Income Data for Policy Analysis: A Comparative Assessment of Eight Surveys

Final Report

December 23, 2008

John L. Czajka
Gabrielle Denmead

Submitted to:

U.S. Department of Health and Human Services
Office of the Secretary
Assistant Secretary for Planning and Evaluation
Suite 436E.1, 200 Independence Ave., SW
Washington, DC 20201

Project Officer:
Joan Turek

Submitted by:

Mathematica Policy Research, Inc.
600 Maryland Ave., SW, Suite 550
Washington, DC 20024-2512
Telephone: (202) 484-9220
Facsimile: (202) 863-1763

Project Director:
John Czajka

ACKNOWLEDGMENTS

CONTENTS

CONTENTS *(continued)*

CONTENTS *(continued)*

CONTENTS *(continued)*

TABLES

TABLES *(continued)*

TABLES *(continued)*

TABLES *(continued)*

EXECUTIVE SUMMARY

A. STUDY OVERVIEW

Income is a critical classification variable for policy-related analyses, and together with poverty status is often key in the development of public policy. Most federal household surveys collect some income data and provide measures of poverty status. Yet income is difficult to measure in household surveys, and poverty status depends on how a family is defined, which differs markedly across surveys. Despite many similarities, there are also many differences in the income and poverty concepts used, and different surveys provide markedly differing estimates of income and poverty.

Under contract to the Office of the Assistant Secretary for Planning and Evaluation (ASPE), Department of Health and Human Services (HHS), Mathematica Policy Research, Inc. (MPR) and its subcontractor, Denmead Services & Consulting, have conducted a comprehensive and systematic assessment of the income data and their utility for policy-related analyses in eight major surveys: the Survey of Income and Program Participation (SIPP); the Annual Social and Economic Supplement to the Current Population Survey (CPS); the American Community Survey (ACS); the Household Component of the Medical Expenditure Panel Survey (MEPS); the National Health Interview Survey (NHIS); the Medicare Current Beneficiary Survey Cost and Use files (MCBS); the Health and Retirement Study (HRS); and the Panel Study of Income Dynamics (PSID).

The assessment focuses on three issues:

- The quality and usability of each survey's income and poverty data for policy-related analyses

- The overall impact of different design and methodological approaches

- Specific design and processing choices that may be related to the quality and utility of income and poverty data in each survey

The assessment is both descriptive and empirical. The lengthy descriptive component provides great detail on survey design and methodology and on income data and poverty measures for persons and families in each survey for the files and years used in the study. It includes overall design, timing, recall, reference period, family definition, poverty measurement, content on income and policy-related covariates, income data processing, and public availability and accessibility of income and poverty data. Additionally, it includes an annotated bibliography of literature relevant to the project.

The empirical portion addresses income, poverty, and program participation using the same income measures, definitions, units of analysis, and time period for each survey, to the extent possible, in standardized tabulations. Additional tabulations address methodological issues,

specific survey attributes, and questions raised by the detailed information gathered for the descriptive component.

A Technical Advisory Group (TAG) representing each survey and the policy research community provided input to the project. TAG members reviewed and commented on drafts of the workplan, the annotated bibliography, the analysis plan, the outline of the final report, the detailed survey descriptions, and the final report.

TAG members, Census Bureau staff, and PSID staff at the University of Michigan provided extensive assistance in obtaining documentation not readily available from published sources or public web sites, and the Census Bureau also performed a major series of tabulations pro bono on the internal files of monthly ACS data.

B. POLICY ANALYSIS CONTEXT

Data requirements for policy analysis are not the same as those for more general research—they are both different and more extensive. Whatever issue is addressed, good income information for policy work is likely to require the following:

- *Actual Numbers.* Income is often used to determine potential eligibility; benefits or charges may vary with income; and impact at different points in the income distribution is important—policy work needs actual amounts, not broad income intervals

- *Comparability with Official Poverty Statistics.* Poverty status is important in policy evaluation and in public debate and must be on the same basis as official statistics

- *Other Relevant Variables.* Work on health usually requires data on health insurance status and utilization, and work on policies concerning the elderly requires data on current retirement contributions and coverage, as examples

- *Flexibility on Filing Units.* Policy analysis may deal with individuals or part of a family, and may compare different rules for constructing filing units, which requires income data for each person

- *Credibility and Reliability.* Weaknesses in data underlying policy proposals and cost estimates bring the validity of an initiative into question; significant inconsistencies within a survey or failure to match known population totals lead to challenges to the estimates and proposals themselves

- *Transfer or Other Program Participation Data.* Efficient policy design requires detail on benefits or insurance coverage already in place, and the administrative systems with which persons already interact

- *Immediate Accessability and Speed of Use.* Typically the policy process has tight time frames, and unexpected developments when a proposal is being actively considered require new analyses with very quick turnaround times

Accuracy of income data at the lower end of the income distribution is more important than accuracy at the upper end; measures of particular significance include the number and composition of the poor and near-poor, relative importance of key income sources, and insurance status. In addition to income, employment has consistently been an area of policy concern, as a source of self-support and of health insurance coverage, so that accuracy in its measurement is also key. Lastly, randomness as measured by standard errors is not nearly as important as possible bias. The findings of policy analysis, and budget estimates, are presented as point estimates without standard errors, while bias leads to consistent over- or under-estimates.

C. STUDY SURVEYS

The surveys differ greatly in overall design and purpose. Five major Federal surveys—SIPP, CPS, ACS, MEPS and NHIS—cover the civilian non-institutionalized population (although ACS excluded group quarters until 2006) but differ in various respects:

- ***Timing and Reference Period.*** ACS and NHIS have rolling samples (non-overlapping samples spread across a year), SIPP visits each sample household at strict 4-month intervals, and CPS interviews primarily in March. All but SIPP and ACS get calendar year income; SIPP gets monthly income. ACS gets income for the 12 months prior to the interview; for a given calendar year the ACS income data combine 12 different reference periods.

- ***Income Detail and Income for Persons.*** Income detail ranges from the dozens of items collected in SIPP to a single family level variable in NHIS. All but NHIS get income information for every person over 14 years, but NHIS gets only earnings for each person over 17 years and a family income total.

- ***Family Definition and Poverty Measure.*** SIPP, CPS, ACS and MEPS have a poverty measure based on the family definition used in official poverty statistics. NHIS uses only a broader definition that treats unmarried partners as married and includes foster children; this affects poverty rates. MEPS provides a second coding of family composition based on this broader definition, which can be used to construct an alternative poverty measure. Due to the difference in reference periods, ACS poverty measures are not comparable to CPS.

- ***Family Composition Lag.*** The surveys differ in the timing of family composition used for annual poverty measures. Family composition for poverty estimates is measured December 31 of the income year in MEPS, the month after the income year for ACS, usually March after the income year for CPS, and ranges from January to December after the income year for NHIS. With SIPP, analysts can select the timing used in poverty measures.

None of the other three surveys cover the general population. PSID is a unique survey that has followed the same families and their descendants for 40 years. It has detailed income data that are limited to the head and wife or partner, treats unmarried partners as spouses, has no person totals, and uses a contemporaneous poverty measure. HRS is restricted to persons age 51

or over, treats unmarried partners as spouses, and has detailed income data but no person totals. MCBS covers Medicare enrollees but not their families, asks one income question, and is used primarily to collect information on non-covered services to add to Medicare claims data.

D. METHODOLOGY

The descriptive component of the study simply required gathering and verifying a great deal of information about each survey and using uniform and consistent terminology to describe key features of the eight surveys. The descriptions apply to the files used in the study and are not necessarily applicable in all detail to other years, since survey content, procedures, sampling, and data may change from year to year. The empirical component was more complex.

The study uses income data for 2002 (HRS and MCBS income for 2003 were deflated with the CPI-U) and applied CPS definitions wherever possible. Survey samples were restricted to approximately the same universe by removing any military and their families, unrelated children under 15, persons institutionalized or deceased by the end of the year, and persons residing outside the fifty States and the District of Columbia. Excluded students were restored to families in PSID. On advice of the TAG, analysis of the MCBS was restricted to the population age 65 or over. In conformity with CPS income definitions, lump sums and irregular payments were removed where included in survey income. However, a number of relatively small differences remain among the surveys in universe, relationship information, income definitions, time lag, and treatment of college students, as well as the larger differences in ACS due to the prior 12 months reference period (rolling reference period) as compared to calendar year in all other surveys and the exclusion of group quarters in 2002.

Work was done on public use files with three exceptions. MCBS has no public use files, but allows protected off-site use with approval and has a standing agreement with ASPE, under which this study operated. NHIS income dollar amounts are available only on an internal file that may not be taken off-site and requires prior approval and usage fees, which the study obtained and paid. ACS interview month is available only on internal Census Bureau files, and the Bureau performed a set of analyses on these files without charge that enabled the study to assess the ACS rolling sample, rolling reference period and price level adjustments, and resulted in other important although serendipitous findings.

Standard Tables. Tabulations were done at the person level, with persons classified by family income using the CPS family definition. A simulation model was built for NHIS to divide family income when CPS families were created from 5.8 million non-CPS families. Sensitivity tests of the model measured the highest and lowest possible impact on poverty rates. A simpler version of the model was used for the PSID, which contains substantial person-level income information, and persons currently living with relatives were included in these families.

Standardized tabulations of persons and family income were performed on each survey by demographic group and income level. Family income was classified by poverty relatives— whether the ratio of family income to poverty thresholds was under 100 percent, 100 to under 200 percent, 200 to under 400 percent, or 400 percent or over—and by family income quintiles—quintiles of persons ranked by family income. Tabulations were repeated for population sub-groups such as persons receiving Supplemental Security Income (SSI), and by

health insurance and Medicaid status. Standardized tabulations were also performed for persons with earnings and amount earned, and persons with wages and salaries and wage and salary amounts, reflecting the importance of earned income (82 to 86 percent of total income) and wages and salaries in overall income. Comparison tables were created on other surveys for persons age 51 or over and age 65 or over for comparison to HRS and MCBS, and with demographic and other information restricted to the family head and his wife for comparison to the PSID.

Allocation. Standardized tabulations of persons with income allocations were performed to determine the number of persons and the share of income allocated or imputed, by major income source and family income level. These tabulations were done on each survey containing allocation markers.

Special Analyses. Numerous special tabulations of greater and lesser complexity were performed to address specific methodological issues, including the ACS tabulations described above. The impact of different survey timing of family composition used for annual poverty measures was examined using monthly SIPP data on income and family composition; the use of a single data set ensures that findings are purely methodological and do not reflect differences in data. Comparisons in NHIS and MEPS measured the impact of different family definitions on family and poverty counts. Other special tabulations included the degree of rounding or approximation in income reporting, the impact of including withdrawals from tax-advantaged retirement accounts, and the size and impact of inconsistencies in several surveys where consistency was not ensured by the question sequence or subsequent editing.

E. MAJOR FINDINGS

There are three groups of study findings: important methodological results that could apply to any survey collecting income data; findings on issues specific to individual surveys; and empirical results of comparisons across surveys.

1. Survey Methodology

Two methodological findings result from analyses that were part of the study design, but the third was an unanticipated result of tabulations examining monthly ACS data.

Within-Year Variations in Response Rates. ACS monthly sample data on over 45,000 households per month show significantly higher non-response and allocation rates for March, April, May and June than for other months.

- Allocations rose from 19.0 percent of total income in February to 22.8 percent in March and 24.6 percent in April, and all differentials were highly significant

- Differentials for these months were found for five of seven income sources and were statistically significant for wages and salaries, Social Security, asset income, and pensions, although not for self-employment

- The elevation in non-response rates did not occur for SSI or public assistance

- The differentials for March, April, May and June were statistically significant for all quintiles and for all income subgroups above the poverty level

- The amounts by which non-response rates rise during these months increased with income, although average non-response rates decreased with income

The strong pattern in income non-response has implications for overall survey design. The association with tax-filing months and with income levels and income sources usually subject to income taxation is certainly suggestive but requires further study.

Dynamics of Family Composition. Measuring family size and composition at different points in time to calculate poverty rates from the same income for the same year, in SIPP longitudinal data, shows that poverty rates rise as the time increases between measurement of income and measurement of family size and composition.

- Poverty rates are lowest when income and family composition are measured at the same time, or contemporaneously, in monthly data

- As the interval increases between the income reference period and the fixed date at which family size and composition are determined, the number of persons incorrectly classified as poor increases faster than the number of persons incorrectly classified as not poor, and poverty estimates are mildly biased upwards

- The total number of persons incorrectly classified either as poor or not poor greatly exceeds the net change in the number classified as poor and the poverty estimate

- An average of the poverty calculations for each of the 12 months of the next year (NHIS) will yield more poor than calculations based on the next March (CPS), and both will be higher than calculations based on December 31 (MEPS)

- Larger differences are found for minorities, single parents with children, welfare and Food Stamps recipients, and Medicaid enrollees, as the time lag increases

This finding is purely methodological and is based on SIPP data with very detailed income information, a maximum recall of five months, and an average recall of three months. With this data, the poverty rate based on a March family was 0.6 percentage points above a contemporaneous measure, and the poverty rate based on a December (of next year) family was 0.6 percentage points above the rate based on a January (of next year) family. Other surveys have less or no income detail compared to SIPP, and have long recall intervals that average 12 ½ to 18 months and can be as much as 23 months. In surveys with less income detail and longer recall intervals the impact could well be larger, and standardized tabulations cannot adjust for these differences.

Family Definition. Poverty calculations with NHIS and MEPS data show that a broad family definition—including unmarried partners and their relatives in families—reduces the

number of poor compared to the conventional family definition in CPS. The different definitions also give different pictures of family arrangements. MEPS provides both family definitions and reports income at the person level, so family income and poverty can be constructed for either definition. NHIS codes only the broad definition and reports a single family income total, so the study simulated CPS families for 17 million people.

- In both NHIS and MEPS, when we used the broad or NHIS family definition to calculate poverty rates the number of poor declined by 2.6 million and the overall poverty rate by 0.9 percentage points—the estimated declines in NHIS are plus or minus 230,000 persons, or less than one-tenth of a percentage point

- In MEPS the poverty rate for children declined by 1.7 percentage points, and the poverty rates for single parents and their children declined by well over five percentage points each under the NHIS family definition

- Poverty rates for the elderly were unchanged when the definitions were compared

- In both NHIS and MEPS, quintile bounds all shifted upwards by $1,000 to $2,000

A number of surveys use broader family definitions treating unmarried partners as families. Broader definitions reduce both the number and demographic composition of the poor and change the overall picture of family structure.

2. Survey-Specific Issues

Many issues or procedures are unique to one or two surveys, and one purpose of the study was to identify and describe such issues, and measure their impact if possible.

Design Features. A few design features can be examined empirically, but most can only be described as a context for interpreting the results of standardized tabulations.

- MEPS is designed to piggyback on the NHIS sample, sampling from successful NHIS interviews; only persons selected from NHIS and those who later join MEPS families but were not in scope for the NHIS sample are assigned person weights.

- MEPS respondents who are not eligible for person weights may be eligible for family weights, but not everyone who receives a person weight receives a family weight. This means that the samples for person-level and family-level analysis do not overlap completely. Specifically, 10.4 million persons (weighted) with CPS family weights and 13.0 million with MEPS family weights have no person weights, and 6.1 million persons with person weights but one or more non-interviewed family members have no family weights. This design feature is unique to MEPS among the eight surveys.

- MEPS adjusts (post-stratifies) person weights to ensure that the MEPS public use file yields the same poverty rates by demographic groups as the CPS; MEPS also adjusts

(post-stratifies) family weights to ensure that the MEPS public use file yields the same counts by family size and family type as the CPS.

- ACS income data combine 12 reference periods for a given year that on average lag the calendar year by six months; income is adjusted to the calendar year level for inflation but cannot be adjusted for productivity, unemployment or other factors, nor will it fully reflect economic shocks during the year such as sharp changes in energy costs, food prices, or credit availability.

- The ACS rolling sample, rolling reference period and inflation adjustments were examined through tabulations for each separate month, with and without inflation adjustments, and across income levels, but no discernable patterns were found.

- PSID is a panel survey following the same families and their descendants for 40 years, designed for longitudinal rather than cross-sectional work; responding families may no longer be representative and weighting is done at a family rather than person level.

- Preliminary PSID weights use CPS counts of primary families and primary individuals as control totals, excluding unrelated subfamilies and secondary individuals, and do not fully reflect definitional and universe differences between PSID and CPS. PSID weights to 261.5 million persons, compared to 282.6 million in CPS; excluded groups account for 8.1 million of the 21.1 million person difference.

Editing and Consistency. Income data processing typically includes overall consistency checks, such as whether workers have earnings, those with earnings report working, or whether the type of employment—working for others or self-employment—matches the type of earnings reported. MEPS collects employment and dollars of earnings in separate sections of the instrument (and collects the employment data three times per year but dollars of earnings only once a year). In order to maintain the independent information provided by the responses, which sometimes disagree, MEPS does not impose consistency edits. Here and elsewhere, where edits were not made, the study measured the impact.

- In NHIS, 4.3 million persons reported receiving wage and salary or self-employment income for the year but have no work activity or amounts earned in the same year, and another 4.0 million persons reported working, and amounts earned, but no receipt of wage and salary or self-employment income for the same time period.

- In MEPS, 6.6 million persons reported wage and salary or self-employment income for the year but no work activity on the detailed JOBS file of employment for the same time period.

- In MEPS, 2.6 million persons reported details of one or more jobs working for others or themselves during the year but no wage and salary or self-employment income for the same time period.

- In MEPS, 16.5 million persons with only self-employment for the year on the detailed JOBS file reported $620.2 billion of wages and salaries for the same time period. Re-

classifying the entire amount as self-employment income would give MEPS more than any other survey whereas MEPS shows little self-employment income otherwise.

- SIPP skipped around questions on net profits for 2.0 million self-employed in sole-proprietorships and some partnerships when no monthly draw was reported; this omission of some self-employment income was corrected in the 2004 panel.

- SIPP does not edit or impute monthly work activity against monthly earnings or monthly earnings against monthly work activity, yet finds less than one-half million persons with either work activity but no earnings or earnings but no work activity on an annual basis, compared to 8.3 million in NHIS and 9.2 million in MEPS.

NHIS Family Income Consistency. Most household surveys don't require consistency checks on family income, since it is a calculated sum of income across sources and across persons. NHIS gets family income, and earnings (never negative) for persons, but does not determine whether total earnings in a family exceed the family's income.

- For 61.7 million persons and 9.9 million poor, family earnings exceed family income; family earnings are over $10,000 above family income for 27.6 million people and over $20,000 higher for 15.4 million, with the excess totaling about $290 billion.

- Using higher family earnings to determine poverty reduces the poverty rate 1.4 percentage points on either the CPS or NHIS family definition, and the number of poor by 3.9 or 4.0 million for the CPS and NHIS family definitions, respectively.

- Using higher family earnings improves poverty status for another 12.3 million by shifting them from 100 to 200 percent of poverty to above 200 percent of poverty, or from 200 to 400 percent of poverty to above 400 percent of poverty.

- Earnings and/or family income were imputed for most NHIS families with total earnings in excess of total income; they were imputed for 71 percent of all persons with family earnings greater than family income, 83 percent of those whose poverty status changes and 88 percent of those with a difference of more than $20,000.

These excess earnings were excluded when non-CPS families were split to meet CPS family definitions for the study's standardized tabulations. Instead, the combined income of split-off CPS families was constrained to equal the income of the original NHIS family for which the data had been collected.

Income Definition. The CPS income definition used in the study excludes non-periodic or lump sum withdrawals from tax-advantaged retirement accounts, that are likely in the long term to substantially replace pension income based on defined benefit plans. Tabulations to assess the impact of these withdrawals were done in SIPP and MEPS; other differences remain that cannot be assessed.

- Standard tabulations included $3.3 billion of periodic IRA, Keogh or 401(k) payments in CPS and $18.7 billion in SIPP; non-periodic withdrawals of $12.7 billion were restored to income in SIPP but had no significant impacts

- Taxable IRA withdrawals of $65.6 billion were restored to income in MEPS and reduced the overall poverty rate by 0.1 percentage points and the poverty rate for the elderly by 0.5 percentage points

- MEPS uses Internal Revenue Service definitions that exclude contributions to tax-deferred retirement accounts such as 401(k)s from wages, treat income from self-employment other than a sole proprietorship or farm as rents, royalties or estate income, and exclude interest and dividends from tax exempt municipals—these definitional differences cannot be removed and their impact cannot be measured

- None of the surveys collect information on defined contribution retirement benefits comparable to data on income from traditional pension plans

Relationship Detail. Surveys differ in the information collected on relationships within households or families, whether to the reference person or among other household or family members; this may limit information on family structure and reduce flexibility in constructing potential filing units. Surveys also differ in treatment of college students.

- ACS has no information on relationships among persons not related to the household reference person, so that unrelated subfamilies cannot be identified and their members are treated as unrelated individuals; treating the 1.2 million persons in unrelated subfamilies in CPS as unrelated individuals reduces the number of poor by 173,000 and excludes almost 220,000 poor children under 15 from the poverty universe

- SIPP and CPS only identify parental or marital relationships among persons not related to the household reference person, so that only husband-wife and parent-child unrelated subfamilies can be identified, not other related subfamilies, e.g., siblings

- MEPS identifies members and the reference person of CPS-defined families, and while relationships are coded only relative to the MEPS family reference person, there are virtually no cases where the relationship to the CPS family reference person cannot be discerned

- MEPS sample members with person weights but no family weights have family members who are not on the public use file; these sample members represent 6.1 million persons in families of "undefined size"; 2.4 million are in families with no reference person on the public use file

- Persons in MEPS families of "undefined size" have a poverty rate of 34.5 percent and are disproportionately minority, female, children, and single-parents, but less likely than average to be uninsured, on Medicaid, or on welfare or Food Stamps

- SIPP, CPS and MEPS include college students in the parental family and CPS does not interview in dormitories; NHIS and ACS include students where they currently

reside, so those in student housing in the interview month in NHIS become single individuals and in ACS are omitted until 2006; and HRS and PSID treat students away from home as "institutionalized"

- ACS excludes group quarters until 2006; group quarters in CPS have 205,000 residents of whom 115,000 are poor, but CPS includes over two million residents of college or university housing in parental families that the ACS includes in group quarters; for 2006 and later, students living in dormitories are excluded from the ACS poverty universe, but if included could increase ACS poverty rates up to 0.7 percentage points

- PSID retains separate family status for persons—usually grown children or aging parents—previously living on their own but currently living with a related family

Availability and Utility. Most of the surveys have public use files with dollar amounts for income by source for a month or year for every person above some age. The absence of any of these attributes compromises the usefulness of survey income information for policy work.

- NHIS has no actual dollar amounts on public use files, and MCBS has no public use files; MCBS files are available for off-site use with appropriate confidentiality protections but NHIS files with dollar amounts may not be taken off-site, and users obtain and retain only tabular or analytic output

- ACS income data on public use files (which are samples of the internal files) have neither the month of data collection nor month-specific inflation adjustments; an average of the 12 monthly adjustment factors is provided on the public use file but it under-adjusts months early in the year and over-adjusts months later in the year

- NHIS has no person-level income totals and gets family income only on the NHIS family definition, which is not comparable to official statistics; it required complex modeling to create CPS families, and income estimates for any other filing units would be problematic, especially without files available for off-site work

- PSID has a great deal of income detail for the family head and spouse (or partner) but has no income totals for persons nor income by source for other family members

- ACS income amounts on public use files have been rounded (after top-coding) with items below $1,000 rounded to the nearest $10, those from $1,000 to $50,000 rounded to the nearest $100, and above $50,000 rounded to the nearest $1,000

3. Comparisons Across Surveys

Empirical findings using CPS income and family definitions show major differences among the eight surveys, including varying measures of total income, the distribution of income, earnings and earners, number and demographic composition of poor, poverty rates, program participation, uninsured and low-income uninsured. Additional findings on response rates, allocation and imputation rates and rounding provide information on the quality and reliability of

income data. However, standardization cannot adjust for many design features, including the ACS reference period, post-stratification in MEPS, ACS lack of group quarters in 2002, significantly lower population totals in PSID, person-level income data restricted to the family head and wife in PSID, and the contemporaneous poverty measure embedded in PSID. Other survey differences relate to unrelated subfamilies, timing of family composition, treatment of students, and differences in defining income. Most empirical comparisons involve the five large general population surveys and PSID, although the small PSID sample prevents reliable comparisons for small sub-populations.

Total Income and Income Distribution. The largest difference among surveys is a lower total or aggregate income in SIPP, affecting the upper part of the income distribution. Administrative data matches have shown the difference is not due to an underrepresentation of higher-income families in SIPP, and it is possible that the lower SIPP estimates are an artifact of monthly income reporting and shorter recall intervals.

- Excluding PSID, aggregate income ranges from $5.77 trillion in SIPP to $6.47 trillion in CPS, a difference of $702 billion and over 10 percent; the difference is more than accounted for by $884 billion less wages and salaries in SIPP compared to CPS

- Aggregate income is $6.35 trillion in ACS, $6.26 trillion in MEPS, and $6.12 trillion in NHIS; NHIS is $6.41 trillion if earnings are used for families whose earnings exceed income

- PSID, despite a weighted population of 21 million fewer persons than CPS, has the highest aggregate income at $6.72 trillion

- SIPP has the least inequality in income distribution, and NHIS the most, with ACS and PSID close to CPS; NHIS is also close to CPS if earnings are used for families whose earnings exceed income

Earnings and Earners. In all surveys, earnings (wages and salaries plus self-employment income) account for 82 to 86 percent of aggregate income. Numbers of earners and average earnings differ somewhat among surveys but differences among numbers of self-employed or working for others and among amounts earned from wages and salaries and self-employment are much larger.

- Number of earners ranges from 147.4 million in NHIS to 160.4 million in MEPS, with 151.9 million in ACS, 150.4 million in CPS and 154.1 million in SIPP

- Average earnings per worker vary from $30,899 in SIPP and $32,813 in MEPS to $35,707 in NHIS and $35,591 in CPS; ACS is $34,279

- If those reporting work activity in MEPS or receipt of earned income in NHIS, and those skipped around self-employment income questions in SIPP are included, the range on number of earners changes to 150.4 million in CPS to 163.0 million in MEPS, with 151.7 million in NHIS and 156.0 million in SIPP; ACS does not change

- Number of wage and salary workers, reported for the three Census Bureau surveys, has a narrow range, from 140.4 million in SIPP to 142.4 million in ACS; however, SIPP finds more self-employed than either of the other surveys

- Average wages and salaries per worker are lowest in SIPP at $29,514 and highest in CPS at $35, 514, with ACS mid-way between

- PSID gets earnings only for the family head and wife; comparisons with similarly restricted counts in CPS, SIPP and MEPS find higher proportion of earners and higher average earnings in PSID than the other surveys

- Comparisons between PSID and other surveys for wages and salaries follow the same pattern—PSID has the highest proportions of wages and salary workers and higher average wages and salaries per worker than the other surveys

Number of Poor and Poverty Rates. Standardized comparisons of poor and poverty rates show a wide range. Measures for ACS are affected by its lack of group quarters and treatment of unrelated subfamilies, but these factors may have offset each other.

- Total poor and poverty rates (excluding the contemporaneous PSID measure) vary from 33.2 million and 11.8 percent in SIPP to 41.6 million and 14.7 percent in NHIS—a range of 8.4 million people and 2.9 percentage points

- CPS, ACS and MEPS poverty counts and rates are similar to each other, at 34.4 million and 12.2 percent in CPS, 34.6 million and 12.5 percent in ACS, and 35.3 million and 12.5 percent in MEPS; MEPS is post-stratified to match CPS but adjustments for comparability produced differences

- Poverty rates in PSID are even lower than those in SIPP when both are measured on the same contemporaneous basis—9.8 percent compared to 10.6 percent for all ages—and are also lower for age 65 or over, children, whites and blacks

- SIPP finds fewer poor age 65 or over than the other surveys except PSID, and more poor children than other surveys except NHIS; NHIS has 2.3 million more poor children than CPS, 1.4 million of them living in husband-wife families

- Total numbers and percentages below 200 percent of poverty range from 83.9 million and 30.2 percent in ACS to 95.5 million and 33.7 percent in NHIS—a range of 11.6 million persons and 3.5 percentage points

- CPS and MEPS counts and rates of those below 200 percent of poverty are similar to each other, at 86.2 million and 30.5 percent in CPS and 87.5 million and 30.9 percent in ACS; SIPP is somewhat higher at 89.5 million and 31.8 percent

- The rates below 200 percent of poverty in PSID are also lower than those in SIPP measured on the same basis, 25.5 percent for all ages compared to 29.9 percent

Program Participation. Counts of persons with SSI, welfare, on Medicaid, or living in a family receiving welfare and/or Food Stamps vary sharply among surveys, sometimes by a ratio of two to one. Generally, SIPP has the highest levels of program participation, and CPS and PSID frequently have the lowest.

- SIPP finds 3.4 million persons who ever received welfare during the year, compared to 2.9 million in ACS, 2.2 million in CPS and 1.8 million in MEPS

- SIPP finds 8.4 million persons who ever received SSI during the year, compared to 6.4 million in MEPS, 5.5 million in NHIS, 4.9 million in CPS and 4.5 million in ACS

- SIPP finds 31.4 million persons in families receiving welfare and/or Foods Stamps during the year, compared to 24.3 million in ACS, 22.0 million in NHIS, 20.5 million in CPS and 20.2 million in MEPS

- PSID measures receipt of SSI, welfare or Food Stamps only for the family head and wife; comparisons with similarly restricted counts finds 0.9 percent of persons received SSI during the year in PSID, CPS and ACS, and 1.6 percent of persons in SIPP

- PSID and the comparable count in CPS find 7.3 percent of persons living in families whose head or wife received welfare or Food Stamps during the year, and comparable counts find 8.8 percent in ACS and 11.2 percent in SIPP

- SIPP finds 48.1 million persons ever enrolled in Medicaid during the year, compared to 41.2 million in MEPS and 32.9 million in CPS; PSID has little more than half the number in CPS

- MEPS finds 35.0 million persons currently enrolled in Medicaid, compared to 33.3 million in SIPP and 29.9 million in NHIS

Uninsured. Five surveys contain information on who had health insurance coverage during the last year, and for these surveys the uninsured are persons never covered during the year. Three surveys have information on who is currently uninsured. Counts of uninsured differ greatly, in part because uninsured are a residual after positive responses on health coverage, so that low measures of e.g. Medicaid participation can translate into high counts of uninsured.

- CPS finds the highest level of uninsured last calendar year at 41.8 million persons, compared to 33.3 million in MEPS, 27.5 million in NHIS and 22.9 million in SIPP

- PSID, with 21 million fewer persons than CPS, finds 35.5 million persons uninsured last calendar year or 13.6 percent—slightly below the 14.8 percent in CPS but higher than 11.8 percent in MEPS, 9.7 percent in NHIS and 8.2 percent in SIPP

- Uninsured last calendar year with income under 200 percent of poverty range from 22.9 million in CPS, through 18.2 million in MEPS and 17.8 million in NHIS, to 14.2

million in SIPP; PSID has fewer low income persons but finds 22.9 million are uninsured

- Uninsured children last calendar year under 200 percent of poverty range from 5.2 million in CPS and 4.6 million in PSID to 2.6 to 2.9 million in MEPS, NHIS and SIPP

- Counts of persons currently uninsured, a measure not contained in CPS, are much closer—47.5 million in MEPS, 42.9 million in SIPP and 41.3 million in NHIS

- The ratio of current uninsured to never insured last calendar year in the two surveys with both measures is 1.87 in SIPP and 1.42 in MEPS; the ratio is a measure of turnover and a proxy for duration of uninsurance—higher ratios indicate shorter spells of uninsurance

- Counts of currently uninsured below 200 percent of poverty are very close, and number 25.1 million in NHIS, 24.9 million in SIPP and 24.7 million in MEPS; children account for 6.7 million of these in SIPP, 4.8 million in NHIS and 4.7 million in MEPS

- For the uninsured below 200 percent of poverty, turnover rates are lower, suggesting longer spells of uninsurance—the ratio of current uninsured to never insured last calendar year is 1.76 in SIPP and 1.36 in MEPS

Restricted Populations. Two of the surveys cover subsets of the general population—persons age 51 or over, and Medicare enrollees—with limited information and significant differences from other surveys. Tabulations of income and demographics were done on major surveys as comparably as possible for comparison, using the RAND file for HRS.

- Comparisons of persons 51 or over in CPS, SIPP and ACS with the same population in HRS found those in HRS a little more likely to be living with other relatives and less likely to be living alone; comparisons also found higher family incomes in HRS than for comparable persons in CPS, SIPP and ACS, with HRS incomes 20 to 30 percent higher than CPS and SIPP and about 15 percent higher than ACS.

- Comparisons of persons 65 or over in CPS, SIPP and ACS with Medicare enrollees 65 or over in MCBS found little or no differences in living arrangements but substantially more income, $940 billion for 32.0 million persons 65 or over in MCBS compared to $683 billion for 34.0 million elderly in SIPP, $730 billion for 34.2 million elderly in CPS, and $796 billion for 33.6 million elderly in ACS.

- In CPS, SIPP and ACS, average income per person 65 or over living with a spouse is very similar to that of elderly living alone; in MCBS, average income of enrollees living with a spouse is almost double that of enrollees living alone. The MCBS gets income of the enrollee and spouse for married sample persons, although the MCBS sample frame consists of individual enrollees; income of spouses also enrolled in Medicare is represented by other sample persons and is thus double-counted.

Non-Response and Item Non-Response. Non-response in household surveys is a serious issue. High initial rates of refusal (survey non-response) may lead to non-response bias; longitudinal attrition is a lesser issue given the availability of data from earlier interviews. Replacing missing income information (item non-response) through allocation introduces a stochastic element. In addition, methods vary and may also lead to bias. Both the variability and potential bias are reduced when allocations incorporate partial information supplied by respondents, such as bracketed amounts (collected from respondents who would not provide dollar amounts), wage rates and hour worked, and, for panel surveys, amounts reported in earlier waves. We include as allocations our own pro-rating of part-year income in SIPP to create an annual amount. Allocation rates could not be computed for MCBS or HRS but are reported for the other surveys.

- Initial response rates range from over 97 percent for ACS, the only mandatory survey, to 70 percent for MEPS; SIPP and NHIS are 88 and 89 percent, and CPS is 92 percent for the underlying monthly survey, but about 11 percent of persons with income in CPS are whole imputes who have refused to answer the ASEC supplement; the initial response in 1967 for the major component of the PSID sample was 79 percent.

- Allocation rates range from 17.6 percent of total income in ACS to 42.7 percent in MEPS; SIPP, CPS and NHIS have similar rates from 32.4 to 34.2 percent, including whole-person imputes in CPS and pro-rated income for persons present only part of the year in SIPP.

- When allocations based on partial information supplied by the respondent are excluded, allocation rates range from 6.9 percent of total income in SIPP and 7.1 percent in MEPS to 30.2 percent in NHIS. Allocations in the CPS and ACS do not make use of partial information (as defined here).

- In the five major surveys, allocation rates (as percentages of income from that source) are highest for asset and self-employment income; other income sources may have high allocation rates in one survey but not another.

- Nonetheless, allocated earnings account for 77 to 85 percent of allocated income in the major surveys, and allocations of income from other sources range from minimal to less than ten percent of all allocated income in any survey.

- As shares of total income, allocated earnings (with or without partial information) range from 14.5 percent in ACS to 36.4 percent of income in MEPS; in SIPP, CPS and NHIS allocated earnings have similar shares of 25 to 27 percent of total income.

Rounding. Round numbers suggest inexact reporting or approximations, but the percent of persons with income amounts exactly divisible by $5,000 or $10,000 varies with the number of questions, type of income, and allocation method. If many income amounts are summed, rounded totals are less likely, and hot-deck but not regression-based allocations carry rounding over from donor records. The rounding tests were restricted to amounts below $52,500.

- In SIPP, with detailed income questions and monthly data, virtually no one has rounded income amounts, whether reported or allocated

- In NHIS, with single annual amounts, 40 percent of earners and 36 percent of families report amounts divisible by $5,000, and 23 percent of earners and 21 percent of families report amounts divisible by $10,000; no rounding is found in allocations, which are regression-based

- In CPS and ACS, 28 to 30 percent of earners report amounts divisible by $5,000, and 16 to 17 percent report amounts divisible by $10,000; allocations have similar levels of rounding in CPS but are one-third lower in ACS

- PSID and MEPS have less rounding—19 to 23 percent of earners report amounts divisible by $5,000, and 10 to 12 percent report amounts divisible by $10,000; in PSID allocations are higher but in MEPS allocations are one-third lower

- In contrast to earnings, Social Security and retirement income have little rounding—less than 10 percent of recipients of either reported amounts divisible by $5,000 in CPS, SIPP, ACS or MEPS

- PSID has almost no rounding of family Social Security or transfer income of the head and wife—less than 5 percent of families reported amounts divisible by $5,000

4. Conclusions

Many of the study findings address ways in which survey design and methodology impact the utility of survey income data for policy analysis, although some findings suggest simple and feasible improvements. It is clear that the quality of income data varies substantially. In large part this is a reflection of the different purposes of the various surveys. But we also find that design features adopted to enhance the quality of income data do not always work as intended.

It was not within the scope of this study to make recommendations. However, the study provides the groundwork for both a discussion of future directions and work on issues in individual surveys and, hopefully, will be a solid starting place and perhaps the basis for recommendations on survey improvements and future innovations.

I. INTRODUCTION

A. PURPOSE AND NEED FOR THE STUDY

Income is a critical variable in policy-related analyses. Many public programs are designed to address the consequences of inadequate resources. Others address needs that are conditioned by or correlated with low income. Consequently, income, together with poverty status, often plays a key role in the development of public policy. For these reasons, most federal household surveys collect at least some data on income and provide measures of poverty status. Yet income is exceedingly difficult to measure well in household surveys, and poverty status depends not only on the quality of measured income but on how a family is defined, which differs across surveys. Despite many similarities, there are also many differences in the income and poverty concepts used in major federal and federally-sponsored surveys, and different surveys provide markedly differing estimates of income and poverty.

Under contract to the Office of the Assistant Secretary for Planning and Evaluation (ASPE), Department of Health and Human Services (HHS), Mathematica Policy Research, Inc. (MPR) and its subcontractor, Denmead Services & Consulting, have conducted a comprehensive and systematic assessment of the income data and their utility for policy-related analyses in eight major surveys: the Survey of Income and Program Participation (SIPP); the Annual Social and Economic (ASEC) Supplement to the Current Population Survey (CPS); the American Community Survey (ACS); the Household Component of the Medical Expenditure Panel Survey (MEPS); the National Health Interview Survey (NHIS); the Medicare Current Beneficiary Survey Cost and Use files (MCBS); the Health and Retirement Study (HRS); and the Panel Study of Income Dynamics (PSID). This assessment extends the work of the HHS Data Council,

which is summarized in the HHS working paper, "Measuring Income on Surveys: Content and Quality: An Overview."

B. OBJECTIVES OF THE STUDY

The assessment focuses on three issues:

- The quality and usability of each survey's income and poverty data for policy-related analyses
- The overall impact of different design and methodological approaches
- Specific design and processing choices that may be related to the quality and utility of income and poverty data in each survey

We discuss these three issues in turn and then highlight the methods used in the study.

1. Income Data in a Policy-Analytic Context

The data requirements for policy analysis are not the same as the requirements for more general research. They are both different and more extensive. Whatever the issue being addressed, good income information for policy work is likely to require several additional qualities, outlined below.

In policy analyses, income is often used to determine potential eligibility for a new or existing program. Program benefits or charges to the participant may vary with income. Understanding the impact at different points in the income distribution is important. As a result, policy work requires income expressed in dollar amounts, not fixed income brackets.

The concept of poverty has an official definition and an official source of measurement in the CPS. Poverty status is important in policy evaluation and in public debate and, therefore, must be expressed on the same basis as is done in official statistics. Departures from official concepts may be useful for a variety of purposes, but they need to be tied back to the official statistics.

Policy analysis of health issues usually requires data on health insurance status and utilization of health care. Analyses of means-tested programs frequently require estimates of earnings separately from total income, as earned income is often treated differently than unearned income. Policy analysis on issues affecting the elderly requires data on current retirement contributions and pension coverage.

Policy analysis may deal with individuals or part of a family, and may compare different filing units. To analyze units below the family level requires income data for each person. Health insurance provides a relevant example—the filing unit may be children up to a certain age based on family income, as in the State Children's Health Insurance Program (SCHIP); or a worker, spouse, and dependent children, under a private health insurance plan; or it may include children outside the household in the case of divorce. Frequently, the construction of filing or eligibility units is one of the most challenging aspects of policy analysis and one that is exacerbated by limitations of the data.

Weaknesses in the data underlying policy proposals and cost estimates bring the validity of an initiative into question, even though such weaknesses may not be directly relevant to the estimates. Significant inconsistencies within a survey provide the basis for challenges to the proposals themselves based on the unreliability of the estimates. Differences with alternative estimates of totals on which surveys should seemingly agree invite challenges as well. For example, while the nature of population estimation provides some leeway (the Census Bureau revises historical population estimates each year), significant differences in population totals can hurt credibility.

Efficient policy design requires detail on benefits or insurance coverage already in place and, along with it, the administrative systems with which persons already interact. Participation

in Social Security, Medicare, Supplemental Security Income (SSI), Food Stamps, Medicaid, and welfare is particularly important.

Typically the policy process has tight time frames, particularly when legislation is being written or negotiated or a vote is impending. Unexpected developments growing out of the need to secure votes or satisfy specific constituencies may require new analyses—sometimes with substantial changes—with very quick turnaround times. Immediate access to the data on which the analyses are based is critical, as is the ability to conduct needed analyses without restriction.

With regard to the income data specifically, accuracy at the lower end of the income distribution is more important than accuracy at the upper end. Income measures that are of particular significance include the number and composition of the poor and near-poor and the magnitudes of key income sources, such as earnings and program benefits. In addition to income, employment has consistently been an area of policy concern, both as a source of self-support and the source of most health insurance coverage. Accuracy in the measurement of employment is also critical. Lastly, randomness as measured by standard errors is not nearly as important as possible bias. The findings of policy analysis, and budget estimates, are presented as point estimates without standard errors. Bias, on the other hand, leads to consistent over- or under-estimates.

2.   Survey Design and Methodology

The single biggest design difference across the eight surveys with respect to income data collection contrasts the subannual approach used in the SIPP and, to a limited degree, the PSID, with the retrospective annual approach used in the other surveys. SIPP collects monthly income data from interviews conducted at four-month intervals. Users of SIPP data may cumulate monthly incomes in any way they wish, and we demonstrate this in constructing estimates of annual income from the SIPP. But the SIPP approach is distinctly different from directly asking

people their annual incomes. With the Census Bureau engaged in a redesign of the SIPP that is focused on replacing the current three interviews per year with a single annual interview, the merits of this particular design feature carry significance beyond its methodological interest.

The second biggest design difference among these surveys is the range in the number of questions used to capture total income. Detailed questions on income serve an important purpose beyond whether they lead to better estimates of total income or not, and one would not discard detailed questions from a survey whose major purpose is to capture the breadth and variety of income. But the issue of what level of questioning is needed to capture adequate income is very relevant to surveys that collect policy-relevant or simply analytically important data on other topics but whose users would benefit from the availability of a reasonably good measure of total income.

The use of a rolling versus fixed sample is also a major design difference represented among the eight surveys. A rolling sample consists of non-overlapping subsamples spread across a year (or other time period) and designed to be interviewed sequentially. Within the context of a rolling sample there is an additional difference with respect to the reference period for which annual income data are collected. The ACS uses a rolling reference period, asking respondents their income for the past 12 months while NHIS uses a fixed reference period, asking respondents their income for the previous calendar year. A corollary of the NHIS approach is a variable recall interval, where longer length of recall may bring higher nonresponse and lower-quality data.

Four of the remaining five surveys are longitudinal. They involve repeated interviews with the same individuals over multiple years.[1] Attrition and limited representation of additions to the

---

[1] The CPS design is longitudinal as well, in that sample addresses are revisited seven additional times after the first interview. Unlike the other longitudinal surveys, however, the CPS does not follow sample members who move to new addresses.

population will tend to make individual panels less representative of the U.S. population over time.[2] This is a serious concern with respect to the PSID, which has followed an initial sample of households for 40 years. Another aspect of longitudinal surveys involves the impact of being interviewed multiple times with variations on the same instrument. It is conceivable that the repetition of the income questions—particularly in the SIPP, where the interval between interviews is brief and all of the income questions recur—may improve the quality of the responses over time as respondents learn what to expect. However, our analyses do not explicitly address these features of longitudinal surveys.

3.   Additional Design Elements and Post-Survey Processing

In addition to the fundamental survey design features discussed in the preceding section, there are a number of additional design elements that may affect the data collected on income and poverty. Components of the post-survey processing of survey data may have important effects as well. All of these elements are relevant regardless of the overall survey design. These elements include:

- Family definition, which determines whose income is aggregated and what poverty threshold is used to determine poverty status

- Contemporaneous versus fixed family composition and income for poverty measurement—that is, whether family composition and income reflect changes in composition over the reference period or whether family composition is measured at a fixed point in time and income collected for the members of this fixed family

- Interview month, which affects recall intervals, family composition, the lag between a fixed family composition and the income reference period, response rates, and the quality of income data

- Choice of imputation methodology, including its impact on the distribution of imputed values and their consistency with reported values

---

[2] Both MEPS and MCBS employ overlapping panels, so annual estimates should be less susceptible to these influences.

6

- Application of consistency checks between related items collected at different places in the questionnaire

- Application of inflation adjustments when income reference periods differ

- Post-stratification in general and post-stratification on income in particular

Each of these can affect the quality of the income data that are ultimately released to users and how the income and poverty data compare among surveys.

4. Study Methods

The assessment presented in this report includes both descriptive and empirical components. The descriptive component, which is presented in Chapter II, provides extensive detail on survey design and methodology as well as on income data and poverty measures for persons and families in each of the eight surveys. The information presented in parallel for the eight surveys includes overall design, timing, recall, reference period, family definition, poverty measurement, content on income and policy-related covariates, income data processing, and public availability and accessibility of income and poverty data. An annotated bibliography of literature relevant to the collection and evaluation of income data was assembled separately from the descriptive component and is presented in Appendix A.

The empirical portion of the report presents findings from comparative tabulations, following a standardized format, that addresses income, poverty, and program participation. These estimates were prepared using the same income measures, definitions and units of analysis for each survey, to the extent that this was possible. Additional findings address methodological issues, specific survey attributes, and questions raised by the detailed information gathered for the descriptive component. These findings focus on the implications of particular design choices.

The empirical analysis does not include any effort to compare the survey estimates with independent benchmarks, which would require a separate study in and of itself. Benchmark

construction is difficult because administrative data that are often used to produce benchmarks rarely allow the same degree of flexibility in matching universes and units that we were able to achieve with the survey data alone. Administrative record matches to survey data offer a more promising avenue of research, but they are constrained by legal restrictions on access to administrative data and are very expensive to conduct. A small number of studies using benchmarks or matched survey and administrative records are cited in the annotated bibliography.

Neither do we view any of the surveys as a gold standard against which we can judge the quality of the other surveys. We find it informative to compare the other surveys to the CPS ASEC supplement, given this survey's status as the official source of income and poverty statistics for the U.S., but such comparisons may be just as informative about the CPS income data as they are instructive about other surveys.

The scope of work for this project specifically excludes recommendations. Rather, the project hopefully provides the material for a separate, independent review that would focus explicitly on recommendations, perhaps including some additional, targeted research as well. The conclusions presented in Chapter VII focus on factual findings and documentation of similarities and differences among the eight surveys.

Finally, a Technical Advisory Group (TAG) representing each survey and the policy research community provided input to the project. TAG members reviewed and commented on drafts of the workplan, the annotated bibliography, the analysis plan, the outline of the final report, the detailed survey descriptions, and the final report. TAG members, Census Bureau staff, and PSID staff at the University of Michigan also provided extensive assistance in obtaining documentation not readily available from published sources or public web sites. In addition, the Census Bureau also performed a major series of tabulations pro bono on the internal files of

monthly ACS data. These tabulations provided valuable information that could not be obtained from public use files.

C. OVERVIEW OF SURVEYS AND SIMILARITIES AND DIFFERENCES

In addition to the design features already discussed, there are additional features of the eight surveys that should be noted.

While the CPS is the official source of monthly data on the labor force and employment, the survey has also collected income data for almost 60 years. The ASEC supplement, sponsored by the Census Bureau, is the source of official estimates of income and poverty, and is widely used for policy analysis and legislative cost estimates. The CPS collects detailed annual income information for the prior calendar year once a year. The basic purpose of the CPS—labor force information—suggests that it will be most accurate in the areas of wages and salaries and earned income generally.

SIPP is a longitudinal survey sponsored by the Census Bureau that collects a broad range of information relevant to public policy formulation for income security, retirement and health programs, including within-year patterns of income and program participation. SIPP was designed to address a wide range of policy-analytic needs, including estimation of persons eligible for means-tested programs. Panel households are interviewed three times a year at strict four-month intervals to collect month-by-month information for each person. In SIPP, annual income is obtained by adding up 12 months of data for each person. SIPP questionnaires and field methods are intended to maximize the accuracy of income data, especially for lower income persons with intermittent or irregular income sources, and persons with public program benefits. SIPP is unique among the eight surveys in supporting detailed analysis of short-term behavioral dynamics. This project is especially timely for SIPP, which is undergoing a major redesign that may produce a substantially altered design by early in the next decade.

The ACS, which is also conducted by the Census Bureau, was designed to replace the decennial census long form by collecting the same type of data on a rolling basis rather than only once every ten years. As of 2005 the ACS collects data from 2 million households each year, with an annual sample of group quarters added in 2006. Like the long form the ACS will make available a common set of variables—mandated by law—down to very small levels of geography. The ACS will provide annual estimates for states and the largest counties and municipalities plus three-year and five-year rolling averages for smaller areas of geography.

MEPS is an annual longitudinal survey sponsored by the Agency for Healthcare Research and Quality (AHRQ) with field work conducted by Westat; it replaces earlier one-time longitudinal surveys to provide detailed information on health status, health care, and health care costs. The MEPS sample frame consists of households that participated in the prior year NHIS. MEPS collects annual income information for the prior calendar year once a year, and the Full Year files combine contemporaneous health and income data from overlapping two-year panels for cross-section analysis. MEPS is designed for policy analysis requiring income data, as well as data on health care costs, health insurance coverage, and third-party payments.

The NHIS is a cross-section survey sponsored by the National Center for Health Statistics (NCHS) with field work conducted by the Census Bureau. It is the primary source of information on health status and health care in the United States and is widely used for health-related analysis—particularly of trends. The NHIS is in the field continuously during the year, with an annual sample (consisting of four, nonoverlapping representative panels) that is assigned, first, to four calendar quarters and then, within quarters, to individual weeks. Each weekly subsample is representative of the target population. From this rolling sample the NHIS collects summary annual income information for the prior calendar year. Historically, the NHIS has collected only limited information on personal and family income.

The PSID is sponsored by ASPE and the National Science Foundation and is conducted by the Survey Research Center of the Institute for Social Research at the University of Michigan. The PSID was initiated in 1968 with a sample of approximately 5,000 families selected from two sample frames. Members of this initial sample and all of the families that they have created or joined have been followed continuously, with annual interviews through 1997 and biennial interviews starting in 1999. A Latino supplement was added in 1990 to help compensate for the survey's under-representation of part of the immigrant population. This supplement was later dropped, due to insufficient funding, but a new and more broadly representative sample of immigrants was added in 1997. Where the SIPP was designed to support analysis of short-term dynamics of income, program participation, and related characteristics, the PSID was designed to study long-term dynamics.

The HRS, which is also a panel survey, began with a sample of households containing at least one individual born between 1931 and 1941. Sample members were first interviewed in 1992 and have been reinterviewed every two years since then. A second cohort of "war babies," born 1942 to 1947, was added in 1998. A companion survey, the Asset and Health Dynamics Among the Oldest Old Survey (AHEAD), was started in 1993 with a sample of persons born in 1923 and earlier. A third HRS cohort of "children of the depression," born from 1924 through 1930, was introduced in 1998 to fill the gap, and all of the cohorts have since been shifted to the same interview schedule to facilitate pooling of the data across cohorts. With these additions the HRS/AHEAD sample became representative of the U.S. resident population born before 1948—that is, 51 and older by the end of 1998. A new cohort was added in 2004 representing persons born between 1948 and 1954. Sample members are interviewed every two years. The HRS has employed a number of survey methodological innovations with respect to the collection of data

11

on income and wealth. The income detail that it collects falls between that of the ACS and the CPS, so the HRS demonstrates what can be accomplished with a moderate number of questions.

