresponse files had an identical file structure: one contained records for all individuals who were members of PSID family units interviewed in the most recent interviewing wave, while the other contained information for all individuals who were members of families interviewed in the past but who had attrited by the most recent wave. The Cross-Year Family-Individual File stored both individual- and family-level variables collected in the most current wave and in past waves. The cross-year family file contained only family-level variables.

Beginning with the 1990 data, the record format of the cross-year files exceeded the maximum allowed on most computing systems and, consequently, a new file structure for the PSID data was developed. For 1993, this new file format consists of 26 separate, single-year files with family-level data collected in each wave 1968–1993, and one cross-year individual file with individual-level data collected from 1968–1993.

A family file contains one record for each family interviewed in the specified year. The records in each file are identified by the Family Interview (ID) Number for that year, are sorted by that variable, and contain the family-level variables collected in that year. The cross-year individual file contains one record for each person ever in a PSID family through 1993. The records in the cross-year individual file are identified by 1968 Family Interview (ID) and Person Number and are sorted by these variables. The file also contains the

Family Interview (ID) Number of the family with which the person was associated in each year. The cross-year individual file contains all individual-level variables for 1968 through 1993.

With the new file structure, a moderate amount of data management is required to merge the family files with the individual file to create a traditional PSID cross-year family-individual file. The advantage of this new file format is that the files require the minimum amount of storage space. Since each file is considerably smaller than the traditional cross-year family-individual file, the PSID data in this new file format are less demanding of computing resources. This new file structure also allows users to extract a subsample of individuals or families and the variables of interest to create a substantially smaller file to work with from the beginning of the data analysis process.

In addition to the family and individual files, four supplemental files with information from the 1993 wave are available. Those files are detailed in Parts 6 through 8 above.

All these data files are available from our internet site (http://www.umich.edu/~psid/) where you can download entire data files. We recommend that you check out our "What's New" item to see what new data we are releasing and other things of note.

In addition, the "main" data files are available at our internet site through the Data Center, a user-friendly system for creating custom subsets from the PSID single-year family files and the cross-year individual file. Subsets are based on either (1) a single-year family file or (2) the cross-year individual file with, optionally, data from one or more family-year files. Both final-release and early-release data are available. Analysts can specify a subset of variables and/or a subset of individual records. If multi-year final-release data is selected, variables may be selected from the alphabetic cross-year indices. These indices provide a listing of comparable variables across all the years of the study. If this option is picked, it may be helpful to consult the printable version of these indices which is available at our website (pick the "Documentation" option).

The data files are also available from the Inter-university Consortium for Political and Social Research (ICPSR). Contact the ICPSR representative at your institution for further information. To obtain data directly from ICPSR, contact: ICPSR User Support, 1-734-763-5010, netmail@icpsr.umich.edu.

Refer to ICPSR study number 7439, and please specify which datasets you need. The ICPSR also has a website (http://www.icpsr.umich.edu).

CD-ROMs with the 1968-1987, 1968-1988, 1968-1989 and 1968-1992 data and documentation have been produced. For information and an order form for the most recent CD-ROM, the 1968-1992, see our website. Plans for the production of another CD-ROM are indefinite as of this writing.

Documentation Files

The following machine-readable documentation files are provided for the designated 1993 data files.

- Guide to 1993 Family and 1968-1993 Individual Codebooks (this document)
- 1993 Family Codebook
    - 1968-1993 Alphabetical Index of Family-Level Variables
    - 1993 Numerical Index of Family-Level Variables
- 1968-1993 Individual Codebook
    - 1968-1993 Alphabetical Index of Individual-Level Variables
    - 1993 Numerical Index of Individual-Level Variables
- 1985-1993 Marriage History Codebook
- 1985-1993 Childbirth and Adoption History Codebook
- 1993 OFUM Income Detail Codebook

The five codebooks provide variable-level documentation of the variables in specific data files. See Section II of this document for a little more information about the family and individual codebooks.

In addition, two types of variable indexes are available for the family and individual files. The indexes provide a list of comparable variables from 1968 through the current year, 1993. See Section III of this document for more information.

All these documentation files and a facsimile of the 1993 questionnaire are available at our website.