The MCBS is sponsored by the Centers for Medicare and Medicaid Services (CMS) and is a longitudinal survey of Medicare beneficiaries. A new sample is drawn every year, and sample members are interviewed 12 times over a four-year period. MCBS data are released in annual files that pool four consecutive cohorts. MCBS is unique in that the survey data are not the final product. Cost and utilization data from Medicare claims files are added to the survey data along with information on non-covered medical services. Income data are limited to a single total.

Only two of the surveys—the SIPP and the ASEC Supplement to the CPS—were designed explicitly to measure income, but income is also a major focus of the data collection in both the PSID and HRS. The ACS income data are much more limited than what is collected in the CPS or the SIPP, but income is still considered one of the most important characteristics collected by the survey. By contrast, the measurement of income in the MEPS, the NHIS, and the MCBS is decidedly secondary to the main objectives of each survey. MEPS, nevertheless, collects more detailed income data than the ACS while NHIS collects just total family income and personal earnings (along with receipt of multiple sources) and MCBS collects only the sample member's total income, including that of a spouse.

Five of the eight surveys can be described as general population surveys. But while all five cover essentially the same universe—the full civilian, noninstitutionalized population resident in the United States—no two surveys represent this population at the same point in time. In fact, only the ASEC supplement comes close to capturing the population at a *single* point in time. CPS-ASEC respondents are interviewed primarily in mid-March of each year, but some supplemental interviews—part of a 2001 sample expansion—are conducted in mid-February and mid-April. The survey is weighted to March 1 population controls. The SIPP fully represents the

population only in the first wave of each panel. Over the length of a SIPP panel, people who leave the survey universe are no longer represented, and new entrants through birth are almost fully represented, but immigrants, people returning from abroad, and people released from institutions and the military are represented only if they move in with persons who were included in the SIPP universe at the start of a panel. For cross-sectional estimates, the SIPP is weighted to the full civilian noninstitutionalized population in each month, but this becomes a less accurate reflection of the survey's true universe with each passing month.

As noted, the ACS and the NHIS both use a rolling sample that covers the entire year. For simplicity, the ACS is weighted to mid-year (July 1) population controls while the NHIS is weighted to quarterly population controls to enable users to estimate disease prevalence at different times of the year. The MEPS is a subsample of the NHIS, drawn from completed interviews; MEPS respondents are interviewed multiple times over a two-year period to provide data for the two calendar years following the NHIS survey year from which they were drawn. A single MEPS panel represents the survivors of the population represented by the NHIS sample from which they were drawn—plus births to this population. However, AHRQ also releases annual files that pool two adjacent MEPS panels; the combined sample is weighted to population totals for that calendar year.

Cross-sectional estimation is not the purpose of the PSID, so concerns about how well it continues to represent the general population after 40 years detract only marginally from its value. They do require caution, however, whenever comparisons with other surveys are used to draw inferences about the quality of data in either the PSID or the other surveys. The PSID is included in this project in large part because certain features of its collection of data on income and program participation are being considered in the redesign of the SIPP, but only PSID's use of an annual interview to collect monthly data proved relevant to our findings, and those findings

do not provide any insights into the effective capture of monthly information with an annual interview.

The remaining two surveys, the MCBS and the HRS, represent restricted populations—that is, subsets of the general population—and will be used in this project to help assess the quality of income data on persons 65 and older and persons 51 and older, respectively.

D. ORGANIZATION OF THE REPORT

The survey descriptive portion of the study is presented in Chapter II, which provides detailed, side-by-side descriptions of the surveys along 14 broad dimensions. The empirical analysis is presented in Chapters III through VI. Chapter III describes the methodology, including the steps taken to generate comparable estimates across the surveys, the specification of a set of standardized tabulations, and the design of specialized tabulations addressing a range of specific survey design and definitional issues and exploring internal consistency within individual surveys. Chapter IV presents findings based on the standardized empirical comparisons. Chapter V provides the results of comparisons across survey design, definitional and methodological issues. Chapter VI presents findings with respect to income allocation, approximation and rounding. Chapter VII provides a synthesis of our findings, integrating the descriptive and empirical analyses. Appendix A contains an annotated bibliography of published and unpublished literature on income data while Appendix B provides references and links to questionnaires, data dictionaries, and documentation.

II. DETAILED DESCRIPTIVE ANALYSIS

This chapter presents detailed descriptions of the eight surveys in the study, including overall survey design and methodology, universe, timing, data collection, key definitions, questionnaire content related to income and other policy-relevant topics, processing, and public availability of income and other data. To facilitate comparisons and minimize length while presenting precise information, the substantive content has been arranged in side-by-side descriptions of the eight surveys across 14 domains.

The specific descriptions apply to the files used in the study—the 2001 SIPP panel, 2002 files for ACS and MEPS, 2003 files for CPS, NHIS, MCBS and PSID, and 2004 files for HRS— and are not necessarily applicable in all detail to other years. Surveys are not static, and survey content, procedures, sampling and data may change from year to year. In addition, the specific descriptions apply to data available on public use files unless otherwise noted. Public use files may contain less detail than internal files, and less detail than shown on questionnaires, since data are frequently aggregated, limited or partially suppressed for confidentiality, quality, or other reasons before public release. NHIS tabulations used the internal file since the public use file has no income amounts, $5,000- and $10,000-wide brackets, and both the public use and internal files are described where they differ. MCBS has no public use file and all descriptions apply to the internal files.

The terminology used in these descriptions has been standardized across surveys and often differs from descriptions in survey documentation. The review of survey materials made clear that various surveys apply the same term to somewhat different concepts or measures, and/or use different names for the same concepts or measures. Using the same terminology for all surveys

was the only way to present accurate descriptions that enable readers to determine whether surveys are in fact identical or differ across the characteristics and procedures being compared.

The standard terminology in this chapter employs Census Bureau and CPS definitions, e.g., household refers to all persons residing in a housing unit or group quarter, whether or not they are related. The descriptions note when usage of these terms in specific surveys departs significantly from CPS terminology, e.g., when a family may contain persons not related by blood, marriage or adoption, or when the term household refers to families, or when family income may include amounts not part of pre-tax money income for CPS. We have tried to include the CPS definitions of terms either in the descriptions themselves, or in the content summaries below. For any terms not defined in one or the other location, several sets of definitions are available on the Census Bureau web site.[3]

A. CONTENT SUMMARIES

Within each of the 14 domains, from five to 13 aspects of each of the eight surveys are described. These domains and their aspects are as follows:

Table 1. Background and Overview provides brief thumbnail descriptions of survey purpose, design, history, file availability and organizational responsibilities.

Table 2. Survey and Sample Design summarizes sampling frames, units, oversamples and response definitions and rates. Housing units, as distinct from group quarters, have cooking facilities and separate entrances. Response thresholds are the criteria that must be met for an interview to be deemed successful rather than non-response from a household or family that

---

[3] See <http://www.census.gov/population/www/cps/cpsdef.html>; the Technical Documentation pp. 9-1 through 9-12 available at <http://www.census.gov/apsd/techdoc/cps/cpsmar03.pdf>; and Technical Paper 66 Chapter 5 available at <http://www.census.gov/prod/2006pubs/tp-66.pdf>.

could have been interviewed. Initial response rate is the response rate for one-time surveys or the response rate at the first interview for longitudinal surveys.

Table 3. Universe Definitions, Inclusions and Exclusions specifies precise universes, geographic coverage, definition and inclusion or exclusion of specific types of group quarters, treatment of college students living away at school, and of active military, institutionalized and decedents, plus any exclusions not already specified. Institutions are always group quarters, but many group quarters are non-institutional. Military barracks are non-institutional group quarters but excluded—except for the ACS from 2006 forward—as not civilian. College dormitories are non-institutional group quarters but are treated differently in the various surveys.

Table 4. Timing and Fieldwork describes design and fieldwork time frames and timing, rotation patterns, duration in sample for longitudinal surveys and the monthly survey underlying the CPS (ASEC supplement), who is interviewed, and how. Three surveys—SIPP, CPS and ACS—collect data for a household (all persons dwelling in the housing unit or group quarter) and the others for persons in a family (or narrower) unit. This table also describes the elaborate follow-up process for the mail-out ACS.

Table 5. Longitudinal Inclusion and Follow Rules for the five longitudinal surveys that follow persons over time, summarizes the complex rules on inclusion, exclusion and retention in sample over time, and describes when data is collected for persons no longer in sample, re-contact efforts, and attrition. The monthly survey underlying the CPS ASEC, although it returns to the same addresses repeatedly, is not included since it does not obtain longitudinal information on specific persons–persons moving from the sample address leave the survey.

Table 6. Family Definitions specifies the meaning of the terms "family" and "spouse" for each survey, and differences from the CPS definitions for surveys using these terms differently. This table also provides the definitions of related and unrelated subfamilies for surveys that use

those concepts, and summarizes the information available in each survey on relationships, subfamilies, marriage and parents. When descriptions say an item, e.g., parent or legal spouse or sub-family, is "identified", it means there is a separate variable or marker on the file with that information. Surveys that interview at the family (rather than household) level either exclude unrelated subfamilies (HRS and PSID) or treat them as a separate primary family (MEPS and NHIS). When treated as a separate primary family, a family reference person is identified and the same information is obtained as for the household's primary family. PSID and HRS (neither of which use a CPS family definition) sometimes use the term household interchangeably with family, and HRS uses the term household for a one- or two-person unit that may be part of a family.

Table 7. Work Activity and Earnings provides short descriptions of employment and labor force information available for the income reference year, level of detail on industry and occupation, and for what persons. This table also describes employment and labor force information available for other reference periods, and whether employment data and earned income data are cross-edited for consistency. CPS definitions of labor force status (used in official statistics) draw a clear distinction between unemployment and not in the labor force. A person must be both available for work and have been actively looking for a job in the past four weeks to be classified as unemployed. In labor force statistics, full time work is 35 hours per week. Class of worker as used in CPS and labor force data is a brief categorization of a person's employment as either private, armed forces, federal, state or local government, incorporated or unincorporated self-employment, working without pay or not working.

Table 8. Pre-Tax Money Income describes the detail, reference periods, differences from CPS definitions, population covered, recall interval, and person as compared to family level of income data available for each survey, and how it is collected. Descriptions are based on data

files, not on questionnaires. The aspect "Screeners" describes whether yes/no or other questions are used to identify persons receiving income from specific sources so non-recipients can be skipped around questions on amounts. The aspect "Brackets" describes data on dollar ranges when respondents don't know or refuse queries on exact dollar amounts. The entry "brackets" indicates an offer of a number of ranges from which to choose. The entry "unfolding brackets" indicates a less direct method of determining a dollar range, where the respondent is asked if the amount exceeds some (entry) level, then asked whether (depending on the response) it exceeds or is less than a succession of steps until both upper and lower bounds, e.g., a bracket, have been established.

Note that HRS, and RAND materials on the HRS, apply the term "household" to the age-eligible person and spouse or partner, regardless of the other related or unrelated persons with whom they may be living. RAND "household income" refers only to the income of the surveyed individual or couple.

Table 9. Income Allocation and Top-Coding on Public Use Files summarizes some of the changes made in processing raw survey data to fill in blanks, improve quality and/or protect confidentiality, especially top-coding and suppressions, and whether changes in income data to protect confidentiality prevent tabulations on public use files from matching published totals.

Table 10. Poverty Status describes the poverty status (ratio of family income to the poverty threshold used in official statistics for families of that size and composition) that has been and/or can be calculated, and how the universe, family, income and/or timing differ from the official poverty measure contained in the CPS.

Note that poverty status calculations for the PSID have a different measurement structure than in the CPS and can only be replicated in SIPP. In the CPS (and other surveys), poverty thresholds based on family composition as of a fixed date are compared with prior year income

of members of the family as of that date. In the PSID, an average annual poverty threshold that reflects changes in family size or composition during the year (is a weighted average of the thresholds appropriate for different part-year compositions) is compared with prior calendar year family income calculated by including part-year amounts for persons there only part of the year.

Table 11. Non-Cash Benefits and Health Insurance summarizes information available, and for which persons, on Food Stamps, other nutrition, housing, energy and welfare to work assistance, and the detail and timing (current coverage, ever-covered prior year, or month-by-month in the prior or current year) of information on public and private health insurance including coverage from or to persons outside the household or family.

Table 12. Person-Level Health and Health Care Utilization describes information on health status, disability, health care services utilization, health conditions and whether conditions associated with disability and/or limitations in activities and/or utilization of health care services are identified, informal care, providers and types of services, and payments, costs, and sources of payment. The event or encounter level information available in MEPS and MCBS constitute a group of very large separate files comparable to (and for MCBS based on) medical claims or bill files.

Table 13. Weights and Control Totals provides an overview of weighting strategy and post-stratification, what weights are available for cross-section and longitudinal analysis, and the relationship of person weights and universes to family or household weights and universes.

Table 14. Ease of Access describes the availability of files with income data, including cost, approval or other barriers or limitations on use, complexity of files and degree of variable construction or assembly needed for calendar year analyses, and availability, accessibility, comprehensiveness and content of survey and file descriptions, questionnaires, data dictionaries,

interviewer instructions, technical descriptions of sample design and weights, glossaries and technical assistance.

B. IMPORTANT DIFFERENCES

Besides obvious differences in purpose, sample size, response rates, number of income items and interview frequency, there are many important differences among the surveys that affect the quality and utility of their income data for policy analysis. The immense quantity of information in the tables about the design, definitions, fieldwork, content and processing of each survey is not easily summarized. As a guide to users, some survey features which are unique or not well publicized, yet have significant impact on the potential utility of a survey's income data for policy analysis, are noted here.

ACS Universe. While most of the surveys are described as covering the resident civilian non-institutional population, that is not precisely the case for any of them. The largest difference is for ACS, which was designed to replace the decennial census long form and uses decennial census definitions, including "current residence" rather than "usual residence".

- Until 2006, ACS excluded all residents of group quarters, whether institutional or non-institutional, including over two million students in college and university dormitories, but included all active duty military not living in group quarters such as barracks

- Current ACS data for 2006 and beyond covers the U.S. resident population whether active duty military or not, and in 2006 has 4.1 million persons in institutional group quarters, including 2.1 million persons in adult correctional facilities and 1.8 million in nursing homes and skilled nursing facilities

- ACS data for 2006 also include 3.9 million persons in non-institutional group quarters that include active duty military, mostly in barracks, and 2.3 million students residing in college and university dormitories and treated as unrelated individuals

Universe for Other Surveys. The other survey universes also differ from the resident civilian non-institutional population, although not usually as much as ACS.

- CPS and SIPP both include active duty military living with one or more related civilians age 15 or over, on or off base

- MEPS and NHIS both include civilians living with active duty military on or off base, and the income of the active duty person, but give that individual a zero person weight

- MEPS and NHIS both exclude unrelated minors age 15 or over, who are included and have income data in SIPP, CPS and ACS

- SIPP, HRS and PSID include persons who later join sample households, and they are assigned both person weights and household weights for the months during which they live with sample persons; MEPS piggybacks on the NHIS sample, sampling from successful NHIS interviews, and persons who later join MEPS families are not assigned person weights for cross-section analysis unless they were out of scope for the NHIS

- MCBS includes all Medicare enrollees including those in institutional as well as non-institutional group quarters, and also includes Puerto Rico

- HRS includes all sample persons including those in institutional as well as non-institutional group quarters and those living in other countries, but in the RAND-HRS files there are no person weights for those in nursing homes

- HRS includes all persons and their income who are current or former spouses or partners of sample persons, but gives those under age 51 a zero person weight

- PSID currently includes some persons living in other countries as well as military and institutionalized persons under some circumstances, but excludes most students living in college housing

Students. There is a broad range in treatment of college students, as noted above for ACS, revolving around the treatment of college dormitories and their residents.

- NHIS, like ACS, treats college and university housing as group quarters and those in student housing in the interview month become single individuals

- SIPP, CPS and MEPS include students in the parental family and CPS does not interview in dormitories at all

- HRS and PSID treat students away from home as "institutionalized" – HRS always excludes them and PSID usually excludes them

ACS Reference Period. Unlike the other seven surveys, ACS income data do not cover the same time period for everyone on the file. ACS gets income for the 12 months prior to the

interview, and for a given calendar year the ACS income data are a combination of 12 different 12-month time periods, depending on the month of data collection. Published data and on-line tables based on internal files have been adjusted to the same real dollars using the CPI-U through 2005, and the CPI-U-RS for 2006 and later years.

Timing of Family and Poverty Measures. The surveys differ in the how family composition is measured and how poverty status is calculated. MCBS does not get family information and thus has no poverty measures. Five of the surveys use full-year income for family composition as of a fixed date to calculate poverty measures. PSID takes a different approach that is not definitionally equivalent to official poverty statistics.

- Family composition is measured December 31 of the income year in MEPS, the month after the income year for ACS, usually March after the income year for CPS, and ranges from January to December after the income year for NHIS, and all use a full year's income for each person in the family as of that date

- PSID contains part-year income for part-year family members and a family poverty threshold based on month-by-month family composition to calculate a contemporaneous poverty status for the PSID family

- SIPP monthly data allow analysts to select the timing used in poverty measures and allows both full-year fixed-date and contemporaneous measures

Family Definition and Poverty. Official poverty statistics incorporate the family definition of the official source of such statistics, the CPS. A different family definition changes who is included or excluded from the family, which affects not only family size but who contributes to family income, and so can change the family's poverty status.

- NHIS and MEPS define a family to include unmarried partners of either sex and children or other relatives of the partner, and foster relationships. Partnerships of any duration are treated as marriage.

- NHIS uses only the broader family definition; MEPS uses both definitions.

- HRS defines a family in the same way but restricts income data to the age-eligible sample person and spouse or partner. Although the age-eligible sample person and

spouse or partner are referred to as a household, they may be members of a larger CPS family and not contain the CPS family householder.

- PSID defines a family to include unmarried partners but only of the opposite sex, as well as children or other relatives of the partner, foster relationships, and any unrelated persons who are identified as part of the family (which may include same-sex partners). Partnerships of at least one year's duration are treated as marriage.

Identifying Families. Various of the surveys also have limitations in how families are identified.

- ACS has no information on relationships among persons not related to the household reference person, so that no unrelated subfamilies can be identified

- SIPP and CPS identify only parental or marital relationships among persons not related to the household reference person, so that only husband-wife and parent-child unrelated subfamilies can be identified, not other related subfamilies, e.g., siblings

- MEPS identifies members and the reference person of CPS-defined families, but not their relationship to the CPS family reference person; however, while relationships are coded only relative to the MEPS family reference person, there are virtually no cases where the relationship between a CPS family member and the CPS family reference person cannot be discerned

- Some MEPS families have members not on the public use file—these are designated as families of "undefined size" and close to half have no reference person on the public use file

- PSID retains separate family status for persons—usually grown children or aging parents—previously living on their own but currently living with a related family

Income by Source by Person. Most of the surveys, like the CPS, have public use files with dollar amounts of income from different sources for a specific time period for each person above some age. Income for each person is the sum of income by source, and family income is the sum of income by person. There are some exceptions.

- NHIS provides only brackets and not dollar income amounts on public use files. The highest family income bracket was set in 1997 at $75,000 or more and included 28 percent of persons on the file for 2003 (the 2007 redesign of NHIS raised this bracket to $100,000 or more).

- NHIS internal (non-identifier) files containing dollar amounts are never available for off-site use.

- NHIS has no income data by person, although there are earnings amounts for each person age 18 or over.

- NHIS has total income only for the NHIS family, and complex modeling is required to create income for a CPS family to calculate poverty rates equivalent to official poverty statistics. Creating income estimates for policy purposes for other filing units—such as parent(s) and own children under age 22—would be even more difficult.

- PSID contains total family income and income data by source excluding Social Security for the family head and his spouse or partner (family heads are the male in a couple), but not income by person or by source for other family members, and summary recodes combine the income of the head and spouse (excluding Social Security). Social Security is available as one total amount for the family.

- MCBS income amounts for married sample persons include the income of the spouse, although the sample is enrollee-based and spouses also enrolled in Medicare are separately represented in the sample.

- ACS income data on public use files (which are samples from the full internal file for a year) do not contain the month of data collection nor a month-specific inflation adjustment. The price adjustment variable on the file is an average of the 12 monthly adjustment factors, and under-adjusts early months and over-adjusts later months.

- ACS income amounts on public use files have been rounded (after top-coding). Income amounts from $10 to $1,000 are rounded to the nearest $10, amounts from $1,000 to $50,000 are rounded to the nearest $100, and amounts above $50,000 are rounded to the nearest $1,000.

Income Definition. Official poverty statistics incorporate the income definition of the official source of such statistics, the CPS. A different income definition can change family income and so can change whether the family is counted as poor. Small differences may affect few persons, but large differences or ones affecting many persons in an important demographic group can alter results. There are some significant differences.

- MEPS income questions use Internal Revenue Service definitions, since the questions reference specific lines on the personal income tax return. Some of these definitions differ significantly from CPS definitions, e.g. taxable wages exclude tax-deferred contributions to retirement accounts (such as 401(k)s, traditional IRAs, 403(b)s and the Federal Thrift Plan) or to Health Savings Accounts; self-employment income refers only to sole proprietorships and farms—other self-employment income from partnerships or S corporations is included with other Schedule E income from rents, royalties and estates; and interest and dividends exclude payments from tax exempt municipals.

- SIPP, MEPS and HRS include non-periodic (lump-sum) withdrawals from tax-deferred retirement accounts (such as 401(k)s, traditional IRAs, 403(b)s and the Federal Thrift Plan) and (except for MEPS) from tax-advantaged Roth IRAs, which are increasingly important sources of income for the elderly. These sources are likely to substantially replace pension income based on a defined benefit plan in the long term, but are not included in CPS money income.

- HRS rental income is gross rent before deduction of expenses such as mortgage or tax payments. HRS income for 2003 excludes several CPS income sources including alimony, child support, income from trust funds and royalties and financial assistance from family or friends; however, HRS income exclusions have varied from year to year.

- RAND's total income variables for HRS include Food Stamps, although the RAND poverty status variable has been calculated based on income excluding Food Stamps.

Internal Consistency. Most of the surveys, like the CPS, do edits or consistency checks between important variables such as whether a person worked and has earned income, or whether a person receives income from a given source and has an income amount for that source. Most surveys also ensure that total money income for each person equals the sum of income by source for the person, and that total family income equals the sum of incomes for persons in the family. There are some exceptions.

- NHIS does not address consistency of family earnings (the sum of earnings for persons in the family) and family income to ensure that family income is at least as large as family earnings, even when earnings and income are imputed

- NHIS does not edit or impute work activity or earnings amounts against reported receipt of wage and salary income or self-employment income for the same time period

- MEPS does not edit or impute reported wage and salary or self-employment income for the year against work activity reported for the same time period

- MEPS does not edit or impute work activity against reported wage and salary or self-employment income for the same time period

- MEPS does not edit or impute type of earned income (wages and salaries vs. self-employment) against data on work activity

The MEPS practices reflect an explicit decision to preserve independent information collected in different sections of the survey instrument (and, to a large extent, at different times of the year), even when discrepancies exist.

Weighting. All surveys calculate weights in stages, starting with person,[4] family or household weights based on selection probabilities adjusted for non-response. The resulting first-stage weights are then post-stratified to Census Bureau, CPS, or Medicare enrollment-based control totals by age, race/ethnicity, and (except for PSID) sex of the person, family head, or family reference person. Some surveys use additional demographic information such as type of household, state or county of residence, or Census region and MSA/non-MSA status. Some surveys include a family equalization process to ensure that husbands and wives or partners have the same weights within overall control totals. After person or family weights are calculated, in most of the surveys the family or household weight is set equal to the person weight of the reference person or another family member. One survey uses additional and unique control totals.

- MEPS post-stratifies persons on the public use file to match CPS poverty rates by age, sex, race/ethnicity, Census region and MSA/non-MSA status for CPS-type families (as of December 31 of the year), thus ensuring that the MEPS public use file yields the same poverty rates by demographic groups as the CPS

- MEPS also post-stratifies families on the public use file to match CPS counts of families by family size and family type (couple, male head no spouse present, female head no spouse present)

---

[4] In MCBS these are individual enrollees.

PSID constructs family weights based on selection probabilities adjusted for non-response and attrition, post-stratifies the resulting family weights to CPS-based control totals,[5] and then sets person weights equal to the family weight.

- PSID control totals use CPS counts of primary families and primary individuals by family sizes 1, 2 or 3 or more; they exclude CPS unrelated subfamilies and secondary individuals, and make no adjustments for different treatment in PSID and CPS of unmarried partners and their relatives, students, active duty military, persons living in other countries, some related subfamilies and some institutionalized persons

Other Differences. It should be noted that, in addition to the differences listed above, there are a myriad of other differences among the surveys in almost every aspect in every domain described in the tables. These differences in details may also affect outcomes of policy analysis.

---

[5] The description applies to the preliminary cross-section weights available when this report was prepared and may not apply fully to the final weights.

TABLE II.1A. BACKGROUND AND OVERVIEW

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Purpose	Specifically designed to provide information on income, populations at risk, and needs and utilization of government programs, for public policy formulation Provides within-year patterns of income and program participation and a broad range of data relevant to health, retirement and income security programs Design and field methods intended to maximize accuracy of income and program participation data for lower income populations and those whose income varies within the year	Primary source of detailed information on income and work experience in the United States Source of official income and poverty estimates, and most widely used estimates of the uninsured Underlying monthly survey (CPS-1) is the source of official labor force, unemployment and wage rate estimates Widely used for policy analysis and legislative cost estimates, and as basis for major micro-simulation models such as TRIM
Design Summary	Longitudinal panel survey collecting 2½ to 4 years of detailed monthly income and other data through interviews at 4 month intervals (8 to 12 interviews) for all persons in initial sample households and persons added through household formation or change, with most recent panels started in 2001, 2004 and 2008	Annual cross-section household survey at a fixed point in time collecting detailed prior calendar year income and employment data and current demographic and labor force data
File Summary	Multiple public use files contain person-month data from full panel	Public use file contains household, family and person data
History	Based on the Income Survey Development Program, a joint effort of HEW/ASPE, HEW/SSA and Census from 1977 to 1981 First panel fielded in October, 1983 Has been redesigned, expanded and contracted as budget varied Presently being re-engineered to be implemented in 2011 or 2012	Income questions first asked in April, 1948 Supplement expanded and redesigned many times Non-cash benefits added in 1980; health insurance questions changed significantly in 1988 and revised in 1995 and 2001 Sample expanded in 2001 to improve State estimates of children in low-income families without health insurance
Responsibilities	Survey and questionnaire design: Census Bureau Field work conducted by Census Bureau All processing done at Census Bureau	Survey and questionnaire design: Census Bureau Field work conducted by Census Bureau All processing done at Census Bureau

TABLE II.1B. BACKGROUND AND OVERVIEW

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Purpose	Small area demographic, income and poverty data, with emphasis on demographics and local area data As of 2010 will replace the Decennial Census Long Form, the traditional source of small area demographic and income data	Detailed information on health conditions, use of medical services, cost and source of payments in the United States
Design Summary	Monthly cross-section household survey collecting prior 12 months income (rolling reference period) and current demographics continuously during the year (rolling sample) Income in internal files and published data has month-by-month inflation adjustments to calendar year price levels Unlike other surveys, participation is mandatory not voluntary	Longitudinal panel survey collecting 2 years of event-level health care services and cost information through interviews at 6 month intervals (5 interviews), and prior calendar year income information once per year, for all persons in initial sample households and persons added through household formation or change, with new panel starting every year
File Summary	Public use file contains person-level data but data collection month and reference period timing are suppressed, income is rounded, and income has no inflation adjustment File is geographically oriented based on Decennial Long Form files	Cross-section public use files contain person-wave data for a calendar year from two overlapping annual panels Public use file includes same-year income information collected in subsequent year
History	After pretests, data collection began in 36 counties in 1999 and National sample of 800,000 households in 2000 Full implementation with annual sample of 3 million households covering all counties and county-equivalents began in 2005 Expanded to include non-household population (institutionalized and group quarters) in 2006	Based on National Center for Health Services Research (NCHSR) 1977 National Medical Care Expenditure Surveys and 1987 National Medical Expenditure Survey First panel fielded in 1996
Responsibilities	Survey and questionnaire design: Census Bureau Field work conducted by Census Bureau All processing done at Census Bureau	Survey and questionnaire design: Agency for Healthcare Research and Quality (AHRQ) Field work conducted by Westat under contract with AHRQ Processing done at AHRQ, Westat, and some years Social and Scientific Systems (SSS)

TABLE II.1C. BACKGROUND AND OVERVIEW

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Purpose	Primary source of information on health conditions, access to care and use of medical services in the United States Widely used for policy analysis	Completion of cost and utilization data for persons ever enrolled in Medicare during each calendar year by linking administrative and Medicare A and B claims data to survey-reported events to add information on non-covered medical services Event-level information on diagnoses, services, providers, charges, payments, and sources of payment
Design Summary	Annual cross-section household survey collecting prior month and year health information, current demographics, and prior calendar year family income continuously during the year (rolling sample)	Longitudinal panel enrollee survey collecting 3 years of event-level utilization and other data through interviews at 4 month intervals over 4 years (12 interviews) to supplement Medicare claims data merged with administrative records, with new panel started annually and prior year total income questions in summer and fall
File Summary	Public use file contains household, family and person data Public use file income information limited to $5,000- and $10,000-wide brackets Files with actual income amounts may not be taken off-site and are available only if research plan approved and daily access fees paid	No public use files – data available if research plan approved and files purchased Multiple limited access files contain person, service and event data for enrollees ever-on during calendar year including information collected in subsequent years, merged, unduplicated and validated with administrative records and complete Medicare bill files
History	Initially fielded in 1957 Redesigned at 10-year intervals 1997 redesign removed person-level income amounts	First panel fielded in 1991 In 1993 redesigned to have 4-year limit on participation and annual new panels adding new enrollees, replacing participants expected to be lost or rotated out and maintaining age stratification
Responsibilities	Survey and questionnaire design: National Center for Health Statistics (NCHS) Field work conducted by Census Bureau under interagency agreement with NCHS Processing done at Census Bureau and at NCHS	Survey and questionnaire design: Center for Medicare and Medicaid Services (CMS) Field work conducted by Westat under contract with CMS Processing done by Westat and CMS

TABLE II.1D. BACKGROUND AND OVERVIEW

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Purpose	Longitudinal data on economic circumstances, health and social support of older population including retirement decisions, income, assets, health, family affiliations and support structure as persons age, including institutional care and final illness	Longitudinal data on complete family life cycles including family formation and dissolution, changes in employment, income, wealth, housing, fertility and use of transfers, and intergenerational transfer of behavior such as welfare use
Design Summary	Longitudinal cohort survey collecting demographic, financial and health information until death through interviews every two years with sampled persons age 51 or over (rises to 55 or over between cohort additions) and their current and former spouses or partners	Longitudinal cohort survey collecting demographic, behavioral and financial information until death through interviews every two years with initial sampled families and their descendants including those added through household formation or change
File Summary	Public use files available from two sources

HRS: Multiple public use files for each year (37 excluding decedent exit interviews and imputation files) on age-eligible sample, current and former spouses or partners, family, and helpers, from multiple respondents -- longitudinal tracking files allow links across years

RAND: One person-based flat file for each year after cross-section and longitudinal edits and clean-up, with a broad range of new variables having consistent names and content across years | Public use files contain family data with person information for head and current spouse or partner and limited data on other family members

Public use longitudinal files track family relationships over almost 40 years and contain cross-year variables such as gender |
| History | Cohort born 1931-1941 (age 51 to 61) initially fielded in 1992

Cohort born 1923 or earlier (age 70 or more) initially fielded in 1993

Cohorts combined and 2 cohorts added in 1998 to expand sample to persons born 1947 or earlier (age 51 or over) in 1998

New cohorts added every 6 years, starting in 2004 with persons born1948-1953 (age 51 to 56 at time of addition) | Initial cohort of sample families fielded in 1968 and interviewed annually through 1997 then every two years from 1999 forward

Redesigned and supplemented with a sample of recent immigrant families in 1997 |
| Responsibilities | Survey and questionnaire design: Institute for Social Research (ISR), University of Michigan, under a grant from the National Institute on Aging (NIA)

Field work conducted by Survey Research Center (SRC), ISR

All processing done at SRC and ISR

RAND work funded by Social Security Administration | Survey and questionnaire design: Institute for Social Research (ISR), University of Michigan, under multiple and varied sponsorship over the years

Field work conducted by Survey Research Center (SRC), ISR

All processing done at SRC and ISR |

TABLE II.2A. SURVEY AND SAMPLE DESIGN

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Sample Frame	Housing units including housing units on military bases in 50 States and DC Non-institutional group quarters in 50 States and DC Frame and sample selection: Census Bureau	Housing units including housing units on military bases in 50 States and DC Non-institutional group quarters in 50 States and DC excluding college dormitories Frame and sample selection: Census Bureau
Sample Design	Multi-stage sampling design 2001 panel designed for National but not State estimates although 45 States are coded – 2004 panel designed for State estimates Sample designated at beginning of each new panel	Multi-stage sampling design for underlying monthly survey Designed for State estimates with sub-State samples in NY and CA and estimates for MSAs over 500,000 Sample expanded in 2001 to improve State-specific estimates of children in low-income families without health insurance
Sample Unit	Household in housing unit or group quarter Includes persons who "usually reside" in unit e.g. most college students in dormitories usually reside with parents May contain multiple families and/or unrelated individuals	Household in housing unit or group quarter Includes persons who "usually reside" in unit e.g. most college students in dormitories usually reside with parents May contain multiple families and/or unrelated individuals
Oversamples	Low income areas	Hispanics, non-Whites, and Whites with children 18 or younger
Response Thresholds	Interview must obtain household roster, relationships, names, demographics, labor force and types of income Files contain actual or imputed data on all persons in responding households	Interview must obtain household roster, relationships, names, demographics and labor force data (CPS-1) File contains actual or imputed data on all persons in responding households
Initial Response Rate	2001 panel initial response: 87.7%	2003 Supplement month response to CPS-1: 92.2% Supplement data is imputed for CPS-1 responders who refuse to answer Supplement

TABLE II.2B. SURVEY AND SAMPLE DESIGN

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Sample Frame	Housing units including housing units on military bases in 50 States and DC Puerto Rico added in 2005 No group quarters through 2005 After 2005, all group quarters, including barracks and institutions Frame and sample selection: Census Bureau	Families or individuals with successful prior year NHIS interview NHIS frame consists of housing units including housing units on military bases and non-institutional group quarters in 50 States and DC Frame and sample selection: AHRQ and SSS for MEPS subsample of NHIS responders
Sample Design	Multi-stage sampling design at the county level Designed for State and sub-State estimates with estimates for areas of 65,000 or more annually and 3-year estimates for areas of 20,000 or more and 5-year estimates for geographic areas regardless of size including census tracts	Samples from families and unrelated individuals interviewed for NHIS the previous year, except those in college dormitories Designed for National but not State estimates
Sample Unit	Household in housing unit (or group quarter after 2005) Includes persons who currently (not "usually") reside in unit e.g. college students in student housing are single individuals Only one family is identified but may contain multiple unrelated individuals	Family or individual in housing unit or group quarter Includes persons who "usually reside" in unit College students in student housing always merged back into parental family
Oversamples	Small governmental units	Blacks, Hispanics, Asians, and families predicted to have income less than 200% of poverty
Response Threshold	Interview must obtain two data elements for each person e.g. name and age or age and sex File contains actual or imputed data on all persons in responding households	Interview must ask all survey questions of all family members Policy is for non-response to be assigned at family level but file contains partial families
Initial Response Rate	97.3% for 2002 (See "Contact Method" in Table II.4B on how response rate calculated)	Including non-response to prior-year NHIS used as sample frame: averages 70%

TABLE II.2C. SURVEY AND SAMPLE DESIGN

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Sample Frame	Housing units including housing units on military bases in 50 States and DC Non-institutional group quarters in 50 States and DC Frame and sample selection: Census Bureau	Medicare enrollees (Part A and/or Part B) as of January 1 of each year located in 50 States, DC, and Puerto Rico Housing units, non-institutional and institutional group quarters except prisons and facilities for the criminally insane Frame and sample selection: CMS and Westat
Sample Design	Multi-stage sampling design Designed for National but not State estimates 4 separate quarterly samples	Multi-stage sampling design Designed for National but not State estimates Sample stratified by age New 4-year panel started annually to add new enrollees, replace participants expected to be lost or rotated out, and designed to maintain age stratification
Sample Unit	Household in housing unit or group quarter Includes persons who "usually reside" in unit except students away at college College students in student housing are single individuals May contain multiple families and/or unrelated individuals	Individual Medicare enrollee
Oversamples	Blacks and Hispanics	Over-representation of enrollees under 45 and over 80 years old
Response Threshold	Interview must obtain household roster, relationships, demographics, and all health data and demographics through education for at least one respondent in a family File contains actual or imputed data on all persons in responding families	Data on utilization and charges must be present for 2/3 of the Medicare-enrolled days in the year, or be missing for less than 60 days, for sample person to be included in calendar year Cost and Use File File contains utilization and cost data for all responding persons
Initial Response Rate	89.2% for 2003	Roughly 80%

TABLE II.2D. SURVEY AND SAMPLE DESIGN

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Sample Frame	Housing units excluding housing units on military bases in 48 contiguous States and DC in 1992 for 1992 and 1993 cohorts, 1993 cohort also used 1993 Medicare enrollment file as frame 1992 frame (not updated) used for younger 1998 cohort 1998 Medicare enrollment file used for older 1998 cohort Housing units excluding housing units on military bases in 48 contiguous States and DC in 2004 Frame and sample selection: SRC	SRC component: Housing units in 48 contiguous States and DC in 1968 SEO component: low income families in 48 contiguous States and DC with successful interview in 1967 Survey of Economic Opportunity and who signed data release agreement Housing units (recent immigrant component) in 48 contiguous States and DC in 1997 for post-1968 immigrants with spouses not in the US in 1968 Frame and sample selection: SRC
Sample Design	Multi-stage sampling design Designed for National but not State estimates Cohort samples drawn in 1992 (born 1931-1941), 1993 (born 1923 or earlier), 1998 (born 1924-1930 and 1942-1947) and 2004 (born 1948-1953) 1993 and 1998 cohorts (not 2004) adjusted for representation in previously selected samples	SRC component: Multi-stage sampling design SEO component: "Low" income families with head under 60 years old in metropolitan PSUs where SRC could field interviewers and in a sample of non-metropolitan PSUs in the South (90 total PSUs from 357 in SEO) 1997 redesign removed all non-Black SEO component original families and descendants, reduced total SEO component by 2/3 and added recent immigrant families
Sample Unit	Age-eligible person and spouse or partner of any age	Families originally selected plus split-off families containing persons or descendants of persons in originally selected families
Oversamples	Blacks, Hispanics, and residents of Florida	Blacks, low income and urban families (in SEO component)
Response Threshold	Interview must update family roster and relationships, and obtain demographics and health data Files contain actual or imputed data on responding age-eligible persons, current spouse or partner, and former (interviewed) spouses or partners	Interview must update family roster and relationships, and obtain housing, food cost, labor force and income data File contains actual or imputed data on all responding families
Initial Response Rate	81.4% for 1992 cohort, 80.4% for 1993 cohort, 72.5% and 70.0% for 1998 cohorts and 75.6% for 2004 cohort	79% for SRC component and 71% for selected eligible names/addresses sent to SRC for the SEO component

TABLE II.3A. UNIVERSE DEFINITIONS, INCLUSIONS AND EXCLUSIONS

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Survey Universe and Area	Resident civilian noninstitutionalized population of the US plus military living with civilian family members on or off base 50 States and DC	Resident civilian noninstitutionalized population of the US plus military living with civilian family members on or off base 50 States and DC
Non-Institutional Group Quarters	Living arrangements other than housing units whose occupants are free to come and go Includes retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, convents and monasteries, shelters and group homes, and college dormitories if "usual residence" Military barracks are non-institutional but excluded as not civilian May include staff of non-institutional or institutional group quarters who do not live in housing units	Living arrangements other than housing units whose occupants are free to come and go Includes retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, convents and monasteries, shelters and group homes, but not college dormitories Military barracks are non-institutional but excluded as not civilian May include staff of non-institutional or institutional group quarters, who do not live in housing units
Institutions	Group quarters whose occupants are not free to come and go Excluded locations include prisons, nursing homes, juvenile detention facilities and residential mental hospitals Sample persons not interviewed in institutions but return to interview status if they rejoin family unit or establish their own households	Group quarters whose occupants are not free to come and go Excluded locations include prisons, nursing homes, juvenile detention facilities and residential mental hospitals
Students	College students normally included in parental family	College students included in parental family
Active Military	Person and income included if living with one or more civilian family members age 15 or over on or off base	Person and income included if living with one or more civilian family members age 15 or over on or off base
Institutionalized	Data on person and income while institutionalized not available	Person and prior calendar year income excluded
Decedents	Data on person and income available through wave prior to death if not institutionalized at death	Person and prior calendar year income excluded
Other Exclusions	Unrelated children under age 15 in group quarters such as shelters	Unrelated children under age 15 in group quarters such as shelters

TABLE II.3B. UNIVERSE DEFINITIONS, INCLUSIONS AND EXCLUSIONS

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Universe and Geographic Area	Resident household population of the US through 2004 with Puerto Rico added in 2005 Resident population of the US and Puerto Rico (same as Decennial) for 2006 and subsequent years	Resident civilian noninstitutionalized population of the US as of NHIS interview Persons institutionalized subsequent to NHIS interview included 50 States and DC
Non-Institutional Group Quarters	Living arrangements other than housing units No distinction between two types of group quarters All group quarters excluded through 2005 Includes dormitories, barracks, retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, convents and monasteries, shelters and group homes, prisons, nursing homes, juvenile detention facilities and residential mental hospitals	Living arrangements other than housing units whose occupants are free to come and go Includes dormitories, retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, convents and monasteries, shelters and group homes Military barracks are non-institutional but excluded as not civilian May include staff of non-institutional or institutional group quarters, who do not live in housing units
Institutions		Group quarters whose occupants are not free to come and go Excluded locations include prisons, nursing homes, juvenile detention facilities and residential mental hospitals
Students	Students in dormitories excluded through 2005 Students included where they currently reside -- if in student housing during interview month are single individuals	College students are interviewed in dormitories but included in parental family
Active Military	Military in barracks excluded through 2005	Military living with one or more civilian family members on or off base not given person weights but income included
Institutionalized	Person and prior 12-month income excluded through 2005	Person excluded but some data and calendar year income while institutionalized available
Decedents	Person and prior 12-month income excluded	Data on person and income available until death
Other Exclusions	None	Unrelated minors (usually under age 18) in households or group quarter if not foster children

TABLE II.3C. UNIVERSE DEFINITIONS, INCLUSIONS AND EXCLUSIONS

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Universe and Geographic Area	Resident civilian noninstitutionalized population of the US 50 States and DC	Current Medicare enrollees in the US and Puerto Rico regardless of living arrangements 50 States, DC, and Puerto Rico
Non-Institutional Group Quarters	Living arrangements other than housing units whose occupants are free to come and go Includes dormitories, retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, convents and monasteries, shelters and group homes Military barracks are non-institutional but excluded as not civilian May include staff of non-institutional or institutional group quarters, who do not live in housing units	No hard distinction between types of group quarters Community interviews conducted with persons in housing units and group quarters that are not skilled nursing homes or otherwise require a facility interview May include retirement homes, assisted living facilities, personal or residential care homes and other group living arrangements
Institutions	Group quarters whose occupants are not free to come and go Excluded locations include prisons, nursing homes, juvenile detention facilities and residential mental hospitals Non-institutional shelters for battered women also excluded	Facility interviews with staff proxies conducted with persons in nursing homes, and other residential facilities where sample persons cannot be directly contacted Excluded locations: prisons and facilities for the criminally insane Other residential facilities may include retirement homes, assisted living facilities, personal or residential care homes and other group living arrangements
Students	Students included where they currently reside -- if in student housing during interview month are single individuals	Current student status not ascertained
Active Military	Military living with one or more civilian family members on or off base not given person weights but income included	Current military status not ascertained
Institutionalized	Person and prior calendar year income excluded	Person and prior 12-month income included
Decedents	Person and prior calendar year income excluded	Data on person and income available until death
Other Exclusions	Unrelated minors (usually under age 18) in households or group quarter if not foster children	None

TABLE II.3D. UNIVERSE DEFINITIONS, INCLUSIONS AND EXCLUSIONS

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Universe and Geographic Area	Age-eligible resident household population of the US excluding households on military bases at time of selection, and spouses or partners regardless of age or whether institutionalized Sample frames restricted to 48 contiguous States and DC but sample persons followed and interviewed wherever they move, including other countries	Members of original 1968 sample and 1997 recent immigrant sample of resident household population, and their descendants Sample frames restricted to 48 contiguous States and DC but sample persons followed and interviewed wherever they move, including other countries
Non-Institutional Group Quarters	Living arrangements other than housing units No distinction between two types of group quarters At time of selection, sample persons including those ages 70 or over (1993 cohort) and 74 or over (older 1998 cohort) limited to residents of housing units and exclude residents of group quarters	Living arrangements other than housing units No distinction between two types of group quarters Excluded locations include retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, prisons, nursing homes, and residential mental hospitals
Institutions	Excluded locations include retirement homes, assisted living facilities, personal or residential care homes, rooming or boarding houses, prisons, nursing homes, and residential mental hospitals Sample persons followed and interviewed regardless of subsequent living arrangements or institutionalization	Sample persons in group quarters interviewed if no sample persons remain in family they left Sample persons moving to group quarters from family that still contains sample persons excluded, but return to interview status if they rejoin family or establish their own households
Students	Students living at school excluded from family	Treatment follows group quarter rules
Active Military	Included if sampled but not identified on public use files	Treatment follows group quarter rules
Institutionalized	Data on person and prior calendar year income available	Treatment follows group quarter rules
Decedents	Data on person until death does not include income If decedent was alive at 2002 interview, prior year (CY2001) income data is available from that interview, otherwise last income data is CY1999 RAND person-based files do not include close-out interviews containing data on size and disposition of estate	Data on person and income available until death
Other Exclusions	None	None

TABLE II.4A. TIMING AND FIELDWORK

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
New Samples	Panels started in 2001, 2004 and 2008 Panels have 4 rotation groups which start in 4 consecutive months	Uses the underlying monthly survey (CPS-1) sample New rotation group (1/8 of CPS-1 sample) starts each month each year
Duration in Sample	Persons in sample household in survey for duration of panel Persons joining sample household or in households joined by sample persons in survey until no longer living with sample persons	Sample address in sample and its occupants interviewed 4 consecutive months for CPS-1 and supplements, not in survey next 8 months ("resting"), interviewed next 4 consecutive months then retired or rotated out (4-8-4 rotation pattern) Occupants (same or different) at sample address interviewed
Interview Timing	Each rotation group interviewed in a separate month Fieldwork is continuous during the year Data collected at a fixed interval of 4 months – three times a year – with 4 one-month reference periods for core questions Topical modules range from annual to once per panel Total number of interviews varies panel to panel	Interview conducted in the week containing the 19th of the month ASEC data collected in February through April, with bulk of data collection in March – prior to 2002 collected only in March ASEC data collected once per year with prior calendar year reference period Half of sample addresses in same month of consecutive years are the same
Sample Size	28,000 interviewed households in 2001 panel after wave 2 reduction	78,300 interviewed households in 2003 (2002 income)
Contact Method	CAPI (computer assisted personal interviewing) and CATI (computer assisted telephone interviewing)	CAPI (computer assisted personal interviewing) and CATI (computer assisted telephone interviewing)
Respondents	Persons 15 years or over each respond for self	Householder (person who owns or rents housing unit) or a knowledgeable adult household member
Proxies	Proxy respondents for 39% of individual interviews	One person responds for household
Field Work	Census Bureau	Census Bureau

41

TABLE II.4B. TIMING AND FIELDWORK

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
New Samples	Independent sample (1/12 of annual sample) starts each month each year	New panel starts every year Drawn from prior year NHIS respondents for first 2 or 3 quarters
Duration in Sample	Sample household in survey once	Persons in sample family in survey for duration of panel Persons joining sample family or in families joined by sample persons in survey until no longer living with sample persons
Interview Timing	Questionnaires are mailed each month Fieldwork is continuous during the year One-time data collection with prior 12-month reference period ending in month prior to month in which survey is answered Mail responses accepted through a 3 month response interval	Interviews spaced at approximately 6 month intervals 2 panels in the field simultaneously Fieldwork is fairly continuous during the year 5 interviews per family with interview to interview reference periods that start 1/1 of first year and end 12/31 of second year
Sample Size	512,768 interviewed households in 2002 (2002 income) Full implementation sample is 3 million households (2005)	14,700 interviewed families in 2002 HC file (2002 income)
Contact Method	Mail out and mail back questionnaire with space for person-level responses on 5 people Followup by CATI (computer assisted telephone interviewing) if not returned after 1 month or is incomplete or more than 5 persons are listed (about 1/3 of sample completed by CATI) Followup by CAPI (computer assisted personal interviewing) after 2 months attempted for 1/3 of remaining non-responders, non-contacts and incompletes during 1 more month, with other 2/3 dropped from sample and from non-response computation	CAPI (computer assisted personal interviewing) of each family and of college students in dormitories possibly supplemented by CATI (computer assisted telephone interviewing)
Respondents	Householder (person who owns or rents housing unit) or knowledgeable adult household member	Person knowledgeable about health of family members
Proxies	One person responds for household	One person responds for family
Field Work	Census Bureau	Westat

TABLE II.4C. TIMING AND FIELDWORK

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
New Samples	Independent sample (1/4 of annual sample) starts every quarter every year	New panel starts every year
Duration in Sample	Sample household is in survey once	Sample enrollee is in survey for duration of panel, about 4 years
Interview Timing	Interviews assigned for each of 52 weeks in year and field staff have 17 days in which to complete each week's interviews Fieldwork is continuous during the year (rolling sample) One-time data collection with prior calendar year reference period Each family in household is separately interviewed	Interviews spaced at approximately 4 month intervals 4 panels in the field simultaneously Fieldwork is fairly continuous during the year 12 interviews per enrollee with interview to interview reference periods
Sample Size	Over 36,000 interviewed families in 2003 (2002 income)	Over 12,000 enrollees in 2003 Cost and Use file Over 16,000 enrollees in 2003 Access to Care file
Contact Method	CAPI (computer assisted personal interviewing)	CAPI (computer assisted personal interviewing) for community interviews CAPI (computer assisted personal interviewing) for facility interviews after 1996
Respondents	Person knowledgeable about health of family members Persons 17 years or over may respond for self if present	Enrollee or family member for community interviews Nursing staff or care-givers and business office staff for facility interviews
Proxies	One person responds for family	Proxy respondents for 19% of community interviews Proxy respondents for all facility interviews
Field Work	Census Bureau	Westat

TABLE II.4D. TIMING AND FIELDWORK

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
New Samples	New birth cohort (and spouses or partners of any age) added every 6 years	Core panel started in 1968 Recent immigrant sample added in 1997
Duration in Sample	Age-eligible sample person in survey for life Spouse or partner of age-eligible person at time of selection in survey for life regardless of age New spouses or partners of age-eligible persons, or of spouse or partner of age-eligible persons at time of selection, in survey until remarried or re-partnered, according to documentation	Sample persons and their descendants in survey for life Persons joining sample family or in families joined by sample persons or their descendants in survey until no longer living with sample persons or their descendants
Interview Timing	Data collected every two years with prior month or prior calendar year reference periods for most income sources, prior 2 years for Food Stamps and for changes in assets	Data collected every two years with prior calendar year reference period for most income sources Interviews were annual for the 30 years 1968 through 1997
Sample Size	13,650 households interviewed in 2004 (2003 income) including cohort added in 2004	Original sample 2,930 families in SRC component and 1,872 families in SEO component Recent immigrant sample (1997) 511 families 7,822 families interviewed in 2003 (2002 income)
Contact Method	CAPI (computer assisted personal interviewing) for most initial (baseline) interviews except CATI (computer assisted telephone interviewing) for 1993 cohort CAPI (computer assisted personal interviewing) and CATI (computer assisted telephone interviewing) for subsequent interviews	CATI (computer assisted telephone interviewing)
Respondents	Multiple respondents – can differ for each of coverscreen, family, financial, sibling, child, helper, and asset-transfer questionnaires	Family head or wife/"wife" (male head of family or legal spouse/unmarried partner)
Proxies	Most questionnaires addressed to proxies	Proxy respondents for 2% of family interviews
Field Work	SRC	SRC

TABLE II.5A-B. LONGITUDINAL INCLUSION AND FOLLOW RULES

	2001 Panel of Survey of Income and Program Participation	2002 Medical Expenditure Panel Survey Household Component
Sample Person Movers	Followed to new locations within 50 States and DC Move-outs create new household and at least one new reference person who need not be a sample person	Followed to new locations within 50 States and DC but may be restricted to sample PSUs Move-outs create new family and at least one new reference person who need not be a sample person File does not contain any non-sample reference persons
Permanent Additions	Children born to sample members	Children born to sample members
Temporary Additions	Move-ins including new spouses while they live with sample persons Interviews may not provide full calendar year income	Move-ins including new spouses or partners while they live with sample persons Move-ins with data on file have calendar year income
Institutionalized	Sample persons followed in and out of institutions but not interviewed and no data while institutionalized	Sample persons followed in and out of institutions but not interviewed and no data while institutionalized Includes those institutionalized after prior year NHIS interview
Active Military	Followed while living with one or more civilian family members age 15 or over on or off base	Followed while living with one or more civilian family members on or off base but not given person weights
Students	College students away at school but "usually residing" with parents included in parental family	College students are interviewed in dormitories but included in parental family
Decedents	No close-out interview Data as of last interview outside an institution	Proxy close-out interview Income information obtained for calendar year of death
Re-contact Efforts	Refusals always re-contacted each wave	No information available
Responding Households	Total decreases over life of panel	Total decreases over life of panel
Attrition	Interview to interview attrition (4 months) averages 6.5% after wave 2 sample reduction	MEPS sample first interviewed in NHIS and MEPS 1st interview equivalent to other surveys' 2nd interview Interview to interview attrition (6 months) averages 2% after 1st interview
2002 Response Rate	Average response for interviews covering 2002: 72.5%	2002 Full Year file: 64.7%

TABLE II.5C-D. LONGITUDINAL INCLUSION AND FOLLOW RULES

	2003 Medicare Current Beneficiary Survey	2004 Health and Retirement Study
Sample Person Movers	Followed to new locations in a sample PSU in the 50 States, DC, and Puerto Rico	Followed to new locations in or out of the US Move-out of spouse or partner creates new sample family
Permanent Additions	No additions	No additions
Temporary Additions	No additions	New spouses or partners of age-eligible person or of former spouse or partner of age-eligible person until remarried or re-partnered, according to documentation If person on file, prior calendar year income available
Institutionalized	Followed in and out of institutions and exact dates obtained Facility interview in institution and an additional community interview on release	Followed in and out of institutions and income data obtained Nursing home residents have weights from 2000 forward
Active Military	Current military status not ascertained	Included if sampled but not identified on public use files
Students	Current student status not ascertained	Non-sample students living at school during school year classified as institutionalized and excluded
Decedents	Proxy close-out interview Does not obtain income information for year of death	Proxy close-out interview includes information on settled estate Does not obtain income information for year of death RAND person-based files do not include close-out interviews
Re-contact Efforts	Attempts made to convert refusals	To date, refusals re-contacted unless request otherwise
Responding Households	Remains fairly constant unless sample size is changed New panels start annually to add new enrollees, replace participants expected to be lost or rotated out, and maintain sample size and age stratification	Total decreases over time until new cohort added
Attrition	Interview to interview attrition (4 months) averages 4%	Interview to interview attrition (2 years) averages 7%
2002 Response Rate	2003 Cost and Use file: 69.5%	Not applicable

TABLE II.5D. LONGITUDINAL INCLUSION AND FOLLOW RULES

	2003 Panel Study of Income Dynamics
Sample Person Movers	Followed to new locations in or out of the US
	Move-outs create new families with new reference persons who need not be sample persons – through 1992, move-outs under 18 not followed unless established own household or moved with adult sample person but followed from 1993 forward
	Followed in and out of group quarters such as rooming houses, boarding houses, shelters and group homes but not interviewed while in group quarters if family they left still contains sample members
	If no sample members remain in family, sample person in institution or group quarters interviewed as a single person family
Permanent Additions	Children by birth or adoption (if not a stepchild), grandchildren and any other descendants of sample members
Temporary Additions	Persons joining families for period of time they live with sample members or their descendants
	Income information may not cover full calendar
Institutionalized	Followed in and out of institutions but not interviewed while institutionalized unless no sample members remain in family they left
Active Military	Active military in barracks not interviewed while living in barracks if family they left still contains sample members
	Person and income included if usually resides with sample family, or is a sample person not living in barracks, or is a sample person living in barracks and no sample persons remain in family they left
Students	Students living at school not interviewed while living at school if family they left still contains sample members
	If no sample members left in family, student will be interviewed as a single person family at school
Decedents	Same treatment as move-outs
	Income information obtained for portion of reference year until death
Re-contact Efforts	No attempts to convert prior refusals until 1990 but major efforts in 1992, 1993 and 1994 restored many prior dropouts
	All prior wave non-responses now re-contacted each year
Responding Households	Total increases over time
Attrition	Interview to interview attrition (2 years) averages 3%
2002 Response Rate	Not applicable

TABLE II.6A. FAMILY DEFINITIONS

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Family Definition	Persons related by blood, marriage or adoption	Persons related by blood, marriage or adoption
Household Relationships	Relationship of every person to household reference person at each interview Person who owns or rents housing unit at each interview is household reference person and may be either husband or wife Relationship of every person in household to every other person in household identified in 2nd interview	Relationship of every person to household reference person Person who owns or rents housing unit is household reference person (householder) and may be either husband or wife
Family Relationships	Relationship of every person to family reference person at each interview	Relationship of every person to family reference person
Related Subfamilies	Married couple with or without never-married children under 18 and single parents with never-married children under 18 Persons and subfamily identified	Married couple with or without never-married children under 18 and single parents with never-married children under 18 Persons and subfamily identified
Unrelated Subfamilies	Married couple with or without never-married children under 18 and single parents with never-married children under 18 Persons and subfamily identified	Married couple with or without never-married children under 18 and single parents with never-married children under 18 Persons and subfamily identified
Marriage	Legal marriage Legal spouse identified if present for all persons Cohabiting (unrelated) partner of reference person identified	Legal marriage Legal spouse identified if present for all persons Cohabiting (unrelated) partner of reference person identified
Parents	Mother and father each identified if present as well as type of relationship (biological, step, adopted) for all persons	One parent identified if present for all persons through 2006 Mother and father each identified if present as well as type of relationship (biological, step, adopted) for all persons after 2006
Alternative Definitions	Any required family unit can be constructed	
Differences From CPS	No differences	

TABLE II.6B. FAMILY DEFINITIONS

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Family Definition	Persons related by blood, marriage or adoption	Persons related by blood, marriage or adoption, including foster relationships and unmarried (opposite or same sex) partners
Household Relationships	Relationship of every person to household reference person Person who owns or rents housing unit is household reference person (householder) and may be either husband or wife	Concept not used
Family Relationships	Only when family reference person is household reference person (householder)	Relationship of every person to family reference person at each interview
Related Subfamilies	Married couple with or without never-married children under 18 and single parents with never-married children under 18 Persons and subfamily identified	Concept not used Related subfamilies can be constructed
Unrelated Subfamilies	Concept not used and cannot be identified Members treated as unrelated individuals	Concept not used and interview is at family level
Marriage	Legal marriage Presence of legal spouse for all persons but spouse not identified except spouse of reference person Cohabiting (unrelated) partner of reference person identified	Legal marriage or self-identified cohabitation Legal spouse identified if present for all persons Cohabiting partner of reference person or of reference person's parents, and children of partner, identified in relationship codes
Parents	Presence of mother and/or father for all children but parent(s) not identified	Mother and father each identified if present for all persons
Alternative Definitions	Cannot be constructed	Identifies members of CPS-defined families as of December 31 but not relationships if CPS family has different reference person than MEPS family
Differences From CPS	No unrelated subfamilies, so that members of unrelated subfamilies all become unrelated individuals	Unmarried (opposite or same sex) partners, relatives of partner and foster relationships treated as blood or marital relationships

49

TABLE II.6C. FAMILY DEFINITIONS

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Family Definition	Persons related by blood, marriage or adoption, including foster relationships and unmarried (opposite or same sex) partners	Families not identified
Household Relationships	Relationship of every person to household reference person Person who owns or rents housing unit is household reference person (householder) and may be either husband or wife	Householder concept not used Number of persons in household and some information on relationships obtained in community interview
Family Relationships	Relationship of every person to family reference person	Number of persons related to the enrollee and some information on relationships obtained in community interview
Related Subfamilies	Concept not used Related subfamilies can be constructed	Not identified
Unrelated Subfamilies	Concept not used and interview is at family level	Not identified
Marriage	Legal marriage or self-identified cohabitation Presence of legal spouse or self-identified partner for all persons but spouse not identified Cohabiting partner of reference person and children of partner identified in relationship codes	Legal marriage Presence of legal spouse
Parents	Mother and father each identified if present as well as type of relationship (biological, step, adopted) for all persons	Not ascertained
Alternative Definitions	Income data not available if CPS family constructed	Family structure not ascertained
Differences From CPS	Unmarried (opposite or same sex) partners, relatives of partner and foster relationships treated as blood or marital relationships	Families not identified

TABLE II.6D. FAMILY DEFINITIONS

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Family Definition	Persons related by blood, marriage or adoption, including foster relationships and unmarried (opposite or same sex) partners	Persons related by blood, marriage or adoption, including foster relationships, unmarried (opposite sex) partners, and unrelated persons (may be same-sex partners) identified as part of family
Household Relationships	Terms household and family used interchangeably	Concept not used
Family Relationships	Relationship of every person in household/family to age-eligible person

Relationship of every person in household/family to spouse or partner of age-eligible person

Information obtained on siblings and children not in household | Relationship of every person in household/family to family head

Family head is always male spouse or partner when present

Legal spouse of head is wife, and unmarried partner of at least one year is "wife" |
| Related Subfamilies | Concept not used | Concept not used

Related subfamilies can be constructed |
| Unrelated Subfamilies | Concept not used and interview is at family level | Concept not used and interview is at family level |
| Marriage | Legal marriage or self-identified cohabitation

Legal spouse identified in relationship codes for each reference person's parents, siblings, children and grandchildren

Cohabiting partners identified in relationship codes for each reference person's parents, siblings, children and grandchildren | Legal marriage or self-identified (opposite sex) partner of a year or more

Cohabiting partner of reference person of any duration, and children, siblings or parents of partner identified in relationship codes |
| Parents | Mother and father each identified for age-eligible person and spouse or partner and for their children if present | Mother and father and type of relationship identified for most persons ever in survey since 1968 in separate Parent Identification and Childbirth and Adoption History Files |
| Alternative Definitions | Income data not available if CPS family constructed | Income data not available if CPS family constructed |
| Differences From CPS | Unmarried (opposite or same sex) partners treated as married, relatives of partner treated as relatives of householder, and foster relationships treated as blood relationships | Unmarried (opposite sex) partners, relatives of partner, foster relationships, and some unrelated persons (may be same-sex partners) treated as blood or marital relationships

Related subfamilies that previously split off from but rejoined primary family kept as separate families |

51

TABLE II.7A. WORK ACTIVITY AND EARNINGS

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Persons Covered	All persons age 15 or over	All persons age 15 or over
Reference Interval	Monthly and weekly periods in prior 4 months	Prior calendar year
Number of Jobs	Detail on up to 2 separate jobs and 2 businesses	Detail on longest job and longest business
Job or Business Information For Reference Interval	Start and end dates and weeks worked 3-digit industry and occupation codes for each job and major industry and 3-digit occupation codes for each business Class of worker, wage rate and usual hours per week Monthly earnings or monthly draw from business, and net profit	Weeks worked full time, part time and total Detailed and major industry and occupation codes Class of worker and usual hours per week Wages and salary from longest job, from other work, and total, and self-employment earnings
Unemployment	Weekly employment status (e.g. employed, unemployed, not in labor force)	Weeks unemployed (seeking work), weeks not in labor force
Job or Business Information For Other Intervals	Work history obtained in topical module Spells of unemployment or not in labor force can be constructed	Current employment status (e.g. employed, unemployed, not in labor force) and duration of current spell of unemployment last week Current job: 4-digit and major industry and occupation codes Current job: Class of worker, usual hours and earnings per week, hours last week -- wage rate asked for 1/4 of sample (4th and 8th interviews of each rotation group)
Industry/Occupation	3-digit (236-group) and major (14-group) industry codes based on 1987 North American Industrial Classification System (NAICS) used in 1990 Census 3-digit (501-group) occupation codes, based on 1990 Standard Occupational Classification (SOC) Manual used in 1990 Census	4-digit (270-group), detailed (52-group) and major (14-group) industry codes based on 2002 NAICS 4-digit (509-group), detailed (23-group) and major (11-group) occupation codes based on 2000 SOC Manual
Consistency Edits	Hours, pay rate and earnings consistency checked in interview Monthly earnings and work activity not cross-edited	Hours, pay rate and earnings consistency checked in interview Earnings always imputed if work activity

TABLE II.7B. WORK ACTIVITY AND EARNINGS

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Persons Covered	All persons age 15 or over	All persons age 16 or over
Reference Interval	Prior 12 months	Interview to interview, approximately six months
Number of Jobs	Not included	Detail on current main job at each interview
Job or Business Information For Reference Interval	Weeks worked Usual hours per week Wages and salary, and self-employment earnings	Start and end dates Major industry codes and collapsed major occupation codes Usual hours per week and wage rate Self-employed in current main job
Unemployment	Weeks unemployed or not in labor force (not differentiated)	Unemployed or not in labor force (not differentiated) at interview
Job or Business Information For Other Intervals	Current employment status (e.g. employed, unemployed, not in labor force) last week Whether worked in last 5 years Current main job or most recent job: 3-digit and major industry and occupation codes Current main job or most recent job: Class of worker	Detail on all jobs held at each interview and/or during interview reference period in separate unedited research file, JOBS
Industry/Occupation	3-digit (265-group) industry codes based on 1997 NAICS, used in 2000 Census 3-digit (509-group) occupation codes based on 2000 SOC Manual, used in 2000 Census	Major (14-group) industry codes based on 2002 NAICS Collapsed major (9 not 11-group) occupation codes based on 2000 SOC Manual
Consistency Edits	Pre-edits to correct issues related to incorrectly filled-out forms Work activity edited or imputed if earnings present Earnings edited or imputed if work activity reported	Neither edits of work activity nor imputation based on earnings Earnings imputed based on work activity for non-response but negative values not edited based on work activity Type of earnings (wages and salaries vs. self-employment) not edited based on type of work activity

TABLE II.7C. WORK ACTIVITY AND EARNINGS

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Persons Covered	All persons age 18 or over	All enrollees age 16 or over living in community
Reference Interval	Prior calendar year	At interview or "current"
Number of Jobs	Not included	Not included
Job or Business Information For Reference Interval	Whether worked and number of months worked Total amount (internal file) or bracket (public use file) of earnings – wages and salaries not differentiated from self-employment Brackets on public use file are $5,000 wide below $25,000 and $10,000 wide from $25,000 to $75,000 Recipiency questions on wages and salary and self-employment income	Whether working at a job or business
Unemployment	Whether unemployed or not in labor force (not differentiated) for all of prior calendar year	Not included
Job or Business Information For Other Intervals	Current employment status (e.g. employed, unemployed, not in labor force) last week Current job: Usual hours per week or hours last week	Not included
Industry/Occupation	Not included	Not included
Consistency Edits	Recipiency of wages and salary or self-employment income not used to edit or impute work activity or earnings Earnings imputed if work activity reported or imputed File contains persons reporting receipt of wages and salary or self-employment income without work activity or earnings	Earned income not included

54

TABLE II.7D. WORK ACTIVITY AND EARNINGS

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Persons Covered	Age-eligible sample person and spouse or partner but not other family members	Head and wife/"wife" but not other family members
Reference Interval	Prior calendar year	Prior calendar year
Number of Jobs	No detail on jobs or businesses	Detail on up to 4 separate jobs or businesses
Job or Business Information For Reference Interval	Start and end dates and months worked Wages and salary, and self-employment earnings	Start and end dates and weeks worked 3-digit industry and occupation codes Class of worker, wage rate and usual hours per week
Unemployment	Months unemployed or not in labor force (not differentiated)	Weeks unemployed and weeks not in labor force
Job or Business Information For Other Intervals	Current employment status (e.g. employed, unemployed, not in labor force) Current job: Wage rate and usual hours per week	Current employment status (e.g. employed, unemployed, not in labor force) Current main job: 3-digit industry and occupation codes Current main job: Class of worker and wage rate
Industry/Occupation	Not on HRS or RAND public use files Suppressed codes described as major industry and occupation codes (13- and 18-group) used in 1980 Census	3-digit (265-group) industry codes based on 1997 NAICS, used in 2000 Census 3-digit (509-group) occupation codes based on 2000 SOC Manual, used in 2000 Census
Consistency Edits	No information available	Hours, pay rate and earnings consistency checked in interview Work activity and earnings consistency checked in interview Work activity, pay rates, hours and earnings manually edited for consistency

TABLE II.8A. PRE-TAX MONEY INCOME

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Reference Period	Monthly and 4-month periods in prior 4 months Monthly and 4-month amounts	Prior calendar year Allows choice of up to 6 reporting intervals from which annual amounts are calculated
Recall Length	Average 3 months maximum 5 months	Average 14 ½ months maximum 15 ½ months
Definitions Differ From CPS	Self-employment is monthly draw plus net profit (cash basis) Net profit not asked for sole proprietors or most partnerships not taking a monthly draw	
Non-CPS Source Included	Lump-sum or non-periodic payments such as IRA withdrawals	
CPS Source Excluded	Educational benefits	
Amount Detail For Persons	Up to 60 sources and amounts of income	Over 50 sources and up to 24 amounts of income
Screeners and Brackets	Multiple screeners and skip patterns No brackets	Identifies which of multiple possible sources for an income type (e.g. survivors benefits) to screen into amount questions No brackets
Persons Covered	All persons for Social Security, SSI and TANF Persons 15 or over for all other income sources	All persons for Social Security, SSI and TANF Persons 15 or over for all other income sources
Income Reassigned	For persons under 15, Social Security, SSI and TANF assigned to a representative payee or guardian	For persons under 15, Social Security, SSI and TANF assigned to a representative payee or guardian
Person and Family Totals	Sum of detail for persons and family	Sum of detail for persons and family

TABLE II.8B. PRE-TAX MONEY INCOME

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Reference Period	12 months prior to interview (rolling reference period) with average of July prior year to June current year Income on internal files (for publications and on-line tables) but not on public use file inflated to calendar year price levels by ratio of average annual CPI to average CPI over reference period – CPI-U through 2005 and 2006 forward by CPI-U-RS	Calendar year Income information collected in subsequent year is added to same-year file Annual amount for 12 income sources, monthly amount and months received for 4 sources
Recall Length	12 ½ months	Average 15 months maximum 18 months
Definitions Differ From CPS	Depends on respondent interpretation of summary descriptions of income sources on mail questionnaire Reference period -- income cannot be adjusted for differences in productivity, unemployment or other factors from the CPS' calendar year reference period	Internal Revenue Service definitions used for tax filers[1] Wages omit "above the line" items that are not subject to income taxes such as 401(k) contributions Self-employment earnings other than sole proprietorships and farm reported with rents, royalties, estates and trusts
Non-CPS Source Included	None	Lump-sum payments from retirement accounts
CPS Source Excluded	None	Tax exempt interest for tax filers
Amount Detail for Persons	Up to 8 sources and amounts of income	Up to 16 sources and amounts of income Taxable income sources not person-level for joint return filers -- primary filer allocates amounts between self and spouse
Screeners and Brackets	No screeners or skips except age No brackets	Type of tax form used as screener for specific income sources – 1040A short form skips self-employment and other items "Don't knows" (DKs) offered 10 annual brackets (to $100,000 or more) for most sources, 5 monthly brackets for 4 sources
Persons Covered	Persons age 15 or over	Persons 16 or over, and persons under 16 who report filing a tax return, for taxable income source[2] All persons for all other income sources
Income Reassigned	No reassignment – income not asked if person under 15	No reassignment
Person and Family Totals	Sum of detail for persons and family	Sum of detail for persons and family

TABLE II.8C. PRE-TAX MONEY INCOME

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Reference Period	Prior calendar year Annual amounts	Summer: Prior calendar year amount (on Cost and Use file) Fall: Prior 12 months bracket (on Access to Care file) Allows choice of monthly reporting interval from which annual amount or bracket is calculated
Recall Length	Average 18 months maximum 23 months	18 months (summer) or 12 ½ months (fall)
Definitions Differ From CPS	Earnings amount includes net income from rental property and unemployment or worker's compensation Recipiency data groupings conform to CPS definitions[3]	Depends on respondent interpretation of "total income before taxes"
Non-CPS Source Included	None	None
CPS Source Excluded	None	None
Amount Detail For Persons	One amount (internal file) or bracket (public use file) for total earnings from all sources (See Table II.14 on Ease of Access) Recipiency but no amounts for up to 10 other sources – file has no imputations of recipiency	One amount or bracket for total income of enrollee or enrollee and spouse from all sources (See Table II.14 on Ease of Access) 13 recipiency items asked in summer but not on file
Screeners and Brackets	No screener, skip or brackets on earnings No screener on total family income -- DKs and refusals asked "over or under $20,000" then offered 24 or 20 brackets	No screeners or skips Summer: DKs and refusals asked simple unfolding brackets (2 steps) with entry and steps for enrollees with spouse present twice as large as for single enrollees Fall: "Over or under $25,000" then offered 6 or 5 brackets
Persons Covered	Persons 18 or over for earnings	All enrollees
Income Reassigned	No reassignment since no person-level income	No reassignment
Person and Family Totals	One amount (internal file) or bracket (public use file) for total income of NHIS-type family (See Table II.14 on Ease of Access) File contains families with total income less than total earnings	One amount or bracket for enrollee or enrollee and spouse

TABLE II.8D. PRE-TAX MONEY INCOME

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Reference Period	Prior calendar year Annual amounts for non-retirement income, amount last month for retirement income and SSI	Prior calendar year Allows choice of up to 6 reporting intervals from which annual amounts are calculated
Recall Length	Maximum probably 24 months	Average 18 months maximum 24 months
Definitions Differ From CPS	Rental income is gross rent before deducting expenses RAND income totals and groupings do not conform to CPS[4]	None VA benefit variable includes military retirement
Non-CPS Source Included	HRS: No total income re-code but appears to include IRA withdrawals and lump sums such as inheritances RAND: Lump sums such as inheritances, and Food Stamps	None
CPS Source Excluded	Alimony, child support, income from trust funds and royalties, and financial assistance from family or friends[5]	None
Amount Detail For Persons	18 non-asset and 8 joint asset sources and amounts for age-eligible person and spouse or partner Non-self-employment earnings for other family members but no other sources or person-level amounts -- one catchall total	Up to 31 sources and amounts of income excluding Social Security for head and wife/"wife" and which months received No person-level sources or amounts for other family members and 2002 Social Security only a family total
Screeners and Brackets	Multiple screeners and skip patterns DKs and refusals asked unfolding brackets with item-specific entry points and steps – some items have randomly selected entry points	Multiple screeners and skip patterns No brackets for income – unfolding brackets used for some asset values and expenses (e.g. medical)
Persons Covered	Age-eligible person and spouse or partner	Head and wife/"wife"
Income Reassigned	No reassignment	No reassignment
Person and Family Totals	For age-eligible person or spouse or partner separately, sum of detail excluding asset income or self-employment, and as couple, sum of detail For family, sum of detail for age-eligible person and spouse or partner plus total of other family member income	Sum of detail for head and wife/"wife" excluding Social Security – summary recodes combine income of head and wife/"wife" For family, sum of detail for head and wife/"wife" plus total family Social Security plus total other family member taxable[6] income and transfer[7] income excluding Social Security

TABLE II.9A. INCOME ALLOCATION AND TOP-CODING ON PUBLIC USE FILES

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Allocations and Edits	Consistency, out-of-range, and logical edits built into CAPI and CATI and then repeated in processing data file	Consistency, out-of-range, and logical edits built into CAPI and CATI and then repeated in processing data file
Imputations	Statistical match (hot deck) and logical imputations, plus use of prior wave data	Statistical match (hot deck) and logical imputations, including Supplement refusals
Rounding	Income data not rounded	Income data not rounded
Earnings Top-Codes	Monthly earnings amounts from a source (main job, other wages and salary, or self-employment including farm) top-coded at $12,500 (equivalent to $150,000 annually) if 4-month sum from that source exceeds $50,000	Wages and salary from longest job top-coded at $200,000 Other wage and salary, self-employment and farm earnings top-coded at $35,000, $50,000 and $25,000 respectively
Value When Top-Coded (Earnings)	Average across top-coded records in each of 12 demographic cells for each earnings source (some cells empty or collapsed) Tabulations on public use file will add to published totals	Average across top-coded records in each of 12 demographic cells for each earnings source (some cells empty or collapsed) Tabulations on public use file will add to published totals
Other Income Top-Codes	No top-codes for Social Security, SSI, TANF, unemployment benefits, Workers Compensation and Veterans payments Larger of 97th percentile of dollar values or 99.5th percentile of persons 15 or over (whether or not have income)	No top-codes for Social Security, SSI, TANF, unemployment benefits, Workers Compensation and Veterans payments Larger of 97th percentile of dollar values or 99.5th percentile of persons 15 or over (whether or not have income)
Value When Top-Coded (Other Income)	Top-code (no average across top-coded records) Tabulations on public use file may not add to published totals	Average across top-coded records Tabulations on public use file will add to published totals
Person and Family Totals	Not separately top-coded	Not separately top-coded
Other Suppressions	Age top-coded at 88 Geographic and event timing (e.g. month of birth) suppression based on disclosure analysis	Age top-coded at 80 Geographic and some suppression of detailed race, occupation and country of birth based on disclosure analysis

TABLE II.9B. INCOME ALLOCATION AND TOP-CODING ON PUBLIC USE FILES

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Allocations and Edits	Consistency, out-of-range, and logical edits	Some out-of-range, and logical edits built into CAPI Brackets converted to amounts through hot decks Most wage and salary allocations based on JOBS file but job-specific data in JOBS not used for edits
Imputations	Statistical match (hot deck) nearest neighbor imputations	Statistical match (hot deck) and logical imputations
Rounding	$10 to $1,000 to nearest $10, $1,000 to $50,000 to nearest $100, and above $50,000 to nearest $1,000, after top-coding	Income data not rounded
Earnings Top-Codes	2002 wages and salary and self-employment each top-coded at 99.5th percentile nationally – $200,000 and $78,751 – with State-specific top-codes for 2003 and subsequent years	Wages and salary and self-employment each top-coded at 99th percentile – amount not documented
Value When Top-Coded (Earnings)	State-specific average across top-coded records (uninflated) Tabulations on public use file will not add to published totals	Top-codes replaced with a "smeared" or randomized value Tabulations on public use file will add to published totals
Other Income Top-Codes	99.5th percentile nationally for 2002 and State-specific top-codes after 2002 for all other income sources through 2005 No top-codes for Social Security, SSI and TANF after 2005	99th percentile nationally
Value When Top-Coded (Other Income)	State-specific average across top-coded records (uninflated) Tabulations on public use file will not add to published totals	Top-codes replaced with a "smeared" or randomized value Tabulations on public use file will add to published totals
Person and Family Totals	Not separately top-coded No tabulations on public use file can match published totals because public use file lacks inflation adjustments	Person totals separately top-coded at 99th percentile and replaced with a "smeared" value Tabulations on public use file will add to published totals
Other Suppressions	Interview month and inflation adjustment amount suppressed Age top-coded at 90 and replaced by State-specific average across top-coded records Geographic re-codes to aggregate data into micro-areas with populations of approximately 100,000 or more	Random replacements or exchanges of income values Geographic and possibly other suppression based on disclosure analysis

TABLE II.9C. INCOME ALLOCATION AND TOP-CODING ON PUBLIC USE FILES

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Allocations and Edits	Earnings imputed only if work activity reported or imputed	Amounts from previous year or from Fall brackets
Imputations	Multiple sequential regressions for total family earnings and income -- imputed family earnings divided among persons but not constrained to be less than or equal to family income NCHS creates 5 files of imputations and recommends that analyses be performed 5 times and results averaged	Summer: Statistical match (hot deck) based on regression analysis (on Cost and Use file) Fall: Pro-rated among brackets (on Access to Care file) No imputation flag for income on 2003 Cost and Use file
Rounding	No specific rounding of income data	No specific rounding of income data
Earnings Top-Codes	Total earnings top-coded at $75,000 for public use file brackets Total earnings top-coded at $999,995 for internal file amounts Brackets on public use file are $5,000 wide below $25,000 and $10,000 wide from $25,000 to $75,000	
Value When Top-Coded	Top-code (no average across top-coded records)	
Other Income Top-Codes		
Value When Top-Coded		
Person and Family Totals	No person totals Family income top-coded at $75,000 for brackets on public use file with no average across top-coded records – 28% of persons on public use file are in top-coded families Family income top-coded at $999,995 for internal file amounts Brackets on public use file are $5,000 wide below $25,000 and $10,000 wide from $25,000 to $75,000	No top-codes
Other Suppressions	Imputed covariates for total family earnings and income including numerous health, insurance coverage, socio-demographic and income recipiency variables suppressed and may not be on internal file Major geographic suppressions	Some geographic suppressions

TABLE II.9D. INCOME ALLOCATION AND TOP-CODING ON PUBLIC USE FILES

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Allocations and Edits	HRS: Information not available RAND performs major consistency, out-of-range, and logical cross-section edits and multiple longitudinal consistency edits	Consistency, out-of-range, and logical cross-section edits
Rounding	Income data not rounded	Income data not rounded
Imputations	HRS: Statistical match (hot deck) including conversions of brackets to amounts – bracket data provided as well so users can perform their own imputations RAND: Statistical match (hot deck) based on case-specific regression models	Statistical match (hot deck) and logical imputations, plus use of prior wave data
Earnings Top-Codes	No apparent top-codes Values of $2,000,000 are on file and earlier waves have values over $3,500,000	Maximum values of $9,999,999 are de facto top-codes
Value When Top-Coded (Earnings)	Probably maximum value	Maximum value Tabulations on public use file will add to published totals
Other Income Top-Codes (Other Income)	No apparent top-codes File contains pension income values over $2,700,000 and asset income values over $3,500,000 -- earlier waves have asset income values over $7,000,000	Maximum values, usually $9,999,999, are de facto top-code Maximum value for e.g. total family Social Security and total non-taxable income of other family members, is $999,999
Value When Top-Coded	Probably maximum value	Maximum value Tabulations on public use file will add to published totals
Person and Family Totals	No apparent top-codes	Maximum values of $9,999,999 are de facto top-codes
Other Suppressions	Major geographic suppressions Industry, occupation and month and day of birth suppressed to preserve confidentiality	Major geographic suppressions to preserve confidentiality

TABLE II.10A. POVERTY STATUS

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Income Period Covered	Constructed intervals that can be monthly or part-year or calendar year for analysis of spells of poverty	Prior calendar year
Family Definition Used	CPS family but other definitions can be constructed	CPS family
Coverage	Persons in families with related subfamilies folded in Persons in unrelated subfamilies Unrelated individuals	Persons in families with related subfamilies folded in Persons in unrelated subfamilies Unrelated individuals
Exclusions From Survey Universe For Poverty Status	Unrelated children under 15	Unrelated children under 15 Family counts exclude subfamilies (related and unrelated) but all persons in subfamilies included
Difference From CPS	None	
Family Unit Timing	Can choose any month for construction of family units and consequent family composition and income amounts	Interview month (usually March) an average of 3 months after income reference year
Difference From CPS	Family composition and thus income need not match CPS	
Income for Computation	Total income of CPS family	Total income of CPS family
Difference From CPS	Minor income differences noted in Table II.8: educational benefits, lump-sum or non-periodic payments such as IRA withdrawals, measurement of self-employment income	
Poverty Status On File	Monthly poverty thresholds – 1/12 of annual threshold inflated by CPI-U to that month -- for each family and subfamily for composition as of that month, but ratio not calculated	Calendar year poverty status of CPS family as of the following March (usually) based on CPS income for year
Calculation of Alternatives	User can construct monthly, part-year or annual poverty measures, and can use alternate family compositions and/or timing of family composition	Official definition and official poverty statistics

TABLE II.10B. POVERTY STATUS

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Income Period Covered	12 months prior to interview (rolling reference period)	Calendar year of data file (data collected the following year)
Family Definition Used	CPS-type family but unrelated subfamilies not identified	NHIS-type family for most data including relationship codes CPS-type family for poverty status
Coverage	Persons in families with related subfamilies folded in Unrelated individuals (includes all persons in unrelated subfamilies)	Persons in families with related subfamilies folded in Unrelated individuals
Exclusions From Survey Universe For Poverty Status	Unrelated children under 15 Persons in group quarters through 2005 After 2005 institutionalized, military in barracks and college students in dormitories	Unrelated minors (usually under age 18) if not foster children Active military living with civilian family members (but included to determine poverty status of civilian family members)
Difference From CPS	Students away at school and persons in unrelated subfamilies are unrelated individuals, with any children under 15 excluded Civilians in non-institutional group quarters excluded until 2006	Military and unrelated minors age 15 or over excluded
Family Unit Timing	Interview month of rolling sample	December 31 of income reference year
Difference From CPS	Family composition does not lag income reference period	Family composition does not lag income reference period
Income for Computation	Total income of CPS-type family for prior 12 months	Pre-tax money income of CPS-type family for calendar year
Difference From CPS	Definitional difference in timing of income but not in income Poverty status of CPS-type families for the prior 12 months measured during each sample month (rolling reference period) and effectively averaged for the year (rolling sample) Differs from calendar year poverty status	Income differences noted in Table II.8 due to use of Internal Revenue Service definitions: No "above the line" earnings, or tax-exempt interest and includes taxable IRA withdrawals
Poverty Status On File	Ratio of unadjusted rounded income to adjusted thresholds (inflated by CPI-U to 12-month reference period price levels)	Calendar year poverty status of CPS-type family as of December 31 based on pre-tax money income for year[8]
Calculation of Alternatives	No replication or alternatives possible due to suppression of sample month and rounding	Can calculate status of NHIS-type family

TABLE II.10C. POVERTY STATUS

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Income Period Covered	Prior calendar year	Prior 12 months
Family Definition Used	NHIS-type family	Families not identified
Coverage	Persons in families with related subfamilies folded in Unrelated individuals	Families not identified
Exclusions From Survey Universe For Poverty Status	Unrelated minors (usually under age 18) if not foster children Active military living with civilian family members (but included to determine poverty status of civilian family members)	Poverty status not calculated
Difference From CPS	Military and unrelated minors age 15 or over excluded Unmarried (opposite or same sex) partners, relatives of partner and foster relationships treated as part of family Students away at school are unrelated individuals	Poverty status not calculated
Family Unit Timing	1 to 12 months after income reference year (rolling sample)	Families not identified
Difference From CPS	Family composition lags income reference period by 1 to 12 months compared to an average of 3 months for CPS	Families not identified
Income for Computation	Total income of NHIS-type family for prior calendar year	Family income not ascertained
Difference From CPS	No definitional differences in income	Family income not ascertained
Poverty Status On File	Ratio or bracket for calendar year poverty status of NHIS-type family as of interview month based on pre-tax money income for year (See Table II.14 on Ease of Access) Brackets on public use file are 25% wide below 200% of poverty and 50% wide from 200 to 500% of poverty	Poverty status not calculated
Calculation of Alternatives	None can be calculated since only income amount is total income for NHIS family Replication or validation possible only via on-site tabulations (See Table II.14 on Ease of Access)	Cannot be calculated since family not identified and family income not ascertained

TABLE II.10D. POVERTY STATUS

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Income Period Covered	Prior calendar year	Prior calendar year
Family Definition Used	RAND: NHIS-type family including related persons other than age-eligible person and spouse or partner	NHIS-type family except same-sex partners not identified, and unrelated persons (may be same-sex partners) identified as part of family included
Coverage	Persons in families with related subfamilies folded in Unrelated individual	Persons in families with related subfamilies folded in unless the related subfamily had split off but rejoined primary family Unrelated individual
Exclusions From Survey Universe For Poverty Status	Students away at school	Institutionalized, military and students away at school unless no sample persons remain in family they left
Difference From CPS	Unmarried (opposite or same sex) partners, relatives of partner and foster relationships treated as part of family but students away at school excluded from family	Unmarried (opposite sex) partners, relatives of partner, foster relationships, and some unrelated persons treated as part of family but related subfamilies that had split off but rejoined primary family remain separate families
Family Unit Timing	From 1 to 12 months after end of income reference year	Average composition during the income reference year
Difference From CPS	Family composition lags income reference period by 1 to 12 months compared to an average of 3 months for CPS	Family composition is contemporaneous with monthly reference periods
Income for Computation	RAND: Total income of age-eligible person and spouse or partner (re-code) with Food Stamps excluded plus earnings and other income of other household members	Total income of PSID family for prior calendar year with part-year not full year income for part-year family members
Difference From CPS	Income differences noted in Table II.8: Includes lump sums such as inheritances, some sources such as child support excluded, and rental income is gross of expenses	No definitional differences in income but part-year family members have only part-year income, not full year
Poverty Status On File	RAND: Calendar year poverty status of NHIS-type family as of interview month based on pre-tax money income for year	Calendar year threshold for PSID family on file For families with changes in composition during the year (with part-year family members) threshold is weighted average of thresholds for the various family compositions during the year -- consistent with part-year treatment of income
Calculation of Alternatives	Cannot be calculated since no person-level income for family members other than age-eligible person and spouse or partner	Cannot be calculated since no person-level income for family members other than head and wife/"wife" and part-year family members have only part-year income

TABLE II.11A. NON-CASH BENEFITS AND HEALTH INSURANCE

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Food Stamps	Monthly recipiency and amount for each person Start date and benefit history -- first interview	Number of persons and months received in prior calendar year Total amount in prior calendar year
Other Nutrition	Free or reduced-price School Lunch or Breakfast: number of children receiving and which program in prior 4 months Monthly WIC recipiency and amount for each person	Free or reduced-price School Lunches: number of children receiving in prior calendar year WIC in prior calendar year for each person
Housing And Energy	Current public housing, or other housing assistance if renting Energy Assistance recipiency and amount in prior 4 months	Current public housing, or other housing assistance if renting Energy Assistance after October 1 of last year and amount
Welfare To Work	Multiple types non-cash welfare assistance (e.g. education, child care, job search, job training) in prior 4 months for each person	7 types non-cash welfare assistance (e.g. education, child care, job search, job training) in prior calendar year for each person
Insurance Information	Coverage of each type in each month for each person Policyholder and coverage unit for up to 4 plans Insurance coverage is contemporaneous with income	Coverage of each type in prior calendar year for each person Insurance coverage is contemporaneous with income year
Medicaid	Starting month and year of coverage -- first interview	Number of months in prior calendar year (each person)
SCHIP	Children under 20	Children under 19 without Medicaid
Medicare and Other Public	Medicare, TRICARE/CHAMPUS, CHAMPVA, VA or military health care, other public	Medicare, CHAMPUS, CHAMPVA, VA or military health care, Indian Health Service, or other government
Work-Related	Policyholder, source of coverage and if part of premium paid	Policyholder, source of coverage and if part of premium paid
Coverage Outside Household	Identifies persons with coverage from outside household and age and relationship of anyone covered outside household	Identifies persons with coverage from outside household and whether anyone outside household is covered
Other Private	Private or direct purchase	Private or direct purchase
Periods of Uninsurance	Spells of uninsurance can be constructed	Uninsured are those never covered in prior calendar year

TABLE II.11B. NON-CASH BENEFITS AND HEALTH INSURANCE

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Food Stamps	Received in prior 12 months by anyone in household Total amount in prior 12 months	Number of months household received in prior calendar year Monthly amount paid (purchase requirement) and monthly value
Other Nutrition	Free or reduced-price School Lunch or Breakfast received in prior 12 months by anyone in household	Not included
Housing And Energy	Current public housing, Section 8 or other housing assistance Energy Assistance in prior 12 months	Not included
Welfare To Work	Not included	Not included
Insurance Information	Not included	Month-by-month coverage for each person from over 10 sources Whether HMO or gatekeeper for many sources and other plan and managed care attributes for private coverage Monthly family cost for private plans Insurance coverage is contemporaneous with income year
Medicaid	Not included	No distinction between Medicaid and SCHIP
SCHIP	Not included	No distinction between Medicaid and SCHIP
Medicare and Other Public	Not included	Medicare, TRICARE, 2 other public sources, and other State programs
Work-Related	Not included	Policyholder, source of coverage and if part of premium paid
Coverage Outside Household	Not included	Identifies persons with coverage from outside household
Other Private	Not included	Other group, non-group, or source unknown – policyholder
Periods of Uninsurance	Not included	Whether uninsured in prior 2 calendar years for all persons and when last insured for uninsured

TABLE II.11C. NON-CASH BENEFITS AND HEALTH INSURANCE

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Food Stamps	Number of months in prior calendar year for each person	Not included
Other Nutrition	WIC in prior calendar year for each person	Not included
Housing And Energy	Current housing assistance if renting	Not included
Welfare To Work	Non-cash welfare assistance (e.g. job placement, job training, education, child care) in prior calendar year for each person	Not included
Insurance Information	Current coverage of each type for each person and number of policies for private plans Annual family cost for private plans Whether managed care and type of restrictions Insurance coverage is 1 to 12 months after end of income year	Month-by-month coverage from up to 5 sources of any type Annualized premium for each plan Whether HMO for each plan Policyholder relationship for each plan
Medicaid	All persons	Month-by-month administrative data with exact coverage type
SCHIP	All persons	Not included
Medicare and Other Public	Medicare Parts A or B or both, TRICARE/CHAMPUS/CHAMP-VA, military health care/VA, Indian Health Service, State-sponsored, or other government	Medicare Parts A or B or both
Work-Related	Policyholder, source of coverage and if part of premium paid	Current or former employer, policyholder, source of coverage and 2-digit industry
Coverage Outside Household	Identifies persons with coverage from outside household	Not included
Other Private	Direct purchase, through a public program, Medi-Gap or Single Service	Direct purchase, Medi-Gap or AARP
Periods of Uninsurance	Persons covered only by Indian Health Service defined as uninsured Duration of current spell of uninsurance for uninsured persons Whether uninsured in prior 12 months for each person and number of months uninsured	Not applicable

TABLE II.11D. NON-CASH BENEFITS AND HEALTH INSURANCE

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Food Stamps	Monthly household recipiency since last interview (last 2 years) and amount in last month received	Monthly household recipiency January 2001 to interview, total amount each calendar year and current number of recipients
Other Nutrition	Any free or subsidized delivered meals ("meals on wheels") for age-eligible person or spouse or partner since last interview	Free or reduced-price meals for elderly, Free or Reduced-price School Lunch, Free or Reduced-price School Breakfast, or WIC in prior calendar year for each person
Housing And Energy	Current public or subsidized housing if renting	Current public or subsidized housing if renting Energy Assistance last winter and amount
Welfare To Work	Not included	Not included
Insurance Information	Current coverage of specified types for age-eligible person and spouse or partner and number of policies for private plans Whether HMO for Medicare/Medicaid (not differentiated) For up to 3 private plans, whether managed care, who else covered, number of years in plan and monthly insurance cost Insurance coverage is 1 to 12 months after end of income year	Up to 4 sources of coverage during prior 2 years for each person Reference period for each source of coverage includes income year but sources during income year not determined
Medicaid	Any coverage in last 2 years	Any coverage in last 2 years for each person
SCHIP		Not separately identified
Medicare and Other Public	Medicare and if Part B, or TRICARE/CHAMPUS/CHAMPVA or other military – no other public, e.g. Indian Health Service or VA	Medicare, TRICARE/CHAMPUS/CHAMP-VA, military health care/VA, Indian Health Service, State-sponsored, or other government in last 2 years for each person
Work-Related	Whether part of premium paid Selective screens skip some sources of coverage	Any coverage in last 2 years for each person
Coverage Outside Household	Not included	Not included
Other Private	Some source of coverage and whether part of premium paid	Direct purchase or Medi-Gap in last 2 years for each person
Periods of Uninsurance	If ever uninsured in prior 2 years for persons under 65	Months uninsured in each of prior 2 years for each person

TABLE II.12A. PERSON-LEVEL HEALTH AND HEALTH CARE UTILIZATION

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Health Status	4 times in panel, at least once per year Health status during third and fourth quarters of income year	Month of interview Health status is 2 to 4 months after end of income year
Work Disability	Persons 15 or over each interview and work disability history in wave 2	Persons 15 or over
Disability	Detailed functional limitations – ADLs and IADLs – twice in panel Conditions causing limitations or fair/poor health, and duration Disability days in prior 12 months	Not included
Informal Care	Identifies relationship, if in household, and if paid, for up to 2 helpers, and amount paid last month	Not included
Inpatient Utilization	Total inpatient days in prior 12 months -- annually	Not included
Ambulatory Care	Number of home or office visits or phone consultations in prior 12 months -- annually Number of dental care visits in prior 12 months -- annually	Not included
Other Medical Services	Summary questions	Not included
Prescription Drugs	Summary question	Not included
Cost of Insurance	Cost per person or policy in prior 12 months – annually	Not included
Out-Of-Pocket Costs	Total in prior 12 months including dental care and prescriptions but not cost of insurance -- annually	Not included
Charge or Reimbursement For Covered Services	Not included	Not included
Total payments	Not included	Not included
Sources of payments	Not included	Not included

TABLE II.12B. PERSON-LEVEL HEALTH AND HEALTH CARE UTILIZATION

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Health Status	Not included	Each interview Health status is contemporaneous with income year
Work Disability	Persons 16 or over	Work disability each reference period
Disability	Summary data on functional limitations for persons 5 or over	Detailed functional limitations annually – ADLs and IADLs each interview Detailed conditions and duration each interview and whether cause limitations or disability days or utilization of health care Disability days each reference period
Informal Care	Not included	Whether receives help
Inpatient Utilization	Not included	8 event level files for home health, office-based providers, outpatient hospital, emergency room, inpatient hospital, other medical expenses, dental, and prescriptions
Ambulatory Care	Not included	Data include dates, condition and procedure codes, provider type, medical or ancillary services, tests, medical supplies, DME, location, total payment, and source of payment including out-of-pocket per day, service or item reported in survey data
Other Medical Services	Not included	Summary variables contain annual utilization, charges, source of payment including out-of-pocket, and expenditure by service type
Prescription Drugs	Not included	
Cost of Insurance	Not included	By month by policy
Out-Of-Pocket Costs	Not included	Event level and totals by service type, provider and location
Charge or Reimbursement For Covered Services	Not included	Event level and totals by service type, provider, location and source of payment
Total payments	Not included	Event level and totals by service type, provider, location and source of payment
Sources of payments	Not included	Event level and totals by service type, provider, and location

73

TABLE II.12C. PERSON-LEVEL HEALTH AND HEALTH CARE UTILIZATION

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Health Status	Month of interview Health status is 1 to 12 months after end of income year	Current health status annually Health status is during income year
Work Disability	Persons 18 or over	Detail for persons under 65
Disability	Detailed functional limitations – ADLs and IADLs Conditions causing limitations, duration, and whether chronic	Detailed functional limitations – ADLs and IADLs – annually Diagnosis, condition and procedure codes for covered medical events from combined administrative and survey data, and from survey data for at-risk plans and uncovered events
Informal Care	Not included	Whether receives help and number of helpers
Inpatient Utilization	Number of stays and total inpatient days in prior 12 months	7 event level files for medical providers, outpatient hospital, dental, inpatient hospital, facility, institutional and prescriptions Data include dates, diagnosis, condition and procedure codes, provider type, location, medical or ancillary services, tests, medical supplies, DME, total payments, costs and source of payment per day, service or item from administrative and survey data, also a facility time line
Ambulatory Care	Home or office visits or phone consultations in last 2 weeks (dental care specifically excluded) Whether 10 or more medical provider visits in prior 12 months	
Other Medical Services	Not included	Summary files contain total utilization and expenditure for the 7 service types plus home health and hospice
Prescription Drugs	Not included	
Cost of Insurance	Family cost for each policy	By month by policy
Out-Of-Pocket Costs	Family total (no service detail) including dental care and prescriptions but excluding cost of insurance	Event level and totals by service type, provider and location
Charge or Reimbursement For Covered Services	Not included	Event level and totals by service type, diagnosis, condition, procedure, provider, location and source of payment
Total payments	Not included	By service type, provider, location and source of payment
Sources of payments	Not included	By service type, provider, location and event

TABLE II.12D. PERSON-LEVEL HEALTH AND HEALTH CARE UTILIZATION

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Health Status	Month of interview (age-eligible person and spouse or partner) Health status is 1 to 12 months after end of income year	Month of interview (head and wife/"wife") Health status is 1 to 12 months after end of income year
Work Disability	Age-eligible person and spouse or partner	Head and wife/"wife"
Disability	Detailed functional limitations – ADLs and IADLs Conditions (not linked to limitations) Disability days last month	Detailed functional limitations – ADLs and IADLs and whether caused by health problem (head and wife/"wife") Conditions, whether cause limitations, and duration (head and wife/"wife")
Informal Care	Detailed information on amounts and types of assistance, sources of assistance, whether paid and cost last month	Whether receives help for each ADL and IADL (head and wife/"wife")
Inpatient Utilization	Number of stays and total inpatient days in prior 2 years	Total inpatient days in prior 2 years (head and wife/"wife")
Ambulatory Care	Number of physician contacts in prior 2 years	Not included
Other Medical Services	Nursing home: Number of stays and total days in prior 2 years Summary questions on other services including dental care	Not included
Prescription Drugs	Summary question	Not included
Cost of Insurance	Current monthly cost for Medicare/Medicaid HMO and up to 3 private policies	Cost in prior 2 years (combined) for all coverages for family
Out-Of-Pocket Costs	Cost in last 2 years separately for inpatient, nursing home, outpatient surgery, physician, dental, home health and other (e.g. social worker) services Cost in last month for prescriptions	Cost in prior 2 years (combined) for inpatient and nursing home, for doctor, outpatient surgery and dental, and for prescriptions, in-home medical care, special facilities and other services
Charge or Reimbursement For Covered Services	Not included	Not included
Total payments	Not included	Cost in prior 2 years (combined) for all family medical care – out-of-pocket expenses plus reimbursed (covered) services
Sources of payments	Not included	Not included

TABLE II.13A. WEIGHTS AND CONTROL TOTALS

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
Basic Schema For Person Weights	Selection probabilities adjusted for non-response, attrition and movers, and post-stratified to control totals derived from the CPS	Selection probabilities adjusted for non-response and post-stratified to independent monthly control totals developed by the Census Bureau
Cross-Section Weights	Person, family and household weights	Person, family and household weights
Cross-Section Weight Timing	Each month and calendar year	March after reference year for all data collection months
Family and/or Household Weight Calculation	Person weight of reference person after family equalization process that ensures husbands and wives have the same weights while overall age, sex, and race/ethnicity control totals are maintained Family equalization averages the weights of the householder and spouse	Person weight of reference person after family equalization process that ensures husbands and wives or partners have the same weights while overall age, sex, and race/ethnicity control totals are maintained Method of family equalization depends on household composition and sex of reference person -- householder weight is used for spouse or partner, or the weights of the householder and spouse or partner are averaged, or a separate ratio adjustment is calculated
Longitudinal Weights	Person, family and household weights Longitudinal weights for panel	
Person and Family Universes	Anyone with person weight has family and household weights – 3 universes the same	Anyone with person weight has family and household weights – 3 universes the same
Person Control Totals	Age, sex, race/ethnicity, and marital and family status of householder, by month and rotation group	Age, sex, race/ethnicity, and State of residence
Family Control Totals	None	None
Income Control Totals	None	None