Part 10: Creating Datasets from the Single-Year Family and the Cross-Year Individual Files

The PSID no longer releases merged cross-year files. The traditional cross-year family-individual files have been replaced by single-year family files and a cross-year individual file. The twenty-six single-year family files (one for each year of the study from 1968 through 1993) contain all of the family-level variables collected in each wave. Each of these family files has one record for each *family* interviewed in that wave.

The cross-year individual file contains all twenty-six years of individual-level variables collected from 1968 to 1993, and each *individual* has his own record. Both current-year response and nonresponse persons are included. The file contains an interview number for the family to which each person has belonged in each wave, together with information unique to that person. That is, each member of a family has a family Interview (ID) Number whose value is identical with the values of that data item for all the other family members in that family. In addition, each individual is assigned a unique sequence number, which indicates the person's position and status for any given year's list of family members. Thus, the first person listed, always the Head of the family, is 01, the second person listed is 02, and so on.

Creating a Cross-Year Family-Individual File

As mentioned above, each single-year FAMILY file contains one record for each family interviewed in the specified year. The records in each file are identified by the family Interview Number for that year, in sort order by that variable, and contain the rest of the family-level variables for that year.

The cross-year INDIVIDUAL file contains all individual-level variables for 1968 through 1993 and includes one record for each person ever in a PSID family up to and including the current wave. The records in this file are identified by 1968 family ID (V30001) and Person Number (V30002) and are in sort order by these variables. The file contains the Interview Number of the family with which the person was associated in each year after 1968, as well.

Few users will want to analyze the full data file for all persons ever in the study, and so your first step is to decide which variables, individuals and waves of data interest you.

The root principle in any merge of family data with individuals involves a match of the two files using yearly Interview Numbers for the wave(s) in which the chosen family variables were collected, and so these variables for Interview Number must be retained as part of any subsetted data, either family or individual. Table 14 shows the annual variable numbers for the single-year family and cross-year individual files.

Table 14
FAMILY AND INDIVIDUAL INTERVIEW NUMBERS

Wave	Family Variable Number	Individual Variable Number
1968	V3	V30001
1969	V442	V30020
1970	V1102	V30043
1971	V1802	V30067
1972	V2402	V30091
1973	V3002	V30117
1974	V3402	V30138
1975	V3802	V30160
1976	V4302	V30188
1977	V5202	V30217
1978	V5702	V30246
1979	V6302	V30283
1980	V6902	V30313
1981	V7502	V30343
1982	V8202	V30373
1983	V8802	V30399
1984	V10002	V30429
1985	V11102	V30463
1986	V12502	V30498
1987	V13702	V30535
1988	V14802	V30570
1989	V16302	V30606
1990	V17702	V30642
1991	V19002	V30689
1992	V20302	V30733
1993	V21602	V30806

Not all cases in the cross-year individual file have a matching record in a given single-year family file. This happens when an individual who was part of a responding family has moved away or died and is no longer associated with any family in the study; the person is said to be nonresponse. The nonresponse person's Interview Number in the cross-year individual file is filled with 0s (as are the other variables) for a wave in which no data were collected about him or her.

We can think of several approaches to creating a cross-year family-individual file from the components. Two good ones are described below.

Method 1. First select individuals and variables from the cross-year individual file and then match those data, using a one-to-many match, with the desired variables from a single-year family file. (Remember to retain the yearly Interview Numbers from all files when subsetting.) Next, match the resulting file (which now contains selected variables from the cross-year individual file and the first family file) with a second family file. Repeat with additional single-year family files until all required family data are obtained and merged with the cross-year individual data. See SPSS or SAS examples at our website for an illustration of this approach using three years of family data.

Method 2. Alternatively, first do a series of one-to-many matches of a single-year family file and the cross-year individual file matching on that wave's Interview Number. Be sure to retain the 1968 Interview Number (V30001) and Person Number (V30002) from the individual file on each family-individual output

48

file. The resulting single-year family-individual files are then merged in a one-to-one match using the 1968 Interview Number and Person Number. Detailed steps are noted below.

> **Step 1**: Subset annual family Interview Number and other selected variables and select cases from cross-year individual file.
>
> **Step 2a**: Subset selected variables from the year-n family file.
>
> **Step 2b**: Sort subsetted year-n family file from Step 2a by year-n family Interview Number.
>
> **Step 2c**: Sort subsetted cross-year individual file from Step 1 by year-n family Interview Number.
>
> **Step 2d**: Merge sorted cross-year individual file from Step 2c with sorted year-n subsetted family file from 2b (a one-to-many, family-to-individual, match) matching on the year-n family Interview Number.
>
> **Step 2e**: Sort resulting year-n family-individual file from Step 2d by the individual identifiers, 1968 family Interview Number (V30001) and Person Number (V30002).
>
> ... **Repeat Steps 2a-2e** for all other years.
>
> **Step 3**: Merge family-individual files from Step 2e by the individual identifiers, 1968 family Interview Number (V30001) and Person Number (V30002).

See SPSS or SAS examples at our website for an illustration of this approach using 26 years of family data.

Creating a Cross-Year Family File

As we have already said, each member of a family has a family ID number for each wave with a value identical to the values of that data item for all the other family members in that family that year. In addition, each individual is annually assigned a unique sequence number, which indicates the person's position and status for any given year's list of family members. Thus, the first person listed, always the Head of the family, is 01, the second person listed is 02, and so on. To create a current cross-year family-level file, select from the cross- year individual file those cases where V30807 (1993 Sequence Number) is equal to 01, since each family must have at least one member, although it may or may not have more. Then merge data from the single-year family files using the yearly ID numbers to match as described in Method 1 or 2 above. These instructions create a merged 1968-1993 family-level file for *currently responding* families.

For other years' cross-year family-level files, the Sequence Number variable for the latest desired year of data should be used and merges done with the appropriate single-year family files. Again, this produces a file of families who were response through the latest year and eliminates families who had already become nonresponding.

Single Year Files

Producing single-year family files for cross-sectional analysis is simplicity itself. Just use the single-year file. Single-year family-individual files are also relatively simple. Select all individuals whose Sequence Number for the desired year is nonzero and match the Interview (ID) Number for that year from the individual file with the Interview Number from the corresponding family file. The Interview (ID) Numbers for family and individual files are listed in "Creating a Cross-Year Family-Individual File" above.

Part 11: User Guide for the PSID

Although the PSID staff completed a *User Guide* to the panel study in 1984 that was designed to assist users in struggling through the complexities of the study, this guide has not been significantly revised since that time. Some topics covered in the old *Guide* are still relevant, but others are obsolete. The chapters on, e.g., PSID history, sample composition and weighting, and study content are still quite accurate, if dated, but the major drawbacks to the G*uide* in its present incarnation are that it does not take into account the inclusion of

data for currently nonresponse individuals nor does it deal with the current structure of many single-year family files and the cross-year individual file. It also does not include any information regarding the past decade of the PSID. We recommend instead that users consult Martha S. Hill's *The Panel Study of Income Dynamics: A User's Guide*. This book is the second in Sage Publications' series of guides to major social science databases. (To order, contact Sage Publications, Inc., 2455 Teller Road, Newbury Park, CA 91320-2218. Telephone: 805-499-0721, Order numbers: 46090 for hardbound copy ($18.95) and 42303 for paperback copy ($9.25).) Although Hill's book was written while merged cross-year family-individual files were still available, this is not a major drawback to its use. For an update on file structure, see "About PSID" on our website and Part 9 of this section.

SECTION II
CODEBOOK FOR WAVE XXVI

Part 1: Twenty-sixth-Year Family-Level Codebook

The codebook for the twenty-sixth wave of family-level data from the interview schedule is available on the website. The variable numbers refer to those on the 1993 cross-year file.

The distributions for the following variables are computed based on all *families* interviewed in 1993. Minimum and maximum values, unweighted means, and standard deviations are included for relevant variables.

Codebook Information

The example below illustrates the information contained in this codebook for a typical variable. The numbers in brackets do not appear in the codebook, but refer to the explanations following this example.

```
[1] V21607        [2] MODE OF 1993 INTERVIEW        [3] Reference: 7

[4] Type: Num  [5] Width: 1  [6] Decimals: 0  [7] Location: 16


[8] Mode of Interview in 1993

   [9]  [10]   [11]
   257   0     Personal interview
  9711   1     Telephone interview
     1   2     Mail interview
     8   9     NA


        [12]
  Missing Data: 9

[13]--                         [14]--                  [15]--
[16]--                          [17]--
```

[1] Indicates the variable number. A variable number is assigned to each item in the study. (See the introduction to the numerical index, Section III, Part 1, for a list of the range of variable numbers specific to each year.)