TABLE II.13B. WEIGHTS AND CONTROL TOTALS

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
Basic Schema For Person Weights	Selection probabilities adjusted for non-response and post-stratified to independent monthly control totals developed by the Census Bureau	Selection probabilities adjusted for non-response and attrition and post-stratified to control totals derived from CPS by AHRQ staff
Cross-Section Weights	Person and household weights	Person weights for restricted universe (see below) Two family weights -- for CPS-type and NHIS-type families
Cross-Section Weight Timing	July 1 of survey year for all data collection months	December of calendar year
Family and/or Household Weight Calculation	No family weight Household weight is person weight of female spouse of householder, or householder if not married, to prevent over-representation of husband-wife households	Person weight of reference person – same weights used in CPS-type families and NHIS-type families – then post-stratified to control totals[9]
Longitudinal Weights		Available for persons in individual 2-year panels
Person and Family Universes	Anyone with person weight has household weight – 2 universes the same	Person universe restricted to original NHIS sample persons and move-ins who were out-of-scope for original NHIS CPS-type family universe excludes part of person universe but includes move-ins related by blood or marriage (meeting CPS family definition) Broader NHIS-type family universe adds unmarried partner move-ins and others (meeting NHIS family definition) to CPS-type family universe
Person Control Totals	Age, sex, race/ethnicity and county	Age, sex, race/ethnicity, Census Region, and MSA/non-MSA (and income)
Family Control Totals	None	Family type (spouse present or not), family size, age, sex, and race/ethnicity of reference person, MSA/non-MSA, and region for CPS-type families CPS has no family control totals so family counts depend on CPS method of calculating family weights
Income Control Totals	None	CPS poverty rates for persons in CPS-type families as of December 31 – crossed with demographic control totals when person weights calculated

TABLE II.13C. WEIGHTS AND CONTROL TOTALS

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
Basic Schema For Person Weights	Selection probabilities adjusted for non-response and post-stratified to control totals derived from CPS and provided by Census Bureau	Selection probabilities adjusted for non-response and attrition and post-stratified to control totals from Medicare administrative files
Cross-Section Weights	Person, family and household weights	Person weights
Cross-Section Weight Timing	Weights separately calculated for each calendar quarter for control totals as of February 1, May 1, August 1 and November 1 Four quarter average is effectively mid-June of survey year	Ever enrolled during calendar year – not a point in time
Family and/or Household Weight Calculation	Person-weight of family member with smallest post-stratification adjustment	
Longitudinal Weights		Multi-year person weight
Person and Family Universes	Anyone with person weight has family and household weights – 3 universes the same Some families with household weights are refusals but family weights not adjusted to compensate	No family weight or universe
Person Control Totals	Age, sex, and race/ethnicity Weights prior to post-stratification also on file	Age, sex, race/ethnicity, MSA/non-MSA, region, and new/existing enrollee status
Family Control Totals	None	None
Income Control Totals	None	None

TABLE II.13D. WEIGHTS AND CONTROL TOTALS

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
Basic Schema For Person Weights	Selection probabilities adjusted for non-response and post-stratified to control totals derived from current March CPS Nursing home residents have separate weights in 2002 and subsequent years -- decedents have zero weights Person weights derived from family ("household") weights Summary description lacks key information	Longitudinal -- Selection probabilities adjusted for attrition and non-response, scaled to arbitrary totals, combining 1968 core and 1997 recent immigrant weights at 93:7 ratio Cross-section – longitudinal family weights trimmed and post-stratified to family control totals derived from March CPS excluding unrelated subfamilies and secondary individuals – used as person and family weights[10] Summary description lacks key information
Cross-Section Weights	Person and "married or partnered" (family) weights	Person and family weights for all interviewed persons, that user may scale or post-stratify
Cross-Section Weight Timing	March after reference year	March after reference year
Family and/or Household Weight Calculation	Relation to person weights not described	Longitudinal – family weight is average of person weights Cross-section – family weight is used for all persons
Longitudinal Weights	Not specifically described	Updated periodically for attrition and recently revised to reflect sample restorations from re-contact efforts
Person and Family Universes	Non-institutionalized with person weights have family weights – 2 universes the same	Anyone with person weight has family weight – 2 universes the same but longitudinal and cross-section universes differ
Person Control Totals	Age, sex, race/ethnicity, and "married or partnered" status – unmarried opposite-sex persons in same CPS households counted as partners if ages within 20 years CPS group quarters may be in or out of totals	None User may scale weights to current CPS or use more detailed demographic control totals
Family Control Totals	Age, sex, race/ethnicity, and "married or partnered" status – unmarried opposite-sex persons in same CPS household with age within 20 years counted as partners to create CPS "married or partnered" control total CPS group quarters may be in or out of totals	Age and race of head, region, and family size (1, 2 or 3+) for primary families and primary individuals excluding unrelated subfamilies and secondary individuals[10] CPS has no family control totals so family counts depend on CPS method of calculating family weights
Income Control Totals	None	None

79

TABLE II.14A. EASE OF ACCESS

	2001 Panel of Survey of Income and Program Participation	2003 Current Population Survey Annual Social and Economic Supplement
File Availability	All files on-line for download Can be subset with DataFerret before downloading or used on-line with DataFerret without downloading	All files on-line for download Can be subset with DataFerret before downloading or used on-line with DataFerret without downloading On-line table-creator
Files and Structure	All files are person-based, include family and household data, but contain 4 months of data that cover different months for each of the 4 rotation groups 9 core files each contain one 4-month wave and separate files each topical modules in various waves	One file with household record followed by family and primary individual record(s) followed by person record(s)
Variable Construction and Calendar Year Data	Many summary variables and recodes on files All calendar year data must be constructed for each person from monthly data in multiple core files	Summary variables and recodes on files – few variable need to be constructed
Survey and File Descriptions	Not fully updated from 1996 panel	Extensive and detailed technical write-up
Questionnaires	On-line and downloadable for core and all modules in easy to read format	On-line and downloadable in easy to read format
Data Dictionaries	Include alphabetical variable listings Each variable has short description with question wording and universe description including screen-ins and -outs	Include alphabetical variable listings Each variable has short description with question wording and universe description including screen-ins and -outs
Interviewer Instructions	Not available	Clear and spells out content item by item but geared to CAPI
Sample Design and Weights	Technical write-up but use is very complex	Extensive and detailed technical write-up
Technical Assistance	By phone or e-mail for simple and some technical questions	By phone or e-mail for simple and technical questions
Glossary	Short glossary is part of documentation More detailed CPS glossary is applicable	Detailed glossary is part of each year's documentation
Typical File Timing	2 to 3 years after fieldwork is complete	5 months after fieldwork is complete

TABLE II.14B. EASE OF ACCESS

	2002 American Community Survey	2002 Medical Expenditure Panel Survey Household Component
File Availability	Sub-sample public use files on-line for download lack sample month and inflation adjustment, income amounts are rounded and poverty status on file based on rounded income Over 700 detail tables using internal files for over 7,000 areas available on-line as well as more summary profiles and tables	All files on-line for download
Files and Structure	Household record followed by family and primary individual record(s) followed by person record(s)	Person-based file with round-specific data plus separate files with round- and event- specific data on employment, private insurance and health conditions, plus event-level medical services files, and linking files
Variable Construction and Calendar Year Data	Rounding and lack of either inflation adjustments or sample month data limit utility of public use files	Many variables round-specific and calendar year data must be constructed (not always possible)
Survey and File Descriptions	Geared to non-technical general public and to the on-line data products based on internal files Minimal description of public use file content or limitations including lack of inflation adjustments and rounding Extensive step-by-step descriptions of procedures of survey, and of processing, editing and preparation of internal files	Extensive information geared to technical user for health, insurance, utilization and expenditures-related files and variables Summary information on income data
Questionnaires	On-line and downloadable in easy to read format	On-line and downloadable (42 sections per year)
Data Dictionaries	Code lists for the relatively small number of variables with very abbreviated descriptions, in alphabetical not logical order (in 2 main groups)	Include alphabetical variable listings Very abbreviated variable descriptors with no question wording, universe description or screen-ins and -outs
Interviewer Instructions	Instruction brochure for mail survey but not CATI guide	Not available
Sample Design and Weights	Extensive step-by-step descriptions	Summary description lacks key information, e.g. impact or validation of post-stratification based on income
Technical Assistance	By phone or e-mail for simple and some technical questions	Not readily available
Glossary	Lengthy, comprehensive, detailed and clear	Lengthy glossary devoted almost entirely to medical and health-related terms, e.g. does not define family, or earnings
Typical File Timing	6 months after fieldwork is complete	1½ years after fieldwork is complete

TABLE II.14C. EASE OF ACCESS

	2003 National Health Interview Survey Family Core	2003 Medicare Current Beneficiary Survey
File Availability	Public use file on-line for download has income information limited to $5,000- and $10,000-wide brackets Access to internal files with actual income amounts requires approved analytic plan but files may never be taken off-site Users obtain and retain only tabular or regression output Cost was $500 plus $200 per day (or part) on site at RDC	No public use files on-line for download Files (Access to Care and Cost and Use) are Limited Data Sets but protected off-site use of files allowed Data use agreement and approved analytic plan required Cost: $480 per data set includes all claims files
Files and Structure	Separate household, family and person files with 5 separate files containing alternative imputation values	9 person-based files with survey or administrative record data, plus facility characteristics, residence time line, person summary and service summary files plus 7 event-level files – 7 bill files are also available
Variable Construction and Calendar Year Data	Many summary variables and recodes on files	Survey and administrative data from multiple years and sources have already been combined, unduplicated, imputed and placed on a calendar-year basis
Survey and File Descriptions	Extensive technical write-up	Extensive information geared to technical user that clearly lays out sample, survey, file and data construction
Questionnaires	On-line and downloadable in easy to read format	On-line and downloadable in easy to read format
Data Dictionaries	Public use file -- short variable descriptions with universe, question wording and screen-ins and -outs Internal file – not available during access application process	For Limited Data Sets -- available on-line for download Abbreviated variable descriptors with universe, question number, years available and screen-ins and -outs
Interviewer Instructions	Clear and spells out content item by item but geared to CAPI	Clear and spells out content item by item but geared to CAPI
Sample Design and Weights	Technical write-up and file contains weights prior to post-stratification as well as final weights	Technical write-up
Technical Assistance	By phone or e-mail for simple questions	By phone or e-mail
Glossary	No separate glossary, only NCHS definitions mostly of health and medical terms, with some incorrect NHIS information[11]	No separate glossary
Typical File Timing	6 months after fieldwork is complete	1 year after fieldwork is complete for Access to Care and 2 years for Cost and Use

TABLE II.14D. EASE OF ACCESS

	2004 Health and Retirement Study	2003 Panel Study of Income Dynamics
File Availability	HRS: All files on-line for download RAND: All files available on-line to download	All files on-line for download Easy to subset files and automatically link years to download
Files and Structure	HRS: Survey year has 37 files for living sample persons plus separate imputation files and files on decedents Cross-year tracker file for longitudinal work RAND: Single person-based flat file for each year	Survey year has family file on head and wife/"wife" and family (most data), individual file (has insurance coverage) -- cross-year individual file also available Various longitudinal files track events and family structure over time, e.g. parent identifier file
Variable Construction and Calendar Year Data	HRS: Essentially raw data RAND: Files have had major consistency checks and edits -- including longitudinal -- with new and (relatively) consistent summary variables created and reconciled for all years	Many summary variables and recodes on files and most data on calendar year basis and/or month-by-month Comprehensive topical index drills down to question and data dictionary entry for all variables in all years 1968 forward
Survey and File Descriptions	Little systematic or technical information beyond grant application information and publications -- some obsolete material (e.g. original incorrect weight calculations) remains	Extensive information geared to technical user that clearly tracks survey evolution, changes and supplemental data collections over its 40 year span
Questionnaires	On-line and downloadable (37 per year) with CAPI code	On-line and downloadable in easy to read format
Data Dictionaries	HRS: 37 each year for sample persons and 37 for decedents, each with abbreviated variable descriptors and raw counts RAND: single data dictionary covers all sample persons for all years, crosswalks constructed variables to HRS names, and documents changes or inconsistencies	For files, years and variables selected for download, custom data dictionary automatically created for selected variables with codes, raw counts and variable list Each variable entry has text of question, universe for code of "inapplicable", years available and variable name each year
Interviewer Instructions	Not available	Clear, comprehensive, and spells out content item by item
Sample Design and Weights	Summary descriptions lack key information	Summary description lacks key information
Technical Assistance	Not readily available for non-academic users	By phone or e-mail for simple and technical questions
Glossary	Minimal 2-page glossary	No separate glossary but data dictionary and interviewer instructions provide definitions "as you go"
Typical File Timing	2 years after fieldwork is complete	12 to 14 months after fieldwork is complete

Endnotes to Tables

[1] Filers of tax form 1040EZ skip self-employment income questions -- sole proprietorship or farm (Schedules C and F) or other business arrangements (Schedule E).

[2] Prior to 2002, non-filers, and refusals/"don't knows" for specific tax form (1040, 1040A or 1040EZ), skipped all questions on taxable income amounts which were then allocated or imputed. Persons filing 1040EZ skipped many questions on taxable income amounts, including Social Security and pensions.

[3] In recipiency (no amounts) data, rental income is grouped with dividends, estates and trusts, and unemployment and worker's compensation are grouped with Veterans payments and family contributions in "other"

[4] There are numerous differences. Except for the household income used to calculate RAND's poverty measure, RAND total income includes Food Stamps. Earnings exclude self-employment, which is included with asset income. Pensions and annuities include not only private but public retirement systems – except military retirement – but with no distinction between disability, retirement and survivors benefits. Other government transfers include Veterans' benefits -- including "military pensions" -- welfare and Food Stamps. The variable SSDI combines SSI and the disability component of Social Security. The Social Security variable includes the old age, survivor and dependents components but not disability. And Unemployment and Workers Compensation is a combination that may also include Workers Compensation survivor benefits.

[5] HRS income exclusions vary from year to year. The thousand-page RAND HRS Data Documentation (Version G) includes careful descriptions of wave-to-wave differences and as much of a concordance as is possible.

[6] PSID groups earnings from all sources and asset income from all sources under the term "taxable income".

[7] PSID groups all government and private transfers except Social Security under the term "transfer income". This includes AFDC/TANF, SSI, other welfare, Veterans benefits including military retirement, Unemployment and Workers Compensation, all retirement, pension, annuity and periodic IRA income, child support, alimony and contributions from relatives or friends.

[8] An apparent error in the algorithm calculating poverty status appears to use the annual (ever-on) family composition and income rather than that as of December 31. AHRQ staff have been informed of this problem.

[9] The impact of the additional sequential post-stratification of family weight can be seen in the following: For persons from the NHIS sample (key) who were respondents and in the universe (in-scope) on December 31 in CPS-type families of size one, the sum of person weights is 3.0 million higher than the sum of their CPS-type family weights. Even when 11 cases with a person-weight but no CPS-type family weight are removed, the difference is still 2.7 million.

[10]The description applies to the preliminary cross-section weights available when this report was prepared and may not apply fully to the final weights.

[11] The on-line NCHS definition of family income says, "For purposes of the National Health Interview Survey (NHIS) and National Health and Nutrition Examination Survey (NHANES), all people within a household related to each other by blood, marriage, or adoption constitute a family." See <http://www.cdc.gov/nchs/datawh/nchsdefs/familyincome.htm>.

III. METHODOLOGY

The principal goal of our empirical methodology was to produce survey estimates of income that reflected, to the extent feasible, comparably defined universes, income, and families across the surveys. Comparability was based on CPS definitions of these concepts, as the CPS is the official source of statistics on family income and poverty for the United States. With survey-specific adjustments designed to achieve this, we created a standard set of income tabulations for four of the five general population surveys—CPS, ACS, SIPP, and MEPS. More limited tables were produced for NHIS and PSID because neither survey collects total personal income for all adults. A smaller set of tables was created for HRS and MCBS because of their restricted universes and specific limitations of their data. We prepared additional tabulations to examine specific survey design issues. These tabulations were based on individual surveys so that we could simulate different design features while holding constant all other aspects of the survey estimates across the simulations. This chapter documents the methodology for creating comparably defined universes, income, and families across the surveys and reviews the specification of the full range of tables on which the analyses reported in Chapters IV through VI are based.[6]

A. DEVELOPING COMPARABLE ESTIMATES ACROSS SURVEYS

All analyses in the study use income data for 2002 (HRS and MCBS income for 2003 were deflated with the CPI-U), which is defined as the calendar year except for the rolling reference period in ACS, which spans 23 instead of 12 months. In developing comparable estimates across

---

[6] For those wishing to replicate any study calculations on any of the eight surveys, copies of the file extracts and all of the SAS programs used by the study will be delivered to the government on completion of the study and should be requested from ASPE.

the surveys, however, we sought to compensate for differences in the universe, the income concept, and the definition of a family. Our methods and their limitations are reviewed below.

1. Comparable Universe

Even though our estimates of income focus on calendar year 2002, no two surveys among the eight provide this information for populations at the same point in time, so no two sets of estimates refer even nominally to the same universe. We did not attempt to correct for universe differences that were due to survey timing or to the ACS's exclusion of college students living in dormitories. However, we did adjust for universe differences that arose from differential treatment of six specific subpopulations: (1) decedents, (2) persons living abroad, (3) residents of institutions, (4) active duty armed forces, (5) unrelated children under 15, and (6) exclusion of students temporarily away from home in the PSID. Specific procedures and their impact are described below, followed by a discussion of sample selection issues that we encountered in developing the estimates for MEPS.

a. Universe Adjustments

Our income estimates for each survey are restricted to persons who were alive and residing in the U.S. at the time the survey was conducted and not living in an institution. For MEPS this meant that we restricted the sample to persons who were in-scope on 12/31/02. Original sample members who died or entered an institution during the year have sample weights, so it was not sufficient to restrict the MEPS estimates to persons with weights. For SIPP, we restricted our estimates to persons with December 2002 cross-sectional weights. No specific restrictions were required for CPS, ACS, or NHIS, but for PSID and MCBS we had to exclude sample members residing in institutions at the time the survey was fielded, and we also had to remove persons living in Puerto Rico (MCBS) or more generally abroad (HRS and PSID). In addition, for PSID we had to add back students who were away at school.

86

While four of the five general population surveys are described as representing the civilian non-institutional population, and the ACS recently added residents of institutional and non-institutional group quarters, including military barracks and college dormitories, all five surveys include some members of the armed forces on active duty living in housing units on or off base, as detailed in Chapter II. Coverage of this subpopulation differs among the surveys, however. Furthermore, neither the NHIS nor the MEPS assigns weights to sample members on active duty in the armed forces. For these reasons we have removed members of the active duty armed forces from our comparative estimates. We have also removed all members of their families—largely because of the differential coverage of armed forces members across surveys but also because the removal of the latter often took away their families' principal source of income. Rather than misrepresent their families' economic circumstances or attempt to add back their contributions to family income while excluding the members themselves from our estimates, we opted for this simpler solution.

The official definition of poverty excludes unrelated children under 15 because the CPS does not collect income data from such individuals. We have followed suit. Unrelated children under 15 are excluded from all of our estimates. In addition to conforming to the official definition of poverty, this decision to exclude such children from our estimate also reflects the fact that two of the surveys, the NHIS and MEPS, exclude unrelated minors from their sample frames.

As we enumerated in Chapter II, the surveys differ with respect to whether college students who are temporarily away at school are counted where they usually reside (generally at home with their families) or where they are living at the time of the interview. While this will affect estimates of family income and poverty, it does not affect the comparability of survey universes except for the PSID and ACS.

The PSID excludes from the interviewed family any students who were away at school but, unlike NHIS or ACS, does not attempt to interview them separately. They are not counted in the family size used to determine an annual poverty threshold, and their incomes during the reference year are excluded from family income. Nevertheless, records for students are included with those for other family members, so it is possible to add back these students into their respective families and the total population. We did so and increased the size of the population by about 3 million.

The ACS counts students where they live but did not begin to include college dormitories in its sample frame until 2006. For 2002, then, college students who were living in dormitories at the time their families were interviewed are excluded from the ACS universe. Because the ACS uses a rolling sample, the number of students who are excluded from the ACS universe will vary with the survey month. Few students will be excluded in the summer months while many will be excluded during the school year. Over the full calendar year, perhaps three-quarters of the students who attend college and live in dormitories will be excluded from our estimates from the ACS.[7]

The impact of the ACS exclusion of students in college dormitories, along with other residents of non-institutional group quarters, is evident in Table III.1, which reports survey population estimates, before and after the adjustment to a common universe, for the five general population surveys arrayed in chronological order by the calendar date(s) of their respective population controls. Prior to adjustment, the ACS falls short of the next highest population estimate (for SIPP, five months later) by 3.4 million persons. This difference is unchanged by the

---

[7] How the ACS is post-stratified will affect this result. The Census Bureau post-stratifies the CPS to population controls representing the civilian non-institutional population. Residents of non-institutional group quarters must be excluded from such controls to match the ACS universe. The Census Bureau can use group quarters population data for this purpose, but we are not certain how the changing size of the college dormitory population over the calendar year is handled.

exclusion of active duty armed forces members and their families and unrelated children under 15.[8] Population growth between July and December would account for about 1.2 million of the difference, based on Census Bureau estimates of the civilian non-institutional population. This leaves 2.2 million to be attributed to the ACS group quarters exclusion.

TABLE III.1

SURVEY POPULATION ESTIMATES BEFORE AND AFTER ADJUSTMENT TO COMMON UNIVERSE
(1,000s OF PERSONS)

Estimate	ACS	SIPP	MEPS	CPS	NHIS
Population Control Date(s)	**07/01/02**	**12/01/02**	**12/01/02** [a]	**03/01/03**	**Quarterly 2003**[b]
Survey Total Population	280,717	284,101	284,569 [c]	285,933	286,010
Exclusions					
Active duty armed forces and families	1,881	2,595	1,043	2,766	2,055
Unrelated children under 15	1,145	426	230	616	244
Residual Population for Comparisons	277,692	281,080	283,296	282,551	283,711

Source: Mathematica Policy Research, from tabulations of the 2002 ACS, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 CPS ASEC supplement, and the 2003 NHIS.

[a] For post-stratification to population totals, the MEPS sample and family composition were defined as of December 31, 2002. MEPS documentation indicates that the sample weights were controlled to population totals "derived by scaling back the population distribution obtained from the March 2003 CPS to reflect the December 2002 estimated population distribution, employing age and sex data available from the December 2002 CPS." The Census Bureau population estimates used to weight the latter have a reference date of December 1, 2002.

[b] Population controls by calendar quarter refer to February 1, May 1, August 1, and November 1 of 2003. The midpoint of these dates is June 15, 2003.

[c] The population listed for MEPS corresponds to sample persons identified as in-scope on 12/31/02 and with a person weight. Armed forces members with weights add an additional 45 thousand to the population total but are defined as out-of-scope, so their weights would not have been post-stratified to the population controls. The 45 thousand weighted armed forces members are excluded from the count of excluded active duty armed forces and families.

---

[8] The ACS has 0.7 million fewer members of the armed forces and their families but 0.7 million more unrelated children under 15. The greater number of the latter is due to the ACS's not identifying family relationships among persons unrelated to the householder. Children who would be classified as members of unrelated subfamilies in the SIPP and CPS are identified as unrelated individuals in the ACS.

In addition to what it tells us about the ACS, Table III.1 also shows that adjusting the surveys for comparable universes actually increased rather than reduced the variation in population estimates among the remaining four surveys. More specifically, while the CPS and SIPP estimates became more similar to each other, and the MEPS and NHIS estimates did the same, the disparity between the first two and the second two grew larger. Initially, the MEPS population total was 0.47 million greater than SIPP, and the NHIS population was just 0.08 million greater than the CPS. As a result of adjustment, the difference between the MEPS and SIPP populations grew to 2.2 million while the difference between the NHIS and CPS populations grew to 1.2 million.

There is a big difference between the CPS and SIPP, on the one hand, and MEPS and NHIS, on the other, in the number of active duty armed forces members and their families who were removed from their respective populations: 2.6 and 2.8 million for SIPP and CPS versus 1.0 and 2.1 million for MEPS and NHIS. This accounts for the bigger difference in population sizes after rather than before adjustment. Active duty armed forces members do not receive weights in MEPS or NHIS whereas they do receive weights in SIPP and the CPS. The MEPS sample was post-stratified to totals constructed from the CPS. If active duty armed forces members had not been removed from the constructed totals, then the initial MEPS population would have been too high, and removing the families of active duty armed forces members would not have removed enough persons. This could explain the MEPS results relative to the two Census Bureau surveys, and the same phenomenon may be at work in the NHIS as well, but we cannot confirm this in either case without more detailed information on each survey's post-stratification than is readily available.

Because of these population differences, particularly those between the ACS and the other four general population surveys, we must be aware that when we compare estimates of

population subgroups or total dollars, a portion of the difference will be attributable to differences in population size.

Population estimates from the PSID are substantially lower than those reported in Table III.1 for the other surveys. Preliminary PSID cross-sectional weights for 2003 made available for use by the study yield a population estimate of 261.45 million after the exclusion of persons living abroad, in institutions, or in families with active duty armed forces members, and the *addition* of 3 million students temporarily away at school. This is 21.1 million lower than the CPS estimate even though the PSID was post-stratified to controls obtained from this same CPS file.

The shortfall can be attributed to several aspects of how the PSID weights were post-stratified. First, post-stratifying to CPS families rather than persons introduces a downward bias from the outset because CPS family weights underestimate the population by several million persons. Second, CPS unrelated subfamilies and secondary individuals were excluded from the family level controls to which the PSID was post-stratified. Given that PSID families include unmarried partners, who are counted as secondary individuals or unrelated subfamilies in the CPS, it would be appropriate to exclude a portion of these families and individuals from the controls—but no more than 38 percent.[9] Third, because they include unmarried partners, PSID families are somewhat larger than CPS families, so post-stratifying to CPS families by size, with no correction for this size difference, introduces a further downward bias. Fourth, families of size three or greater were combined for post-stratification, so larger families, which are more numerous in the CPS than the PSID, are underestimated in the latter. Fifth, PSID sample members who were outside the CPS universe—specifically, living abroad, in institutions, or in military barracks—were not excluded from post-stratification. When we dropped them from the

---

[9] This estimate is based on a tabulation of the 2003 CPS ASEC supplement.

PSID sample, we reduced the estimated population even further below the CPS. It is possible, too, that the shortfall would be even greater if persons excluded from the PSID universe—students temporarily away at school—had been removed from CPS families when constructing the controls (possibly shifting some CPS families to smaller size categories). Given that PSID staff will be aware of these shortcomings as they work on revisions to the sample weights, it is likely that the 21.1 million person shortfall in the PSID will be reduced when final weights are released in early 2009.

b. Selection of MEPS Records

Of the 37,015 MEPS sample members who are identified as in-scope on December 31, 2002 and have nonzero person weights, 882 were in families from which one or more members had no records in the public use file. For 382 of these sample members, the missing family members included the family head.[10] Despite the missing family members (and their incomes), family incomes and poverty were calculated for the family members who were present, and the resulting ratios of income to poverty thresholds were used to post-stratify the person weights to the distribution of persons by poverty class observed in the March 2003 CPS. Not surprisingly, the members of these "partial families," as we shall term them, show exceptionally high poverty rates, which we attribute in large part to their incomplete family and income data. Weighted, the sample members from these partial families represent 6.1 million persons or 2.15 percent of the December 31, 2002 MEPS population.

We considered alternative ways to deal with the partial family members in constructing MEPS estimates for comparison with the other surveys. One strategy was to exclude the most

---

[10] Among those who were in-scope and had nonzero person weights, sample persons in families with missing members are identified by FMRS1231 < 1, and those with missing family heads are identified by FMRS1231 = -1 or, alternatively, FCSZ1231 or CPS family size = -1 (inapplicable or unknown).

troublesome partial families—those with missing reference persons. Another approach was to exclude *all* partial families. Yet another approach was to use family weights instead of person weights. With the MEPS family weights, missing sample members are not an issue. All of the families with family weights have data on all of their members, and all family members are assigned family weights, regardless of how or when they entered the sample.[11]

Estimates based on these alternative strategies are presented in Table III.2. By retaining all sample members from partial families we end up with a sample of 36,820 persons after dropping unrelated individuals under age 15 and families with armed forces members on active duty. Weighted, this sample represents 283.3 million persons with an aggregate income of 6,257.7 billion dollars. Excluding the partial family members from families with missing reference persons reduces the estimated population by 2.4 million persons and the estimated aggregate income by $25 billion. Excluding all persons from partial families reduces the sample count by another 499 persons, lowers the population estimate by an additional 2.7 million, and removes $74 billion from aggregate income.

Even more striking is the incidence of poverty among members of partial families. With all sample members with nonzero person weights included, the overall poverty rate is 12.48 percent. Persons in partial families have a poverty rate of 34.45 percent, however. Dropping those individuals in partial families with missing reference persons lowers the overall poverty rate to 12.12 percent. From these changes we can calculate that the poverty rate among the excluded subset of persons in families with missing reference persons is 54 percent. Dropping the remaining persons in partial families—those with a reference person—reduces the overall poverty rate to 11.99 percent.

---

[11] By contrast, MEPS person weights are assigned only to original sample members—that is, those who were selected from the respondents to the NHIS—and to new family members (primarily newborns) who were not in the survey universe at the time the sample was selected. See Chapter II.

TABLE III.2

ALTERNATIVE MEPS ESTIMATES BASED ON ALTERNATIVE TREATMENT OF PERSONS
IN PARTIAL FAMILIES
(1,000S OF PERSONS)

Weight and Subsample[a]	Sample Persons	Weighted Persons (Millions)	Aggregate Income ($Billions)	Number Poor (Millions)	Percent Poor	Poverty Rate Among Members Of Partial Families	Percent Of All Poor Who Are in Partial Families
Person Weight							
All sample members with nonzero person weights	36,820	283.30	6,257.7	35.35	12.48	34.45	5.95
Excluding members of partial families with no data on the family reference person	36,465	280.87	6,232.9	34.04	12.12	21.61 [b]	2.34
Excluding all members of partial families	35,966	277.19	6,158.9	33.24	11.99	NA	NA
Family Weight							
All sample members with nonzero family weights	37,347	278.81	6,000.0	35.16	12.61	NA	NA

Source: Mathematica Policy Research, from the 2002 Full-Year Consolidated MEPS HC.

[a] All estimates are restricted to persons who were in scope on 12/31/02. Estimates exclude unrelated individuals under age 15 and persons in families with members of the armed forces on active duty.
[b] This is the poverty rate for partial families after those with missing reference persons are excluded. The poverty rate among members of partial families with missing reference persons is 53.98 percent.

An alternative way of dealing with the partial families is to use family weights instead of person weights. MEPS family weights are assigned only to families with complete data. Unlike the person weights, they are assigned to both original sample members and persons who joined MEPS families after the start of the panel and, for that reason, did not qualify for person weights. Family weights in general are problematic for person-level analysis. None of the surveys with which we are familiar reconciles their family and person weights, which means that population estimates obtained using family weights are not consistent with the population estimates obtained from person weights. As a rule, it appears that applying family weights to individual family members yields too few total persons. The shortfall varies substantially by survey, but in our experience the direction is always the same. This holds true even though the methods used to

develop family weights vary across the surveys. The MEPS results are consistent with this experience. With the family weight the population estimate is 278.81 million or 4.5 million below the person weight total. Furthermore, aggregate income drops to $6,000.0 billion or $258 billion below its maximum value while the poverty rate rises to its highest level, 12.6 percent.

2. Common Income Concept

The definition of income used in the comparative analysis is the same definition that is used in official poverty statistics, which is pretax money income as measured in the CPS. Table II.8 in Chapter II identifies differences between the CPS income concept and the income concepts used in the other seven surveys. For example, SIPP excludes educational benefits that are included in CPS money income, but it includes lump-sum payments from certain retirement accounts that are not counted in CPS money income. MEPS excludes tax exempt interest for tax filers, which is counted in the CPS, but includes taxable lump-sum payments from retirement accounts. In addition, by referring respondents to their tax returns, MEPS implicitly uses tax concepts to define income, which implies that wages may exclude, for example, pre-tax deductions for contributions to 401(k) plans or some health insurance premiums. For other surveys, whether there are differences in the income concepts depends heavily on respondent interpretation of questions asking about broadly-defined sources.

While our intent was to adjust the survey estimates for departures from the CPS income concept, very few adjustments were needed or possible. CPS income includes only regular payments from an IRA, Keogh, or 401(k) plan whereas a single MEPS variable includes both regular and lump-sum payments from this source. Since we needed micro-level data, our two options were to include or exclude the entire amount of the MEPS variable. The regular payments captured by the CPS question totaled only $3.3 billion whereas the MEPS item collected $65.6 billion in both regular and lump-sum payments. Based on these comparative

magnitudes, we concluded that the income captured by the MEPS item was almost entirely outside the CPS income concept. Therefore, we excluded the MEPS variable from MEPS income. But in Chapter V we show the income picked up by this variable and how its inclusion or exclusion affects the number of poor. SIPP also collects lump-sum payments, but they are recorded separately from regular payments. We were able to exclude just the lump-sum payments from the SIPP income estimates. Our analysis in Chapter V compares the MEPS and SIPP amounts of combined regular and lump-sum payments.

NHIS collects total family income in a single question, so there were no sources to add or subtract in order to match the CPS income concept. However, in more than a fifth of NHIS families the sum of reported personal earnings over all family members exceeds the reported total family income. We investigated substituting the sum of reported earnings for total family income when the former exceeded the latter; the results are reported in Chapter IV.

Both the HRS, through a version of the data produced and released by RAND, and the PSID provide a single constructed family income variable. For both surveys this is what we used as family income in our analyses. MCBS collects dollar amounts for only one measure of income, which is the sum of the incomes of the sample member and spouse.

SIPP required a special income adjustment to compensate for income that is not collected in SIPP but is needed to calculate annual income. SIPP is unique among the eight surveys in collecting income month-by-month, four months at a time, rather than asking respondents to report their income for a previous 12-month period. To obtain annual income for a population defined at a point in time, the monthly amounts must be summed over a specified 12-month period. This in itself is not difficult, but because SIPP was not designed to collect retrospective annual income, some respondents are missing one or more months out of a prior 12-month period. For example, to construct annual income for the 2002 calendar year, as we do here, we

sum the reported amounts for January through December 2002 for the sample of respondents with weights for December 2002.[12] Among these weighted respondents, those who joined sample households after January 2002 will have no reported income for the months before they joined these households. Those whose households missed one or more interviews during the year, regardless of when they joined the sample, will be missing up to four months of CY 2002 income data for each missed interview. To produce an estimate of annual income for each such respondent, it is necessary to compensate for the missing months in some way. To create the estimates presented in this report, we applied a simple ratio adjustment to the sum of the reported months, inflating the reported sum by a factor of 12 divided by the number of reported months. This is not a sophisticated imputation strategy, by any means, but it serves the purpose of giving us annual numbers that are consistent with the reported data. It also reflects what a typical user might do.

Two of the surveys—HRS and MCBS—provided income for a 2003 reference year rather than 2002. Following the recommendation of the TAG, we deflated the 2003 incomes to 2002 dollars. This was accomplished by dividing each reported 2003 income by 1.0228, which represents the price increase between calendar years 2002 and 2003 recorded in the CPI-U series.

Income data from the ACS do not correspond to a calendar year or to any single 12-month period. Instead, respondents are asked to report their incomes for the 12 months preceding the interview. Thus the income data collected in the 2002 ACS represent 12 successive 12-month periods ending December 2001 through November 2002 (or starting January 2001 through December 2001).[13] In the Census Bureau's internal files, which are used to produce both

---

[12] These respondents have been weighted to Census Bureau population controls for that month.

[13] The distribution of reference months over the 12 interview months resembles a pyramid, with the middle month, December 2001, occurring in all 12 reference periods and the outermost months, January 2001 and November 2002, occurring in only one reference period each.

published and on-line tables, income from the 12 different reference periods is inflation-adjusted to reflect price levels during a fixed period corresponding to the calendar year of the survey. The public use files contain only unadjusted income and an average of the 12 adjustment factors, and they do not include the interview month. With these data it is not possible to replicate the adjusted incomes that appear in the Census Bureau's internal files. To prepare the estimates of ACS income presented in this report, we inflated the reported incomes by the average adjustment factor. This *under*-adjusts incomes collected *early* in the survey year and *over*-adjusts incomes collected *late* in the survey year. To prepare estimates of ACS poverty status, we used the ratio of income to poverty reported on the public use file, which incorporates the Census Bureau's inflation adjustments by interview month.[14]

3. **Common Family Definition**

Official poverty statistics incorporate the definition of a family that is used in the CPS, and we apply this same definition to compare estimates of income relative to poverty across surveys. A family in the CPS consists of two or more persons living in the same household and related by blood, marriage, or adoption. A CPS family does not include unmarried partners or foster children, but such persons are included in the family definitions of some of the other surveys.[15]

For two of the eight surveys—NHIS and PSID—we created CPS families within a subset of families that reflected a broader family concept. In each case the family members were

---

[14] To determine poverty status in the ACS, the Census Bureau compares *unadjusted* incomes to *adjusted* thresholds. This is mathematically equivalent to comparing incomes that have been adjusted to reflect price levels during the calendar year to thresholds for that same calendar year as long as the same inflation index is used for both.

[15] There is one additional wrinkle in the CPS family definition, but we do not apply it to the other surveys even though it has implications for poverty measurement. If a family does not include the householder (or, by implication, relatives of the householder), making it an "unrelated subfamily" in Census Bureau terminology, its membership is restricted to the family reference person, a spouse, and/or one or more never-married children under 18. The restrictive composition of an unrelated subfamily is a function of limitations in the relationship data collected in the CPS.

reassembled into two or more CPS families, and the income of the original family was apportioned among the new families. These procedures are detailed below. A third survey, MEPS, uses both the CPS family concept and a broader family concept (the same as NHIS), and both are coded on the public use file. For MEPS, then, it was not necessary to create CPS families from more inclusive families; we could use the family data coded on the file.

A fourth survey, HRS, also includes unmarried partners as members of the same family. This affects the family income variable on the RAND-HRS file, which we elected to use for our comparative analysis. Unmarried partners are much less common in the older population that the HRS represents than in the general population.[16] Furthermore, our comparative analysis with the HRS data was designed to be much more limited than the analysis involving the general population surveys. For these reasons and because the RAND file lacked suitable personal income variables on which to base a decomposition, we elected to proceed with the HRS analysis without attempting to separate unmarried partners from the family.

a. Creating CPS Families in NHIS

The family is the basic data collection unit in NHIS. A family respondent provides much of the information obtained from the family, including the family's total income for the prior calendar year. The concept of family used in NHIS is more inclusive than the CPS family concept. The NHIS family encompasses unmarried partners of the reference person or (in a few cases) of a child or parent of the reference person, whereas the CPS would treat the partner as an unrelated individual. Most commonly, an NHIS family that departs from the CPS family concept will include just a reference person and the reference person's partner. Most of the rest include

---

[16] Based on our own tabulations, fewer than 15 percent of the unmarried partners of householders or unrelated subfamily reference persons in the 2003 CPS ASEC supplement were 50 and older. Two-thirds of the unmarried partners were under 40.

children of the reference person or partner but no other adults.[17] In addition to including unmarried partners, the NHIS family also includes foster children, who would be treated as unrelated individuals in the CPS, regardless of their age. In all, an estimated 5.8 million or 4.9 percent of NHIS families included unmarried partners or foster children. We designated these "non-CPS" families.[18] In order to generate income and poverty statistics from the NHIS that were comparable to the CPS and the other surveys, it was necessary to break up these non-CPS families to form new families that were consistent with the CPS family concept.

Operationally, we achieved this as follows. First, we created two new families (or, in four cases, we created three new families) from each non-CPS family. An unmarried partner of the reference person was assigned to one new family along with anyone identified as that partner's child. The remaining family members were assigned to the second new family. Some 100,000 foster children age 15 or over were assigned to new, one-person families. Some 200,000 foster children under age 15 were dropped from the sample, as was done for unrelated children under 15 in all surveys. For the families from which these children were dropped, family size was reduced to calculate poverty status, but family income was not changed.

Next, the total family income of each non-CPS family had to be distributed among its subsidiary CPS families. Because there was any number of ways to do this, we elected to apply two alternative algorithms, described in some detail below, in order to determine the possible range of impacts on the poverty count. One algorithm, yielding a *lower bound*, would distribute

---

[17] About 40 percent of the minor children in these families are biological children of both the reference person and partner, but their relationship codes identify them as children of the reference person. In the relatively few cases where a child is identified in the relationship code as a child of the partner, typically that child belongs to the partner but not the reference person. In the CPS, until recently, a biological child of two unmarried partners would be identified as the child of only one partner. In NHIS both biological parents are identified as such.

[18] A family qualified as a non-CPS family if any member was identified as the unmarried partner of another member or if any member was identified as a foster child with no biological or adoptive parent present. That is, in a two-parent family the child had to be the foster child of both parents since nearly half of the children in husband-wife families who were coded as the foster child of one parent were coded as the biological child of the other parent.

the family income in a manner that would produce the *fewest* number of poor persons. The second algorithm, yielding an *upper bound*, would distribute the family income in a manner that would produce the *most* poor persons. We designed and applied the two algorithms and determined that the range between their additions to the poverty count was 430,000 or just 0.15 percent or total persons. Given the small magnitude of the range, we decided to use the average of the lower and upper bounds for each family as a point estimate. That is, for each new family, we calculated two alternative family incomes and assigned the average.

In applying this approach, we made use of personal earnings, which was reported, potentially, for each person 18 and older. We calculated the sum of personal earnings over all members of the NHIS family, calling it family earnings, and compared the result to the total family income. Three scenarios were possible: (1) family earnings and total family income were identical, (2) family earnings exceeded total family income, or (3) total family income was greater than family earnings. What we did next depended on which scenario applied.

If family earnings and family income were equal, then no additional distribution of income was necessary. We assigned each person the amount of his or her own personal earnings and then summed these amounts over the members of each subsidiary CPS family to obtain CPS family incomes that summed to the NHIS family income. The lower and upper bounds were identical.

If NHIS family income was less than NHIS family earnings, we multiplied each person's earnings by the ratio of family income to family earnings. This was done to maintain the original family's total income (and aggregate family income in the population). This reduced the sum of earnings over all NHIS family members to the amount of total family income. We then summed the reduced earnings over the members of each subsidiary CPS family in order to obtain a family income for each CPS family. Here, too, the lower and upper bounds were identical.

101

If total family income exceeded family earnings, and the earnings were not zero, we calculated the excess of family income over family earnings and then distributed the excess among the subsidiary CPS families in two alternative ways, representing the lower and upper bounds. With either alternative, each person started with his or her full earnings. For the lower-bound estimate, we assigned the excess family income to the adult with the *lowest* earnings.[19] For the upper-bound estimate, we assigned the excess family income to the adult with the *highest* earnings. Incomes were then aggregated over the members of each CPS family within the larger NHIS family to create both lower- and upper-bound estimates of family income for each CPS family. The average of the two estimates for each CPS family was then assigned as the family's income.

For the small number of families (under 400,000 or less than a third of a percent of all families) with no NHIS family earnings, the NHIS family income was apportioned among adults as follows. For the lower bound, we split the family income equally among the adults. For the upper bound, we assigned twice as much income to each adult male as to each adult female, approximating the typical ratio of Social Security benefits between husband and wife, where the spousal benefit is 50 percent of the retiree's benefit. If the adults were the same sex and there were only two, we assigned two-thirds of the income to the older adult, with the other adult receiving one-third. If there were more than two adults, we assigned twice as much income to the oldest adult as to the rest. As above, we then aggregated each alternative set of incomes over the members of each CPS family to create both lower- and upper-bound estimates of family income for each CPS family. We assigned the average of the two estimates for each CPS family as the family's income.

---

[19] In the vast majority of cases, as noted above, there were only two adults.

Poverty thresholds for all of the new CPS families were determined based on the new family size, number of related children under 18, and whether the family included anyone 65 or older. Estimates of the impact of using the NHIS family concept to assign poverty status are reported in Chapter IV.

b.   Creating CPS Families in the PSID

Like the NHIS, the PSID includes unmarried partners in the same family, except that it does so only for partners of the opposite sex, and in husband-wife or unmarried-partner families the male is always identified as the family head. Relatives of both partners living in the same household are included as well, as are foster children and, in some circumstances, other persons identified as non-relatives of the family head.[20] Another departure from the CPS family definition involves families that separated but later reunited (that is, moved back together). Where the CPS would count these as subfamilies within a single family, the PSID continues to treat them as separate families. The family incomes and poverty thresholds for these previously separated families do not reflect their common family membership.

To create CPS families from PSID families that did not conform to a CPS family definition, we had to separate the unmarried partners and combine the related subfamilies. We also had to divide or combine their family incomes and calculate new poverty thresholds that reflected the membership of each family. In addition, we had to remove foster children and other non-relatives.

Operationally, we achieved this as follows. First, we created two or more new families from each non-CPS family. An unmarried partner of the reference person was assigned to one new

---

[20] The inclusion of these additional non-relatives does not follow a fixed set of rules, although survey documentation suggests that there may be patterns. For example, an unmarried, same-sex partner is included in the family as a non-relative.

103

family along with anyone identified as the partner's relative. The remaining family members were assigned to the second new family. Foster children and other non-relatives of the family reference person were dropped from the sample, rather than assigned to separate families, because their records contained no personal income data.

Next, the total family income of each non-CPS family had to be distributed among its subsidiary CPS families. Because of restrictions on the income data available, there was little choice about how to do this. The PSID provides some person-level income data for the family head and wife/partner, but certain other components are shared between them. In addition, incomes for all other family members are combined while Social Security is reported as a single amount for the entire family. For the family head, the income components reported are farm income, labor income from unincorporated businesses, asset income from unincorporated businesses, and labor income from employers. For the wife/partner the components are labor income from unincorporated businesses, asset income from unincorporated businesses, and labor income from employers. We assigned the head's income to the head and the wife/partner's income to the partner.

The combined asset income of the head and wife/partner from sources other than their respective unincorporated businesses can be calculated by subtracting their individual incomes, as we have just described them, from an amount identified as the taxable income of head and wife/partner. Their combined transfer income, except for Social Security, is reported in a single variable as well. We divided these two sources evenly between the head and partner.

The total taxable income and transfer income (except for Social Security) of all other family members is reported in two additional fields. If one of the two partners had no family members while the other had at least one, then the partner with the family member received all of the

income reported for other family members.[21] Otherwise, we divided this additional income in proportion to the number of other family members in each family. Thus if the unmarried partner had one other family member while the family head had two, then the family head received two-thirds of the income recorded for other family members.

Lastly, as we have noted, the combined Social Security income of all family members is reported in a single field. If one and only one partner was 62 or older, we assigned all of the Social Security income to that partner. If both partners or neither partner was 62 or older, we divided the Social Security income evenly between them. This completed the apportionment of total family income between the family of the head and the family of the partner.

Related subfamilies living in the same household but treated as separate families can be identified by fields on their respective records. When combining two or three separate families into a single family, we designated the head of the family with the largest total family income as the head of the combined family.[22] If the incomes of the separate families were identical, we designated the head of the family with the smaller (or smallest) family ID as the head of the combined family. We summed the family incomes of the two or three separate families to create a family income for the combined family.

If a family member was present for only part of the income reference year, or if another person who was no longer with the family at the time of the interview was present for part of the income reference year, the poverty threshold for that family will reflect the number of months

---

[21] We relied on the relationship code to determine which other family members were related to the head and which family members were related to the "wife" (or partner). A shared child would have been assigned in this code to either the head or wife.

[22] Designating one family head as the combined family head was necessary for subsequent tabulations that counted only heads of families.

that those persons were present. Likewise, their incomes will be included in the family's annual income only for those months that they lived with the family.

To account for part-year family members when separating or combining families, we first determined for every family the difference between the poverty threshold reported on the file and the poverty threshold that we would obtain using the reported family size, the number of related children under 18, and whether the head was 65 or older. We defined this difference as the contribution of part-year members to the family poverty threshold. If we separated the families of a head and partner, we assigned this difference to the family of the head. If we combined two or more related families, we summed the values of this difference over the families. When we determined the poverty threshold for a new CPS family, then, we added the value of this difference to the result. Any income received by part-year family members during their period of co-residence with a PSID family would have been included in one of the components discussed above, so there was no need to estimate it separately.

c. Comparison of Living Arrangements

Even with the application of a common family concept across the five general population surveys, we find differences in the distribution of living arrangements, which are difficult to explain.

After breaking up the non-CPS families in NHIS, we obtain a total of 123.21 million families, which is about 0.7 million more than the CPS (Table III.3).[23] Because NHIS counts college students where they are living at the time of the interview, the difference ought to be even greater. Earlier we attributed a 2.2 million shortfall of persons in the ACS to the exclusion of college dormitories and other non-institutional group quarters from the sample frame. NHIS,

---

[23] As noted in the table, families include individuals living alone or with non-relatives.

on the other hand, includes college dormitories in its frame and should be counting the residents of such facilities as unrelated individuals for nine months out of the year, whereas the CPS counts them as members of their parents' families. However, even with the splitting of unmarried partners we find 1.7 million fewer adult singles (18 and older) in NHIS than in the CPS. This is offset by 5.3 million more married persons in NHIS than the CPS, yet the numbers of married persons ought to be very similar between the two surveys. We are not able to explain this divergence. Rather, we can only suggest that it may stem from differences in the nonresponse adjustments or, more generally, the weighting procedures applied in the two surveys. For example, the NHIS weights do not incorporate a direct adjustment for nonresponding families in responding households, and we suspect that the missing families are primarily single young adults. Post-stratification to population totals may shift the family composition by compensating for too few young adults.

The similar family counts among the CPS, NHIS and ACS suggest that SIPP, with 120.3 million, is at least 2 million too low while MEPS, at 130.90 million, is more than 8 million too high. The excess families in MEPS are especially baffling, as its sample is drawn from responding families in NHIS. We see that large difference between MEPS and the CPS occur in the number of singles, where MEPS is 3.5 million higher than CPS (and 5.2 million higher than NHIS); the number of married childless persons, where MEPS is 3.2 million higher than CPS (but 0.9 million lower than NHIS); and the number of married persons with children, where MEPS is 3.4 million higher than CPS (but only 0.4 million higher than NHIS). In Chapters IV and V we raise the possibility that post-stratification of the MEPS person weights to the CPS poverty distribution may play a role. Differences in the numbers of families would not exist with the MEPS family weights, which are post-stratified to CPS family counts, but as noted earlier, our analysis requires the estimation of person-level characteristics and, therefore, person weights.

TABLE III.3

LIVING ARRANGEMENTS OF PERSONS: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Millions of Families		
All Families	122.48	122.66	120.33	130.90	123.21
			Millions of Persons		
All Persons	282.55	277.69	281.08	283.30	283.71
Living Arrangements					
Single, 18 or older[a]	46.91	47.72	45.24	50.40	45.19
Married, childless	63.85	63.27	63.75	66.46	67.33
Single parent	12.36	12.75	12.79	12.76	12.22
Children of single parents	20.00	20.73	21.16	19.98	19.31
Married, with children	51.89	50.75	51.73	54.12	53.70
Children of married couples	48.89	46.66	48.23	50.18	49.59
All other	38.66	35.83	38.16	29.39	36.37
			Percent of the Population		
All Persons	100.0	100.0	100.0	100.0	100.0
Living Arrangements					
Single, 18 or older[a]	16.6	17.2	16.1	17.8	15.9
Married, childless	22.6	22.8	22.7	23.5	23.7
Single parent	4.4	4.6	4.6	4.5	4.3
Children of single parent	7.1	7.5	7.5	7.1	6.8
Married, with children	18.4	18.3	18.4	19.1	18.9
Children of married couples	17.3	16.8	17.2	17.7	17.5
All other	13.7	12.9	13.6	10.4	12.8

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2002 ACS, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: In the Census Bureau surveys, families include primary families, nonfamily householders, unrelated subfamilies and unrelated (secondary) individuals. Children are under 18.