[2] Indicates the abbreviated variable name (maximum of 40 characters) used in the SAS and SPSS statements to identify the variable for the user. Refer to the following list of abbreviations for help in translating the names into sensible English.

[3] Indicates the order of the variable within the file.

[4] Indicates whether data are character or numeric.

[5] Indicates the width of the variable.

[6] Indicates the number of decimals defined in the SAS or SPSS statements.

[7] Indicates the starting location and ending location for this variable.

[8] Indicates the full question number that was used in the questionnaire, as well as the exact wording of the questionnaire item; for variables not coded directly from the questionnaire, such as generated data, an appropriate title appears here.

[9] Indicates the *unweighted* family-level N for each code value. Blanks indicate that no cases have this value.

[10] Indicates the code values occurring in the data for this variable. For variables containing field amounts, refer to the notes appearing directly below items [13] through [16] for the range of data values.

[11] Indicates the textual definitions of the codes. Abbreviations commonly used in the code definitions are "DK" (Don't Know), "NA" (Not Ascertained), and "Inap." (Inappropriate).

[12] Indicates the code value for missing data. In this example, code values equalling nine are missing data. In cases where nothing is printed in this space, missing data are not permitted for the variable; either values were assigned for such cases or missing data were impossible.

[13] Some analysis software packages require that certain types of data which the user desires to exclude from analysis can be designated as "missing data," e.g., inappropriate, unascertained, or ambiguous data categories. Although these codes have been defined by the PSID staff as missing data categories, this does not mean that the user should not or cannot use them in a substantive role if so desired.

[14] Indicates the number of cases with nonmissing values, where specified.

[15] Indicates the mimimum value for this variable, where specified.

[16] Indicates the maximum value for this variable, where specified.

[17] Indicates the *unweighted* mean for this variable, where specified, using family-level case counts.

[18] Indicates the standard deviation for this variable, where specified.

Part 2: Twenty-Six Year Individual-Level Codebook

The individual codebook (see our website) contains the code values and corresponding documentation for all twenty-six years of individual-level variables. Note that the variable numbers for individual-level variables apply *only* to the Wave XXII (1968-1989) and later individual files; see pp. 434-455 of Volume I of the Wave XXII (1989) documentation for details.

File positions for comparable variables on the 1968-1993 cross-year individual file differ from those for all other years' cross-year individual files, including early release files through 1996, because all year variables, e.g., year of birth and year moved in or out, were converted to four digits representing the actual year instead of just the last two digits of the year.

See Section III, Part 2 for a wave-by-wave list of individual-level variable numbers.

Although the summary variables (V31994-V32049) are included on each cross-year individual file, the values for these variables may change for an individual because they are updated with each additional wave of data collection to reflect conditions from 1968 through the most recent year covered by the file.

The user is encouraged to read the chapter on family composition in Martha S. Hill's *User Guide* (see Section I, Part 11) for more details about individual-level data.

A few terms used in the individual codebook may need explanation. The phrase "*main family nonresponse*" means that both the individual and his or her family have at that time become lost to our study, although either or both may reappear in the study in subsequent waves. In the wave just *prior* to becoming nonresponse, the individual was connected with a family interviewed by our study; thus, both family and individual data are available for that prior year, and the individual's Sequence Number at that time was 01-59.[5] However, data were collected for neither the individual nor his or her family *in the nonresponse wave*. The data for the wave in which nonresponse occurs (and all subsequent waves if and until the individual reappears as a member of a responding family unit, including a recontact family) are zeroes excepting the variables for type of individual record and reason for nonresponse, and if an individual was selected for recontact, follow status and reason for following the individual.

In contrast, *mover-out nonresponse* individuals have left a family that was still in the study. Since such individuals were usually present in that family for at least part of the calendar year preceding nonresponse,

---

[5] Sequence Number in the range 51-59 indicates that such an individual was in an institution in that year. If the individual entered an institution since the previous interview, there are some income data available at the individual level. If, however, the individual had remained in an institution for a wave of data or more, only sequence number, age, relation to head, and sex are available, although family-level data are connected with such an individual's record.

they have some additional nonzero data for the wave in which they became nonresponse, such as part-year income information. In later waves, mover-out nonresponse individuals are treated in two ways, depending on why they left the family. Those *who moved out to institutions* have several variables (sequence number, age, sex, relationship to head, type of individual and reason for nonresponse) with nonzero values, although income, housework, and other individual-level variables are filled with zeroes. Eventually, such an individual may (a) become response by moving into a family or by becoming a splitoff, (b) move from the institution and remain mover-out nonresponse (shown when Sequence Number=71-89), or (c) become main family nonresponse because the family itself became nonresponse. (See the preceding paragraph for an explanation of main family nonresponse data records.) The *other type of mover-out nonresponse* individual has either moved out, but not to an institution, or died. Later waves of data contain zeroes, as described above for main family nonresponse, unless they subsequently rejoined a responding family or were selected for recontact.

The data are released as one file, which includes not only those individuals with nonzero data records in 1993 (i.e., current response plus 1993 mover-out nonresponse), but also all other individuals—those who have zero data records for 1993 (i.e., 1993 main family nonresponse and all nonresponse of either kind from 1992 and earlier waves.

Statistical information, such as means, Ns, standard deviations, minimum and maximum values, etc. are printed for individual-level variables, just as described in Part 1 of this section for family-level variables (see above). *All such information is unweighted.*

SECTION III
INDEXES FOR WAVE XXVI

Four variable indexes that attempt to organize the data for easy reference are included on our website. The first two are (a) a topical alphabetic arrangement of family-level data, listing pertinent variables from all waves, and (b) a list of the 1993 family-level data in order by 1993 variable name with comparable 1968-1992 variables included for each item.

The other two indexes are parallel with the family indexes described above but include just individual data for 1968-1993. Please note that all individual-level variable numbers are valid only for files from Wave XXII (1989) onward. Refer to the introduction on pp. 434-455 of Volume I of the Wave XXII (1989) Documentation for changes in individual-level variable numbers between the 1968-1988 and 1968-1989 and later cross-year files.

Due to limitations in paper width, we use two sets of lines for each data item in all four indexes. The first set of variable numbers is for 1968-1977, and the second is for 1978-1993. Each set lists the variable numbers with comparability annotations in the form of lower-case alphabetic characters following, where applicable.

We no longer produce the employment index, which compared questions from questionnaire Sections B and C (Head's employment), and Sections D and E (Wife's/"Wife's" employment).

Although we have spent much more time on these indexes than is prudent for sound mental health, errors remain. Since this index is used as the basis for all future versions of the documentation volumes, we would appreciate information on such errors. Please send your discoveries by e-mail to PSID_staff@umich.edu.

All variables listed under the same heading should be comparable for analysis purposes with, perhaps, some recoding. Again, despite careful checking, errors may exist, and the user should consult variable descriptions in the codebooks before doing analysis.

Part 1: Family-Level Indexes

Family-level variables are associated with interviewing years as follows:*

> Variables 1-439: 1968 family data
> Variables 441-909: 1969 family data that precede change variables (includes raw data and most generated
>   variables)
> Variables 910-989: 1968-69 change variables**
> Variables 990-998, 1008-1016: 1969 family data that follow change variables (includes deciles and a few other
>   variables)
> Variables 999-1007: calculated in 1969 for responding 1969 families only. Values are for their 1968 data.
> Variables 1101-1624, 1766-1767: 1970 family data
> Variables 1625-1627: 1969 labor market data, based on 1970 sample!; i.e., gathered in 1970 for responding 1970
>   families, but values are for their 1969 data
> Variables 1701-1748: 1970 Census data on county. See Wave VII (1974) documentation, pp. 154-161, for
>   details.
> Variables 1749-1763: consumer price indexes for each area for each year 1968-1972, based on 1974 sample, i.e.,
>   calculated in 1974 for responding 1974 families but values are for their 1968-1972 data. See Wave VII
>   (1974) documentation, pp. 162-166, for details.
> Variables 1801-2346: 1971 family data
> Variables 2401-2980: 1972 family data
> Variables 3001-3311: 1973 family data