[a] Single means living with no relatives. Persons classified as single may be living with non-relatives, including an unmarried partner.

These differences among the surveys in their estimates of living arrangements raise an important point. If major household surveys cannot agree on something as fundamental as the number of people living alone or with only non-relatives or the number living with spouses, what does this say about their comparative estimates of more complex phenomena? Furthermore, how do differences in the distribution of living arrangements affect estimates of other characteristics?

Standardizing on the distribution of living arrangements was not a part of the design of our study, but differences across surveys may have implications for estimates of the poor, or the uninsured, or other subpopulations of policy interest.

4. Limitations

While the goal of these efforts was to make the survey estimates of income as comparable as possible, there remain a number of differences due to design or methodological features for which we can make no adjustment. Estimates of the potential impact of some of these differences are presented in Chapter V.

First, our adjustments do not compensate for the fact that the surveys represent populations at different times. The MEPS and SIPP estimates represent populations in December 2002 while the CPS represents a March 2003 population. Both the ACS and NHIS represent an average of populations over a calendar year. The 2002 ACS is weighted to July 1, 2002 while the four segments of the 2003 NHIS are weighted, separately, to February 1, May 1, August 1, and November 1 of that year. Weights for the four segments are combined to create a single annual weight on the public use file, with an effective reference date of mid-June 2003. The PSID interviewed families between March and November 2003, with most of the interviews conducted in April and May. When families were interviewed will determine who was in the universe, but the data were weighted to CPS population controls for March 1, 2003.

Differences in survey timing affect population estimates as follows. If survey B is conducted three months later than survey A, then survey B will exclude people who died in the interval between them, and it will exclude people who, while still living, have left the survey universe as we have defined it by moving outside the country, becoming institutionalized, joining the armed forces, or, if under 15, moved into a household of unrelated persons. Conversely, survey B will include people who have been born or otherwise joined the common universe—for example, by

moving (back) into the country, being released from an institution, being discharged from the military, or, for those who were unrelated children under 15 at the time of survey A, have turned 15, been adopted, or otherwise joined a related family.

Second, our adjustments do not compensate for differences in the source of the population controls that were applied or how they were applied, including possible inconsistencies between how the survey post-strata and the post-stratum totals were defined. Some surveys use Census Bureau population estimates directly; others (except for MCBS) use CPS population estimates, or perform their own calculations of control totals based on CPS data. These alternative controls do not agree completely. The different sources of control totals (Census Bureau versus staff or survey contractor calculations) may explain why the MEPS population prior to the application of our adjustments exceeds the December 2002 SIPP population by nearly 500,000 even though MEPS was post-stratified to December 2002 controls.

Third, the adjustments do not correct for the differential treatment of college students living away from home while attending school. These differences affect the size of the ACS population and the composition of ACS, NHIS, and PSID families.

Fourth, the adjustments to the income concept do not address differences on how respondents interpreted what they were to include or not include in the income they reported.

Fifth, the adjustments to family composition involve assumptions about the allocation of family income among the families created by dividing each non-CPS family.

B. STANDARD TABULATIONS AND ANALYSES

After making the adjustments for comparability detailed above, we produced a set of standard tabulations across the surveys that provide the basis for the comparative analysis of income data presented in Chapter IV. Separate sets of standard tables dealing with income

allocation and rounding provide the material for a separate analysis of these topics in Chapter VI.[24]

1. Standard Tabulations by Family Income

A common set of tabulations by categories of family income was prepared for each of the five general populations surveys and the PSID. Tables III.4 and III.5 depict the first two tables from the standard tabulations by family income. The pair of tables illustrates the tabulations for the full universe of persons classified, first, by poverty relative (III.4) and, in the second table, by quintile of family income (III.5). This pair of tables was repeated for eight subpopulations:

- Persons receiving SSI

- Persons in families with welfare and/or Food Stamps

- Persons enrolled in Medicaid or SCHIP in the prior calendar year

- Persons currently enrolled in Medicaid or SCHIP (or, for SIPP and MEPS, persons enrolled in a specified month)

- Persons never insured in the prior calendar year

- Persons currently uninsured

- Persons with earned income in the prior calendar year

- Persons with wage and salary income in the prior calendar year

For the NHIS and PSID, which do not provide total income for each person, the tabulation of aggregate income in each pair of tables was replaced by a single line tabulation of aggregate family income.

---

[24] The NHIS has five files of independently calculated income imputations. We performed each analysis five times, using each of the five files in turn, and averaged the results. This procedure was followed for all tabulations on NHIS.

TABLE III.4

TABLE SHELL, POVERTY RELATIVES: ALL PERSONS

| | Sample Size | Millions of Persons by Family Income as % of Poverty | | | | |
		<100 %	100-<200%	200-<400%	400%+	Total
All Persons						
Gender						
Male						
Female						
Race/Ethnicity						
White, non-Hispanic						
Black, non-Hispanic						
Hispanic						
Age						
<18						
18-64						
65+						
62+						
Family composition						
Singles (age 18 or older)						
Childless couples[a]						
Single parents with children[b]						
Children in single-parent families						
Husband-wife families with children[a, b]						
Children in husband-wife families						
Health status fair or poor						
With inpatient stay						

Source: Mathematica Policy Research.

[a] Tabulations count only the heads of families, with related subfamilies counted separately from primary families.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

TABLE III.4 *(continued)*

	Sample Size	CY 2002 Income ($Billions) by Family Poverty Level				
		<100 %	100-<200%	200-<400%	400%+	Total

All Persons

Gender
  Male
  Female

Race/Ethnicity
  White, non-Hispanic
  Black, non-Hispanic
  Hispanic

Age
  <18
  18-64
  65+
  62+

Family composition
  Singles (age 18 or older)
  Childless couples[a]
  Single parents with children[a, b]
    Children in single-parent families
  Husband-wife families with children[a, b]
    Children in husband-wife families

Health status fair or poor

With inpatient stay

Source: Mathematica Policy Research.

Note: Income by gender, race/ethnicity, age, and health status is the sum of total personal income. Income by family composition is the sum of total family income with related subfamily income included only in the primary family.

[a] Family income is tabulated only for heads of families, with heads of related subfamilies excluded.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

TABLE III.5

TABLE SHELL, QUINTILES: ALL PERSONS

	Sample Size	Millions of Persons by Family Income Quintile					
		Lowest	Second	Third	Fourth	Highest	Total
All Persons							
Gender							
Male							
Female							
Race/Ethnicity							
White, non-Hispanic							
Black, non-Hispanic							
Hispanic							
Age							
<18							
18-64							
65+							
62+							
Family composition							
Singles (age 18 or older)							
Childless couples[a]							
Single parents with children[b]							
Children in single-parent families							
Husband-wife families with children[a, b]							
Children in husband-wife families							
Health status fair or poor							
With inpatient stay							

Source: Mathematica Policy Research.

[a] Tabulations count only the heads of families, with related subfamilies counted separately from primary families.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

114

TABLE III.5 *(continued)*

	Sample Size	CY 2002 Income ($Billions) by Family Income Quintile					
		Lowest	Second	Third	Fourth	Highest	Total
All Persons							
Gender							
Male							
Female							
Race/Ethnicity							
White, non-Hispanic							
Black, non-Hispanic							
Hispanic							
Age							
<18							
18-64							
65+							
62+							
Family composition							
Singles (age 18 or older)							
Childless couples[a]							
Single parents with children[a, b]							
Children in single-parent families							
Husband-wife families with children[a, b]							
Children in husband-wife families							
Health status fair or poor							
With inpatient stay							

Source: Mathematica Policy Research.

Note: Income by gender, race/ethnicity, age, and health status is the sum of total personal income. Income by family composition is the sum of total family income with related subfamily income included only in the primary family.

[a] Family income is tabulated only for heads of families, with heads of related subfamilies excluded.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

These tabulations focus on total income, whether for the population as a whole or, more importantly, within poverty level or quintile of family income. There are other ways to approach the comparison of survey estimates of income—for example, by highlighting recipiency, or examining other aspects of the distribution of income than those that we have chosen, or even applying each survey to a set of illustrative policy analyses. An alternative approach would have given us information on different aspects of the comparative quality of income data collected in the eight surveys. Ideally, with more time and resources, we would have taken multiple approaches. Given the limits on the study's scope, we feel that our selected approach, which was

supported by the TAG, yielded a broad range of findings that greatly enhance our understanding of income data across surveys.

2. Standard Tabulations for Restricted Populations

Neither the HRS nor the MCBS could support the full set of tabulations, so we prepared a more limited standard tabulation for each.

3. Tabulations of Income Allocation and Rounding

Two sets of standard tables showing the frequency of allocation and its contribution to total dollars were prepared for the five general population surveys, which provided full identification of allocated amounts. The first set of tables presented estimates of allocation by source of income (Tables III.6 and III.7), and the second set presented estimates of allocation by demographic characteristics (Tables III.8 and III.9).

Estimates of the frequency of rounding were prepared for the five general population surveys and the PSID. Table III.10 illustrates the first of six tables. This table presents estimates of rounding for total family income. Column one provides estimates of rounding based on reported amounts while column two provides estimate for allocated amounts. Additional tables were prepared for personal earnings, personal wages and salaries, Social Security, retirement income, and total personal income. NHIS collects only total family income and personal earnings, so the tables for that survey were limited to these two sources. The PSID tables included five sources besides total family income, but they were specific to that survey.

116

TABLE III.6

TABLE SHELL, ALLOCATED INCOME BY FAMILY INCOME AS A PERCENT OF POVERTY BY SOURCE

Source of Income	Sample Size	Persons by Family Income as % of Poverty				
		<100 %	100-<200%	200-<400%	400%+	Total

Persons with Income with Allocations (Thousands)

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

Percentage of Persons with Income with Allocations

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

Total Persons with Income (Thousands)

Total Income
Wages and Salaries
Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Asset Income
  Negative Asset Income
  Non-negative Asset Income
Social Security or Railroad Retirement
SSI
Welfare
Pensions

Source: Mathematica Policy Research.

117

TABLE III.6 *(continued)*

Source of Income	Sample Size	Total Amounts for CY 2002 by Family Income as % of Poverty				
		<100%	100-<200%	200-<400%	400%+	Total

Allocated Amount (Millions)

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

Percentage of CY2002 Income Allocated

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

Total Income (Millions)

Total Income
Wages and Salaries
Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Asset Income
  Negative Asset Income
  Non-negative Asset Income
Social Security or Railroad Retirement
SSI
Welfare
Pensions

Source: Mathematica Policy Research.

TABLE III.7

TABLE SHELL, ALLOCATED INCOME BY FAMILY INCOME QUINTILE BY SOURCE

	Sample Size	Persons by Family Income Quintile					
Source of Income		Lowest	Second	Third	Fourth	Highest	Total

<center>Persons with Income with Allocations (in Thousands)</center>

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

<center>Percentage of Persons with Income with Allocations</center>

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

<center>Total Persons with Income (in Thousands)</center>

Total Income
Wages and Salaries
Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Asset Income
  Negative Asset Income
  Non-negative Asset Income
Social Security or Railroad Retirement
SSI
Welfare
Pensions

Source: Mathematica Policy Research.

TABLE III.7 *(continued)*

Source of Income	Sample Size	Total Amounts for CY 2002 by Family Income Quintile					
		Lowest	Second	Third	Fourth	Highest	Total

Allocated Amount (Millions)

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

Percentage of CY2002 Income Allocated

Any Allocated Income (Total Income)
Allocated Wages and Salaries
Allocated Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Allocated Asset Income
  Negative Asset Income
  Non-negative Asset Income
Retirement
Allocated SSI
Allocated Welfare
Allocated Pensions

Total Income (Millions)

Total Income
Wages and Salaries
Self-Employment
  Negative Self-employment Income
  Non-negative Self-employment Income
Asset Income
  Negative Asset Income
  Non-negative Asset Income
Social Security or Railroad Retirement
SSI
Welfare
Pensions

Source: Mathematica Policy Research.

TABLE III.8

TABLE SHELL, ALLOCATED INCOME BY FAMILY INCOME AS A PERCENT OF POVERTY BY CHARACTERISTICS

Characteristic	Sample Size	Persons by Family Income as % of Poverty				
		<100 %	100-<200%	200-<400%	400%+	Total
Persons with Income with Allocations (Thousands)						
All Persons						
White, non-Hispanic						
Black, non-Hispanic						
Hispanic						
Under age 18						
Age 18-64						
Age 65 and older						
Single parents with children[b]						
Husband-wife families with children[a, b]						
Health status fair or poor						
Never insured prior calendar year						
Currently uninsured						
Medicaid/SCHIP in prior calendar year						
Currently covered by Medicaid/SCHIP						
With SSI						
In family with Welfare and/or Food Stamps						
Percentage of Persons with Income Allocations						
All Persons						
White, non-Hispanic						
Black, non-Hispanic						
Hispanic						
Under age 18						
Age 18-64						
Age 65 and older						
Single parents with children[b]						
Husband-wife families with children[a, b]						
Health status fair or poor						
Never insured prior calendar year						
Currently uninsured						
Medicaid/SCHIP in prior calendar year						
Currently covered by Medicaid/SCHIP						
With SSI						
In family with Welfare and/or Food Stamps						
Total Persons with Income (Thousands)						
All Persons						
White, non-Hispanic						
Black, non-Hispanic						
Hispanic						
Under age 18						
Age 18-64						
Age 65 and older						
Single parents with children[b]						
Husband-wife families with children[a, b]						
Health status fair or poor						
Never insured prior calendar year						
Currently uninsured						
Medicaid/SCHIP in prior calendar year						
Currently covered by Medicaid/SCHIP						
With SSI						
In family with Welfare and/or Food Stamps						

Source: Mathematica Policy Research.

[a] Tabulations of persons count only the heads of families, with related subfamilies counted separately from primary families.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

TABLE III.8 *(continued)*

Characteristic	Sample Size	Total Amounts for CY 2001 by Family Income as % of Poverty				
		<100%	100-<200%	200-<400%	400%+	Total

Allocated Amount (Billions)

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Percentage of CY2002 Income Allocated

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Total Income (Billions)

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Source: Mathematica Policy Research.

[a] Tabulations of persons count only the heads of families, with related subfamilies counted separately from primary families.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

TABLE III.9

TABLE SHELL, ALLOCATED INCOME BY FAMILY INCOME QUINTILE BY CHARACTERISTICS

Characteristic	Sample Size	Persons by Family Income Quintile					
		Lowest	Second	Third	Fourth	Highest	Total

Persons with Income with Allocations (Thousands)

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Percentage of Persons with Income with Allocations

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Total Persons with Income (Thousands)

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Source: Mathematica Policy Research.

[a] Tabulations of persons count only the heads of families, with related subfamilies counted separately from primary families.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

TABLE III.9 *(continued)*

Characteristic	Sample Size	Total Amounts for CY 2001 by Family Income Quintile					
		Lowest	Second	Third	Fourth	Highest	Total

Allocated Amount (Billions)

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Percentage of CY2002 Income Allocated

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Total Income (Billions)

All Persons
White, non-Hispanic
Black, non-Hispanic
Hispanic
Under age 18
Age 18-64
Age 65 and older
Single parents with children[b]
Husband-wife families with children[a, b]
Health status fair or poor
Never insured prior calendar year
Currently uninsured
Medicaid/SCHIP in prior calendar year
Currently covered by Medicaid/SCHIP
With SSI
In family with Welfare and/or Food Stamps

Source: Mathematica Policy Research.

[a] Tabulations of persons count only the heads of families, with related subfamilies counted separately from primary families.
[b] Children are restricted to own, never-married children under 18 within the same family or subfamily.

TABLE III.10

TABLE SHELL, ROUNDING OF TOTAL FAMILY INCOME

Income Amount	Weighted Number of Families	
	Reported Income	Allocated Income
Less than zero		
Zero		
> 0 to < $2,500		
$2,500 to < $5,000		
$5,000		
> $5,000 to < $7,500		
$7,500 to < $10,000		
$10,000		
> $10,000 to < $12,500		
$12,500 to < $15,000		
$15,000		
> $15,000 to < $17,750		
$17,500 to < $20,000		
$20,000		
> $20,000 to < $22,500		
$22,500 to < $25,000		
$25,000		
> $25,000 to < $27,750		
$27,500 to < $30,000		
$30,000		
> $30,000 to < $32,500		
$32,500 to < $35,000		
$35,000		
> $35,000 to < $37,750		
$37,500 to < $40,000		
$40,000		
> $40,000 to < $42,500		
$42,500 to < $45,000		
$45,000		
> $45,000 to < $47,750		
$47,500 to < $50,000		
$50,000		
> $50,000 to < $52,500		
Total (excluding <= zero)		
Exactly divisable by $5,000		
Number		
Percent of total		
Exactly divisable by $10,000		
Number		
Percent of total		
Total families with income		
Percent with $1 to < $52,500		

Source: Mathematica Policy Research.

4. Data Sources

The data files used to generate the tabulations described above were the 2003 CPS ASEC supplement, the 2002 ACS, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 NHIS, the 2003 PSID, the 2004 HRS, and the 2003 MCBS Cost and Use File. All but the 2003 MCBS are public use files, but NHIS income amounts are available only on an internal file that is restricted to on-site tabulations with prior approval and usage fees. We submitted a proposal, as required, to obtain access to the NHIS income data. We carried out our work with these data at the NCHS Research Data Center, located at NCHS headquarters in Hyattsville, Maryland, for which the study paid user fees. MCBS has no public use files, but allows protected off-site use with approval and has a standing agreement with ASPE, under which the study operated.

C. SPECIALIZED TABULATIONS

A set of specialized tabulations focused on survey design and definitional issues and on internal consistency. These tabulations were survey specific.

1. Tabulations Addressing Specific Design and Definitional Issues

To measure the impact of definitional differences between other surveys and the CPS, tabulations were produced to examine:

- Including lump sums or irregular retirement account withdrawals in income (SIPP and MEPS)
- Treating unmarried cohabiting individuals as a family rather than as unrelated individuals (MEPS and NHIS)
- Treating unrelated subfamily members as unrelated individuals (performed on the CPS to simulate ACS)

In addition, we used the flexibility afforded by monthly income, family composition and weights in SIPP to examine the impact of differences in the timing of family composition relative to the

income reference period. We replicated the timing of family composition relative to the income reference period in the surveys (contemporaneous, end of year, following March, month by month up to 12 months after end of income year) and measure the resulting differences in the distribution of poverty status. Making these comparisons within a single data set ensures that the results reflect only the differences between methodologies and not differences in data.

2.   Tabulations Addressing Specific Consistency Issues

Internal consistency emerged as an issue during our analysis of employment in MEPS, leading us to specify additional tabulations to examine this issue explicitly. Tabulations focused on work activity and reported earnings (MEPS and SIPP) and on reported receipt versus reported dollars of earned income (NHIS). We also examined consistency between total family income and the sum of personal earnings in the NHIS.

IV. STANDARDIZED EMPIRICAL COMPARISONS

In this chapter we present findings based primarily on the extensive standardized empirical comparisons that were described in the preceding chapter. These findings, which cover all eight surveys, include our principal comparative estimates of income. For these estimates the study uses income data for 2002 (HRS and MCBS income for 2003 were deflated with the CPI-U) that covers a calendar year, except for the rolling reference period in ACS. We compare the survey estimates of income along several dimensions, as no single measure captures the full breadth of what good income data should provide. We look in turn at aggregate income and its distribution by quintile, the location of quintile boundaries, per capita income by quintile, estimates of the poor and near poor, employment and earned income, unearned income, and program participation. These comparisons focus on the five general population surveys that are conducted by the federal government and designed to provide representative estimates of the full civilian noninstitutional population: the CPS, ACS, SIPP, MEPS, and NHIS. More limited comparisons that include the PSID are interspersed among the findings on the five surveys. Comparative estimates of the population without health insurance coverage—the uninsured—and its distribution by income are presented for the CPS, SIPP, MEPS, NHIS, and PSID. Separate analyses of the two surveys of restricted populations—that is, the HRS and MCBS—are included near the end of the chapter and followed by an analysis of internal inconsistencies relevant to the income data in the NHIS, MEPS, and SIPP.

The purpose of the comparisons presented in this chapter is not to establish statistically significant differences or demonstrate that alternative estimates are statistically the same. Our efforts to adjust for differences in universe, income concept, and family definition, and our earlier findings on living arrangements underscore the importance of nonsampling error in

comparative estimates across surveys. Furthermore, the surveys included in the study have large

samples, for the most part, which means that small differences may be statistically significant yet

unimportant from a policy perspective. Given these considerations, we felt that our fixed

resources were better spent in furthering our understanding of the surveys, the differences we

were observing, and the impact of various design features than in calculating statistics that would

provide only marginal value-added at best.

A.   AGGREGATE INCOME

As a summary statistic, the weighted total or aggregate income is appealing for its simplicity

and its use of all the income data collected by each survey, but its value is heavily dependent on

the amount of income captured from the upper end of the income distribution, which holds the

least interest for policy analysis. In presenting estimates of aggregate income, we include a

breakdown by quintile, which enables us to compare the surveys with respect to their collection

of income from different segments of the distribution. We also examine per capita income, which

is calculated by dividing the estimate of aggregate income by population size. This corrects the

aggregate estimates for slight differences in the size of the population represented by each survey

after the adjustment to a common universe described in Chapter III. The estimates of per capita

income are presented by quintile as well.

1.   Aggregate Income by Quintiles

Estimates of aggregate income, for the whole population and broken down by quintile of

family income, are presented in Table IV.1 for the five general population surveys. In addition to

the dollar amounts, the table presents the estimated amounts as a percentage of the corresponding

amounts for the CPS. While the CPS does not represent the gold standard for estimates of

income, and we do not mean to suggest that the CPS estimates are the best, the CPS *is* the

TABLE IV.1

AGGREGATE INCOME BY QUINTILE OF FAMILY INCOME: FIVE SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Billions of Dollars		
Aggregate Income, All Persons	6,468.4	6,346.3	5,766.2	6,257.7	6,116.2
Family Income Quintile					
Lowest	370.5	368.7	391.4	360.0	313.7
Second	774.1	778.4	750.8	808.4	717.7
Third	1,090.2	1,087.4	1,008.8	1,144.7	1,058.4
Fourth	1,446.8	1,415.8	1,307.2	1,461.8	1,420.7
Highest	2,786.7	2,696.0	2,308.0	2,483.0	2,605.8
Sum through Four Quintiles	3,681.7	3,650.3	3,458.2	3,774.7	3,510.4
			Percent of CPS		
Aggregate Income, All Persons	100.0	98.1	89.1	96.7	94.6
Family Income Quintile					
Lowest	100.0	99.5	105.6	97.2	84.7
Second	100.0	100.6	97.0	104.4	92.7
Third	100.0	99.7	92.5	105.0	97.1
Fourth	100.0	97.9	90.3	101.0	98.2
Highest	100.0	96.7	82.8	89.1	93.5
Sum through Four Quintiles	100.0	99.1	93.9	102.5	95.3

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

official source of household income and poverty statistics for the U.S., so expressing other survey estimates of income as a percentage of the CPS provides a useful standardization.

Aggregate income ranges from $5.77 trillion in the SIPP to $6.47 trillion in the CPS—a difference of nearly 11 percent. The other three surveys produce estimates that lie within 2 to 5 percent of the CPS. Aggregate income is $6.35 trillion in the ACS, $6.26 trillion in MEPS, and $6.12 trillion in NHIS. Aggregates in the top quintile may be affected by outliers and by differences in survey practice with respect to the topcoding of public use data, documented in Chapter II. For example, the CPS assigns the means of topcoded values as their respective topcodes, which preserves overall means and totals, but not all surveys do this for all income items. For this reason, we summed the survey aggregates through the bottom four quintiles.[25] For every survey, the four-quintile sum is closer to the CPS estimate than is the full aggregate, with the MEPS total exceeding the CPS by 2.5 percent. The SIPP total moves to within 1.5 percent of the NHIS total but is still 6 percent below the CPS.

When we examine the results by quintile of family income, we find that SIPP obtains the most income from the lowest quintile, at 105.6 percent of the CPS total. SIPP's apparent success in collecting income data from the low end of the income distribution begins to erode noticeably by the second quintile, however. In that quintile, SIPP collects 97 percent as much total income as the CPS. This drops to 92.5 percent by the third quintile, 90.3 percent by the fourth and 82.8 percent in the top quintile. MEPS aggregates exceed the corresponding CPS amounts for quintiles two through four while the ACS aggregates lie within a percent of the CPS aggregates (both above and below) through the first three quintiles before dropping to 98 and 97 percent of the CPS in the fourth and fifth quintiles.

---

[25] Within each survey, each of the five quintiles contains the same number of people (weighted) except when the numbers are affected by heaping at quintile boundaries.

This is only the first of numerous tables, and it examines only one dimension of income, but it presents several striking findings that raise fundamental questions about the collection of income data. One such finding is that with a single question NHIS captures 95 percent as much total income as the CPS, despite the latter's sizable battery of income questions and its status as the official source of income and poverty estimates for the U.S. Second, with far more income questions than any of the other four surveys, SIPP captures 11 percent less total income than the CPS and 6 percent less than the NHIS's single question. Third, with its massive sample size and an instrument that is filled out primarily by respondents working without the assistance of a trained interviewer, the ACS nevertheless manages to approximate the CPS more closely than any other survey. Fourth, the MEPS person weights used to prepare the estimates in Table IV.1 were post-stratified to CPS totals by demographic characteristics and the distribution of income relative to poverty. What impact does this have on the MEPS estimates of aggregate income? Would MEPS, with its SIPP-like panel design, yield SIPP-like income estimates in the absence of this post-stratification, or does the use of retrospective annual versus monthly income questions trump the panel design?

More generally, what do these findings say about the collection of income data? Does the strategy of asking respondents about their incomes over the prior calendar year or even the past twelve months have a bigger impact on the amount of income collected than the level of detail that is incorporated into the questions? It will become clear as we progress through this chapter that the limitations of a single-question approach are indeed numerous, but this is a separate issue from the retrospective approach. We also have to ask if the SIPP approach of collecting income at four-month intervals and compiling annual totals month by month is inherently inferior, or whether the other surveys share a common upward bias that arises from their retrospective approach. These are compelling questions, and as we walk through the rest of the findings in this

chapter it will become apparent that there are areas in which SIPP clearly excels. Nevertheless, we will also see that outside of these exceptions, SIPP's estimates of income are consistently low.

2. The Distribution of Income

The boundaries between quintiles (that is, the dollar values of the 20th, 40th, 60th, and 80th percentiles) are themselves informative about the distribution of total family income in each of the surveys. These percentile points are rather similar for the CPS, ACS, and MEPS, but the SIPP quintile boundaries start above the CPS and decline progressively from there (Table IV.2). The NHIS boundaries remain at 92 to 93 percent of the CPS values through the 60th percentile but then rise to nearly 98 percent for the 80th percentile.

The ratio of the 80th to the 20th percentile provides a measure of inequality across the income distribution. The higher the ratio, the more unequally family income is distributed. Given the similarity of their quintile values, the ratios for the CPS, ACS and MEPS are very similar as well. Ratios for the latter two surveys are 97 percent of the CPS ratio of 4.56. The SIPP ratio is much lower at 3.96 or 87 percent of the CPS ratio, reflecting the progressive decline of the SIPP quintiles relative to the CPS values. The NHIS ratio, however, is 6 percent higher than the CPS at 4.83 because the 80th percentile in the NHIS income distribution is relatively higher than the 20th percentile when compared to the CPS.

We obtain similar but more complex findings if we compare per capita income by quintile across the five surveys. Using the ratio of per capita incomes between the top and bottom quintiles as our measure of income dispersion, we find that ACS is just two percentage points below the CPS with a ratio of 7.44 versus 7.57 (Table IV.3). MEPS is now markedly lower with a ratio of 6.90 or 91 percent of the CPS value. SIPP continues to have the lowest ratio at 5.90 or

TABLE IV.2

FAMILY INCOME QUINTILE BOUNDARIES: FIVE SURVEYS

Quintile Boundaries	CPS	ACS	SIPP	MEPS	NHIS
Percentile Value					
20 %-ile	20,000	20,191	20,672	19,670	18,443
40 %-ile	37,051	37,656	35,870	37,214	34,584
60 %-ile	59,133	58,453	54,328	58,000	55,000
80 %-ile	91,207	89,548	81,785	87,338	89,068
Ratio of 80th to 20th %-ile	4.56	4.44	3.96	4.44	4.83
			Percent of CPS		
Percentile Value					
20 %-ile	100.0	101.0	103.4	98.4	92.2
40 %-ile	100.0	101.6	96.8	100.4	93.3
60 %-ile	100.0	98.9	91.9	98.1	93.0
80 %-ile	100.0	98.2	89.7	95.8	97.7
Ratio of 80th to 20th %-ile	100.0	97.3	86.8	97.4	105.9

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

TABLE IV.3

AVERAGE INCOME PER CAPITA BY QUINTILE OF FAMILY INCOME: FIVE SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS	NHIS
All Persons	22,893	22,854	20,514	22,089	21,558
Family Income Quintile					
Lowest	6,513	6,526	6,962	6,352	5,528
Second	13,789	14,259	13,355	14,269	12,649
Third	19,293	19,576	17,946	20,052	18,493
Fourth	25,604	25,496	23,250	25,976	25,151
Highest	49,316	48,543	41,062	43,855	46,114
Ratio of fourth to lowest	3.93	3.91	3.34	4.09	4.55
Ratio of highest to lowest	7.57	7.44	5.90	6.90	8.34
			Percent of CPS		
All Persons	100.0	99.8	89.6	96.5	94.2
Family Income Quintile					
Lowest	100.0	100.2	106.9	97.5	84.9
Second	100.0	103.4	96.8	103.5	91.7
Third	100.0	101.5	93.0	103.9	95.9
Fourth	100.0	99.6	90.8	101.5	98.2
Highest	100.0	98.4	83.3	88.9	93.5
Ratio of fourth to lowest	100.0	99.4	84.9	104.0	115.7
Ratio of highest to lowest	100.0	98.2	77.9	91.2	110.2

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

136

only 78 percent of the CPS ratio. By contrast, the NHIS ratio of 8.34 is 10 percent above the CPS ratio.

Like aggregate income, per capita income in the top quintile is affected by outliers and topcoding, so we also calculated the ratio of per capita incomes between the fourth and lowest quintiles. Here the patterns are more similar to what we saw with the ratio of the 80th to the 20th percentile, yet there are notable differences. First, the NHIS ratio exceeds the CPS ratio by an even larger amount, being 16 percent higher at 4.55 versus 3.93 for the CPS. In all cases the NHIS results are driven by a very low per capita income in the bottom quintile (and a low 20th percentile). Large ratios result despite the fact that the upper quintiles and percentiles never match the CPS. The MEPS ratio is also higher than the CPS ratio in this case—by 4 percent. The ACS ratio is 99 percent of the CPS ratio, while the SIPP ratio is 85 percent of the CPS ratio.

Overall, then, we see somewhat greater inequality in the income distribution in the NHIS than the CPS and lower inequality in the SIPP. The ACS matches the CPS very closely while the estimates for MEPS show less, about the same, or more inequality than the CPS depending on the ratio we calculate.

3. Income in the PSID

To assess the reporting of income in the PSID in comparison with other surveys, we replicated the tables above for the PSID and the three Census Bureau surveys. As we explained in Chapter III, the application of preliminary cross-sectional weights to the PSID yields an estimated population that falls short of the CPS population by 21 million. In part this is due to an omission of unrelated subfamilies and secondary individuals from the CPS-based control totals that were used to post-stratify the PSID weights. In addition to reducing the weighted number of persons, this omission from the PSID is likely to have an effect on the distribution of income because singles—who tend to have lower income than other family units—will be

underestimated relative to the CPS. Therefore, we created an additional CPS series, labeled CPS-X in the tables, that excludes unrelated subfamilies and all secondary individuals except those who were identified as unmarried partners. In creating CPS-like families from the PSID families with unmarried partners, we separated the unmarried partners into their own families. We needed their counterparts in the CPS.

Despite 21 million fewer persons, as we noted, the PSID captures 3.9 percent more aggregate income than the CPS, or an additional $253.4 billion dollars (Table IV.4). Compared to the CPS-X series with the aforementioned exclusions, the PSID captures an additional $416.5 billion. The PSID also captures more aggregate income than the full CPS in every quintile, with the biggest difference in the top quintile, where the PSID aggregate is 105.5 percent of the full CPS aggregate.

PSID quintile boundaries are also higher than the quintiles from the full CPS, CPS-X, ACS, or SIPP. The biggest difference occurs at the 20th percentile, where the PSID value of $24,200 exceeds the corresponding full CPS value by $4,200 (or 21 percent) and exceeds the corresponding CPS-X value by $3,300 (Table IV.5). At higher percentiles, the PSID values exceed the corresponding full CPS values by 10 to 13 percent. The ratio of the 80th to the 20th percentile is 8 percent lower than that of the full CPS because the PSID exceeds the CPS by a smaller margin at the 80th percentile than the 20th percentile.

Because the PSID obtains more aggregate income than the CPS from a smaller weighted population, the differences in per capita income are even greater than the differences in aggregate income. The overall per capita income in the PSID, $25,710 is 12 percent higher than both the full CPS and CPS-X per capita incomes (Table IV.6). By quintile the differences grow from 10 percent in the lowest quintile to 14 percent in the highest quintile. Ratios of per capita

TABLE IV.4

AGGREGATE INCOME BY QUINTILE OF FAMILY INCOME:
PSID AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
			Billions of Dollars		
Aggregate Income, All Persons	6,468.4	6,346.3	5,766.2	6,721.8	6,305.2
Family Income Quintile					
Lowest	370.5	368.7	391.4	375.8	361.4
Second	774.1	778.4	750.8	798.3	755.4
Third	1,090.2	1,087.4	1,008.8	1,103.7	1,054.6
Fourth	1,446.8	1,415.8	1,307.2	1,504.9	1,414.5
Highest	2,786.7	2,696.0	2,308.0	2,939.0	2,719.3
Sum through Four Quintiles	3,681.7	3,650.3	3,458.2	3,782.8	3,585.9
			Percent of CPS		
Aggregate Income, All Persons	100.0	98.1	89.1	103.9	97.5
Family Income Quintile					
Lowest	100.0	99.5	105.6	101.4	97.6
Second	100.0	100.6	97.0	103.1	97.6
Third	100.0	99.7	92.5	101.2	96.7
Fourth	100.0	97.9	90.3	104.0	97.8
Highest	100.0	96.7	82.8	105.5	97.6
Sum through Four Quintiles	100.0	99.1	93.9	102.7	97.4

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003
CPS ASEC supplement, the 2001 SIPP panel, and the 2003 PSID, and prior 12 months
income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except
unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.5

QUINTILES OF FAMILY INCOME: PSID AND CENSUS BUREAU SURVEYS

Quintile Boundaries	CPS	ACS	SIPP	PSID	CPS-X[a]
	Family Income in Dollars				
Percentile Value					
20 %-ile	20,000	20,191	20,672	24,200	20,900
40 %-ile	37,051	37,656	35,870	42,025	38,410
60 %-ile	59,133	58,453	54,328	64,996	60,162
80 %-ile	91,207	89,548	81,785	101,817	92,500
Ratio of 80th to 20th %-ile	4.56	4.44	3.96	4.21	4.43
	Percent of CPS				
Percentile Value					
20 %-ile	100.0	101.0	103.4	121.0	104.5
40 %-ile	100.0	101.6	96.8	113.4	103.7
60 %-ile	100.0	98.9	91.9	109.9	101.7
80 %-ile	100.0	98.2	89.7	111.6	101.4
Ratio of 80th to 20th %-ile	100.0	97.3	86.8	92.3	97.1

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, and the 2003 PSID, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.6

AVERAGE INCOME PER CAPITA BY QUINTILE OF FAMILY INCOME:
PSID AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
All Persons	22,893	22,854	20,514	25,710	22,975
Family Income Quintile					
Lowest	6,513	6,526	6,962	7,178	6,584
Second	13,789	14,259	13,355	15,261	13,762
Third	19,293	19,576	17,946	21,132	19,204
Fourth	25,604	25,496	23,250	28,785	25,777
Highest	49,316	48,543	41,062	56,220	49,561
Ratio of fourth to lowest	3.93	3.91	3.34	4.01	3.92
Ratio of highest to lowest	7.57	7.44	5.90	7.83	7.53
			Percent of CPS		
All Persons	100.0	99.8	89.6	112.3	100.4
Family Income Quintile					
Lowest	100.0	100.2	106.9	110.2	101.1
Second	100.0	103.4	96.8	110.7	99.8
Third	100.0	101.5	93.0	109.5	99.5
Fourth	100.0	99.6	90.8	112.4	100.7
Highest	100.0	98.4	83.3	114.0	100.5
Ratio of fourth to lowest	100.0	99.4	84.9	102.0	99.6
Ratio of highest to lowest	100.0	98.2	77.9	103.4	99.4

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, and the 2003 PSID, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

141

income between quintiles are only slightly higher than the corresponding CPS ratios, implying that inequality across the income distribution is about the same in the two surveys.

Does the PSID truly capture more income than the CPS or does the PSID sample with its current weights simply overrepresent higher income families? We cannot answer this with the data available to us. We compared distributions of selected characteristics between the PSID and the CPS and found that the PSID had proportionately fewer Hispanics and blacks and slightly more persons with college degrees, but the PSID also had proportionately more persons with less than a high school education, so the comparison was inconclusive.[26] Our conclusion at this point is that incomes in the PSID appear to run higher than in any of the other surveys, but given the nature of the PSID sample, this could easily be due to the PSID being less representative of the U.S. population as a whole than the Census Bureau surveys.

B. THE POOR AND NEAR POOR

Another useful summary statistic, but one that is informative about only the lower end of the income distribution, is the poverty rate—that is, the percentage of persons whose family incomes lie below the official poverty threshold. Estimates of the number of poor and near poor (whom we define as those between 100 and 200 percent of the poverty threshold) are important measures for policy analysis.[27] Marked differences across surveys in estimates of the poor and near poor would be a source of concern among policy analysts and other data users.

---

[26] We suspect that this higher proportion with less than a high school education is incorrect. The people we assigned to this status were identified as a residual and may include persons with missing data or otherwise undefined educational attainment.

[27] Near poor does not have a standard definition. We use the term to give a name to those who are low income (below 200 percent of poverty) but not poor. Elsewhere, near poor is sometimes used to identify persons between 100 and 125 percent of poverty.

1. **Poverty and Near Poverty in the General Population**

SIPP obtains the lowest poverty rate among the five surveys at 11.8 percent, based on an estimate of 33.2 million poor persons (Table IV.7). The CPS, ACS, and MEPS cluster very close to each other and not far from SIPP with poverty rates between 12.2 percent and 12.5 percent. As we have noted, the MEPS sample weights that we are using are post-stratified to the CPS poverty distribution, so the poverty rates for the two surveys should be identical if not for the differential effect of our universe adjustments.[28] At the high end, the NHIS is an outlier with an estimate of 41.6 million poor and a poverty rate of 14.7 percent.[29] The NHIS poverty rate is more than two percentage points higher than any of the other four surveys and nearly three percentage points higher than the SIPP.

Despite having the lowest poverty rate, SIPP exceeds all four of the other surveys in its estimate of the *near* poor. SIPP finds 20.0 percent of the population to be near poor. This is nearly two percentage points above the CPS and MEPS, *more* than two percentage points above the ACS, and one percentage point above NHIS. SIPP's estimate of 56.2 million near poor exceeds the ACS by 7.0 million and surpasses NHIS by 2.3 million.

Combining the estimates of the poor and near poor, which define the low-income population, SIPP is higher than all but NHIS with respect to both the estimated number and percentage of persons who are low-income. For the SIPP, 31.8 percent or 89.5 million persons are low-income compared to 30.5 percent or 86.2 million persons for the CPS. MEPS is somewhat higher than the CPS on both dimensions while the ACS is lower. NHIS finds 33.7 percent of the population or 95.5 million persons to be low-income. The number of persons

---

[28] Also, our independently calculated poverty status differs occasionally from the status on the MEPS public use file due to an apparent error in the algorithm used in creating the recode for the public use file.

[29] Creating CPS-like families by splitting NHIS families with unmarried partners increases the NHIS poverty rate. An estimate of the impact is reported in Chapter V.

TABLE IV.7

ESTIMATES OF THE POOR AND NEAR POOR: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Millions of Persons		
All Persons	282.55	277.69	281.08	283.30	283.71
Poverty Status					
Poor	34.38	34.61	33.25	35.35	41.58
Near Poor	51.81	49.28	56.25	52.14	53.91
Total Low Income	86.19	83.89	89.50	87.48	95.49
			Percent of the Population		
All Persons	100.0	100.0	100.0	100.0	100.0
Poverty Status					
Poor	12.2	12.5	11.8	12.5	14.7
Near Poor	18.3	17.7	20.0	18.4	19.0
Total Low Income	30.5	30.2	31.8	30.9	33.7

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold.

estimated to be low-income in the NHIS exceeds the SIPP by 6.0 million and the CPS by 9.3 million.

2.   Poverty and Near Poverty Among Children and the Elderly

SIPP's comparatively high estimates of the frequency of near-poor and low-income persons in the general population extend to children as well. SIPP finds more near-poor and low-income children than any of the other four surveys. While the estimates of children in low-income families from the CPS, ACS, and MEPS cluster between 27.4 and 28.0 million, or 38.2 to 38.9 percent, SIPP finds 30.5 million low-income children or 42.7 percent of all children (Table IV.8). NHIS is slightly lower than SIPP with 41.4 million low-income children or 29.7 percent. Furthermore, unlike the general population, where SIPP had the lowest estimate of persons in poverty, SIPP's estimate of poor children exceeds those of the ACS, MEPS, and CPS, if only marginally. NHIS finds the most poor children with a child poverty rate that exceeds the other surveys by 2 to 3 percentage points, but NHIS has no more near-poor children than CPS or MEPS. In fact, the estimates of near-poor children vary from only 14.9 to 15.4 million or 21.1 to 21.5 percent across the CPS, ACS, MEPS, and NHIS while SIPP finds 17.7 million or 24.8 percent.

The living arrangements of poor, near-poor, and low-income children are generally similar across the five surveys. Poor children are much more likely to be living in single-parent than husband-wife families while near-poor children are more likely to be living in husband-wife than single-parent families (Table IV.9). All low-income children divide almost equally between the two types of living arrangements in the CPS, SIPP, and MEPS, with single-parent families more prevalent in the ACS and husband-wife families more common in the NHIS.

When a survey shows excessive numbers of poor or near-poor children, this may reflect an income reporting problem, which may affect the distribution of living arrangements among such

TABLE IV.8

ESTIMATES OF POOR AND NEAR-POOR CHILDREN: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Millions of Persons		
All Children under 18	71.67	70.79	71.36	71.80	71.73
Poverty Status					
Poor	12.03	12.51	12.78	12.47	14.29
Near Poor	15.38	14.94	17.72	15.47	15.41
Total Low Income	27.41	27.45	30.50	27.95	29.70
			Percent of the Population		
All Children under 18	100.0	100.0	100.0	100.0	100.0
Poverty Status					
Poor	16.8	17.7	17.9	17.4	19.9
Near Poor	21.5	21.1	24.8	21.5	21.5
Total Low Income	38.2	38.8	42.7	38.9	41.4

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold.

TABLE IV.9

LIVING ARRANGEMENTS OF POOR AND NEAR-POOR CHILDREN: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
	Millions of Children Under 18				
Poor Children	12.03	12.51	12.78	12.47	14.29
In single-parent family	7.02	7.60	7.93	7.96	7.79
In husband-wife family	4.09	3.78	4.19	3.81	5.49
Not living with a parent	0.92	1.13	0.67	0.71	1.01
Near Poor Children	15.38	14.94	17.72	15.47	15.41
In single-parent family	5.87	5.70	6.69	5.45	5.48
In husband-wife family	8.74	8.37	10.49	9.53	9.24
Not living with a parent	0.77	0.87	0.54	0.49	0.68
Total: Low-income Children	27.41	27.45	30.50	27.95	29.70
In single-parent family	12.89	13.29	14.61	13.40	13.27
In husband-wife family	12.83	12.15	14.68	13.35	14.73
Not living with a parent	1.69	2.01	1.21	1.20	1.69

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold. If a child is living with both parents, but they are not married, the child is counted as as living in a single-parent family.

children. This is illustrated by the NHIS, which yields high estimates of poor children relative to the other four surveys. If we compare the living arrangements of poor children in the NHIS with those of poor children in the any of the other surveys, we find that most of the difference is due to children in husband-wife families. For example, compared to the CPS the NHIS has .77 million additional poor children in single-parent families and 1.40 million additional poor children in husband-wife families. If the excess poverty among children in the NHIS is due to the survey's underestimating their families' incomes, such that the excess poor children should really be in a higher poverty bracket, the comparatively high frequency of husband-wife families among the poor children in the NHIS is consistent with the living arrangements of near-poor children. That is, if a near-poor family is misclassified as poor in the NHIS, such a family is more likely to be a husband-wife family than a single-parent family. We see the same phenomenon among near-poor children in the SIPP, which has substantially more of such children than any other survey. Comparing the living arrangements of near-poor children in the SIPP and CPS, we see that most of the excess in the SIPP is due to children in husband-wife families.

SIPP's comparatively high estimates of low-income persons do not extend to the elderly. SIPP finds fewer low-income elderly than the CPS, MEPS, or NHIS at 11.6 million versus 12.9 to 13.6 million, or 34.1 percent versus 37.6 to 39.7 percent (Table IV.10). The ACS finds the fewest low-income elderly at 11.2 million or 33.3 percent, but SIPP finds the fewest poor elderly (3.0 million) and the lowest elderly poverty rate (8.9 percent). However, estimates of the number of poor elderly do not differ greatly among the five surveys, with the range among the CPS, MEPS, and NHIS being only 3.6 to 3.8 million or 10.5 to 11.3 percent.

TABLE IV.10

ESTIMATES OF POOR AND NEAR-POOR ELDERLY: FIVE SURVEYS

Population Subgroup	CPS	ACS	SIPP	MEPS	NHIS
	Millions of Persons				
All Persons 65 and Older	34.22	33.56	33.95	34.15	34.22
Poverty Status					
Poor	3.58	3.20	3.03	3.84	3.76
Near Poor	9.58	7.98	8.56	9.72	9.10
Total Low Income	13.16	11.18	11.59	13.56	12.86
	Percent of the Population				
All Persons 65 and Older	100.0	100.0	100.0	100.0	100.0
Poverty Status					
Poor	10.5	9.5	8.9	11.3	11.0
Near Poor	28.0	23.8	25.2	28.5	26.6
Total Low Income	38.5	33.3	34.1	39.7	37.6

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year
2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-
year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior
12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a
family income at or above the poverty threshold but below twice the poverty threshold.

3. Poverty in the PSID

In all surveys except the PSID, estimated poverty is based solely on who was living with the family at the time of the interview, and annual family income is the sum of the annual incomes of the persons present at the time of the interview, regardless of where they lived during the income reference year. In contrast, PSID family income and poverty thresholds reflect the income and composition of the family during each month of the year, a contemporaneous measure creating an annual poverty threshold and income consistent with changing family composition throughout the income reference year. Essentially, the PSID calculates twelve separate poverty thresholds, one for each month, and sums the values for the year. Similarly, PSID collects information on people who lived with a sample family for just part of the income reference year and the amount of income they received during their period of co-residence, and these part-year contributions of persons who lived with the family for only part of the reference year are included in the family's annual income.

Based on simulations that we conducted with SIPP, and which are discussed in the next chapter, we found that the contemporaneous measurement of income and family composition reduced the estimated poverty rate by 0.6 percentage points relative to a poverty rate calculated with a fixed family composition measured in the third month after the end of the income reference year (the CPS model). Other things being equal, we would expect the PSID to produce a lower poverty rate than the CPS (or any of the other surveys) due to the PSID's contemporaneous measurement of income and family composition. Compared to the CPS, the PSID poverty rate ought to be (very roughly) 0.6 percentage points lower. If the observed difference departs substantially from that expectation, then we would infer that some additional factors are at play. The PSID may be capturing more income or less income than the CPS, or the sample after 40 years may over- or under-represent families in particular ranges of income.

Given the low weighted total for the PSID, we focus on rates rather than numbers. We obtain a poverty rate of 9.8 percent from the PSID, based on a CPS-comparable family concept and universe (Table IV.11). This compares to 12.2 percent for the full CPS and 11.6 percent for the CPS-X, which removes subpopulations that were excluded from the CPS population controls when the PSID weights were post-stratified. If we allow that contemporaneous measurement will depress the PSID poverty rate by roughly 0.6 percentage points, this implies that the remaining gap is perhaps a little over a percentage point. This is not a particularly large difference, but it is consistent with the earlier evidence that the PSID may be capturing more income from families at the lower end of the income distribution than the other surveys or under-representing such families.

We find a somewhat larger difference between the PSID estimate of the near poor (15.6 percent of the population) and the estimates from the other surveys, which range from 17.7 percent for the ACS to 20.5 percent for the SIPP. The CPS-X estimate is 18.1 percent or 2.5 percentage points higher than the PSID estimate. In our SIPP simulations we found no net difference between contemporary measurement of income and family composition and the CPS model with respect to the number of near poor, so it would appear likely that, except for sampling error, all of the 2.5 percentage point difference between the PSID and CPS-X can be attributed to some combination of better income measurement and under-representation of the near poor in the PSID sample.

The PSID estimate of poor children (14.3 percent) is 2.5 percentage points lower than the full CPS and 2.1 percentage points lower than CPS-X (Table IV.12). This is comparable to what we observed for the general population. The PSID estimates of near-poor children, however, are closer to the CPS and ACS estimates than was true of the general population. The PSID estimate

TABLE IV.11

ESTIMATES OF THE POOR AND NEAR POOR:
PSID AND CENSUS BUREAU SURVEYS

Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
			Millions of Persons		
All Persons	282.55	277.69	281.08	261.45	274.44
Poverty Status					
Poor	34.38	34.61	33.25	25.73	31.82
Near Poor	51.81	49.28	56.25	40.85	49.81
Total Low Income	86.19	83.89	89.50	66.58	81.62
			Percent of the Population		
All Persons	100.0	100.0	100.0	100.0	100.0
Poverty Status					
Poor	12.2	12.5	11.8	9.8	11.6
Near Poor	18.3	17.7	20.0	15.6	18.1
Total Low Income	30.5	30.2	31.8	25.5	29.7

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.12

ESTIMATES OF POOR AND NEAR-POOR CHILDREN:
PSID AND CENSUS BUREAU SURVEYS

Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
Millions of Persons					
All Children under 18	71.67	70.79	71.36	67.48	70.82
Poverty Status					
Poor	12.03	12.51	12.78	9.68	11.63
Near Poor	15.38	14.94	17.72	13.40	15.16
Total Low Income	27.41	27.45	30.50	23.08	26.78
Percent of the Population					
All Children under 18	100.0	100.0	100.0	100.0	100.0
Poverty Status					
Poor	16.8	17.7	17.9	14.3	16.4
Near Poor	21.5	21.1	24.8	19.9	21.4
Total Low Income	38.2	38.8	42.7	34.2	37.8

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

of 19.9 percent is just 1.5 percentage point lower than the estimate from CPS-X, 1.6 percentage points lower than the full CPS, and 1.2 percentage points lower than the ACS. SIPP is an outlier.

Among the elderly, the differences between the PSID and other survey estimates reverse the pattern observed for children. The elderly poverty rate estimated by the PSID matches the rate recorded by SIPP and is just 0.6 percentage points lower than the ACS and 1.4 percentage points lower than CPS-X (Table IV.13). However, the PSID identifies substantially fewer elderly than the other surveys as near poor—18.2 percent versus 28.0 percent for CPS-X and the full CPS, 25.2 percent for SIPP, and 23.8 percent for the ACS. This pattern suggests that representativeness may play a greater role than better income measurement within this subpopulation.

C. EMPLOYMENT AND EARNINGS

Employment is both a key policy variable and the principal source of income among families in the United States. Good estimates of employment are critical to policy analysis in their own right and through their impact on the quality of income data. A review of survey data on both employment and earnings—the income from employment—is critical to the goals of this study.

1. Persons with Earned Income

Table IV.14 provides comparative estimates of persons with calendar year 2002 earnings, broken down by the source of earnings: wages and salaries versus self-employment income. A person may have had both wage and salary income and self-employment income during the year, so in addition to showing the number of persons with any amount of either source, the table breaks down the population of earners into those with only wages and salaries, those with only self-employment income, and those with both.

TABLE IV.13

ESTIMATES OF POOR AND NEAR-POOR ELDERLY:
PSID AND CENSUS BUREAU SURVEYS

Population Subgroup	CPS	ACS	SIPP	PSID	CPS-X[a]
			Millions of Persons		
All Persons 65 and Older	34.22	33.56	33.95	29.95	33.94
Poverty Status					
Poor	3.58	3.20	3.03	2.65	3.49
Near Poor	9.58	7.98	8.56	5.44	9.49
Total Low Income	13.16	11.18	11.59	8.10	12.98
			Percent of the Population		
All Persons 65 and Older	100.0	100.0	100.0	100.0	100.0
Poverty Status					
Poor	10.5	9.5	8.9	8.9	10.3
Near Poor	28.0	23.8	25.2	18.2	28.0
Total Low Income	38.5	33.3	34.1	27.0	38.2

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and poverty status in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.14

PERSONS WITH EARNINGS BY SOURCE: FIVE SURVEYS

Employment	CPS	ACS	SIPP	MEPS	NHIS
	Millions of Persons				
Persons with Earnings	150.4	151.9	154.1	160.4	147.4
Persons with Wages and Salaries	141.5	142.4	140.4	NA	NA
Persons with Self-employment Income	13.3	14.5	20.1	NA	NA
Persons with negative SE income	1.2	0.7	0.0	NA	NA
Persons with positive SE income	12.1	13.8	20.1	NA	NA
Persons with Only Wages and Salaries	137.1	137.4	134.0	NA	NA
Persons with Only Self-employment Income	8.9	9.6	13.7	NA	NA
Persons with Both Wages and Salaries and Self-employment Income	4.4	5.0	6.4	NA	NA
	Percent of the Population				
Persons with Earnings	53.2	54.7	54.8	56.6	51.9
Persons with Wages and Salaries	50.1	51.3	49.9	NA	NA
Persons with Self-employment Income	4.7	5.2	7.1	NA	NA
Persons with negative SE income	0.4	0.3	0.0	NA	NA
Persons with positive SE income	4.3	5.0	7.1	NA	NA
Persons with Only Wages and Salaries	48.5	49.5	47.7	NA	NA
Persons with Only Self-employment Income	3.2	3.4	4.9	NA	NA
Persons with Both Wages and Salaries and Self-employment Income	1.6	1.8	2.3	NA	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Because more than half of the total population has income from employment, estimates of persons with earnings are affected by differences in population size across the surveys. In particular, the 5 million fewer persons represented in the estimates from the ACS versus the CPS would imply 2.5 million fewer earners in the ACS than the CPS if the percentages of persons employed were identical in the two surveys. Therefore, the bottom half of Table IV.14 expresses each category of earners as percentages of the total population.

While the CPS is the official source of monthly labor force estimates, the ACS, SIPP, and MEPS all find both numerically and proportionately more persons with reported earnings in 2002 than the CPS. While the CPS identifies 150.4 million persons with earnings, or 53.2 percent of the population, the estimates from these other surveys range from 151.9 million (or 54.7 percent) in the ACS to 154.1 million (or 54.8 percent) in SIPP and 160.4 million (or 56.6 percent) in MEPS. The NHIS finds the fewest, at 147.4 million or 51.9 percent of the population, but this can be attributed to the fact that the NHIS does not collect earnings data from persons under 18. The 150.4 million earners reported in the CPS include 3.4 million who were under 18. Removing these from the CPS would yield 147.0 million earners, which is slightly lower than the NHIS estimate.

Estimates of persons with wage and salary income are very similar across the three Census Bureau surveys, with the ACS about a million above the CPS estimate of 141.5 million and the SIPP about a million below that number.[30] While the three surveys identify very similar numbers

---

[30] The NHIS does not collect wage and salary or self-employment income separately from total earnings. MEPS *does* collect both sources of earnings, and the estimate of persons with earnings in Table IV.14 is based on sample members reporting income from one or both sources. However, employment and annual income are collected in different sections of the MEPS instrument, and a comparison of reported employment and income by source suggests that most self-employed sample members report their business income as wages and salaries. For this reason we do not break down MEPS earnings or persons with earnings by type of employment except to illustrate (in the next section) the impact of reclassifying reported wage and salary income for a those reporting self-employment as their sole work activity.

of persons with both wages and salaries and self-employment income (estimates range from 4.4 to 6.4 million), differences in the numbers with only self-employment income are striking. The CPS and ACS find between 9 and 10 million while SIPP finds 13.7 million.[31]

Average annual earnings per worker range from a low of $30,899 in SIPP to a high of $35,707 in the NHIS, with the CPS just below that at $35,591 (Table IV.15). ACS lies closer to CPS than to SIPP, with average earnings of $34,279 while MEPS is closer to SIPP at $32,813. Estimates of average wages and salaries in the Census Bureau surveys are consistent with this ordering. SIPP is lowest at $29,514, and the CPS is highest at $35,514, with ACS at $33,837. The average wage and salary income for workers in the SIPP is only 83 percent of the CPS average. SIPP finds the highest average annual self-employment income, however, at $30,755 or 25 percent higher than the CPS average of $24,670. The ACS estimate lies between the CPS and SIPP at $26,893.

2. Measurement Issues

The fact that SIPP identifies so many more persons with self-employment income than the CPS can be attributed to the unique way in which SIPP requests such income. SIPP asks business owners to report their monthly draw from the business as part of their monthly earnings from self-employment, and this appears to have a marked impact on the number of business owners reporting nonzero self-employment income. Beginning with the 2004 SIPP panel the Census Bureau has expanded the questions about self-employment in order to obtain distinctly separate reports of draw and net profit or loss. While the impact of these changes has yet to be determined, two possible outcomes are more total income from self-employment and, for the

---

[31] In 2003 (for tax year 2002), taxpayers filed 18.6 million returns with non-farm sole proprietorships, which exclude self-employed farmers and those with partnerships, LLCs, or S corporations. Even the SIPP estimate appears to understate the self-employed by a substantial margin.

TABLE IV.15

AVERAGE EARNINGS, WAGES AND SALARIES, AND SELF-EMPLOYMENT INCOME
OF WORKERS: FIVE SURVEYS

Source of Income	CPS	ACS	SIPP	MEPS	NHIS
			Average Per Worker		
Earnings	35,591	34,279	30,899	32,813	35,707
Wages and Salaries	35,514	33,837	29,514	NA	NA
Self-employment Income	24,670	26,893	30,755	NA	NA
			Percent of CPS		
Earnings	100.0	96.3	86.8	92.2	100.3
Wages and Salaries	100.0	95.3	83.1	NA	NA
Self-employment Income	100.0	109.0	124.7	NA	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

first time, the reporting of negative self-employment income, which the SIPP has historically failed to elicit.

In contrast to SIPP, MEPS shows the impact of collecting annual income data separately from employment and without reference to reported work activity. With income questions designed around the tax return, MEPS obtains only $92.1 billion of business income compared to $617.6 billion for SIPP, $394.3 billion for ACS, and $334.7 billion for CPS (Table IV.16). Using the MEPS "JOBS" file, which contains the detailed employment data collected two to three times a year, we find that 16.5 million persons (weighted) who reported only self-employment as a work activity in 2002 reported only wage and salary earnings for the year. The wage and salary income for these persons totaled $620.2 billion. If this wage and salary income were reclassified as business income, then net self-employment income would reach $712.3 billion in MEPS, exceeding even SIPP by $95 billion and more than doubling the amount reported in the CPS. At the same time, the MEPS wage and salary income would drop to $4,551.5 billion, which is still more than $400 billion higher than SIPP but $475 billion below the CPS.

Estimates of persons with earnings are not affected by the reclassification of some earned income from wages and salaries to self-employment income. While the ACS and SIPP numbers suggest that the CPS may be underestimating the number of persons with annual earnings, the MEPS estimate of 160 million earners (reported in Table IV.14) lies well above these other surveys, exceeding the CPS by 10.0 million. Recalling from Chapter III that MEPS also finds 8.4 million more family heads than the CPS, one has to wonder if the two estimates are related. Going further, could both be an artifact of the post-stratification of MEPS person weights to the distribution of CPS persons by poverty status? This is an intriguing question, and we will return to the topic of post-stratification in Chapter 5, but providing an answer was not possible with available data. Another possible explanation for the substantially greater number of persons with

TABLE IV.16

AGGREGATE EARNED INCOME BY SOURCE: MEPS AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS as Reported	Alternate MEPS[a]
			Billions of Dollars		
Earned Income	5,354.3	5,207.9	4,760.1	5,263.8	5,263.8
Wages and Salaries	5,026.3	4,817.2	4,142.5	5,171.7	4,551.5
Self-employment Income	328.0	390.7	617.6	92.1	712.3
Negative income	-6.7	-3.6	0.0	-19.5	-19.5
Positive income	334.7	394.3	617.6	111.6	731.8
			Percent of CPS		
Earned Income	100.0	97.3	88.9	98.3	98.3
Wages and Salaries	100.0	95.8	82.4	102.9	90.6
Self-employment Income	100.0	119.1	188.3	28.1	217.2
Negative income	100.0	53.2	0.0	289.1	289.1
Positive income	100.0	117.8	184.5	33.3	218.6

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] The alternative MEPS estimates were derived by moving $620.2 billion from wages and salaries to positive self-employment income. This estimate is based on persons with only self-employment reported in the JOBS file and wages and salaries but no self-employment income.

earned income in MEPS versus the other surveys relates, again, to the separate collection of annual income and employment. Perhaps some of the persons reporting wage and salary income in MEPS should have reported it as unearned income instead. Evidence in support of this possibility is presented in Section G below, which compares reported employment and reported income in SIPP, MEPS, and NHIS. Again, no resolution of these reporting issues was possible with the data available to us in this study, but highlighting the issues is informative about the complexity of measuring income in surveys.

3. Contributions of Earned and Unearned Income to Total Income

Across the three Census Bureau surveys, earned income accounts for between 82.1 and 82.8 percent of total income (Table IV.17). MEPS is only slightly higher at 84.1 percent while NHIS is highest at 86.0 percent, but inconsistencies between earnings and total income among NHIS families, discussed later in this chapter, may explain the relatively high share of earnings in total NHIS income. It is particularly striking that the earnings share of total income should be the same in CPS and SIPP despite the fact that total earned income in the SIPP is only 89 percent of the corresponding amount in the CPS. This implies that *unearned* income, the difference between total income and earnings, must be estimated no better than earnings in the SIPP—at least relative to the CPS. This is confirmed in the bottom panel of the table, where we see that aggregate unearned income in the SIPP is just 90 percent of the CPS amount. Continuing with SIPP we see again how much the survey's estimates of self-employment help to offset the SIPP's apparent understatement of wage and salary income. SIPP captures only 82.4 percent as much aggregate wage and salary income as the CPS, but the SIPP's identification of 88 percent more self-employment income raises the SIPP's share of CPS earnings by 6.5 percentage points. Overall, self-employment income in the SIPP is 10.7 percent of total income, which is more than double the CPS share.

TABLE IV.17

CONTRIBUTION OF EARNED AND UNEARNED INCOME TO TOTAL INCOME: FIVE SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Billions of Dollars		
Total Income	6,468.4	6,346.3	5,766.2	6,257.7	6,116.2
Earned Income	5,354.3	5,207.9	4,760.1	5,263.8	5,261.4
Wages and salaries	5,026.3	4,817.2	4,142.5	NA	NA
Self-employment income	328.0	390.7	617.6	NA	NA
Unearned Income	1,114.1	1,138.3	1,006.0	994.0	854.8
			Percent of Total Income		
Total Income	100.0	100.0	100.0	100.0	100.0
Earned Income	82.8	82.1	82.6	84.1	86.0
Wages and salaries	77.7	75.9	71.8	NA	NA
Self-employment income	5.1	6.2	10.7	NA	NA
Unearned Income	17.2	17.9	17.4	15.9	14.0
			Percent of CPS Income by Source		
Total Income	100.0	98.1	89.1	96.7	94.6
Earned Income	100.0	97.3	88.9	98.3	98.3
Wages and salaries	100.0	95.8	82.4	NA	NA
Self-employment income	100.0	119.1	188.3	NA	NA
Unearned Income	100.0	102.2	90.3	89.2	76.7

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS
ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003
NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

The ACS captures 4 percent lower wage and salary income than the CPS but 19 percent more self-employment income, which raises the ACS earned income to 97.3 percent of the CPS total. The ACS also captures slightly more (2.2 percent) unearned income than the CPS, which contributes to an overall total income that is 98.1 percent of the CPS total. MEPS earned income that is 98.3 percent of CPS earned income—the same share as NHIS. Similar to SIPP, MEPS captures 89 percent as much unearned income as the CPS (the aggregates for the two surveys are essentially identical), which lowers its total income to 96.7 percent of the CPS total. The NHIS does not collect unearned income, but the difference between total income and earned income collected in the NHIS implies unearned income that is 77 percent of the CPS total. This implied shortfall is simply an indication that the NHIS does not do as well in obtaining total income with its single question as it does in collecting earned income from all adults.

When we compare survey estimates of earned income by quintile of family income, we find, interestingly, that the ACS, SIPP, MEPS, and NHIS all find more earnings in the lowest quintile of family income than does the CPS (Table IV.18). The additional earnings range from 9 to 17 percent of the CPS total, but the aggregate amounts are small. The ACS and SIPP find progressively less total earnings relative to the CPS as the quintile increases. For MEPS this is true after the first quintile. The NHIS, on the other hand, finds progressively more aggregate earnings relative to the CPS over quintiles two through four.

Unearned income does not show such clear patterns. Overall, the ACS finds slightly more unearned income than the CPS, but unlike earned income, the ACS finds progressively *more* than the CPS as the quintile rises (Table IV.19). In the top quintile, the ACS finds 23 percent more unearned income than the CPS. Through the first four quintiles, SIPP obtains 99.6 percent as much unearned income as the CPS but identifies only 64 percent as much as the CPS in the top quintile. MEPS, on the other hand, falls short of the CPS in every quintile, being closest in

164

TABLE IV.18

AGGREGATE EARNED INCOME BY QUINTILE OF FAMILY INCOME: FIVE SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Billions of Dollars		
Aggregate Earned Income	5,354.3	5,207.9	4,760.1	5,263.8	5,261.4
Family Income Quintile					
Lowest	176.1	206.5	200.5	191.5	196.4
Second	542.9	565.3	528.0	615.5	514.4
Third	889.2	878.8	795.4	950.8	888.2
Fourth	1,255.9	1,225.5	1,119.4	1,288.3	1,301.9
Highest	2,490.2	2,332.0	2,116.7	2,217.7	2,360.5
Sum through Four Quintiles	2,864.1	2,876.0	2,643.5	3,046.1	2,900.9
			Percent of CPS		
Aggregate Income, All Persons	100.0	97.3	88.9	98.3	98.3
Family Income Quintile					
Lowest	100.0	117.3	113.9	108.8	111.6
Second	100.0	104.1	97.3	113.4	94.7
Third	100.0	98.8	89.5	106.9	99.9
Fourth	100.0	97.6	89.1	102.6	103.7
Highest	100.0	93.6	85.0	89.1	94.8
Sum through Four Quintiles	100.0	100.4	92.3	106.4	101.3

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

TABLE IV.19

AGGREGATE UNEARNED INCOME BY QUINTILE OF FAMILY INCOME: FIVE SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Billions of Dollars		
Aggregate Unearned Income	1,114.1	1,138.3	1,006.0	994.0	854.8
Family Income Quintile					
Lowest	194.4	162.2	190.8	168.4	117.3
Second	231.2	213.1	222.7	192.8	203.3
Third	201.0	208.6	213.4	193.9	170.2
Fourth	190.9	190.3	187.7	173.5	118.7
Highest	296.5	364.0	191.3	265.3	245.3
Sum through Four Quintiles	817.6	774.3	814.7	728.7	609.5
			Percent of CPS		
Aggregate Income, All Persons	100.0	102.2	90.3	89.2	76.7
Family Income Quintile					
Lowest	100.0	83.4	98.2	86.6	60.3
Second	100.0	92.2	96.3	83.4	88.0
Third	100.0	103.8	106.2	96.4	84.6
Fourth	100.0	99.7	98.3	90.9	62.2
Highest	100.0	122.8	64.5	89.5	82.7
Sum through Four Quintiles	100.0	94.7	99.6	89.1	74.5

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: Unearned income is the difference between total income, reported in Table IV.1, and earned income, reported in Table IV.5.

the middle quintile. With its unearned income as a residual rather than a reported amount, the NHIS is erratic. The difference between aggregate total and aggregate earned income is as low as 60 percent of the CPS aggregate in one quintile and as high as 88 percent (in the adjacent quintile).

4. Employment and Earnings in the PSID

The PSID collects individual earnings from only the head and wife of each sample family, so in comparing the PSID to the Census Bureau surveys, we restrict the estimates of persons with earnings and the total amount of earnings to the head and spouse of the primary family (including "nonfamily" householders—that is, those living with no relatives).

Because the PSID weights sum to substantially less than the total population, the estimates of persons with earnings are lower than the Census Bureau surveys except when we remove unrelated subfamilies and most secondary individuals (the CPS-X estimates). As a percentage of the population, however, the PSID finds a higher overall share of the population with earnings—47.2 percent versus 45.7 percent for the full CPS and 44.9 percent for CPS-X—and with wage and salary income—45.7 percent versus 42.7 percent for the full CPS and 41.9 percent for CPS-X (Table IV.20). But the PSID finds only 1.5 percent with self-employment income compared to 3.0 percent for both the full CPS and CPS-X and 4.6 percent for SIPP.

Because the PSID obtains higher aggregate earnings than any of the Census Bureau surveys, we can compare their aggregate estimates directly—without taking account of the PSID's smaller weighted population size. We find that the PSID obtains 3 to 5 percent more aggregate earnings than the full CPS in each quintile except the middle quintile, where the PSID aggregate is 3 percent lower than the CPS (Table IV.21). The differences are greater when the PSID is compared to the CPS-X estimates (and essentially identical in the middle quintile). The ACS and

TABLE IV.20

HEADS AND SPOUSES WITH EARNINGS AND WAGE AND SALARY INCOME

Employment	CPS	ACS	SIPP	PSID	CPS-X[a]
			Millions of Persons		
Persons with Earnings	129.01	128.07	126.72	123.31	123.24
Persons with Wages and Salaries	120.63	119.21	113.77	119.49	115.05
Persons with Only Self-Employment	8.38	8.86	12.95	3.81	8.18
			Percent of the Population		
Persons with Earnings	45.7	46.1	45.1	47.2	44.9
Persons with Wages and Salaries	42.7	42.9	40.5	45.7	41.9
Persons with Only Self-Employment	3.0	3.2	4.6	1.5	3.0

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, and the 2003 PSID, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.21

AGGREGATE EARNED INCOME OF FAMILY HEADS AND WIVES BY QUINTILE
OF FAMILY INCOME: PSID AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
			Billions of Dollars		
Aggregate Earned Income	5,043.5	4,803.5	4,367.4	5,178.9	4,898.1
Family Income Quintile					
Lowest	169.4	198.4	192.5	177.6	155.6
Second	513.0	529.6	494.7	535.8	494.5
Third	826.8	806.1	729.2	802.4	799.5
Fourth	1,174.4	1,118.4	1,014.7	1,211.3	1,145.8
Highest	2,359.9	2,150.9	1,936.2	2,451.8	2,302.7
Sum through Four Quintiles	2,683.6	2,652.6	2,431.2	2,727.1	2,595.4
			Percent of CPS		
Aggregate Income, All Persons	100.0	95.2	86.6	102.7	97.1
Family Income Quintile					
Lowest	100.0	117.1	113.6	104.8	91.9
Second	100.0	103.2	96.4	104.4	96.4
Third	100.0	97.5	88.2	97.0	96.7
Fourth	100.0	95.2	86.4	103.1	97.6
Highest	100.0	91.1	82.0	103.9	97.6
Sum through Four Quintiles	100.0	98.8	90.6	101.6	96.7

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003
CPS ASEC supplement, the 2001 SIPP panel, and the 2003 PSID, and prior 12 months
income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except
unmarried partners of the householder) to mimic the population controls applied to the PSID.

SIPP obtain more aggregate earnings from the lowest quintile than the PSID, but the differences in aggregate dollars are small.

Given that the boundaries between quintiles are higher in the PSID than the CPS (recall Table IV.5), the PSID's consistently higher aggregates across quintile could simply reflect the fact that each PSID quintile includes somewhat higher earners than the corresponding CPS quintile. However, we would very likely see the same pattern if the PSID respondents were uniformly reporting more of their income than their CPS counterparts.

D. PROGRAM PARTICIPATION

How fully and accurately a survey identifies the participants in an entitlement or other means-tested program determines how useful that survey may be for policy analysis of that program and related programs. In comparing the surveys with respect to their estimates of program participation, we focus on welfare (cash assistance) and Food Stamps, SSI, and Medicaid.[32] As a rule, surveys underestimate the numbers of participants in means-tested programs, so in comparing estimates of participants across surveys, "more" is generally better.

Differences in estimates of participants in welfare or Food Stamps, SSI, and Medicaid are quite substantial across the five surveys. For each program, SIPP finds the most participants by a wide margin over any other survey (Table IV.22). For example, SIPP finds 31.4 million persons (or 11.2 percent of the population) in families receiving welfare or Food Stamps at any time during 2002. The ACS is second with 24.3 million or 8.8 percent, followed by the CPS and

---

[32] In 2003 the initial CPS question in the section on public assistance asked about "cash assistance from a state or county welfare program such as [state program name]" and the follow-on question asked about payments received on behalf of children. The more detailed SIPP questions mentioned AFDC or TANF, general assistance, payments for foster children, and "other" welfare. We combine estimates of welfare and Food Stamps because most persons receiving income from welfare also receive Food Stamps.

TABLE IV.22

ESTIMATES OF PROGRAM PARTICIPANTS: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
	Millions of Persons				
All Persons	282.55	277.69	281.08	283.30	283.71
Program					
Welfare or Food Stamps	20.50	24.33	31.41	20.23	14.29
SSI	4.88	4.55	8.38	6.40	5.50
Medicaid					
Ever in prior calendar year	32.86	NA	48.11	41.23	NA
Current month	NA	NA	33.28	34.96	29.90
	Percent of the Population				
All Persons	100.0	100.0	100.0	100.0	100.0
Program					
Welfare or Food Stamps	7.3	8.8	11.2	7.1	5.0
SSI	1.7	1.6	3.0	2.3	1.9
Medicaid					
Ever in prior calendar year	11.6	NA	17.1	14.6	NA
Current month	NA	NA	11.8	12.3	10.5

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2002 ACS, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Except where noted, participation is ever during 2002 or the previous 12 months (ACS).

MEPS with 20.5 million (7.3 percent) and 20.3 million (7.1 percent), respectively. The NHIS identifies fewer than half as many as SIPP—just 14.3 million or 5.0 percent of the population.[33]

For SSI, the CPS and ACS find the fewest participants among the five surveys. SIPP finds 8.38 million or 3.0 percent of the population, followed by MEPS with 6.4 million (or 2.3 percent) and NHIS with 5.5 million (or 1.9 percent). The CPS and ACS find 4.9 million (1.7 percent) and 4.6 million (1.6 percent), respectively.

Our comparisons of Medicaid enrollment utilize two different reference periods in order to maximize the possible comparisons across surveys and to make a point about reporting error. The CPS asks respondents if they were ever enrolled in Medicaid during the previous calendar year (2002) while the NHIS asks respondents if they are enrolled at the time of the survey (January through December 2003). Both SIPP and MEPS capture Medicaid enrollment on a monthly basis, so we can compare estimates of persons ever enrolled in 2002 with the CPS and compare enrollment in December 2002 to the NHIS. The ACS did not include a question on Medicaid enrollment until January 2008.

SIPP finds 48.1 million or 17.1 percent of the population ever enrolled in Medicaid during the 2002 calendar year while MEPS finds 41.2 million or 14.6 percent. The CPS is well behind with 32.9 million or 11.6 percent of the population. MEPS finds more Medicaid enrollees in December 2002 than SIPP, with 35.0 million or 12.3 percent of the population compared to SIPP's 33.3 million or 11.8 percent. NHIS finds an average monthly enrollment of 29.9 million or 10.5 percent of the population in 2003. It is noteworthy that the MEPS and SIPP estimates of Medicaid enrollment in December 2002 exceed the CPS estimate of persons who were ever

---

[33] With the notable exception of family income, employment, and earnings, NCHS does not allocate missing values on the NHIS file. Generally, item non-response rates are very low, however. For instance, only about 1 percent of the sample could not be classified as insured or not. We have not compensated for item non-response in any way, so our estimates of the number of program participants and their share of the population will be very slightly lower than if the missing values had been allocated.

enrolled in 2002. This illustrates a well-known problem with CPS estimates of Medicaid—namely, that the survey's estimates of people who were ever enrolled during a year bear a closer resemblance to panel surveys' estimates of persons enrolled at a single point in time than to estimates of persons ever enrolled in a year. A popular interpretation is that CPS respondents are answering the question about their Medicaid enrollment in the prior year with their current enrollment.

For the overlapping populations participating in either welfare or Food Stamps (or both), we compared the five surveys' estimates of participants by quintile of family income. On doing so, we find that the SIPP's margin over the CPS grows as the quintile increases (Table IV.23). While SIPP finds a third more participants than the CPS among persons in the bottom quintile and 46 percent more in the second quintile, SIPP finds twice as many in the third quintile, more than four times as many in the fourth quintile, and nearly six times as many in the top quintile. The ACS shows a similar pattern relative to the CPS, and both SIPP and ACS show progressively more beneficiaries than MEPS or NHIS as the income quintile rises as well. This may be indicative of a problem with the reporting or, more likely, the imputation of welfare and Food Stamp Program benefits in the SIPP and ACS. At the same time, however, if both surveys are more effective at identifying welfare and Food Stamp Program beneficiaries than other surveys, part of their success may lie in eliciting reports of participation from people who might be least inclined to report their participation—such as those who received benefits for a brief period when their incomes were much lower than they were for the rest of the year.

Reports of participation in welfare or Food Stamps or SSI (the latter among family heads only) in the PSID compare to those that were obtained in the CPS (Table IV.24). Focusing on participants expressed as percentages of the population, we find that participants in the PSID, the full CPS, and CPS-X were 7.3 percent of the population compared to SIPP's 11.2 percent.

TABLE IV.23

PERSONS IN FAMILIES WITH WELFARE AND/OR FOOD STAMPS
BY QUINTILE OF FAMILY INCOME: FIVE SURVEYS

Income Estimate	CPS	ACS	SIPP	MEPS	NHIS
			Thousands of Persons		
All Participants	20,496	24,325	31,406	20,226	21,990
Family Income Quintile					
Lowest	13,562	14,879	18,001	13,949	14,783
Second	4,461	4,854	6,498	4,512	4,355
Third	1,748	2,396	3,520	1,189	1,685
Fourth	493	1,273	2,044	393	719
Highest	233	923	1,343	183	447
Sum through Four Quintiles	20,263	23,402	30,063	20,043	21,543
			Percent of CPS		
All Participants	100.0	118.7	153.2	98.7	107.3
Family Income Quintile					
Lowest	100.0	109.7	132.7	102.9	109.0
Second	100.0	108.8	145.7	101.1	97.6
Third	100.0	137.0	201.3	68.0	96.4
Fourth	100.0	258.4	414.8	79.7	145.9
Highest	100.0	396.8	577.2	78.6	192.2
Sum through Four Quintiles	100.0	115.5	148.4	98.9	106.3

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2002
ACS, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Participation is ever during 2002 or the previous 12 months (ACS).

TABLE IV.24

ESTIMATES OF PROGRAM PARTICIPANTS: PSID AND CENSUS BUREAU SURVEYS

Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
	Millions of Persons				
All Persons	282.55	277.69	281.08	261.45	274.44
Program					
Welfare or Food Stamps [b]	20.50	24.32	31.41	19.19	19.91
SSI among family heads	2.44	2.64	4.45	2.45	2.44
Medicaid					
Ever in prior calendar year	32.86	NA	48.11	16.00 [c]	32.01
Current month	NA	NA	33.28	NA	NA
	Percent of the Population				
All Persons	100.0	100.0	100.0	100.0	100.0
Program					
Welfare or Food Stamps [b]	7.3	8.8	11.2	7.3	7.3
SSI among family heads	0.9	1.0	1.6	0.9	0.9
Medicaid					
Ever in prior calendar year	11.6	NA	17.1	6.1	11.7
Current month	NA	NA	11.8	NA	NA

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2002 ACS, the 2001 SIPP panel, and the 2003 PSID.

Note: Except where noted, participation is ever during 2002 or the previous 12 months (ACS).

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.
[b] Persons are counted if the head of the family is receiving welfare or if anyone in the family is receiving food stamps.
[c] Persons with Medicaid and no other coverage in 2001 and 2002 and at least one month of health insurance coverage in 2002.

Likewise, family heads who received SSI were 0.9 percent of the population in the PSID, the full CPS, and CPS-X and 1.0 percent in the ACS. The corresponding participation rate in the SIPP was 1.6 percent.

The PSID asks its respondents about their Medicaid participation over the prior two calendar years and does not obtain separate reports by year. We approximated a measure of ever enrollment in 2002 by identifying persons who were ever enrolled in Medicaid in 2001 or 2002, had no other coverage during that period, and were uninsured for no more than 11 months of 2002. This yielded a very low enrollment estimate—only 16.0 million persons or 6.1 percent of the population or just a third of the SIPP estimate of 48.1 million or 17.1 percent and barely half of the CPS estimate of 11.6 percent. Had we included PSID respondents who reported coverage in addition to Medicaid during the two-year period (and who might have been covered by something other than Medicaid during 2002), we would have increased the PSID estimate by only a small amount.

When reported welfare or Food Stamp Program beneficiaries are distributed by quintile of family income, we find that the PSID estimates fall off more rapidly than the CPS estimates as the quintile increases (Table IV.25). In the lowest quintile the beneficiaries identified in the PSID are 97.8 percent of the number identified in the full CPS (and a larger fraction of those identified in CPS-X). This drops to 93.7 percent in the second quintile, with no change relative to CPS-X, and then 66.1 percent in the third quintile, where the PSID falls relative to CPS-X as well. In the fourth quintile the PSID estimate is comparable to both CPS estimates, but in the highest quintile the PSID estimate is only 51 percent of the full CPS estimate and, not shown, only 56 percent of the CPS-X estimate. We find it interesting and perhaps informative that the PSID should show the same pattern of declining enrollment by quintile when compared to the CPS that we saw when comparing the CPS to SIPP. Once again, this could reflect growing reluctance to report

TABLE IV.25

PERSONS IN FAMILIES WITH WELFARE AND/OR FOOD STAMPS BY QUINTILE
OF FAMILY INCOME: PSID AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	PSID	CPS-X[a]
	Thousands of Persons				
All Participants	20,496	24,325	31,406	19,186	19,906
Family Income Quintile					
Lowest	13,562	14,879	18,001	13,264	13,389
Second	4,461	4,854	6,498	4,182	4,217
Third	1,748	2,396	3,520	1,155	1,606
Fourth	493	1,273	2,044	466	481
Highest	233	923	1,343	119	212
Sum through Four Quintiles	20,263	23,402	30,063	19,066	19,693
	Percent of CPS				
All Participants	100.0	118.7	153.2	93.6	97.1
Family Income Quintile					
Lowest	100.0	109.7	132.7	97.8	98.7
Second	100.0	108.8	145.7	93.7	94.5
Third	100.0	137.0	201.3	66.1	91.9
Fourth	100.0	258.4	414.8	94.6	97.6
Highest	100.0	396.8	577.2	51.2	91.3
Sum through Four Quintiles	100.0	115.5	148.4	94.1	97.2

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the
2002 ACS, the 2001 SIPP panel, and the 2003 PSID.

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except
unmarried partners of the householder) to mimic the population controls applied to the PSID.

177

prior welfare or Food Stamp receipt as income increases, but it could also reflect differences in imputation across the surveys. The PSID does not make use of the hot deck imputation methods employed in the CPS and SIPP and, therefore, the PSID results may reflect actual reporting patterns more closely than they do in the other surveys.

E. THE UNINSURED

The frequency with which people lack health insurance coverage is an important indicator for health policy analysis and one that is strongly associated with income—hence its inclusion in this study. Surveys differ with respect to the reference period used to identify the uninsured. The most commonly used measures define the uninsured at a point in time or over a period of time— typically a year. We examine both measures and their relationship to family income and conclude by examining how the ratio of full-year to point-in-time uninsured varies across the surveys that support both measures.

1. Uninsured at a Point in Time

Three of the surveys provide estimates of health insurance coverage and the uninsured at a point in time. The NHIS measures coverage at the time of the interview, while SIPP and MEPS obtain estimates of coverage by month from interviews that ask about the previous several months.[34] In the tables presented in this section, the NHIS estimates represent an average of respondents' reports over the 2003 calendar year while the SIPP and MEPS estimates refer to December 2002.

Estimates of the proportion of the population that was without health insurance coverage at a point in time range from 14.6 percent for the NHIS to 16.8 percent for MEPS, with SIPP falling

---

[34] SIPP respondents are asked to report coverage in each of the preceding four calendar months. MEPS respondents are asked when coverage started and stopped over a variable reference period, and these reports are used to determine coverage by month.

between these two at 15.3 percent (Table IV.26).[35] Despite differences in the surveys' estimates of the distribution of the population by poverty level and the fact that income is measured over a period of a year while health insurance coverage refers to a point in time at the end of that year (SIPP and MEPS) or 1 to 12 months later (NHIS), uninsured rates across the three surveys show very similar patterns by poverty relative. Rates of 28 to 31 percent among the poor decline to about 5 to 8 percent among those with incomes above 400 percent of poverty. Estimates of the *number* of uninsured persons below 200 percent of poverty are exceedingly close in the three surveys, with a range of 24.7 million (MEPS) to 25.1 million (NHIS).

Policy analysis of the uninsured often excludes the elderly population, whose coverage rates are so high that estimates of their uninsured rates are dominated by measurement error. Estimates of the uninsured rate among the elderly range from 0.5 percent for MEPS to 0.9 percent for SIPP and 1.1 percent for NHIS (Table IV.27). Both SIPP and NHIS show an uninsured rate of about 3 percent among the poor with declining rates by poverty relative while MEPS shows no distinct pattern.

For the nonelderly population as a whole, the patterns among the three surveys are very similar to those for all persons, except that the uninsured rates are about 2 percentage points higher. MEPS is again highest at 19.0 percent, followed by SIPP at 17.2 percent and NHIS at 16.4 percent (Table IV.28). Uninsured rates by poverty relative decline similarly across the three surveys.

When we separate children from nonelderly adults, however, we find a shift in SIPP relative to the other two surveys. SIPP finds the highest uninsured rate among children at 14.7 percent,

---

[35] NCHS does not allocate missing values for the health insurance items, and we have not compensated for non-response on these items. As we noted earlier, however, only about 1 percent of the sample could not be classified as insured or not. Consequently, our estimates of the number of persons and percent of the population without health insurance coverage will be very slightly lower than if the missing values had been allocated.

TABLE IV.26

PERSONS UNINSURED AT A POINT IN TIME
BY POVERTY RELATIVE: SIPP, MEPS, AND NHIS

Population and Poverty Relative	SIPP	MEPS	NHIS
	All Persons (millions)		
All Persons	281.08	283.30	283.71
Poverty Relative			
Under 100%	33.25	35.35	41.58
100% to under 200%	56.25	52.14	53.91
200% to under 400%	98.37	89.80	87.06
400% and over	93.22	106.02	101.16
	Uninsured Persons (millions)		
All Persons	42.88	47.45	41.32
Poverty Relative			
Under 100%	10.36	10.12	11.77
100% to under 200%	14.56	14.60	13.34
200% to under 400%	13.50	14.56	11.37
400% and over	4.46	8.18	4.84
	Percent Uninsured		
All Persons	15.3	16.8	14.6
Poverty Relative			
Under 100%	31.2	28.6	28.3
100% to under 200%	25.9	28.0	24.7
200% to under 400%	13.7	16.2	13.1
400% and over	4.8	7.7	4.8

Source: Mathematica Policy Research, from tabulations of the 2001 SIPP panel,
the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Poverty status is for calendar year 2002, and uninsured status is in
December 2002 or average monthly in 2003 (NHIS).

TABLE IV.27

ELDERLY UNINSURED AT A POINT IN TIME
BY POVERTY RELATIVE: SIPP, MEPS, AND NHIS

Population and Poverty Relative	SIPP	MEPS	NHIS
	All Persons (millions)		
All Persons 65 and Older	33.95	34.15	34.22
Poverty Relative			
Under 100%	3.03	3.84	3.76
100% to under 200%	8.56	9.72	9.10
200% to under 400%	13.52	10.37	12.96
400% and over	8.84	10.22	8.40
	Uninsured Persons (millions)		
Persons 65 and Older	0.31	0.16	0.36
Poverty Relative			
Under 100%	0.10	0.02	0.10
100% to under 200%	0.06	0.07	0.12
200% to under 400%	0.13	0.02	0.09
400% and over	0.02	0.05	0.05
	Percent Uninsured		
Persons 65 and Older	0.9	0.5	1.1
Poverty Relative			
Under 100%	3.2	0.5	2.8
100% to under 200%	0.7	0.8	1.3
200% to under 400%	0.9	0.2	0.7
400% and over	0.3	0.5	0.6

Source: Mathematica Policy Research, from tabulations of the 2001 SIPP panel,
          the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Poverty status is for calendar year 2002, and uninsured status is in
           December 2002 or average monthly in 2003 (NHIS).

TABLE IV.28

NONELDERLY UNINSURED AT A POINT IN TIME
BY POVERTY RELATIVE: SIPP, MEPS, AND NHIS

Population and Poverty Relative	SIPP	MEPS	NHIS
	All Persons (millions)		
All Persons under 65	247.13	249.14	249.49
Poverty Relative			
Under 100%	30.21	31.50	37.82
100% to under 200%	47.69	42.42	44.81
200% to under 400%	84.85	79.42	74.10
400% and over	84.37	95.80	92.76
	Uninsured Persons (millions)		
Persons under 65	42.58	47.30	40.96
Poverty Relative			
Under 100%	10.27	10.10	11.67
100% to under 200%	14.50	14.52	13.22
200% to under 400%	13.37	14.54	11.27
400% and over	4.44	8.13	4.79
	Percent Uninsured		
Persons under 65	17.2	19.0	16.4
Poverty Relative			
Under 100%	34.0	32.1	30.9
100% to under 200%	30.4	34.2	29.5
200% to under 400%	15.8	18.3	15.2
400% and over	5.3	8.5	5.2

Source: Mathematica Policy Research, from tabulations of the 2001 SIPP panel,
         the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Poverty status is for calendar year 2002, and uninsured status is in
         December 2002 or average monthly in 2003 (NHIS).

followed by MEPS at 12.5 percent and NHIS at 9.8 percent (Table IV.29). SIPP differs most from the other two surveys among poor children. SIPP finds an uninsured rate of nearly 23 percent among poor children compared to 14.2 percent for MEPS and 16.7 percent for NHIS. The SIPP finding may be an artifact of how SIPP collects health insurance coverage for children versus adults. Children under 15 are not defined as respondents in the SIPP, and this has implications for how their data are collected. For persons 15 and older, the survey goes through the household person by person to obtain reported coverage. Children's coverage is obtained solely by asking who else in the household has coverage under each reported plan or source. This may lend itself to periodic omissions of individual children from lists of those covered. Whatever the reason, SIPP finds 1.9 million more uninsured children below 200 percent of poverty (6.7 million) than either MEPS or NHIS (4.7 and 4.8 million).

Among nonelderly adults, SIPP's position reverses, with SIPP having the lowest uninsured rate (18.3 percent) among the three surveys (Table IV.30). NHIS is marginally higher at 19.1 percent while MEPS finds 21.6 percent without coverage. Again, however, the three surveys show similar patterns of decline by poverty relative, with MEPS being highest in every category of relative income.

One other point about income and the uninsured should be noted. Among children, MEPS shows a higher uninsured rate between 100 and 200 percent of poverty than below poverty while the other two surveys show marginally higher uninsured rates among the poor. Among nonelderly adults, all three surveys show higher uninsured rates among the poor than among the near poor. The difference between children and adults in all three surveys reflects the impact of public coverage, which benefits children more than nonelderly adults and largely offsets or perhaps more than offsets the adverse effects of declining income on the availability of coverage from private sources. We see this in all three surveys, albeit to different degrees.

TABLE IV.29

CHILDREN UNINSURED AT A POINT IN TIME
BY POVERTY RELATIVE: SIPP, MEPS, AND NHIS

Population and Poverty Relative	SIPP	MEPS	NHIS
	All Persons (millions)		
All Children under 18	71.36	71.80	71.73
Poverty Relative			
Under 100%	12.78	12.47	14.29
100% to under 200%	17.72	15.47	15.41
200% to under 400%	24.58	24.49	21.67
400% and over	16.28	19.36	20.36
	Uninsured Persons (millions)		
Children under 18	10.47	8.98	7.04
Poverty Relative	6.68	4.74	4.77
Under 100%	2.92	1.77	2.38
100% to under 200%	3.76	2.97	2.39
200% to under 400%	2.95	2.90	1.73
400% and over	0.84	1.34	0.54
	Percent Uninsured		
Children under 18	14.7	12.5	9.8
Poverty Relative			
Under 100%	22.8	14.2	16.7
100% to under 200%	21.2	19.2	15.5
200% to under 400%	12.0	11.8	8.0
400% and over	5.2	6.9	2.6

Source: Mathematica Policy Research, from tabulations of the 2001 SIPP panel,
                 the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Poverty status is for calendar year 2002, and uninsured status is in
                 December 2002 or average monthly in 2003 (NHIS).

184

TABLE IV.30

NONELDERLY ADULTS UNINSURED AT A POINT IN TIME
BY POVERTY RELATIVE: SIPP, MEPS, AND NHIS

Population and Poverty Relative	SIPP	MEPS	NHIS
	All Persons (millions)		
All Persons 18 to 64	175.77	177.34	177.76
Poverty Relative			
Under 100%	17.44	19.03	23.53
100% to under 200%	29.97	26.95	29.40
200% to under 400%	60.27	54.93	52.42
400% and over	68.09	76.43	72.40
	Uninsured Persons (millions)		
Persons 18 to 64	32.10	38.32	33.93
Poverty Relative			
Under 100%	7.35	8.33	9.29
100% to under 200%	10.74	11.55	10.84
200% to under 400%	10.42	11.64	9.55
400% and over	3.59	6.79	4.26
	Percent Uninsured		
Persons 18 to 64	18.3	21.6	19.1
Poverty Relative			
Under 100%	42.2	43.8	39.5
100% to under 200%	35.8	42.9	36.9
200% to under 400%	17.3	21.2	18.2
400% and over	5.3	8.9	5.9

Source: Mathematica Policy Research, from tabulations of the 2001 SIPP panel,
the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS.

Note: Poverty status is for calendar year 2002, and uninsured status is in
December 2002 or average monthly in 2003 (NHIS).

2. Uninsured For a Full Year

While the general similarities of point-in-time uninsured estimates and their relationship to relative income across surveys is heartening, point-in-time estimates reflect uninsured spells of varying durations, which differ in their policy priorities. Policymakers give highest priority to long-term uninsured spells, which are frequently defined by durations of a year or more. Five of the six surveys (all but the ACS, which did not start to measure health insurance coverage until January 2008) provide estimates of people without coverage for an entire year. With SIPP and MEPS, estimates of people uninsured for varying durations can be constructed from the monthly estimates discussed earlier, which are based on multiple interviews conducted over a given year. The remaining surveys rely on retrospective questions asking the respondent to think back over the past 12 months or prior calendar year.

As we saw with estimates of program participation over a 12-month period, the survey estimates of persons uninsured for a full year vary widely. For the population as a whole, SIPP is lowest at 8.2 percent while the CPS is highest at 14.8 percent (Table IV.31). The CPS estimate in fact compares closely to two of the three point-in-time estimates (for SIPP and NHIS). This property of the CPS uninsured estimates is well known among health policy researchers, and because of it the CPS uninsured estimates are widely—but not universally—interpreted and analyzed as if they referred to a point in time. The MEPS estimate of 11.8 percent stands midway between the SIPP and CPS estimates despite the longitudinal basis that it shares with the SIPP estimate. The NHIS estimate of 9.7 percent is closest to the SIPP estimate while the PSID estimate of 13.6 percent is closest to the CPS figure (and not far from the reduced sample CPS estimate). The very low estimate of Medicaid participation obtained in the PSID, as reported earlier, may play a role in this high uninsured rate. Whatever the source, the CPS and PSID uninsured rates are strikingly similar by poverty relative, which suggests that the overall PSID

TABLE IV.31

FULL-YEAR UNINSURED PERSONS BY POVERTY RELATIVE

Population and Poverty Relative	CPS	SIPP	MEPS	NHIS	PSID	CPS-X[a]
	Millions of Persons					
All Persons	282.55	281.08	283.30	283.71	261.45	274.44
Poverty Relative						
Under 100%	34.38	33.25	35.35	41.58	25.73	31.82
100% to under 200%	51.81	56.25	52.14	53.91	40.85	49.81
200% to under 400%	89.62	98.37	89.80	87.06	80.00	87.26
400% and over	106.73	93.22	106.02	101.16	114.87	105.56
	Persons Uninsured for the Prior Year (Millions)					
All Persons	41.80	22.91	33.31	27.47	35.55	38.79
Poverty Relative						
Under 100%	10.36	5.97	7.67	8.47	8.01	9.15
100% to under 200%	12.53	8.23	10.51	9.35	10.86	11.67
200% to under 400%	12.34	6.92	10.20	6.99	9.96	11.62
400% and over	6.57	1.78	4.93	2.65	6.71	6.34
	Percent of Persons Uninsured for the Prior Year					
All Persons	14.8	8.2	11.8	9.7	13.6	14.1
Poverty Relative						
Under 100%	30.1	18.0	21.7	20.4	31.1	28.8
100% to under 200%	24.2	14.6	20.2	17.3	26.6	23.4
200% to under 400%	13.8	7.0	11.4	8.0	12.4	13.3
400% and over	6.2	1.9	4.7	2.6	5.8	6.0

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 NHIS, and the 2003 PSID.

Note: Poverty and uninsured status refer to calendar year 2002, except for NHIS (the past 12 months).

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

estimate would match the CPS even more closely if the PSID had more poor persons. The SIPP, MEPS, and NHIS uninsured rates also show similar patterns of decline by poverty relative. In every poverty class, SIPP is the lowest of the three, and MEPS is the highest.

Full-year uninsured rates among the elderly range from 0.4 to 0.8 percent across all of the surveys but the PSID, which finds 2.7 percent of the elderly population without coverage for all of 2002 (Table IV.32). This is still a very small fraction, however, and may reflect a misunderstanding of the question among a small share of respondents.

While the SIPP had a comparatively high point-in-time uninsured rate for children, this does not carry through to estimates of children without coverage for a year. SIPP and NHIS both show a little over 5 percent of children being without coverage for a full year while the CPS and PSID show more than twice this percentage, with MEPS falling between these extremes (Table IV.33). Uninsured rates by poverty relative are quite similar between SIPP and NHIS and between the CPS and PSID. MEPS exhibits the same general pattern as all four other surveys, with higher rates than SIPP in every poverty class, but continues to show a higher uninsured rate among the near poor than among the poor. Across all of the surveys the uninsured rates are fairly similar between these two subpopulations, reflecting, as we noted, the impact of public programs on children's health insurance coverage.

Among nonelderly adults and all nonelderly persons, the PSID and CPS uninsured rates by poverty relative are in close agreement (Tables IV.34 and IV.35). The full-year uninsured rate for nonelderly adults in the CPS, 19.0 percent, nearly matches the NHIS point-in-time uninsured rate reported for this subpopulation in Table A.5. SIPP has the lowest full-year uninsured rate at 10.8 percent, followed by the NHIS at 13.2 percent. At 15.6 percent, MEPS is not much lower than the PSID (16.5 percent), but the MEPS uninsured rates are markedly lower below 200 percent of poverty. Overall, the importance of income in health policy analysis is underscored by

TABLE IV.32

FULL-YEAR UNINSURED ELDERLY PERSONS BY POVERTY RELATIVE

Population and Poverty Relative	CPS	SIPP	MEPS	NHIS	PSID	CPS-X[a]
	Millions of Persons					
All Persons 65 and Older	34.22	33.95	34.15	34.22	29.95	33.94
Poverty Relative						
Under 100%	3.58	3.03	3.84	3.76	2.65	3.49
100% to under 200%	9.58	8.56	9.72	9.10	5.44	9.49
200% to under 400%	12.07	13.52	10.37	12.96	10.88	12.01
400% and over	8.99	8.84	10.22	8.40	10.98	8.96
	Persons Uninsured for the Prior Year (Millions)					
Persons 65 and Older	0.26	0.14	0.16	0.26	0.82	0.25
Poverty Relative						
Under 100%	0.07	0.06	0.02	0.08	0.19	0.06
100% to under 200%	0.08	0.04	0.07	0.09	0.11	0.08
200% to under 400%	0.06	0.03	0.03	0.06	0.33	0.06
400% and over	0.05	0.01	0.03	0.03	0.18	0.05
	Percent of Persons Uninsured for the Prior Year					
Persons 65 and Older	0.8	0.4	0.5	0.8	2.7	0.7
Poverty Relative						
Under 100%	1.9	2.1	0.5	2.1	7.3	1.7
100% to under 200%	0.9	0.4	0.8	1.0	2.1	0.9
200% to under 400%	0.5	0.2	0.3	0.5	3.1	0.5
400% and over	0.6	0.1	0.3	0.3	1.7	0.5

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 NHIS, and the 2003 PSID.

Note: Poverty and uninsured status refer to calendar year 2002, except for NHIS (the past 12 months).

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.33

FULL-YEAR UNINSURED CHILDREN BY POVERTY RELATIVE

Population and Poverty Relative	CPS	SIPP	MEPS	NHIS	PSID	CPS-X[a]
	Millions of Persons					
All Children under 18	71.67	71.36	71.80	71.73	67.48	70.82
Poverty Relative						
Under 100%	12.03	12.78	12.47	14.29	9.68	11.63
100% to under 200%	15.38	17.72	15.47	15.41	13.40	15.16
200% to under 400%	23.19	24.58	24.49	21.67	21.25	23.01
400% and over	21.06	16.28	19.36	20.36	23.15	21.03
	Persons Uninsured for the Prior Year (Millions)					
All Children under 18	8.05	3.81	5.54	3.73	7.69	7.86
Poverty Relative						
Under 100%	2.38	1.03	1.14	1.41	2.12	2.27
100% to under 200%	2.80	1.53	1.74	1.29	2.44	2.75
200% to under 400%	2.05	1.04	1.82	0.78	1.77	2.03
400% and over	0.81	0.21	0.84	0.24	1.36	0.81
	Percent of Persons Uninsured for the Prior Year					
All Children under 18	11.2	5.3	7.7	5.2	11.4	11.1
Poverty Relative						
Under 100%	19.8	8.1	9.2	9.9	21.9	19.6
100% to under 200%	18.2	8.6	11.2	8.4	18.2	18.2
200% to under 400%	8.9	4.2	7.4	3.6	8.3	8.8
400% and over	3.8	1.3	4.3	1.2	5.9	3.8

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 NHIS, and the 2003 PSID.

Note: Poverty and uninsured status refer to calendar year 2002, except for NHIS (the past 12 months).