Variables 3401-3731, 1764-1765: 1974 family data
Variables 3801-4232: 1975 family data
Variables 4301-4707: 1976 Head's interview data
Variables 4708-5027: 1976 Wife's interview data
Variables 5028-5114: 1976 other family data
Variables 5201-5671, 5681-5682: 1977 family data
Variables 5672-5680: change in marital status variables, based on 1977 sample; i.e., their 1968-1976 data
    calculated in 1977 for responding 1977 families but values are for their 1968-1976 data.
Variables 5701-6221: 1978 family data
Variables 6301-6815: 1979 family data
Variables 6901-7457: 1980 family data
Variables 7501-8111: 1981 family data
Variables 8201-8739: 1982 family data
Variables 8801-9433: 1983 family data
Variables 10001-11079: 1984 family data
Variables 11101-12446: 1985 family data
Variables 12501-13687: 1986 family data
Variables 13701-14737: 1987 family data
Variables 14801-16208: 1988 family data
Variables 16301-17612: 1989 family data
Variables 17701-18945: 1990 family data
Variables 19001-20245: 1991 family data
Variables 20301-21549: 1992 family data
Variables 21601-23363: 1993 family data

*   Skips in numerical sequence are due to dummy variables separating each year's data.
**   These variables are described but not listed individually in this index.

Alphabetical Index. This index includes all family-level variables for the entire twenty-six year data collection effort, even though each year's family-level information is stored on its own separate data file. Since this index is a compilation of all variables on the family-level files only, no individual-level variable numbers are included. However, cross references for those variables are listed at the appropriate content headings. All raw data and generated variables are listed in alphabetical order by topic. Each alphabetic entry is accompanied by a list of variable numbers, file positions, and field widths showing when and where comparable data exist for each of the years of the study. Small changes in questions posed or in coding conventions for a given question are noted in footnotes to the variables. The list of footnotes for family-level data is located at the end of the index file.

Headings may have more than one variable number listed for a given year. This generally indicates that the same question was asked of different subgroups within the sample (employed and unemployed respondents, for example). The user should consult the questionnaires or the "0. Inap.;..." codes in the codebook for each year for details on which subgroups have no data on such variables. When in doubt, always refer to the questionnaire. Note that such terms as "head" or "wife" do not necessarily designate the same person in the data from wave to wave.

Numerical Index. This index lists the 1993 family-level variables in numerical order and includes comparable 1968-1992 variables. Note that both the 1976 and 1985 Heads' and Wives'/"Wives'" interview data are listed in the 1976 and 1985 columns respectively.

Where blanks occur, no similar data item from earlier years exists for that particular 1992 variable. Naturally, there are some exceptions—the employment-related variables that were coded only once for all Heads (1968 and 1969 data) or all Wives/"Wives" (1968-1975, 1977 and 1978 data) regardless of employment status. These are listed with the appropriate 1993 variables from the sections asked about employed Heads and Wives/"Wives", but not with the questions for unemployed Heads and Wives/"Wives".

Data items not asked in 1993 are also not included. Thus, this index does not contain all the cross-year family-level variables. Small differences, such as coding formats or field width variations, have been annotated; the lower case alphabetic characters appearing directly beneath the variables to which they refer indicate footnoted differences. These footnotes are identical with those used for the alphabetical family-level index. For convenience, a list of the footnotes is also located at the end of this index.

Part 2: Individual-Level Indexes

Individual-level variable number ranges changed beginning with the 1968-1989 cross-year file, as sex of individual was removed from each year's individual data and is now included with the summary variables (V32000). The introduction to the individual-level codebook on pp. 434-455 of Volume I of the Wave XXII (1989) Documentation lists the old 1968-1988 variable numbers with their corresponding new 1968-1989 variable numbers.

In both individual indexes, the summary variables have been listed in the column for the most recent year's data, even though they are generated using information from all of the years.

Differences between years in coding format among the variables listed here have been annotated. The lower case alphabetic characters indicating such differences appear *directly below* the variables to which they refer. A list of the footnotes appears at the end of each index file.

Below are the ranges for each year's individual portion of the data on the twenty-six-year final cross-year file.