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.34

FULL-YEAR UNINSURED NONELDERLY ADULTS BY POVERTY RELATIVE

Population and Poverty Relative	CPS	SIPP	MEPS	NHIS	PSID	CPS-X[a]
	Millions of Persons					
All Persons under 65	176.66	175.77	177.34	177.76	164.02	169.68
Poverty Relative						
Under 100%	18.77	17.44	19.03	23.53	13.40	16.70
100% to under 200%	26.85	29.97	26.95	29.40	22.01	25.16
200% to under 400%	54.36	60.27	54.93	52.42	47.87	52.25
400% and over	76.67	68.09	76.43	72.40	80.74	75.57
	Persons Uninsured for the Prior Year (Millions)					
All Persons under 65	33.50	18.96	27.61	23.48	27.04	30.68
Poverty Relative						
Under 100%	7.91	4.88	6.51	6.98	5.70	6.82
100% to under 200%	9.64	6.66	8.70	7.97	8.31	8.83
200% to under 400%	10.23	5.85	8.34	6.14	7.86	9.54
400% and over	5.71	1.57	4.06	2.39	5.17	5.49
	Percent of Persons Uninsured for the Prior Year					
All Persons under 65	19.0	10.8	15.6	13.2	16.5	18.1
Poverty Relative						
Under 100%	42.1	28.0	34.2	29.7	42.5	40.8
100% to under 200%	35.9	22.2	32.3	27.1	37.7	35.1
200% to under 400%	18.8	9.7	15.2	11.7	16.4	18.3
400% and over	7.5	2.3	5.3	3.3	6.4	7.3

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 NHIS, and the 2003 PSID.

Note: Poverty and uninsured status refer to calendar year 2002, except for NHIS (the past 12 months).

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

TABLE IV.35

FULL-YEAR UNINSURED NONELDERLY PERSONS BY POVERTY RELATIVE

Population and Poverty Relative	CPS	SIPP	MEPS	NHIS	PSID	CPS-X[a]
	Millions of Persons					
All Persons under 65	248.33	247.13	249.14	249.49	231.50	240.50
Poverty Relative						
Under 100%	30.80	30.21	31.50	37.82	23.08	28.33
100% to under 200%	42.23	47.69	42.42	44.81	35.41	40.32
200% to under 400%	77.56	84.85	79.42	74.10	69.12	75.25
400% and over	97.74	84.37	95.80	92.76	103.89	96.60
	Persons Uninsured for the Prior Year (Millions)					
All Persons under 65	41.54	22.77	33.15	27.20	34.72	38.54
Poverty Relative						
Under 100%	10.29	5.91	7.65	8.39	7.81	9.09
100% to under 200%	12.45	8.19	10.43	9.26	10.75	11.59
200% to under 400%	12.29	6.90	10.17	6.92	9.62	11.57
400% and over	6.52	1.78	4.90	2.63	6.53	6.29
	Percent of Persons Uninsured for the Prior Year					
All Persons under 65	16.7	9.2	13.3	10.9	15.0	16.0
Poverty Relative						
Under 100%	33.4	19.6	24.3	22.2	33.9	32.1
100% to under 200%	29.5	17.2	24.6	20.7	30.4	28.7
200% to under 400%	15.8	8.1	12.8	9.3	13.9	15.4
400% and over	6.7	2.1	5.1	2.8	6.3	6.5

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, the 2003 NHIS, and the 2003 PSID.

Note: Poverty and uninsured status refer to calendar year 2002, except for NHIS (the past 12 months).

[a] The CPS-X estimates exclude all unrelated subfamilies and most secondary individuals (except unmarried partners of the householder) to mimic the population controls applied to the PSID.

the sharp differential that exists between the poor and near poor, on the one hand, and those above 400 percent of poverty on the other. For SIPP and NHIS, the 2 to 3 percent of people above 400 percent of poverty who were uninsured for the full year contrast with the 28 to 30 percent of the poor who were without coverage for an entire year. For MEPS the range is 5 to 34 percent, and for the CPS and PSID it stands at 7 to 42 percent. Nevertheless, the wide range of estimates of full-year uninsured nonelderly, from 9.2 to 16.7 percent, and even the difference between SIPP and MEPS (9.2 versus 13.3 percent), indicate that the measurement of income poses less of a problem for policymakers than the measurement of health insurance coverage.

3. Ratio of Point-in-Time to Full-Year Uninsured

The ratio of persons uninsured at a point in time to persons uninsured for a full year provides a measure of turnover in the uninsured and therefore a proxy for the duration of uninsurance. Higher ratios imply shorter spells of uninsurance. Table IV.36 presents ratios for the SIPP, MEPS, and NHIS for the entire population, children, and nonelderly adults. Across all populations and poverty brackets, SIPP has the highest ratio, with NHIS narrowly larger than MEPS. For example, among all persons the SIPP ratio is 1.87, followed by 1.50 for NHIS and 1.42 for MEPS. The differences among the surveys narrow among nonelderly adults, with SIPP standing at 1.41, NHIS at 1.25, and MEPS at 1.16. Interestingly, within any age group, the ratios for all three surveys grow with the poverty relative, implying that not only do uninsured rates (for any time period) decline with increasing income; the durations of uninsured spells decline as well.

F. SURVEYS OF RESTRICTED POPULATIONS

Two of the eight surveys focus on restricted populations: Medicare beneficiaries for the MCBS and persons 51 and older for the HRS. We examine the estimates of income from these surveys in comparison with estimates of income for approximately the same target populations

TABLE IV.36

RATIO OF POINT-IN-TIME TO FULL-YEAR UNINSURED BY AGE
AND POVERTY RELATIVE: SIPP, MEPS, AND NHIS

Population and Poverty Relative	SIPP	MEPS	NHIS
All Persons	1.87	1.42	1.50
Poverty Relative			
Under 100%	1.73	1.32	1.39
100% to under 200%	1.77	1.39	1.43
200% to under 400%	1.95	1.43	1.63
400% and over	2.50	1.66	1.82
All Children under 18	2.75	1.62	1.89
Poverty Relative			
Under 100%	2.83	1.55	1.69
100% to under 200%	2.46	1.71	1.84
200% to under 400%	2.83	1.59	2.21
400% and over	4.02	1.59	2.21
All Adults 18 to 64	1.41	1.16	1.25
Poverty Relative			
Under 100%	1.24	1.09	1.11
100% to under 200%	1.31	1.11	1.17
200% to under 400%	1.51	1.15	1.38
400% and over	2.02	1.39	1.62

Source: Mathematica Policy Research, from tabulations of the 2001 SIPP
panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003
NHIS.

Note: Poverty and uninsured status refer to calendar year 2002, except
for NHIS (the past 12 months).

from the three Census Bureau surveys. Because of concerns about the accuracy of reporting of Medicare enrollment for persons under 65 in the national surveys—including, in particular, its confusion with Medicaid—we restricted the MCBS comparisons to elderly beneficiaries. In addition, because the 2002 ACS did not ask respondents to report their Medicare coverage, we defined the comparison population for the national surveys as all elderly persons rather than just those who reported Medicare coverage. We present estimates of income for the elderly first and then turn to the broader population of persons 51 and older.

1. The Elderly

The Census Bureau survey estimates of persons 65 and older exceed the MCBS estimates of elderly Medicare beneficiaries by 1.6 to 2.2 million, which can be attributed in large part to including elderly non-beneficiaries in the former (Table IV.37).[36] Despite this small difference in size, the distribution of the MCBS population by sex and race/ethnicity corresponds very closely to what we find in the Census Bureau surveys, as does the frequency of persons living alone or with no relative ("single"). Elderly respondents to the Census Bureau surveys are more likely to be living with a spouse (by 3 to 5 percentage points) and less likely to be living with other relatives. Estimates of Medicaid enrollment in 2002, which in the MCBS are based in large part on administrative data, lie between the CPS and SIPP, which suggests that the SIPP estimate may be high. And while a third or more of the CPS and SIPP respondents reported their health status as fair or poor, this was true of only 21 percent of the MCBS sample.

---

[36] Two additional factors, one from the MCBS and one from the other surveys, may contribute to the observed differences. The identification of non-institutionalized beneficiaries in the MCBS is based on those interviewed with community rather than facility forms. This distinction is frequently a matter of fieldwork convenience and is not equivalent to the Census Bureau's identification of non-institutionalized persons. It is possible, then, that non-institutionalized beneficiaries are understated in the MCBS. For the other surveys, persons 65 and older may be *overstated* due to respondents under 65 rounding up their ages whereas MCBS beneficiaries would have had to document their ages to qualify for Medicare.

TABLE IV.37

CHARACTERISTICS OF PERSONS 65 AND OLDER:
MCBS AND CENSUS BUREAU SURVEYS

Characteristic	CPS	ACS	SIPP	MCBS[b]
Total Persons	34.22	33.56	33.95	31.99
		Percent of Persons		
Sex				
Male	42.4	42.3	42.3	42.9
Female	57.6	57.7	57.7	57.1
Race/Ethnicity				
White, non-Hispanic	81.9	82.4	82.7	81.8
Black, non-Hispanic	8.4	8.1	8.0	8.1
Hispanic	6.0	5.7	5.8	6.1
Other	3.8	3.8	3.4	3.9
Family Composition				
Single[a]	33.1	33.5	32.5	33.7
With a spouse only	47.2	47.3	49.3	45.0
With a spouse and others	9.5	9.7	9.2	8.1
With others only	10.1	9.5	9.0	13.1
With SSI	3.5	4.2	6.0	NA
With Medicaid	9.6	NA	14.2	11.6
Health status fair or poor	35.1	NA	33.4	21.3
With inpatient stay	NA	NA	18.5	21.4

Source: Mathematica Policy Research, Inc., from 2003 CPS ASEC supplement,
the 2002 ACS, the 2001 SIPP panel, and the 2003 MCBS Cost and Use file.

[a] Includes persons living with a non-relative.
[b] Medicare beneficiaries only.

196

The sole MCBS income variable that is reported in dollars represents the income of both the sample beneficiary and spouse, if present. To confirm that spouse incomes were indeed being reported, we calculated per capita income and aggregate income under the assumption that the reported amount applied to the sample beneficiary alone. Under this assumption, the MCBS obtains more aggregate income ($939.8 billion) than any of the Census Bureau surveys, which range from $683.2 billion in SIPP to $796.5 billion in the ACS (Table IV.38). It is readily apparent from the per capita income calculations (aggregate income divided by the number of persons 65 and older) that the MCBS is indeed obtaining income for both the respondent and spouse. The per capita income for persons living with only a spouse is slightly higher than that for singles in the CPS and SIPP but it is nearly double the per capita income for singles in the MCBS: $39,022 versus $20,661.

Given that the MCBS income data include spouses' incomes, the incomes of spouses who are Medicare enrollees 65 and older are represented twice (or double-counted, in effect) when the reported incomes of sample members are aggregated. Because the sample members are weighted to the number of Medicare beneficiaries by age, each such spouse is represented by another sample member, and this is what produces the double counting. The survey could eliminate this problem by requesting only the sample member's income. If the incomes of other family members were collected separately, and the number of other family members were counted as well, then it would also be possible to determine the poverty status of each sample member.

Given the limitations of the MCBS income data, the best way to assess how much income the survey is capturing relative to the Census Bureau surveys is to compare singles across the surveys. From Table IV.38 we see that the per capita income of singles in the MCBS lies between the CPS and SIPP estimates. More specifically, the MCBS estimate is $1,600 above the

TABLE IV.38

DERIVATION OF PER CAPITA INCOME OF PERSONS 65 AND OLDER:
MCBS AND CENSUS BUREAU SURVEYS

	CPS	ACS	SIPP	MCBS[b]
	Millions of Persons			
All Persons	34.22	33.56	33.95	31.99
Family Composition				
Single[a]	11.34	11.24	11.03	10.79
With spouse only	16.16	15.88	16.74	14.40
With spouse and others	3.26	3.26	3.11	2.60
With others only	3.46	3.17	3.06	4.20
	Billions of Dollars			
All Persons	730.1	796.5	683.2	939.8
Family Composition				
Single[a]	242.4	256.0	210.0	222.8
With spouse only	369.0	420.6	366.7	562.0
With spouse and others	65.8	65.3	58.7	85.5
With others only	53.0	54.5	47.9	69.5
	Income Per Capita			
All Persons	21,335	23,732	20,124	29,375
Family Composition				
Single[a]	21,379	22,777	19,033	20,661
With spouse only	22,836	26,479	21,901	39,022
With spouse and others	20,154	20,012	18,844	32,861
With others only	15,301	17,194	15,639	16,530

Source: Mathematica Policy Research, from tabulations of calendar year 2002
income from the 2003 CPS ASEC supplement, the 2001 SIPP panel,
and the 2003 MCBS Cost and Use file, and prior 12 months income,
inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] Includes persons living with a non-relative.
[b] Medicare beneficiaries only. Income, reported for 2003, has been deflated to
2002 dollars by the CPI-U.

SIPP estimate and $700 below the CPS estimate. In addition, the MCBS estimate is $2,100 below the ACS estimate.

A comparison of the four surveys with respect to the distribution of singles' incomes by brackets shows that the MCBS finds somewhat fewer people in the tails ($10,000 or less or $50,001 or more) and somewhat more people in the middle bracket (Table IV.39). For single elderly persons, then, the MCBS income data bear a reasonable resemblance to the data collected in the Census Bureau surveys, but this is a very limited assessment.

2. Persons 51 and Older

Our comparative analysis of income data from the HRS is based on the RAND file, which contains a constructed measure of family income without the value of Food Stamps (included by RAND in constructed income for sample persons). We selected this variable so that we would be able to estimate poverty status. While we cannot aggregate family income because this would double count the incomes of spouses and other family members, we *can* calculate the average family income of persons 51 and older and in so doing obtain comparable estimates across surveys.

One other point about our comparisons should be noted. While the HRS collects data from age-eligible sample members and their spouses, the records of spouses who are not themselves age-eligible are not assigned weights.[37] Furthermore, about half of the youngest age-eligible sample members and spouses—that is, those who were born in 1953—were still 50 at the time of their 2004 interviews while the other half had turned 51. To make the comparison samples comparable on age, we chose to restrict our estimates to persons 51 and older.

---

[37] For the 2004 HRS, an age-eligible sample member or spouse was born before 1954.

TABLE IV.39

DISTRIBUTION OF PERSONAL INCOME AMONG PERSONS 65 AND
OLDER AND LIVING WITH NO RELATIVES:
MCBS AND CENSUS BUREAU SURVEYS

Income	CPS	ACS	SIPP	MCBS[b]
	Millions of Persons			
Single[a]	11.34	11.24	11.03	10.79
Income				
$10,000 or less	3.06	3.11	3.02	2.66
$10,001 to $20,000	4.65	4.01	4.36	4.13
$20,001 to $35,000	2.15	2.41	2.53	2.59
$35,001 to $50,000	0.70	0.87	0.68	0.90
$50,001 or more	0.77	0.84	0.45	0.50
	Percent of Persons			
Single[a]	100.0	100.0	100.0	100.0
Income				
$10,000 or less	27.0	27.7	27.3	24.7
$10,001 to $20,000	41.0	35.7	39.5	38.3
$20,001 to $35,000	18.9	21.5	23.0	24.0
$35,001 to $50,000	6.2	7.7	6.1	8.4
$50,001 or more	6.8	7.4	4.1	4.6

Source: Mathematica Policy Research, from tabulations of calendar year 2002
income from the 2003 CPS ASEC supplement, the 2001 SIPP panel,
and the 2003 MCBS Cost and Use file, and prior 12 months income,
inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] Includes persons living with a non-relative.

[b] Medicare beneficiaries only. Income, reported for 2003, has been deflated to
2002 dollars by the CPI-U.

The weighted total persons 51 and older in the HRS exceeds those of the three Census Bureau surveys by 4.0 to 5.7 million (Table IV.40). Distributions by age, sex, and race/ethnicity are similar across the four surveys. HRS sample members are somewhat less likely to be single and more likely to be living with relatives in addition to a spouse (typically their children). Reported receipt of SSI and welfare or Food Stamps in the HRS is similar to the CPS and ACS but substantially below SIPP—3.2 percent versus 5.3 percent for SSI and 4.5 percent versus 6.3 percent for welfare and/or Food Stamps. The proportion reporting a health status of fair or poor is essentially the same across the HRS, CPS, and SIPP.

The average family income of persons 51 and older in the HRS is 27 percent higher than the comparable figure from the CPS (Table IV.41). At $72,303 the average family income from the HRS exceeds the CPS estimate by $15,500, the ACS estimate by $13,700, and the SIPP estimate by nearly $20,800. The HRS exceeds the other surveys by a somewhat greater margin proportionately among persons living with spouses versus no relatives. Because couples have more than double the family income of singles, the gap between the HRS and the other surveys is much greater for sample members living with a spouse than living alone. Among singles, the HRS average income exceeds the CPS average by $6,000. Among married persons the HRS average family income exceeds the CPS estimate by nearly $18,000.

The quintile boundaries are higher than those of the other surveys (Table IV.42). At the 20th percentile the HRS exceeds the CPS by $3,000. At the 80th percentile the HRS exceeds the CPS by $14,000. The ratio of the 80th to the 20th percentiles, one of the measures of income inequality used earlier in this chapter, is essentially the same in the two surveys, however (5.10 in the HRS compared to 5.18 in the CPS).

Average family income by quintile shows a similar pattern but the gap between the HRS and the CPS jumps to $54,000 in the top quintile (Table IV.43). While the ratio of average family

TABLE IV.40

CHARACTERISTICS OF PERSONS 51 AND OLDER:
HRS AND CENSUS BUREAU SURVEYS

Characteristic	CPS	ACS	SIPP	HRS
Total Persons	76.15	74.44	75.38	80.18
		Percent of Persons		
Sex				
Male	45.6	45.4	45.6	46.0
Female	54.4	54.6	54.4	54.0
Age				
51 to 64	55.1	54.9	55.0	55.8
65 and older	44.9	45.1	45.0	44.2
Race/Ethnicity				
White, non-Hispanic	79.3	79.6	80.4	81.2
Black, non-Hispanic	9.3	9.2	9.0	9.3
Hispanic	7.0	6.7	6.8	6.9
Other	4.4	4.6	3.8	2.6
Family Composition				
Single	25.8	26.8	26.6	22.2
With a spouse/partner only[a]	46.1	46.1	46.0	45.3
With other relatives[b]	28.1	27.2	27.5	32.6
With welfare or food stamps	3.7	5.0	6.3	4.5
With SSI	3.2	3.5	5.3	3.2
Health status fair or poor	26.1	NA	25.9	26.5

Source: Mathematica Policy Research, Inc., from 2003 CPS ASEC supplement,
the 2002 ACS, the 2001 SIPP panel, and the 2004 HRS.

[a] Includes persons living with a spouse or (HRS only) unmarried partner but no
other relatives of either.
[b] Includes both married and unmarried persons living with other relatives.

TABLE IV.41

AVERAGE FAMILY INCOME BY FAMILY COMPOSITION:
HRS AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	HRS
	Family Income in Dollars			
All Persons	56,800	58,625	51,546	72,303
Family Composition				
Single	26,954	28,522	24,713	32,974
With a spouse/partner only[a]	63,156	66,365	57,013	81,039
With other relatives[b]	73,764	75,177	68,336	86,916
	Percent of CPS			
All Persons	100.0	103.2	90.7	127.3
Family Composition				
Single	100.0	105.8	91.7	122.3
With a spouse/partner only[a]	100.0	105.1	90.3	128.3
With other relatives[b]	100.0	101.9	92.6	117.8

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income
from the 2003 CPS ASEC supplement, the 2001 SIPP panel, and the 2004 HRS
(reported for 2003 but deflated to 2002 dollars by the CPI-U) and prior 12 months
income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] Includes persons living with a spouse or (HRS only) unmarried partner but no other
relatives of either.
[b] Includes both married and unmarried persons living with other relatives.

203

TABLE IV.42

QUINTILES OF FAMILY INCOME AMONG PERSONS 51 AND OLDER:
HRS AND CENSUS BUREAU SURVEYS

Quintile Boundaries	CPS	ACS	SIPP	HRS
	Family Income in Dollars			
Percentile Value				
20 %-ile	16,348	17,900	17,892	19,359
40 %-ile	30,600	32,900	31,020	36,200
60 %-ile	50,380	52,400	47,743	58,923
80 %-ile	84,721	85,400	75,087	98,788
Ratio of 80th to 20th %-ile	5.18	4.77	4.20	5.10
	Percent of CPS			
Percentile Value				
20 %-ile	100.0	109.5	109.4	118.4
40 %-ile	100.0	107.5	101.4	118.3
60 %-ile	100.0	104.0	94.8	117.0
80 %-ile	100.0	100.8	88.6	116.6
Ratio of 80th to 20th %-ile	100.0	92.1	81.0	98.5

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income
from the 2003 CPS ASEC supplement, the 2001 SIPP panel, and the 2004 HRS
(reported for 2003 but deflated to 2002 dollars by the CPI-U) and prior 12 months
income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

TABLE IV.43

AVERAGE FAMILY INCOME BY QUINTILE OF FAMILY INCOME:
HRS AND CENSUS BUREAU SURVEYS

Income Estimate	CPS	ACS	SIPP	HRS
All Persons	56,800	58,625	51,546	72,303
Family Income Quintile				
Lowest	9,795	10,439	11,030	11,442
Second	23,271	25,134	24,317	27,428
Third	39,661	42,170	39,047	46,933
Fourth	65,756	67,110	60,069	76,563
Highest	145,530	148,356	123,312	199,246
Ratio of fourth to lowest	6.71	6.43	5.45	6.69
Ratio of highest to lowest	14.86	14.21	11.18	17.41
		Percent of CPS		
All Persons	100.0	103.2	90.7	127.3
Family Income Quintile				
Lowest	100.0	106.6	112.6	116.8
Second	100.0	108.0	104.5	117.9
Third	100.0	106.3	98.5	118.3
Fourth	100.0	102.1	91.4	116.4
Highest	100.0	101.9	84.7	136.9
Ratio of fourth to lowest	100.0	95.8	81.1	99.7
Ratio of highest to lowest	100.0	95.6	75.2	117.2

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, and the 2004 HRS (reported for 2003 but deflated to 2002 dollars by the CPI-U) and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

income between the fourth and lowest quintiles is the same in the two surveys, the ratio between the highest and lowest quintiles exceeds that in the CPS by 17 percent.

Poverty rates in the HRS and the Census Bureau surveys are more similar than we might have guessed from the differences in average family income. The poverty rate of 8.4 percent in the HRS is a full percentage point below the CPS poverty rate, but it lies between the ACS and SIPP poverty rates (Table IV.44). The fraction of persons 51 and older who are near near-poor or low-income in the HRS (15.7 percent and 24.1 percent, respectively) is below that of the other three surveys, however. For the low-income population the differences range from 2.0 to 4.6 percentage points.

Does the HRS simply capture more income than the other surveys, or does it over-represent higher income families? We asked the same question with respect to the PSID, which has run for much longer than the HRS. After 40 years, it is easy to imagine that the PSID would have drifted from its most representative state. Nevertheless, the data we examined did not allow us to answer that question for the PSID. For the HRS, the differences with the other surveys are more substantial, particularly at higher income levels. Yet the comparison of selected characteristics did not reveal anything striking. With respect to those characteristics, the HRS is not markedly different from the other surveys. We are left with the observation that HRS incomes are higher than those of the three Census Bureau surveys, but resolving whether this is due to better measurement or over-representation of higher-income families must be left to future research.

G. INTERNAL CONSISTENCY

Consistency between total income and its sources or between reported employment and reported income from employment is an important indicator of data quality. Internal consistency can be achieved through the design of the survey instrument or through the application of consistency checks in the editing procedures that are invoked in processing the raw survey data.

TABLE IV.44

ESTIMATES OF THE POOR AND NEAR POOR:
HRS AND CENSUS BUREAU SURVEYS

Estimate	CPS	ACS	SIPP	HRS
	Millions of Persons			
All Persons	76.15	74.44	75.38	80.18
Poverty Status				
Poor	7.19	6.56	6.15	6.73
Near Poor	14.68	12.89	14.02	12.60
Total Low Income	21.87	19.45	20.18	19.33
	Percent of the Population			
All Persons	100.0	100.0	100.0	100.0
Poverty Status				
Poor	9.4	8.8	8.2	8.4
Near Poor	19.3	17.3	18.6	15.7
Total Low Income	28.7	26.1	26.8	24.1

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the 2003 CPS ASEC supplement and the 2001 SIPP panel, in calendar year 2003 from the 2004 HRS, and in the prior 12 months, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: The poor have a family income below the poverty threshold. The near poor have a family income at or above the poverty threshold but below twice the poverty threshold.

Earlier (section C.2) we documented discrepant reports of the source of earnings in MEPS income data and the separate JOBS file, noting that these data are collected in separate parts of the instrument, with much of the JOBS data coming from different interviews than the annual income data, and that AHRQ has opted to preserve discrepant reports rather than impose consistency edits that would eliminate the independent information contained in the separate responses. We noted that the JOBS file provided information that could be used, potentially, to reclassify reported wages and salaries that, most likely, should have been reported as business income. Here we examine internal consistency between reported family income and reported earnings in the NHIS and between reported earnings and (1) reported receipt of earnings in NHIS or (2) reported work activity in MEPS and SIPP.

1. Family Income and Earnings

In addition to their total family income, NHIS respondents are asked to report the annual earnings of every family member 18 and older. While earnings can include losses from a business, and the sum of earnings over family members is sometimes negative, the difference between total family income and family earnings should be positive in most cases and no less than zero. In fact, however, family earnings often exceed total family income in the 2003 NHIS internal file. This is true for an estimated 61.7 million persons or 21.7 percent of the population (Table IV.45).[38] Over all families the excess of family earnings over total family income sums to $289 billion, with almost half this amount occurring among families with incomes at least four times the poverty threshold.[39] Nevertheless, family earnings are somewhat more likely to exceed

---

[38] These estimates are based on the NHIS family definition, as this is the family unit for which family income is reported.

[39] This estimate was derived by assigning the excess to the family reference person and summing the amounts using the reference person's weight.

TABLE IV.45

NUMBERS OF PERSONS AND EXCESS OF FAMILY EARNINGS OVER
FAMILY INCOME IN NHIS FAMILIES IN WHICH
FAMILY EARNINGS EXCEED FAMILY INCOME

Povery Relative and Excess Earnings	Number of Persons (1,000s)	Excess of Family Earnings Over Family Income ($Billions)	Percent of Persons by Poverty Relative	Percent of Persons With Either of Family Incomes Allocated
Total Persons	61,673	289.4	21.7	71.5
Poverty Relative Based On Family Income				
Under 100%	9,852	28.1	25.3	82.8
100% to under 200%	11,364	38.3	21.1	79.9
200% to under 400%	18,750	83.8	21.4	70.1
400% and over	21,707	139.1	21.0	63.0
Excess of Family Earnings Over Family Income				
$1,000 or less	7,511	1.4	NA	43.1
$1,001 to $5,000	14,810	15.3	NA	60.4
$5,001 to $10,000	11,727	30.4	NA	68.9
$10,001 to $20,000	12,270	59.3	NA	84.3
$20,001 to $40,000	10,334	90.6	NA	89.0
$40,001 or more	5,021	92.3	NA	85.0

Source: Mathematica Policy Research, from tabulations of calendar year 2002
income from the 2003 NHIS.

Note: Estimates use family composition as reported in the survey (NHIS families).

family income when the latter is below versus above poverty: 25.3 percent versus a little over 21 percent.

In most cases, when family earnings exceed family income, the difference is not small. Nearly half the time the excess is $10,000 or more, and differences in excess of $20,000 account for nearly a quarter of the total instances.

Why does this phenomenon occur? In a number of cases that we reviewed we found that the excess earnings could be attributed to a family member whom the respondent had not included in the family income total—such as a child or an unmarried partner. Clearly, the respondent was defining the family more narrowly than the survey interviewer or simply omitting less salient or less central members. Such cases illustrate one limitation of asking respondents to report family income as a single amount rather than collecting income for each person.

Overall, however, reporting error accounted for less than a third of the instances of family earnings exceeding total family income. Allocation of family income, personal earnings, or both was responsible for 71.5 percent of all cases. Furthermore, allocation grew in importance as the magnitude of the excess of family earnings over family income increased. When the excess was $1,000 or less, allocation accounted for 43 percent of the occurrences. When the excess was greater than $10,000, allocation accounted for more than 84 percent of the instances.

If the family sum of personal earnings were substituted for total family income when the former exceeded the latter—one form of a consistency edit—then 16.3 million or 5.7 percent of all persons would be shifted to a higher poverty bracket (Table IV.46). The estimated number of persons with family incomes below poverty would be reduced by 4 million, and the poverty rate would be reduced by 1.4 percentage points. These effects on persons in poverty are true regardless of whether families are defined with the NHIS or CPS family concept. At the other end of the income distribution, the number of families with incomes above 400 percent of

TABLE IV.46

IMPACT OF SUBSTITUTING FAMILY EARNINGS FOR FAMILY INCOME
WHEN FAMILY EARNINGS ARE LARGER: NHIS AND CPS FAMILIES

| | NHIS Families | | CPS Families | |
| | Millions of Persons | Percent of Persons | Millions of Persons | Percent of Persons |
Population				
	Gross Change: Persons Moved to a Higher Poverty Bracket			
Total	16.31	5.74	[a]	[a]
Poverty Relative Based on Family Income				
Under 100%	3.98	1.40		
100% to under 200%	5.57	1.96		
200% to under 400%	6.76	2.38		
400% and over	0.00	0.00		
	Net Change: Net Loss or Addition to Poverty Bracket			
Total	0.00	0.00	0.00	0.00
Poverty Relative Based on Family Income				
Under 100%	-3.98	-1.40	-3.90	-1.38
100% to under 200%	-2.48	-0.87	-2.57	-0.90
200% to under 400%	-0.83	-0.29	-0.64	-0.23
400% and over	7.29	2.57	7.11	2.51

Source: Mathematica Policy Research, from tabulations of poverty status in
calendar year 2002 from the 2003 NHIS.

[a] Gross change was estimated only for NHIS families.

poverty would be increased by more than 7 million or about 2.5 percentage points. In addition, the ratio of per capita earnings between the top and bottom quintiles would drop from 8.34, which is highest among the five surveys, to 7.57 (data not shown), matching the CPS (compare Table IV.3). These are large impacts, but they also reflect the substantial role of allocation in producing excess earnings. An excess of family earnings over family income may suggest that family income is understated when both amounts were reported by respondents. When one or both amounts were imputed, the implications are more ambiguous. In part for this reason, the study did not make use of excess earnings in assigning incomes to simulated CPS families. Instead, the combined incomes of families that were created by dividing an NHIS family were constrained to equal the reported (or imputed) family income of the original NHIS family.

2. Work Activity and Earnings

In some of the surveys, data on work activity and the income from that activity are collected together. Questions ask respondents about their work activity, and those who report such activity are asked how much income they received from it.[40] With this approach, respondents cannot report earnings without first reporting employment. If they do report employment, then there is an income amount associated with it—or a missing data item to be imputed. In the other surveys, questions on work activity and earnings occur in separate parts of the instrument, which does not preclude respondents from reporting one without the other. If the skip logic in the instrument does not enforce consistency, then edits may be required to ensure that respondents do not end up with earned income without work activity or vice versa.

In our review of survey procedures, we found that among the five major surveys the CPS and ACS include consistency checks in their data processing procedures to ensure that there is

---

[40] Unpaid activity may be addressed separately.

income associated with all work activity, and work activity associated with all earned income reported in the final data file. NHIS collects work activity and annual earnings together, and the latter is always positive when employment in the prior year is indicated. However, respondents are asked elsewhere in the questionnaire whether each adult family member received income from wages and salaries or self-employment in the prior year, and the responses to these questions are not edited for consistency with reported employment or earnings. MEPS collects annual earnings separately from work activity and does not include consistency checks in the editing. SIPP collects work activity and earnings in the same part of the instrument, but because this information is captured monthly, and employment may start or end in the month before the income from that activity, consistency is not forced at that level, either in the skip patterns or subsequent editing. We examine the incidence of inconsistency between the presence of earnings and reported receipt (NHIS) or work activity (MEPS and SIPP).

In the NHIS, of those who were reported as having received income from wages and salaries or self-employment during the year, an estimated 4.3 million persons had no reported earnings (Table IV.47). Similarly, an estimated 4.0 million persons with reported annual earnings totaling $105.3 billion had no reported receipt of income from wages and salaries or self-employment during the same period.

For MEPS the JOBS file contains detailed employment data covering the same period of time as the annual income data collected elsewhere in the instrument.[41] Using these data we find that an estimated 2.6 million persons had one or more jobs working for others or in their own businesses during the year but reported no wage and salary or self-employment income for the same time period (Table IV.48). Another 6.6 million persons reported wage and salary or self-

---

[41] As noted earlier, the JOBS data are derived from interviews conducted two to three times a year whereas the annual income data are from a single interview conducted after the end of a year.

TABLE IV.47

ESTIMATES OF CONSISTENCY BETWEEN REPORTED RECEIPT OF
INCOME FROM WAGES AND SALARIES OR SELF-EMPLOYMENT
AND REPORTED WORK ACTIVITY WITH EARNINGS: NHIS

Estimate	With Reported Receipt But No Reported Earnings	With Reported Earnings But No Reported Receipt	With Both Reported Receipt and Reported Earnings
Millions of Persons	4.35	3.98	143.37
$Billions of Earnings	0.0	105.3	5,156.2

Source: Mathematica Policy Research, from tabulations of work activity and
          earnings in calendar year 2002 from the 2003 NHIS.

TABLE IV.48

ESTIMATES OF CONSISTENCY BETWEEN REPORTED WORK ACTIVITY
AND REPORTED EARNINGS: MEPS AND SIPP

Survey	With Reported Work Activity But No Reported Earnings	With Reported Earnings But No Reported Work Activity	With Both Reported Work Activity And Reported Earnings
	Millions of Persons		
MEPS	2.60	6.61	153.81
SIPP	0.08 [a]	0.38	153.68
	$Billions of Earnings		
MEPS	0.0	99.7	5,164.1
SIPP	0.0	1.1	4,759.0

Source: Mathematica Policy Research, from tabulations of work activity and earnings in calendar year 2002 from the 2001 SIPP panel and the 2002 Full-year Consolidated MEPS-HC.

[a] Initially, we identified 2.06 million SIPP respondents with reported work activity but no reported earnings. Census Bureau staff determined that an error in the questionnaire skip logic accounted for 1.97 million of this number.

employment income totaling $99.7 billion for the year but gave no indication of work activity over that same period.

With SIPP we initially identified an estimated 2.1 million persons with work activity but no reported earnings, but on pursuing this matter with the Census Bureau we learned that nearly all of this entire number—all but 0.08 million—was due to an error in the skip logic that was corrected in the 2004 panel. Because of the error, questions on the income from a business (which could include net profit or loss) were skipped for self-employed persons in sole-proprietorships and some partnerships when no monthly draw (salary paid to oneself) was reported. We found an additional 0.4 million persons who reported earnings but no work activity. Thus, while SIPP does not edit or impute monthly work activity against monthly earnings or monthly earnings against monthly work activity, we identified fewer than 0.5 million persons with either work activity but no earnings or earnings but no work activity on an annual basis. This contrasts with an estimated 9.2 million in MEPS.

If those reporting the receipt of earned income but no dollars in NHIS, or work activity but no dollars in MEPS, and those skipped around the self-employment income questions in SIPP are included in the population of earners, the number of earners would be increased to 151.7 million in NHIS, 163.0 million in MEPS, and 156.0 million in SIPP (Table IV.49). In contrast, the estimated 150.4 million earners in the CPS, which is edited and thus unchanged, would become the smallest estimate while the ACS would remain in the middle with 151.9 million. If, at the same time, the 6.6 million with reported earnings but no reported work activity were removed from the MEPS estimate of earners (on the grounds that they may have misreported the source of their income), the MEPS estimate of total earners would drop to 156.4 or narrowly more than the SIPP estimate.

TABLE IV.49

IMPACT ON ESTIMATES OF PERSONS WITH EARNINGS IF PERSONS REPORTING
WORK ACTIVITY OR RECEIPT OF EARNINGS BUT NO EARNED INCOME ARE INCLUDED

Employment	CPS	ACS	SIPP	MEPS	NHIS
			Millions of Persons		
Persons with Earned Income Reported	150.44	151.93	154.06	160.42	147.35
Additional Persons with Evidence of Earnings	0.00	0.00	1.97	2.60	4.35
Total Persons with Evidence of Earnings	150.44	151.93	156.03	163.02	151.70

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

V. COMPARISONS ACROSS DESIGN, DEFINITIONAL AND METHODOLOGICAL ISSUES

Collectively, the eight surveys reflect a number of different choices with respect to design, definitions, and methodology that have implications, potentially, for the measurement of income. Cross-survey comparisons of the kind presented in Chapter IV tell us little about the impact of such choices because (1) the survey designs differ along multiple dimensions and (2) the impact of differences with respect to field procedures, editing, and other aspects of post-survey processing cannot be removed from the estimates that are being compared. To isolate the impact of individual design, definitional, or methodological features, we turn to simulations conducted within individual surveys. This has the advantage of neutralizing the impact of any design or field differences outside of those being evaluated. In this chapter we use this approach to examine the impact of family definition and relationship detail, how surveys deal with the dynamics of family composition, issues raised by rolling samples, the treatment of retirement income, and the use of income as a component of post-stratification. All five of the general population surveys contribute to these simulations, with different surveys called on to address different survey features.

A. FAMILY DEFINITION AND RELATIONSHIP DETAIL

Five of the eight surveys utilize family definitions that deviate from the CPS family concept that is incorporated into the official measure of poverty in the United States, and we have noted (p. 24) differences among the eight surveys in the level of detail with which they capture family relationships. It is important to understand that broadening the family concept from the CPS concept has an impact on estimates of family income and the incidence of poverty. Here we develop estimates of the impact of deviations from the CPS family definition on estimates of the

poor and their characteristics and on the distribution of family income. Following that we assess the implications of the ACS's not collecting relationship information among persons who are unrelated to the household head.

1. Family Concept

The response unit in the NHIS is the family, and families are defined to include unmarried partners and foster children. Family income is collected as a single amount for the entire family. In developing the NHIS estimates of income for comparison with the other surveys, we separated unmarried partners and foster children from the NHIS family and apportioned family income among the two or more family units created from each NHIS family and which conform to the CPS family definition.[39] By comparing the income and poverty estimates that we prepared using the CPS family definition with estimates obtained from the original data, we can assess the impact of using the NHIS versus CPS family definitions to group individuals for the purposes of estimating family income.

MEPS also uses the family as its response unit and defines the family in the same way as the NHIS. However, in order to post-stratify the sample weights to the CPS poverty distribution, AHRQ (or its MEPS survey contractor) defines CPS families within the broader MEPS families. Income, which is reported at the person level, can be aggregated to either family definition using alternative family identifiers on the public use file. We used the CPS family to prepare the income estimates reported in Chapter IV, but by preparing an alternative set of estimates based on the MEPS/NHIS family definition, we can assess the impact of using one versus the other family definition just as we do with the NHIS.

---

[39] See Chapter III, section A.3.a.

220

Our estimates of the impact of the NHIS versus CPS family definitions based on the NHIS and MEPS are remarkably similar. In both surveys we find that the NHIS family definition reduces the number of persons in poverty by 2.6 million and reduces the poverty rate by 0.9 percentage points (Table V.1).[40] There is no impact in either survey on the percentage of persons between 100 and 200 percent of poverty, which means that the number of people who were moved above the poverty line by the NHIS family concept is offset by the number of people who were moved beyond 200 percent of poverty. Most of the upward shift is observed in the top category—that is, among people above 400 percent of poverty, where the broader family concept adds 2.3 million to the number in the NHIS and 1.4 million to the number in MEPS.

We also assessed the impact of the NHIS family definition by demographic characteristics in both surveys. For the NHIS, the reduction in the poverty rate and the number of poor was about twice as great among women as among men (Table V.2). The reduction in the poverty rate was greatest among children under 18 (1.3 percentage points) and least among the elderly (0.3 percentage points). The reduction in the *number* of poor was greatest among persons 18 to 64 at 1.6 million, as this is the largest age group, but the reduction among children was still 0.9 million. The reduction in the poverty rate was essentially the same across four race/ethnicity groups at around a percentage point.

We find generally similar patterns for MEPS, but the broader family concept appears to produce somewhat bigger declines in the poverty rate among blacks and Hispanics than among whites and others (Table V.3). With MEPS we produced consistent measures of family

---

[40] The estimate from the NHIS has an upper and lower bound. As explained in Chapter III, we developed two alternative allocations of NHIS family income to the CPS families that we created. One alternative allocated family income in excess of total personal earnings in such a way as to maximize the number of poor in the CPS families. The other alternative allocated the excess family income in such a way as to minimize the number of poor in the CPS families. Ultimately, we averaged the two results, but the difference in the total number of poor persons with one alternative versus the other was 460,000, so our estimate of 2.6 million has a range of plus or minus 230,000.

TABLE V.1

COMPARISON OF THE CPS AND NHIS/MEPS FAMILY CONCEPTS
WITH RESPECT TO THE ESTIMATED DISTRIBUTION OF
PERSONS BY INCOME RELATIVE TO POVERTY

Family Income as Percent of Poverrty	NHIS			MEPS		
	CPS Family	NHIS Family	Change	CPS Family	MEPS Family	Change
			Percent of Persons			
Total Percent	100.0	100.0		100.0	100.0	
Under 100%	14.7	13.7	-0.9	12.5	11.5	-0.9
100% to under 200%	19.0	19.0	0.0	18.4	18.4	0.0
200% to under 400%	30.7	30.9	0.2	31.7	32.1	0.4
400% or more	35.7	36.4	0.8	37.4	37.9	0.5
			Number of Persons (millions)			
Total Persons	283.7	283.9	0.2	283.3	283.3	0.0
Under 100%	41.6	39.0	-2.6	35.3	32.7	-2.6
100% to under 200%	53.9	53.8	-0.1	52.1	52.2	0.1
200% to under 400%	87.1	87.7	0.6	89.8	90.9	1.1
400% or more	101.2	103.4	2.3	106.0	107.4	1.4

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002
from the 2003 NHIS and the 2002 Full-year Consolidated MEPS-HC.

TABLE V.2

COMPARISON OF THE CPS AND NHIS FAMILY CONCEPTS
WITH RESPECT TO THE NUMBER AND PERCENT POOR
BY DEMOGRAPHIC CHARACTERISTICS: NHIS

Demographic Characteristic	Percent Poor			Number Poor (millions)		
	CPS Family	NHIS Family	Change	CPS Family	NHIS Family	Change
Gender						
Male	13.3	12.6	-0.6	18.4	17.5	-0.9
Female	16.0	14.8	-1.2	23.2	21.4	-1.7
Age						
Under 18	19.9	18.6	-1.3	14.3	13.4	-0.9
18 to 64	13.2	12.3	-0.9	23.5	21.9	-1.6
65 and older	11.0	10.7	-0.3	3.8	3.7	-0.1
Race/Ethnicity						
White non-Hispanic	9.8	8.9	-0.9	19.1	17.4	-1.7
Black non-Hispanic	26.2	25.0	-1.1	9.1	8.8	-0.4
Hispanic	28.0	26.8	-1.2	11.0	10.6	-0.4
Other	16.3	15.4	-0.8	2.3	2.2	-0.1

Source: Mathematica Policy Research from the 2003 NHIS.

TABLE V.3

COMPARISON OF THE CPS AND MEPS FAMILY CONCEPTS
WITH RESPECT TO THE NUMBER AND PERCENT POOR
BY DEMOGRAPHIC CHARACTERISTICS: NHIS

Demographic Characteristic	Percent Poor			Number Poor (millions)		
	CPS Family	MEPS Family	Change	CPS Family	MEPS Family	Change
Gender						
Male	11.2	10.5	-0.7	15.5	14.5	-1.0
Female	13.7	12.6	-1.1	19.9	18.2	-1.6
Age						
Under 18	17.4	15.7	-1.7	12.5	11.3	-1.2
18 to 64	10.7	9.9	-0.8	19.0	17.6	-1.4
65 and older	11.3	11.2	0.0	3.8	3.8	0.0
Race/Ethnicity						
White non-Hispanic	8.3	7.6	-0.7	15.9	14.7	-1.3
Black non-Hispanic	24.9	23.5	-1.5	8.8	8.3	-0.5
Hispanic	22.4	20.7	-1.8	8.7	8.1	-0.7
Other	11.3	10.4	-0.9	1.9	1.7	-0.1
Family Composition						
Single (18 or older)	17.5	17.0	-0.5	8.8	7.4	-1.5
Childless couple	5.1	5.2	0.0	3.4	3.4	0.0
Single parent	32.7	27.3	-5.4	4.2	3.5	-0.7
Child of single parent	39.8	33.9	-6.0	8.0	6.8	-1.2
Couple with children	6.8	6.8	0.0	3.7	3.7	0.0
Child of couple	7.6	7.6	0.0	3.8	3.8	0.0
Other	11.9	11.5	-0.5	3.5	4.2	0.7

Source: Mathematica Policy Research from the 2002 Full-year Consolidated MEPS-HC.

224

composition for both family concepts, so we were able to examine differential effects of the family concept by family composition. For single parents and their children we see the impact of adding an unmarried partner's income.[41] The poverty rates for single parents and their children decline by five to six percentage points with the NHIS/MEPS family definition.

Finally, by creating a somewhat smaller number of families with a slightly larger average size, the broader family concept increases family incomes across the income distribution. This can be seen by comparing the boundaries between family income quintiles (that is, the 20th, 40th, 60th, and 80th percentiles) between the two family concepts. With the broader family definition, the boundaries between family income quintiles increase by $1,000 to $2,000 in both surveys (Table V.4).

2. Unrelated Subfamilies

While the income data collected in the ACS compare relatively closely to the income data collected in the CPS, the ACS does not identify families among persons unrelated to the householder. That is, the ACS does not identify unrelated subfamilies (see Chapter III). All persons unrelated to the householder must be treated as unrelated (or secondary) individuals when calculating poverty rates with the ACS. As an unrelated individual, if a person's own income is below the poverty threshold for a family of size one, then that individual will be considered as poor. This may result in some persons being classified as poor who would not be considered poor if their subfamily membership were taken into account. It may also result in

---

[41] To show this effect, we do not change the classification of a single parent when we include an unmarried partner in the family. However, the partner who is brought into the family with the broader family concept is generally classified as a single under the CPS family concept and as an "other" person (that is, not belonging to any of the specified categories) under the NHIS/MEPS family concept. With the broader family concept the number of persons classified as "other" grows by several million, which explains why the number of poor in this category increases between the two family concepts while the poverty rate declines.

TABLE V.4

COMPARISON OF THE CPS AND NHIS FAMILY CONCEPTS WITH RESPECT TO
THE BOUNDARIES BETWEEN FAMILY INCOME QUINTILES: MEPS

	NHIS			MEPS		
Quintile Boundaries	CPS Family	NHIS Family	Change	CPS Family	NHIS Family	Change
Percentile Value						
20 %-ile	18,443	20,000	1,557	19,670	21,000	1,330
40 %-ile	34,584	35,801	1,217	37,214	38,791	1,577
60 %-ile	55,000	57,022	2,022	58,000	59,332	1,332
80 %-ile	89,068	90,000	932	87,338	88,313	975

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the
2003 NHIS and the 2002 Full-year Consolidated MEPS-HC.

226

some persons being classified as nonpoor when they would be considered poor as subfamily members.[42]

We used the CPS to estimate the impact of the ACS treatment of unrelated subfamilies. Specifically, we assigned a poverty threshold to each member of an unrelated subfamily, reflecting a family size of one, and we used that person's own total income to calculate a poverty ratio. We then compared the poverty class of each individual when calculated in this way to the poverty class obtained when membership in the unrelated subfamily was taken into account. Unrelated children under 15 are excluded from the universe for the calculation of official poverty rates, and they have been removed from our survey estimates (see Chapter III). However, they may be treated as poor for some policy applications, so we examine the impact of treating unrelated subfamily members as unrelated individuals with and without including children under 15 within the poverty universe.

Table V.5 presents a cross-classification of CPS unrelated subfamily member by their poverty class when their subfamily membership is taken into account (the row variable) and their poverty class when treated as an unrelated individuals (the column variable). Tabulations are presented for all persons, for children under 15, and for persons 15 and older. The key transitions and net results are summarized in Table V.6.

Overall, we find that if we retain all children under 15, then the impact of treating the 1.2 million unrelated subfamily members in the 2003 CPS as unrelated individuals is to increase the number of poor from 413.8 thousand to 812.3 thousand, or close to 400 thousand. If we remove from the universe the 571.7 thousand unrelated children under 15, all of whom would otherwise

---

[42] This will happen, for example, if the principal earner in an unrelated subfamily has an income that is above the poverty threshold for a family of size one but below the poverty threshold for that person's actual family size.

TABLE V.5

POVERTY CLASS OF UNRELATED SUBFAMILY MEMBERS BY POVERTY CLASS
WHEN CLASSIFIED AS UNRELATED (SECONDARY) INDIVIDUALS: CPS

| Age and Original Poverty Class | New Poverty Class When Treated as an Unrelated Individual | | | | Total by Age and Original Poverty Class |
	Below 100%	100% to < 200%	200% to < 400%	400% or More	
All Persons	812,332	121,324	198,719	97,885	1,230,260
Below 100%	364,103	49,716	0	0	413,819
100% to < 200%	225,792	66,804	91,945	0	384,541
200% to < 400%	184,184	716	97,605	63,846	346,352
400% or more	38,253	4,088	9,169	34,038	85,548
Persons by Age					
Under 15	571,745	0	0	0	571,745
Below 100%	218,611	0	0	0	218,611
100% to < 200%	173,450	0	0	0	173,450
200% to < 400%	152,565	0	0	0	152,565
400% or more	27,118	0	0	0	27,118
15 and Older	240,587	121,324	198,719	97,885	658,515
Below 100%	145,491	49,716	0	0	195,207
100% to < 200%	52,342	66,804	91,945	0	211,091
200% to < 400%	31,619	716	97,605	63,846	193,787
400% or more	11,135	4,088	9,169	34,038	58,430

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002 from the
2003 CPS ASEC supplement.

TABLE V.6

NET IMPACT OF RECLASSIFYING UNRELATED SUBFAMILY MEMBERS
AS UNRELATED (SECONDARY) INDIVIDUALS, BY AGE: CPS

Transition Status	Under 15	15 and older	Total
People who transition from:			
Poor to nonpoor	0	49,716	49,716
Nonpoor to poor	353,134	95,096	448,230
People who remain			
Poor	218,611	145,491	364,103
Nonpoor	0	368,211	368,211
People who are:			
Poor as subfamily members	218,611	195,207	413,819
Poor when defined as unrelated individuals	571,745	240,587	812,332
Net change in poor			
If unrelated children under 15 are included	353,134	45,380	398,514
If unrelated children under 15 are excluded	-218,611	45,380	-173,231

Source: Mathematica Policy Research, from tabulations of poverty status in calendar year 2002

be counted as poor, then the number of poor persons drops to 240.6 thousand for an overall *reduction* of 173 thousand.

Given that unrelated children under 15 are excluded from the poverty universe for the ACS, our CPS simulation suggests the net effect of the ACS's treating unrelated subfamily members as unrelated individuals is to reduce the estimated number of "officially" poor persons by about 173 thousand relative to what would be observed if unrelated subfamilies could be identified. This result obtains because more than three-quarters of the unrelated subfamily members who transition from nonpoor to poor when treated as unrelated individuals are children under 15, and they are not included in the official poverty universe. Furthermore, more than 200,000 children who would be classified as poor as members of unrelated subfamilies are dropped from the official poverty universe when they are treated as unrelated subfamily members. If, however, unrelated children under 15 are *included* in the poverty universe, as they would be for some policy analyses, then the net effect of the ACS's treating unrelated subfamily members as unrelated individuals is to *increase* the number of poor by close to 400 thousand.

To make the ACS poverty estimates fully comparable to the CPS, then, we would need to add about 173 thousand to the ACS poor, given that we exclude unrelated children under 15 from the poverty universe. However, policy analysts who use the ACS for applications in which unrelated children under 15 are counted as poor would need to *subtract* about 400 thousand from their estimated number of poor persons in order to correct for the survey's treatment of unrelated subfamily members as unrelated individuals.