The 1968-1993 individual-level variable numbers are as follows:

V30020-V30042	1969 individual data
V30043-V30066	1970 individual data
V30067-V30090	1971 individual data
V30091-V30116	1972 individual data
V30117-V30137	1973 individual data
V30138-V30159	1974 individual data
V30160-V30187	1975 individual data
V30188-V30216	1976 individual data
V30217-V30245	1977 individual data
V30246-V30282	1978 individual data
V30283-V30312	1979 individual data
V30313-V30342	1980 individual data
V30343-V30372	1981 individual data
V30373-V30398	1982 individual data
V30399-V30428	1983 individual data
V30429-V30462	1984 individual data
V30463-V30497	1985 individual data
V30498-V30534	1986 individual data
V30535-V30569	1987 individual data
V30570-V30605	1988 individual data
V30606-V30641	1989 individual data
V30642-V30688	1990 individual data
V30689-V30732	1991 individual data
V30733-V30805	1992 individual data
V30806-V30866	1993 individual data
V31987-V32049	Summary variables

Alphabetical Index. This index lists each individual-level variable by topic, and is modeled after the alphabetical family-level data index described in Part 1 above. The major headings are the same as those listed in the family-level index under INDIVIDUAL DATA. All individual-level variables are included in this index.

Numerical Index. The 1993 individual-level variables are listed in numerical order, with the comparable 1968-1992 variables included. Where blanks occur, no similar data item exists in previous years for that particular 1993 variable. Also, some data items from previous years are excluded from this index since no comparable variable exists for 1993. Thus, this index does not include all the cross-year individual-level variables.

REFERENCES

Brewer, K.R.W. & Mellor, R.W. (1973), "The effect of sample structure on analytical surveys," *Australian Journal of Statistics*, Vol. 15, pp. 145-152.

DuMouchel, W.H., & Duncan, G.J. (1983), "Using sample survey weights in multiple regression analysis of stratified samples," *Journal of the American Statistical Association*, Vol. 78, pp. 535-543.

Hansen, M.H., Madow, W.G., & Tepping, B.J. (1983), "An evaluation of model-dependent and probability-sampling inferences in sample surveys," *Journal of the American Statistical Association*, Vol. 81, pp. 1058-1062.

Hess, Irene (1985), *Sampling for Social Research Surveys, 1947-1980*. Ann Arbor, MI: Institute for Social Research, University of Michigan.

Horvitz, D.G., & Thompson, D.J. (1952), "A generalization of sampling without replacement from a finite universe," *Journal of the American Statistical Association*, Vol. 47, pp. 663-685.

Kalton, G. (1987), "Including nonsample persons in PSID analyses." Ann Arbor, MI: PSID Board of Overseers. Working paper.

Kish, L. (1965), *Survey Sampling*. New York: John Wiley & Sons, Inc.

Lepkowski, James M., Sadosky, Sally A., Couper, Mick P., Chardoul, Stephanie, Carn, Lisa, and Scott, Lesli-Jo. (1995), "A Comparison of Recording Errors Between CATI and Paper-and-Pencil Data Collection Modes," Proceedings of the Section on Survey Research Methods, American Statistical Association, 1995, pp. 521-526.

Little, R. (1989), "Sampling weights in the PSID: Issues and comments." Ann Arbor, MI: PSID Board of Overseers. Working paper.

Little, R.J.A., & Rubin, D. (1987), *Statistical Analysis with Missing Data*. New York: John Wiley & Sons, Inc.

PSID (1984), *User Guide to the Panel Study of Income Dynamics*. Ann Arbor, MI: Institute for Social Research, The University of Michigan

PSID (1984), *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1984 Interviewing Year*, Wave XVII. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

PSID (1989), *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1989 Interviewing Year*, Wave XXII. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

PSID (1990), *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1990 Interviewing Year*, Wave XXIII. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

PSID (1992), *A Panel Study of Income Dynamics: Procedures and Tape Codes, 1992 Interviewing Year*, Wave XXV. Ann Arbor, MI: Institute for Social Research, The University of Michigan.

Shryock, H. S. and Siegel, J. S. 1971. *The Methods and Materials of Demography*. Washington, DC: U.S. Bureau of the Census.

Skinner, C.J., Holt, D., & Smith, T.M.F. (1989). *Analysis of Complex Surveys*. New York: John Wiley & Sons.