B.   FAMILY COMPOSITION DYNAMICS AND POVERTY MEASUREMENT

In CPS, ACS, and MEPS, detailed income data are collected for each person, and annual family income is then constructed by summing these person-level amounts for all members of the family as defined at the time the data were collected (CPS and ACS) or the end of the

previous calendar year (MEPS). This method of aggregating income over family members reflects the definition of family income used in the calculation of official poverty statistics, but it embodies a simplified view of family composition. In reality, the people living together as a family at the time the data were collected may not have lived together for the entire income reference year while other individuals, no longer present, may have lived with the family for some or all of the income reference year. For example, a married couple may have lived together during the income reference year, with the husband providing all of the family's income, but divorced before the interview date in the next calendar year. A family consisting of the former wife would report no income for the reference period and be classified as poor. Conversely, a couple who married shortly before the survey date, with the wife having had very little income during the reference year while the husband earned a substantial amount, would be classified as well above poverty when the wife in fact lived in poverty during the reference year. If such cases balance out, the simplification of family composition used in the official definition of poverty will not introduce any bias into the estimates of persons in poverty, but if either type of case predominates, then there *will* be a bias.

If the fixed family composition used in the official definition of poverty does impart a bias, then the magnitude of the bias will depend on how much the family composition lags the income reference period.[43] With a longer lag, more persons will experience changes in family composition. Both the ACS and MEPS fix family composition at the end of the income reference period while the CPS fixes family composition two-and-a-half months later. This suggests that any bias due to changing family composition will be greater in the CPS than in either of these

---

[43] This issue is relevant not only to poverty estimates but to simulated tax calculations, which must approximate the composition of the family at the end of the tax year—and, for some dependents, based on more elaborate residency and support rules.

other surveys. The NHIS collects family income for the prior calendar year from families interviewed over the course of the next calendar year, so family composition lags the end of the income reference period by one-half to 11-and-a-half months, or 6 months on average.

SIPP collects both income and family composition on a monthly basis, so with SIPP data it is possible to construct an annual poverty measure that takes account of changing family composition over the year and reflects the combined incomes of people when they were actually living together as a family. Below, we will explain how this can be done. However, the SIPP estimates of family income and poverty that were constructed for the cross-survey comparisons in the preceding chapter mimic the official concepts, with family composition fixed in the final month of the reference year and family income summed over these same family members.

The PSID collects income for all persons who lived with the sample family during the reference year, but only for the months that they did so. A poverty threshold is constructed to reflect the changing composition of the sample family over the reference year—just as it is possible to do with the SIPP. This yields estimates of income relative to poverty that reflect a contemporaneous measurement of income and family composition. Unlike the SIPP, however, which collects its data at four-month intervals, the PSID asks respondents to recall who was living with the family and how much income they contributed during the prior calendar year.

Neither the HRS nor the MCBS collects income from family members other than the sample member and spouse (or, for the HRS, partner). This limits the construction of poverty measures, so we do not address the timing of family composition relative to the income reference period for these two surveys.

1. Simulating Poverty Measurement with Alternative Timing of Family Composition

To assess the impact of the timing of family composition in relation to the income reference period net of other survey design features, we used the SIPP to perform a set of simulations using

sample members who were present for all of calendar years 2001 and 2002. Persons were included in this subsample if they had data for all 24 months and a longitudinal weight greater than zero.[44] Weighted, the sample members who met these criteria summed to 267.9 million or 95 percent of the population represented in the SIPP comparative income estimates in Chapter IV.

For this fixed sample of individuals we calculated family income in relation to the poverty threshold for 14 alternative scenarios that reflect family composition measured at different times relative to a 2001 income reference year.[45] The first scenario represents a contemporaneous measurement of family income and family composition, which is what the PSID obtains. For this scenario we defined each sample member's 2001 annual family income as the sum of that individual's 12 monthly *family* incomes for the year. Monthly family incomes appear on each sample person's record for the months that they are present. For a given month, the sample member's family income is the sum of the incomes (in that month) of everyone living with the sample member in that month. The corresponding annual poverty threshold is the sum of 12 monthly poverty thresholds that reflect family composition in each month.[46] While family income may include the incomes of persons outside of our fixed sample (the sample members not utilized because they lack the third longitudinal weight or have no data for one or more months of 2001 or 2002), our simulations tabulate only the fixed sample members. It is their

---

[44] Specifically, we used the Census Bureau's third longitudinal weight, which the Census Bureau created for panel members who were present from the beginning through the end of the panel or until they left the survey universe.

[45] We used 2001 as the income reference year so that we could have a full 12 months of SIPP data following the end of the reference year. The 2001 panel did not collect 12 months of data from 2003 for the entire sample.

[46] Monthly poverty thresholds are reported on the SIPP public use file. They are indexed for inflation, consistent with the official annual thresholds.

poverty status, calculated to reflect differential timing of family composition relative to the reference year, that we seek to compare.

In contrast to this first scenario, which reflects each fixed sample member's actual family composition over the 12 months of the income reference year, the next 13 scenarios employ a fixed family composition, which is defined, in turn, for each of the 13 months from December 2001 through December 2002. For example, to construct each sample member's annual family income based on a fixed family composition for December 2001, we first determined who was in a sample member's family in that month. This may have included other members of our fixed sample as well as additional persons who were outside the fixed sample. We then summed the monthly personal income of each family member over the 12 months of calendar year 2001 to obtain an annual total for every family member. By definition, members of our fixed sample will have had complete data for calendar year 2001, but this may not have been true of the additional sample members (people with longitudinal weights of zero or missing data for one or more months of 2001 and 2002). If a family member had missing data for one or more months of 2001, we created an annual total using a simple ratio adjustment based on the number of months (out of 12) with reported income and the sum of reported income over those months.[47] The 2001 annual family income for this family was then calculated by summing the annual incomes of all the family members. The annual poverty threshold for this scenario was determined from the family composition in December 2001.[48] Both the annual family income and the annual poverty threshold for this family were applied to every member of our fixed sample.

---

[47] For example, if a family member had data for only 10 of the 12 months, we multiplied the sum of that person's monthly incomes over the 10 months by the ratio, 12/10, to obtain a 2001 calendar year income for that person.

[48] A calendar year 2001 annual poverty threshold was used for this purpose.

We followed the same procedures to construct annual family incomes and poverty thresholds for each fixed sample member based on that sample member's family composition in each of the 12 months of 2002. For example, to construct family incomes and poverty thresholds for June 2002, we determined who was in each fixed sample member's family in that month and then summed their monthly incomes for calendar year 2001. In doing so, there was one additional wrinkle that we had to address. For families defined in any month of 2002, a family member who was outside the fixed sample may have had missing data for one or more months of 2001 but reported data for one or months of 2002. Rather than discard the 2002 income data—particularly when we may have had no data at all for 2001—we did the following. If a month of data for 2001 was missing but we had a reported income for the corresponding month in 2002, we substituted the 2002 data (deflated by the increase in the CPI-U between those two months) before applying the ratio adjustment.

2. Contemporaneous versus Fixed Family Composition

Estimates of the impact of fixing family composition at each of the 13 successive months relative to a contemporaneous measurement of income and family composition are reported in Table V.7 for the percentage of persons with annual family incomes below the poverty threshold. Both gross and net differences between each fixed measure and the contemporaneous measure are reported, along with the estimated poverty rate. We see first that fixing family composition at the end of the reference year adds nearly half a percentage point to the estimated poverty rate relative to contemporaneous measurement. Specifically, 0.64 percent of the population who are *not* identified as poor with contemporaneous measurement are classified as poor when family composition is fixed at December 2001 while 0.19 percent who *are* identified as poor with contemporaneous measurement are classified as nonpoor when family composition is fixed at

TABLE V.7

IMPACT OF FIXED FAMILY COMPOSITION ON ESTIMATED PERCENT POOR
BASED ON CY 2001 INCOME: SIPP SIMULATION

| Simulated Timing of Family Composition | Difference in Percentage Classified as Poor With Fixed Composition | | | | Poverty Rate |
	Gross Addition[a]	Gross Reduction[b]	Sum	Difference	
Contemporaneous	0.00	0.00	0.00	0.00	10.64
Fixed at:					
Dec 2001	0.64	0.19	0.83	0.45	11.09
Jan 2002	0.72	0.23	0.95	0.49	11.13
Feb 2002	0.82	0.24	1.06	0.58	11.22
Mar 2002	0.93	0.29	1.22	0.64	11.27
Apr 2002	0.99	0.36	1.35	0.63	11.27
May 2002	1.08	0.41	1.49	0.68	11.32
Jun 2002	1.16	0.48	1.64	0.68	11.32
Jul 2002	1.24	0.55	1.79	0.69	11.33
Aug 2002	1.35	0.54	1.89	0.81	11.45
Sep 2002	1.43	0.57	2.00	0.87	11.51
Oct 2002	1.51	0.58	2.09	0.93	11.57
Nov 2002	1.52	0.59	2.11	0.93	11.57
Dec 2002	1.57	0.63	2.20	0.95	11.59
Avg. Jan-Dec 2002	1.19	0.46	1.65	0.74	11.38

Source: Mathematica Policy Research, from 2001 SIPP panel.

Note: See text for description of simulation.

[a] Percentage of population classified as poor when family composition is fixed in time but nonpoor when family composition is contemporaneous with income.
[b] Percentage of population classified as nonpoor when family composition is fixed in time but poor when family composition is contemporaneous with income.

December 2001. The net difference of 0.45 percent is reflected in the higher poverty rate with fixed versus contemporaneous measurement.

Both the gross and net differences between the contemporary and fixed measures increase as the timing of family composition moves farther from the income reference period. Between December 2001 and December 2002 the gross additions (persons classified as poor by the fixed measure but not by the contemporaneous measure) increase from 0.64 percent to 1.57 percent of the population. The gross reductions (persons classified as poor by the contemporaneous measure but not the fixed measure) increase from 0.19 percent to 0.63 percent. The sum of the gross addition and gross reduction in each row is the percentage of persons who are classified differently with a fixed family composition versus contemporaneous measurement. This fraction grows from 0.83 percent to 2.20 percent between December 2001 and December 2002.

Because the gross difference grows in both directions, the net difference grows less rapidly. Nevertheless, the net difference doubles between December 2001 and December 2002, increasing from 0.45 percent to 0.95 percent. That is, fixing family composition at nearly a year after the end of the income reference period adds almost a full percentage point to the estimated poverty rate.

While contemporaneous measurement of income and family composition is arguably more appropriate than fixing family composition at the end of the income reference period or even some months later, this is not the official approach to measuring poverty; nor is it feasible for most surveys. Moreover, in light of our use of the CPS as a baseline for income measurement, we are interested in how the timing of family composition in the alternative surveys affects their poverty estimates relative to the CPS. The results in Table V.7 suggest that the impact on the poverty rate for the population as a whole is rather small. Fixing family composition at the end of the reference year (SIPP and MEPS) lowers the poverty rate by 0.18 percentage points compared

to fixing family composition three months later (or March, as done in the CPS, on average). Fixing family composition in the month following the income reference period, as the Census Bureau interprets the ACS as doing, reduces the poverty rate by 0.14 percent relative to the CPS (if the CPS also had a rolling sample). Defining family composition over the 12 months following the end of the income reference period, as the NHIS does, *increases* the poverty rate by 0.11 percentage points, on average, although the impact ranges from a reduction of 0.14 percentage points to an increase of 0.32 percentage points, depending on the survey month.

The timing of family composition in relation to the income reference period has a bigger impact on estimated poverty rates for selected subpopulations than for the population as a whole. Differences by gender are negligible, but racial and ethnic differentials are more pronounced. Timing has a greater effect on the poverty rates observed for black non-Hispanics than for white non-Hispanics, and the impact is even greater for Hispanics, although the pattern is surprising (Table V.8). We find no difference between the CPS and NHIS simulations for Hispanics despite an average lag of 3.5 months between family composition and the income reference year, yet the simulations that reflect the ACS and SIPP/MEPS timing yield poverty rates that are 0.58 to 0.71 percentage points lower than the CPS. Contemporaneous measurement produces an Hispanic poverty rate that is 1.64 percentage points below the CPS simulation. The elderly show negligible differences by timing, consistent with their low rates of change in family composition, whereas children show larger differences than nonelderly adults.

Differences in the impact of timing are most pronounced across subpopulations defined by family composition. Childless couples show essentially no variation in poverty rates by timing while single parents and children in single-parent families show exceedingly strong variation. Within both subgroups the poverty rates for the SIPP and MEPS simulations are a percentage point lower than for the CPS simulation while the poverty rates obtained with contemporaneous

TABLE V.8

DIFFERENCE IN PERCENT POOR BY SIMULATED TIMING OF FAMILY COMPOSITION RELATIVE TO THE CY 2001
INCOME REFERENCE PERIOD, BY DEMOGRAPHIC CHARACTERISTICS: SIPP

Demographic Characteristic	Percent Poor with Family Composition Fixed Mar 2002 (CPS)	Difference in Percent Poor with:				
		Contemporaneous Measurement (PSID)	Family Composition Fixed Dec 2001 (SIPP/MEPS)	Family Composition Fixed Jan 2002 (ACS)[a]	Family Composition Fixed Jan - Dec 2002 (NHIS)	Family Composition Fixed Dec 2002
All Persons	11.27	-0.64	-0.18	-0.15	0.11	0.31
Gender						
Male	9.60	-0.52	-0.20	-0.15	0.10	0.30
Female	12.84	-0.74	-0.17	-0.14	0.11	0.33
Race/Ethnicity						
White, non-Hispanic	7.44	-0.46	-0.11	-0.07	0.14	0.37
Black, non-Hispanic	24.09	-0.75	-0.32	-0.32	0.12	0.31
Hispanic	20.34	-1.64	-0.71	-0.58	-0.04	0.07
Age						
<18	16.75	-0.98	-0.46	-0.27	0.10	0.36
18-64	9.48	-0.61	-0.13	-0.15	0.16	0.44
65+	9.38	-0.28	-0.08	0.00	0.13	0.31
Family composition						
Singles (age 18 or older)	18.42	-1.13	0.11	-0.01	0.22	0.63
Childless couples	3.11	0.08	-0.03	-0.03	0.03	0.05
Single parents with children	31.05	-2.70	-1.03	-0.79	0.71	1.92
Children in single-parent families	36.21	-2.54	-1.12	-0.62	0.13	0.54
Husband-wife families with children	6.27	-0.46	-0.18	-0.14	0.11	0.16
Children in husband-wife families	8.48	-0.42	-0.25	-0.19	0.06	0.22
Current Program Participants						
Welfare or Food Stamps	51.06	-2.00	-0.75	-0.45	-0.04	0.19
Medicaid or SCHIP	47.14	-2.31	-1.16	-0.42	-0.58	-0.34

Source: Mathematica Policy Research, from 2001 SIPP panel.

Note: See text for description of simulation.

[a] We identify January 2002 as reflecting the ACS lag because, regardless of the interview month, the lag between the ACS family composition and the end of the income reference period is one month. This treatment implicitly assumes real incomes and demographic composition of the population are unchanged.

measurement are 2.54 to 2.70 percentage points lower than for the CPS. The magnitudes of the timing effects indicate that single parents and their children are substantially more likely than the other family types to have experienced recent changes in composition that affected their economic well-being. Husband-wife couples and children in two-parent families show timing effects that are more typical of all persons while singles show very modest effects across the fixed composition scenarios but more than a percentage point decline in poverty with contemporaneous measurement.

Persons who received welfare or Food Stamps in the simulated survey month have poverty rates 2 percentage points lower with a contemporaneous measure than with family composition fixed in March, but fixing family composition later than March does not appear to increase the poverty rate relative to March. Persons enrolled in Medicaid or SCHIP in the simulated survey month show the same pattern, except that their poverty rates with family composition fixed later than March are, if anything, slightly lower than what we observe with composition fixed in March.

The timing of family composition in relation to the income reference period also affects the estimated percentage of the population below 200 percent of poverty—a population commonly defined as low-income. Fixing family composition at the end of the reference year produces a net increase of 0.64 percent in the fraction of the population classified as low income (Table V.9). The low-income population grows by only 0.08 additional percentage points when family composition is fixed in March. When family composition is distributed over 2002 the average increase in the estimated size of the low-income population is a full percentage point (1.03), but this is less than a third of a percentage point higher than fixing family composition in March. Over the 12 months the increase relative to contemporaneous measurement varies from 0.68 percent to 1.50 percent. Compared to fixing family composition in March, the impact on the

TABLE V.9

IMPACT OF FIXED FAMILY COMPOSITION ON ESTIMATED PERCENT BELOW
200% OF POVERTY, BASED ON CY 2001 INCOME: SIPP SIMULATION

Simulated Timing of Family Composition	Difference in Percentage below 200% of Poverty with Fixed Composition				Percent Below 200% of Poverty
	Gross Addition[a]	Gross Reduction[b]	Sum	Difference	
Contemporaneous	0.00	0.00	0.00	0.00	29.86
Fixed at:					
Dec 2001	0.99	0.36	1.35	0.64	30.50
Jan 2002	1.13	0.45	1.58	0.68	30.54
Feb 2002	1.21	0.52	1.73	0.70	30.56
Mar 2002	1.32	0.59	1.91	0.72	30.58
Apr 2002	1.42	0.63	2.05	0.78	30.64
May 2002	1.54	0.66	2.20	0.88	30.74
Jun 2002	1.71	0.73	2.44	0.98	30.84
Jul 2002	1.83	0.79	2.62	1.03	30.89
Aug 2002	1.99	0.86	2.85	1.13	30.99
Sep 2002	2.18	0.91	3.09	1.27	31.13
Oct 2002	2.24	0.94	3.18	1.30	31.16
Nov 2002	2.31	0.97	3.28	1.34	31.20
Dec 2002	2.45	0.95	3.40	1.50	31.36
Avg. Jan-Dec 2002	1.78	0.75	2.53	1.03	30.89

Source: Mathematica Policy Research, from 2001 SIPP panel.

Note: See text for description of simulation.

[a] Percentage of population classified as below 200% of poverty when family composition is fixed in time but not below 200% of poverty when family composition is contemporaneous with income.
[b] Percentage of population classified as below 200% of poverty when family composition is contemporaneous with income but not below 200% of poverty when family composition is fixed in time.

estimated size of the low-income population ranges from a reduction of 0.04 percentage points to an increase of 0.78 percentage points.

Fixing family composition at a point in time does not have the same impact on the upper tail of the distribution that it does in the lower tail. That is, it does not yield more high-income families, which it would do if its overall impact were to move more people to the tails of the distribution. Instead, it produces a small reduction in the proportion of persons identified as high-income. Compared to contemporaneous measurement, fixing family composition at the end of the income reference year reduces the fraction of the population at or above 500 percent of poverty by a quarter of a percentage point (Table V.10). This effect grows to half a percentage point (0.53) as family composition is moved to 12 months later. The magnitudes of these effects are smaller than what we observed at the lower end of the income distribution, but in conjunction with what we saw earlier they indicate that the overall affect of fixed versus contemporaneous measurement of family composition and income is to produce a downward shift in the ratio of family income to the poverty threshold.

We stress that this is a purely methodological exercise, and as such it has limitations. In particular, it reflects the design features of SIPP, with extensive income questions and a recall period of one to four months prior to the interview month. If the CPS ASEC supplement, for example, were conducted in June instead of primarily March, we would not necessarily expect to see the estimated poverty rate rise by the amount that our simulations indicate. The actual impact might be larger, or it might be smaller. Nevertheless, these results are important in demonstrating that the simplification implied by a fixed family composition and the lag between the end of the income reference year and the timing of family composition do tend to bias poverty estimates in an upward direction.

TABLE V.10

IMPACT OF FIXED FAMILY COMPOSITION ON ESTIMATED PERCENT AT OR
ABOVE 500% OF POVERTY, BASED ON CY 2001 INCOME: SIPP SIMULATION

| Simulated Timing of Family Composition | Difference in Percentage at or above 500% Of Poverty with Fixed Composition | | | | Percent At or Above 500% of Poverty |
	Gross Addition[a]	Gross Reduction[b]	Sum	Difference	
Contemporaneous	0.00	0.00	0.00	0.00	22.39
Fixed at:					
Dec 2001	0.32	0.57	0.89	-0.25	22.14
Jan 2002	0.38	0.66	1.04	-0.27	22.12
Feb 2002	0.42	0.71	1.13	-0.29	22.10
Mar 2002	0.46	0.80	1.26	-0.34	22.05
Apr 2002	0.54	0.88	1.42	-0.34	22.05
May 2002	0.57	0.95	1.52	-0.38	22.01
Jun 2002	0.62	1.00	1.62	-0.38	22.01
Jul 2002	0.68	1.05	1.73	-0.36	22.03
Aug 2002	0.73	1.13	1.86	-0.40	21.99
Sep 2002	0.78	1.23	2.01	-0.44	21.95
Oct 2002	0.83	1.30	2.13	-0.47	21.92
Nov 2002	0.84	1.37	2.21	-0.52	21.87
Dec 2002	0.86	1.39	2.25	-0.53	21.86
Avg. Jan-Dec 2002	0.64	1.04	1.68	-0.39	22.00

Source: Mathematica Policy Research, from 2001 SIPP panel.

Note: See text for description of simulation.

[a] Percentage of population classified as at or above 500% of poverty when family composition is fixed in time, but below 500% of poverty when family composition is contemporaneous with income.
[b] Percentage of population classified as below 500% of poverty when family composition is fixed in time but at or above 500% of poverty when family composition is contemporaneous with income.

C. ROLLING SAMPLES

Both the ACS and the NHIS utilize a rolling sample. In each case, an annual sample is distributed systematically over the year. For the ACS, with an annual sample of 3 million households when fully implemented, distributing the workload over the year is an operational necessity. For the NHIS, operational considerations may be important as well, but another factor in the design is the seasonality and possible trend in some of the health measures that the survey collects. Rolling samples raise questions about the best approach to measuring characteristics that can vary over time, and the ACS and NHIS illustrate two different approaches to measuring annual income. The ACS asks respondents to report their income for the past 12 months, which is defined as "the period from today's date one year ago up through today."[49] This represents a rolling reference period with a non-varying recall interval. The NHIS asks respondents, regardless of when they are interviewed, to report their incomes for the previous calendar year. This yields a fixed reference period but with a varying recall interval. Choices such as this one carry implications for the interpretation of estimates and may ultimately affect the quality of the data collected. In this case, is one choice clearly better than the other? In attempting to answer this question, we begin by examining some of the issues raised by the use of rolling samples to collect data for policy analysis. We then turn to empirical analyses bearing, first, on rolling reference periods and, then, on varying recall intervals. We conclude this discussion by looking at the within-year inflation adjustments developed for the ACS in order to put the data collected with a rolling reference period into the same real dollars across the different 12-month intervals used as reference periods.

---

[49] This text is from the ACS questionnaire. The Census Bureau tabulations on the internal file, as well as the ADJUST factor and POVPIP measure on the public use file, assign income to the 12 months prior to the month in which the questionnaire is completed.

1. Issues Raised by Rolling Samples

The different routes to income measurement taken by the ACS and NHIS raise questions about policy relevance and data quality. From a policy perspective, it would be desirable for the income reference period to align as closely as possible with the reference period for other policy-relevant variables, such as health insurance coverage, health status, health care utilization, and program participation. In the NHIS, key health policy variables refer to the time of the survey or the past 12 months, for the most part. However, from the perspective of data quality it would be better to ask the annual income question for whatever reference period respondents can more easily address, and for policy uses it is better to have a reference period that is aligned with official poverty estimates. Faced with a difficult task, respondents may give lower quality responses or mentally change the question to something they can more readily answer. It has been suggested, for example, that the poor measurement of health insurance coverage in the CPS arises from the difficulty of the task that respondents are being asked to perform. With respect to income measurement explicitly, the fact that the statements of annual income supplied by financial institutions refer to the previous calendar year suggests that respondents would find it easier to report their incomes for the prior calendar year than the past 12 months. Conventional wisdom has suggested that respondents are most aware of their income for the prior calendar year when they are engaged in pulling together the financial records needed to prepare their tax returns (for those who file). But for how long might the prior calendar year income remain salient? Will respondents be able to recall this income as easily in December of the following year as in the early part of the year?

One way to approach assessing the difficulty that respondents face in dealing with a rolling reference period or a fixed reference period but varying recall interval is to examine patterns of non-response. If respondents find it easier to report their incomes for the previous calendar year

than for the past 12 months, then we ought to see a decline in response rates to the income questions as the interview date moves farther from the end of the calendar year. We explored this possibility with ACS data and obtained unexpected findings, which are reported in Chapter VI. Similarly, if respondents are challenged by a growing recall interval, then response rates to the income questions in NHIS ought to decline over the course of the survey year. We explored this question as well but found only a modest decline in response rates.

2. Rolling Reference Period

If respondents to the ACS are reporting their incomes for the past 12 months, as requested, then we ought to see evidence of growth in reported incomes as the interview month moves from January through December. After all, compensating for such growth is one of the objectives of the inflation adjustment that is applied to the ACS income data. On the other hand, if income grew very little over the calendar year or even declined, then even highly accurate responses may not show the expected pattern.

Table V.11 shows the aggregate income reported by respondents to the 2003 ACS, by calendar month and family income quintile.[50] We see no indication, either within any quintile or across all quintiles, that respondents interviewed later in the year reported more income than respondents interviewed earlier in the year. Does this suggest, then, that respondents are reporting their incomes for the prior calendar year? Certainly, the case that respondents were in fact giving their income for the past 12 months would be stronger if the reported incomes did grow by interview month. Later in this chapter, however, we look at other evidence of change in

---

[50] These estimates were prepared by the U.S. Census Bureau using the 2003 ACS because the monthly samples in the 2002 ACS were not of uniform size.

TABLE V.11

AGGREGATE INCOME IN PREVIOUS 12 MONTHS BY FAMILY INCOME QUINTILE
WITH NO ADJUSTMENT FOR INFLATION: 2003 ACS

| Month | Quintile of Family Income | | | | | Total |
	Lowest	Second	Third	Fourth	Highest	
Jan	30.04	63.88	91.38	119.24	229.87	534.41
Feb	29.09	65.03	91.16	115.82	222.85	523.96
Mar	30.10	63.14	89.54	112.73	217.96	513.47
Apr	29.62	63.74	88.64	111.89	220.92	514.81
May	29.44	63.45	88.55	119.30	229.98	530.72
Jun	29.63	65.55	90.26	121.62	220.80	527.85
Jul	29.65	64.50	90.11	118.57	235.30	538.14
Aug	29.37	63.44	88.63	118.43	232.17	532.04
Sep	29.36	65.43	89.39	113.17	237.12	534.46
Oct	30.42	64.88	89.18	115.13	232.33	531.95
Nov	30.02	64.84	90.61	113.57	234.72	533.76
Dec	29.38	64.62	90.53	115.96	230.58	531.08

Source: U.S. Census Bureau, Housing and Household Economic Statistics Division, special
tabulations.

Note: The estimates for each month are based on households interviewed in that month. Aggregate
amounts are 1/12 what they would be if all sample households were interviewed in each month.

reported income over time that suggests that the amount of real change in incomes over this period may have been too small to show up in respondents' survey reports.

3. Varying Recall Interval

Conversely to what we explored with the ACS, the varying recall interval in the NHIS creates a possibility that as the survey year progresses, respondents might give responses influenced by their current incomes. If respondents were reporting their prior calendar year incomes as requested, then we would expect to see fairly uniform distributions of income over the survey year, although population change might influence the pattern to some degree. Even though respondents are being asked to report their incomes for the same period, the composition of the population is not constant over time, and families change as well. We saw earlier that with a growing lag between the end of the income reference year and the measurement of family composition, the estimated poverty rate increased. This would apply to the NHIS income measures in a way that it does not apply to the ACS, and in so doing it might obscure any evidence that respondents later in the year were reporting more income than respondents earlier in the year.

The distribution of family income in the NHIS by calendar quarter shows no evidence of change over time (Table V.12). For the reasons discussed above, we find this inconclusive with respect to respondents' compliance with the task of reporting their family incomes for the previous calendar year, as there may be confounding factors. If there is any influence of current income on reported income, however, it would have to be small.

We do find a statistically significant increase between the first and fourth quarters in the proportion of family income allocated, which grows from 29.9 percent to 31.9 percent (Table V.13). This could suggest that respondents are having more difficulty reporting their prior calendar year incomes as the recall interval increases. But the increased non-response is very

TABLE V.12

FAMILY INCOME OF PERSONS BY INTERVIEW QUARTER: NHIS

Family Income ($)	Interview Quarter			
	1	2	3	4
	Percent Distribution			
0 - 4999	3.10	3.24	3.44	4.03
5000 - 9999	4.64	4.86	4.40	4.72
10000-14999	5.68	5.62	5.99	5.56
15000-19999	5.78	6.01	5.67	5.67
20000-24999	6.69	6.54	6.67	6.51
25000-34999	11.94	11.36	11.76	12.71
35000-44999	11.05	10.44	10.66	9.77
45000-54999	9.29	9.18	9.01	9.01
55000-64999	7.81	7.85	7.69	7.54
65000-74999	5.89	6.57	6.64	6.73
75000 and over	28.13	28.31	28.07	27.75
Total	100.00	100.00	100.00	100.00

Source: Mathematica Policy Research, from tabulations of calendar year
2002 income from the 2003 NHIS.

TABLE V.13

FAMILY INCOME ALLOCATION BY INTERVIEW QUARTER: NHIS

Family Income Allocation	Interview Quarter			
	1	2	3	4
	Percent Distribution			
Reported	70.10	68.99	68.83	68.09 *
Allocated	29.89	31.01	31.17	31.91 *
Total	100.00	100.00	100.00	100.00

Source: Mathematica Policy Research, from the 2003 NHIS.

* Estimate is significantly different from quarter 1 at the .05 level.

modest and does not suggest that respondents in the fourth quarter are having a serious problem with the reporting of their income for the prior calendar year or that the quality of the data may be compromised.

4.   Within-Year Inflation Adjustments

While the rolling reference period for income data in the ACS means that the annual incomes that are collected represent an average of 12 different 12-month intervals centered around December of the prior year (which appears in every interval), the Census Bureau applies an inflation adjustment in order to convert the responses to constant dollars for the survey year. For the ACS income data collected in 2003, the reported incomes were adjusted by survey month based on an average of monthly values of the CPI-U. Income data collected in December 2003 received the smallest adjustment while income data collected in January were adjusted for a full year of price inflation.

After application of the inflation adjustment to the data underlying Table V.11, we still find no evident time trend in the distribution of aggregate dollars either within or across income quintiles (Table V.14). We can draw no insights into what the respondents may have reported in response to questions to provide their incomes for the past 12 months.

Another way to look at the inflation adjustment is to compare its effect on per capita income by quintile with the actual growth in per capita income as measured in the ACS between 2002 and 2003. The observed change between 2002 and 2003 will reflect a 12-month increase in income rather than the six-and-a-half month increase that is the goal of the inflation adjustment. In addition, actual annual growth in per capita income will incorporate the net effect of population change—that is, births, deaths, and net migration. These affect not only the size of the population but its income as well, and they are not taken into account in the ACS inflation adjustment.

TABLE V.14

AGGREGATE INCOME IN PREVIOUS 12 MONTHS BY FAMILY INCOME QUINTILE
WITH AMOUNTS ADJUSTED FOR INFLATION: 2003 ACS

Month	Quintile of Family Income					Total
	Lowest	Second	Third	Fourth	Highest	
Jan	30.24	63.07	94.12	122.08	237.04	546.54
Feb	29.29	64.54	93.44	118.78	228.66	534.71
Mar	30.47	64.04	89.71	116.11	222.40	522.73
Apr	29.95	64.77	89.76	113.72	224.59	522.79
May	29.79	64.44	89.68	120.81	233.24	537.95
Jun	29.94	66.51	91.38	126.40	219.91	534.14
Jul	29.98	65.51	90.97	123.83	233.31	543.61
Aug	29.73	64.50	89.48	123.04	229.76	536.51
Sep	29.68	66.37	90.36	117.81	233.77	537.99
Oct	30.74	66.03	89.83	118.78	229.07	534.44
Nov	30.50	65.92	91.04	117.53	230.36	535.36
Dec	29.95	65.49	90.55	121.53	224.37	531.90

Source: U.S. Census Bureau, Housing and Household Economic Statistics Division, special
tabulations.

Note: The estimates for each month are based on households interviewed in that month.
Aggregate amounts are 1/12 what they would be if all sample households were
interviewed in each month.

252

The upper portion of Table V.15 presents estimates of total persons and total income by quintile of family income for the 2002 ACS, with and without adjustment, and for the 2003 ACS without adjustment. The adjustment is based on the application of the income adjustment factor provided on the ACS public use file, which represents an average of the 12 monthly adjustment factors that the Census Bureau applies to reported income on its internal file. The next panel of the table presents estimates of per capita income derived by dividing the aggregate income by the number of persons, by quintile, for the 2002 ACS (with and without adjustment) and the 2003 ACS. The final panel shows the percentage growth in annual per capita income based on comparing both the adjusted 2002 ACS estimates and the unadjusted 2003 ACS estimates with the unadjusted 2002 ACS estimates.

While the application of the ACS adjustment yields a uniform increase of about 0.93 percent in per capita income across the five quintiles (and for the population as a whole), we see a rather different pattern in the *actual* growth of per capita income over the full year. The amount of growth in per capita income increases over the income quintiles, beginning with negative growth in the first two quintiles (-0.58 and -0.35 percent respectively), followed by growth of 0.86 percent and 2.17 percent in the next two quintiles. Growth in the top quintile is slightly lower than in the fourth quintile at 2.02 percent. Over the population as a whole the increase is 1.32 percent.

What these patterns suggest is that income does not grow uniformly by quintile. The application of uniform price adjustments to convert ACS income to constant dollars for the survey year may have the unintended consequence of putting too much income at the low end of the distribution, where immigration and other aspects of population dynamics may function to depress or at least hide growth. In a year with a more substantial inflation than 2002 and 2003, this aspect of the price adjustment is likely to be even more evident.

TABLE V.15

COMPARISON OF ACS INCOME ADJUSTMENT WITH ANNUAL GROWTH IN INCOME
BY QUINTILE, 2002 TO 2003

Estimate	Family Income Quintile					
	Lowest	Second	Third	Fourth	Highest	Total
Millions of Persons						
ACS 2002 unadjusted	56.57	54.61	55.48	55.52	55.50	277.69
ACS 2002 adjusted	56.49	54.59	55.55	55.53	55.54	277.69
ACS 2003 unadjusted	57.04	55.21	55.92	56.37	55.74	280.28
Billions of Dollars						
ACS 2002 unadjusted	365.8	772.2	1,076.2	1,402.5	2,669.5	6,286.2
ACS 2002 (final adjusted)	368.7	778.4	1,087.4	1,415.8	2,696.0	6,346.3
ACS 2003 (original unadjusted)	366.7	777.9	1,094.0	1,454.9	2,734.9	6,428.4
Per Capita Income						
ACS 2002 unadjusted	6,466	14,140	19,396	25,263	48,096	22,637
ACS 2002 (final adjusted)	6,526	14,259	19,576	25,496	48,543	22,854
ACS 2003 (original unadjusted)	6,429	14,090	19,564	25,810	49,065	22,936
Percentage Increase in Per Capita Income						
Adjustment to ACS 2002	0.93	0.84	0.93	0.92	0.93	0.96
Growth from ACS 2002 to ACS 2003	-0.58	-0.35	0.86	2.17	2.02	1.32

Source: Mathematica Policy Research, from tabulations of the 2002 and 2003 ACS.

D. RETIREMENT INCOME

Traditional employer-provided pension plans, known as defined benefit plans, are giving way to other forms of retirement plans, in which an employer may pay a pre-tax contribution to an employee retirement account (a defined contribution plan) or match an employee's own contributions. In addition, a number of retirement savings vehicles have been established by Congress to allow individuals to provide for their retirement separately from what their employers may provide. While many of these non-traditional plans have been around for decades, surveys that collect income data have been slow to develop ways to capture income from such plans. It is notable, for example, that none of the eight surveys collects information on defined contribution retirement benefits that compares with the information collected on income received from traditional pension plans. In part this can be traced to divided opinions among economists on how to treat the deferred income that retirees will obtain from these sources.

The CPS income concept includes regular withdrawals from Individual Retirement Accounts (IRAs) as well as Keogh and 401(k) accounts but excludes lump-sum payments from these or other types of retirement plans.[51] The CPS captures a modest $3.3 billion from this source. SIPP collects both regular and lump-sum payments from IRA, Keogh, and 401(k) accounts in separate items and also collects lump-sum retirement payments. We have included only the regular payments in SIPP income for comparative purposes, but it is possible to examine how much additional income would be added if lump-sum payments were included and how this income would affect the distribution of persons by poverty class.

---

[51] The CPS instrument asks very generally about "pension or retirement income from a previous employer or union, or any other type of retirement income other than Social Security or VA benefits" but includes among the sources that a respondent may identify "regular payments from IRA, Keogh, or 401(k) accounts."

MEPS requests income from payments from IRA, Keogh, or 401(k) accounts without differentiating between regular and lump-sum payments. We did not include this component of MEPS income in our comparative analysis because the amount of income captured by the MEPS variable was nearly 20 times the $3.3 billion in regular IRA withdrawals captured in the CPS, suggesting that nearly all of the income captured in the MEPS variable was outside the CPS concept. However, we can examine how much additional income would be added if we included this additional source and, like SIPP, how it would affect the distribution of persons by poverty class.

The regular IRA, Keogh, and 401(k) payments picked up by SIPP and which we include in SIPP income add $18.7 billion to the total (Table V.16). Adding lump-sum payments from these same sources would add another $12.6 billion. Adding lump-sum payments from other pension or retirement plans would add only $4.1 billion. Including regular IRA, Keogh, and 401(k) payments has a very small effect on the number of poor. The number of poor is reduced by 30,000 (.03 million) compared to the number we would observe if this source were excluded. Were we to include lump-sum payments from IRA, Keogh, and 401(k) accounts as well as pension and retirement plans in SIPP income, the number of poor would be reduced by only another 30,000 while the number of people at 400 percent of poverty or more would be increased by 400,000.

MEPS captures more than twice as much income from IRA, Keogh, and 401(k) accounts as SIPP: $65.6 billion (Table V.17). If income from this source were to be included in MEPS income, the estimated number of poor would be reduced by 410,000 while the number of persons with family incomes above 400 percent of poverty would be increased by 1.83 million. These effects are substantially larger than what we estimated for SIPP. Nevertheless, the estimated poverty rate would be reduced by only 0.1 percent.

TABLE V.16

IMPACT OF INCLUDING NON-REGULAR IRA AND LUMP-SUM PENSION INCOME IN TOTAL INCOME: SIPP

Income Definition	< 100%	100% to < 200%	200% to < 400%	400% or More	Total
		Number of Persons (Millions)			
Excluding all IRA/Keogh/401(k) Income	33.28	56.53	98.57	92.70	281.08
With Regular IRA/Keogh/401(k) Payments[a]	33.25	56.25	98.37	93.22	281.08
Adding Remaining IRA/Keogh/401(k) Payments	33.24	56.11	98.21	93.52	281.08
Adding Lump Sum Pension/Retirement Income[b]	33.22	55.99	98.25	93.62	281.08
		Incremental Impact on Number of Persons (Millions)			
Excluding all IRA/Keogh/401(k) Income	0.00	0.00	0.00	0.00	0.00
With Regular IRA/Keogh/401(k) Payments[a]	-0.03	-0.28	-0.21	0.52	0.00
Adding Remaining IRA/Keogh/401(k) Payments	-0.01	-0.14	-0.15	0.30	0.00
Adding Lump Sum Pension/Retirement Income[b]	-0.02	-0.12	0.03	0.10	0.00
		Total Income ($Billions)			
Excluding all IRA/Keogh/401(k) Income	113.8	482.6	1,608.2	3,542.8	5,747.5
With Regular IRA/Keogh/401(k) Payments[a]	113.7	479.9	1,605.6	3,566.9	5,766.2
Adding Remaining IRA/Keogh/401(k) Payments	113.7	478.6	1,603.9	3,582.7	5,778.8
Adding Lump Sum Pension/Retirement Income[b]	113.6	477.6	1,603.4	3,588.3	5,782.9
		Incremental Impact on Total Income ($Billions)			
Excluding all IRA/Keogh/401(k) Income	0.0	0.0	0.0	0.0	0.0
With Regular IRA/Keogh/401(k) Payments[a]	-0.13	-2.70	-2.60	24.14	18.71
Adding Remaining IRA/Keogh/401(k) Payments	-0.04	-1.32	-1.71	15.72	12.64
Adding Lump Sum Pension/Retirement Income[b]	-0.07	-0.97	-0.51	5.68	4.12

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2001 SIPP panel.

[a] Corresponds to SIPP estimates in comparative tables.
[b] Regular pension and retirement income is included in all of the estimates.

TABLE V.17

IMPACT OF ADDITION OF IRA INCOME TO TOTAL INCOME: MEPS

Estimate	< 100%	100% to < 200%	200% to < 400%	400% or More	Total
Number of Persons (millions)					
Without IRA Income	35.35	52.14	89.80	106.02	283.30
With IRA Income	34.93	51.31	89.20	107.85	283.30
Difference	-0.41	-0.82	-0.60	1.83	0.00
Percent of Persons					
Without IRA Income	12.5	18.4	31.7	37.4	100.0
With IRA Income	12.3	18.1	31.5	38.1	100.0
Difference	-0.1	-0.3	-0.2	0.6	0.0
Billions of Dollars					
Without IRA Income	110.0	458.7	1,473.2	4,215.8	6,257.7
With IRA Income	108.6	451.1	1,463.3	4,300.3	6,323.4
Difference	-1.4	-7.5	-9.9	84.5	65.6

Source: Mathematica Policy Research from tabulations of calendar year 2002 income from the 2002
Full-year Consolidated MEPS-HC.

E. INCOME POST-STRATIFICATION

As we have noted previously, the MEPS survey weights are post-stratified to poverty distributions observed in the CPS. Post-stratification to population totals by age, sex, and race/ethnicity is widely used as a means to correct for undercoverage and differential non-response, but as we demonstrated in Chapter III, it is important to ensure that the survey totals to be adjusted and the post-stratum totals to which they are adjusted reflect the same universe. When an income or poverty distribution obtained from one survey is used to post-stratify the weights for another survey, it is important that the concepts of income or poverty used in the two surveys agree. The survey descriptions presented in Chapter II underscore how difficult it may be to achieve such agreement, and the empirical findings presented in Chapter IV show how survey measures of income that are similar in some respects may be quite different in others.

In Chapter IV we also speculated that a portion of the difference between CPS and MEPS estimates of total families and people with earnings could have arisen from the post-stratification to the CPS poverty distribution. While this is *not* something that we can evaluate with a simulation, the MEPS survey contractors who perform the post-stratification have access to the requisite data to assess the impact of including the CPS poverty distribution among the post-stratum totals. For such an assessment the preliminary MEPS weights prior to post-stratification would have to be post-stratified to CPS control totals that exclude the poverty distribution. Estimates of total income, total earners, total families, and other characteristics could then be prepared using these alternative weights and the results compared to estimates using the person weights on the public use file. In our view, this could provide an extremely interesting methodological study that could shed light on the full range of consequences of post-stratifying the MEPS weights to the CPS poverty distribution. Such a study would be enhanced if the post-stratification itself were altered experimentally to test the impact of alternative refinements to the

MEPS poverty estimates and the survey universe, including, in particular, the treatment of sample members in families with missing data on one or more family members.

VI. INCOME ALLOCATION, APPROXIMATION AND ROUNDING

Two ways in which respondents can diminish the effectiveness of even very well designed income questions are by providing no answers at all or, which may be worse, inaccurate answers. It is well known that income questions generate some of the highest item non-response rates in surveys generally.[51] Frequently, this results in large amounts of missing income data. Unless the data producers choose to leave such missing data for their users to address, they must apply one or more methods of allocation to fill in the missing data.[52] When the data producers elect to allocate their missing income data, high rates of non-response are likely to mean that large fractions of the income data that they provide to their users will have been created by the data producers rather than supplied by their respondents. This makes the quality of the income data dependent not just on the completeness and accuracy of the reported amounts but the quality of the methods used to generate allocated amounts.[53]

We can quantify the amount of income data that are allocated in a survey and, in so doing, measure the magnitude of non-response and its potential impact on data quality. Income allocation is the principal focus of this chapter. We examine, in successive sections, the overall frequency of income allocation across the five general population surveys, the methods of

---

[51] For surveys in which the collection of income data is not a major objective, this can become a rationale for limiting the content of their income questions or relegating them to the end of interview, which minimizes their potential impact on the response rates to other questions or to respondents breaking off their interviews. However, relegating them to the end of the interview may further reduce their responses rates and adversely affect the quality of the responses, as appears to be the case with the CPS health insurance questions.

[52] We follow the Census Bureau in using the term allocation in a generic sense to describe any method of replacing a missing value with a generated value. For all three Census Bureau surveys included in this study, the flags indicating how missing responses were filled in are identified in the survey documentation as allocation flags. Frequently, the methods of allocation listed in the survey documentation are characterized as types of imputation. In other respects, however, the terms allocation and imputation appear to be used interchangeably.

[53] The quality of allocation methods can be defined in terms of unbiasedness and their addition of minimal "noise" or variance to the variables being allocated.

allocation used, differences in allocation across the income distribution and by source, differences by interview month, and issues in using allocation.

We cannot assess in any direct way the accuracy of survey responses to income questions. However, one way in which respondents may reduce the accuracy of their responses is to use a high level of approximation—for example, by reporting a salary of $50,000 when the true salary lies somewhere between $45,000 and $55,000. When a significant number of respondents round their responses in this way, it distorts the distribution by creating spikes at the rounded values. In fact, rounding is a commonly used technique for protecting the confidentiality of income data in public use files.[54] The frequency of round responses can be quantified, and we do so for selected income sources for the five general population surveys and the PSID in the next to last section of the chapter.

Rounding does not lead to bias, but underreporting among persons who provide dollar amounts is evident from comparisons of survey aggregates and administrative totals. While allocation is widely used to compensate for non-response, underreporting is less amenable to correction because individual underreported amounts cannot be identified without additional information—such as linked administrative records. Some agencies substitute their own administrative records for reported data, but, in general, such data cannot be released to outside users. Other than noting such practice, we do not assess the use or effectiveness of strategies to compensate for underreporting of dollar amounts among respondents who report both recipiency and income for a given source.

After presenting our findings on rounding we discuss two issues regarding the application of allocation that have emerged from our analysis of income data. All of our estimates in this

---

[54] Limiting the number of significant digits in reported incomes reduces their uniqueness, making them less identifiable. The ACS uses a very well-defined rounding rule described in Chapter II.

chapter are based on income for 2002, which, as we have noted, covers a calendar year except for the rolling reference period in ACS.

A. FREQUENCY OF ALLOCATION

All eight surveys that are included in this study employ one or more methods of allocation to fill in missing values for income. We can quantify how much of the income data are allocated, how this varies by source, and how the amount of imputation varies across the income distribution, and we do so for the five general population surveys in this section.[55] We can also quantify the alternative types of allocation used, and we do so in the next section, albeit very broadly. Except to a very limited degree, however, we do not attempt to quantify the *quality* of the allocations as this is well outside the scope of this project.

In addition, we note that we do not count zero amounts as allocations for any of the surveys, regardless of how they were obtained, because they do not contribute to total income. Our assessment of the relative magnitudes of allocation across surveys utilizes estimates of the percentage of persons with income who had any portion of their income allocated and the percentage of total dollars that was allocated. Therefore, the allocation rates in the tables presented in this section should not be interpreted as overall non-response rates for the indicated items.

Estimates of the frequency of allocation across the five surveys demonstrate wide variation. After presenting our findings for total income, we turn to differences in allocation rates across the income distribution and by source of income.

---

[55] Income allocations for the 2003 PSID were either limited to earnings or not fully flagged. Income allocations in the 2004 HRS were performed at the source level, but we cannot assess their contribution to total family income from the variables provided in the RAND data release used in our analysis. In the MCBS, allocated amounts were not flagged.

1. Total Income

Among persons (or, for the NHIS, families) with income either reported or allocated, the percentage with any portion of their income allocated varies from a low of 21.2 percent for the ACS to a high of 80.5 percent for the SIPP (Table VI.1). The CPS and MEPS fall in the middle of this range, with allocation frequencies of 52 to 54 percent while the NHIS, with its single family income question, has an allocation rate of one-third. The SIPP's vastly greater number of income questions than any of the other surveys, spread over three to four interviews, undoubtedly contributes to the exceedingly high allocation rate estimated for this survey. A respondent who was able and willing to provide a dollar amount for all but one of these questions is counted among the 80 percent with at least some of their annual total income allocated.

For this reason we find it more useful to look at the proportion of total dollars that was allocated, and here we find that the SIPP was undifferentiated from the CPS and the NHIS, with about one-third of total income being allocated. The ACS had just over half that proportion of total income allocated (17.6 percent) while MEPS had about 10 percentage points more than the CPS, SIPP, and NHIS at 42.7 percent.

2. Differences Across the Income Distribution

Allocation rates by quintile of family income reveal curiously different patterns across the surveys. In the CPS, SIPP and MEPS the percentage of persons with any of their income allocated rises with the level of income (that is, from the lowest to the highest quintile), but it declines slightly in the ACS and shows no clear trend in the NHIS (Table VI.2).

Turning from people to dollars, however, we find that the percentage of dollars allocated shows no trend by quintile in the CPS, SIPP, and NHIS whereas the trend is distinctly downward in the ACS but clearly upward in MEPS (Table VI.3). The nearly identical results in the SIPP

TABLE VI.1

ALLOCATION FREQUENCY FOR TOTAL INCOME: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
Persons with Income (millions)	200.61	195.21	206.21	202.22	117.4 [a]
Percent with Any Allocated Income	52.2	21.2	80.5	54.3	33.3
Amount of Total Income ($billions)	6,468.4	6,346.3	5,766.2	6,257.7	6,115.2
Percent Allocated	34.2	17.6	32.4	42.7	32.4

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] NHIS estimates are millions of families and unrelated individuals.

TABLE VI.2

PERCENT OF PERSONS WITH ANY ALLOCATED INCOME BY QUINTILE: FIVE SURVEYS

Family Income Quintile	CPS	ACS	SIPP	MEPS	NHIS
Quintile	Millions of Persons with Income[a]				
Lowest	37.92	37.74	39.21	38.54	32.26
Second	39.71	38.93	40.93	40.94	24.54
Third	40.55	39.36	41.38	40.55	22.29
Fourth	40.72	39.44	41.58	40.13	20.53
Highest	41.71	39.73	43.12	42.06	17.80
Quintile	Percent with Any Allocated Income				
Lowest	46.3	24.1	71.8	44.4	35.0
Second	48.7	22.0	78.7	49.4	35.6
Third	50.3	20.6	81.2	53.9	29.6
Fourth	54.5	19.4	83.3	58.1	28.9
Highest	60.6	19.9	86.9	64.9	36.5

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

[a] NHIS estimates are millions of families and unrelated individuals.

TABLE VI.3

PERCENT OF TOTAL INCOME ALLOCATED BY QUINTILE: FIVE SURVEYS

Family Income Quintile	CPS	ACS	SIPP	MEPS	NHIS
Quintile	Total Income in Billions of Dollars				
Lowest	370.5	368.7	391.4	360.0	356.0
Second	774.1	778.4	750.8	808.4	687.1
Third	1,090.2	1,087.4	1,008.8	1,144.7	1,020.3
Fourth	1,446.8	1,415.8	1,307.2	1,461.8	1,479.1
Highest	2,786.7	2,696.0	2,308.0	2,483.0	2,572.7
Quintile	Percent of Dollars Allocated				
Lowest	35.1	21.8	33.3	36.1	34.0
Second	33.6	20.1	33.1	38.8	34.6
Third	32.9	18.7	32.3	41.1	29.4
Fourth	32.5	17.2	31.9	43.4	28.9
Highest	35.6	16.1	32.3	45.2	34.9

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003
CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC,
and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002,
from the 2002 ACS.

and NHIS are especially noteworthy because these two surveys have the most income questions (SIPP) and the fewest income questions (NHIS) by far. Clearly, the level of income detail requested of respondents is unrelated to whether allocation rates rise, decline or remain invariant across the income distribution.

3. Differences by Source of Income

Not surprisingly, allocation rates differ more widely by source than by income level, and they are more consistent across surveys. Nevertheless, there are important differences by survey as well. The percentage of persons with any allocation is highest for asset income in the CPS, SIPP and MEPS, but asset income is no worse than wage and salary income in the ACS, where allocation rates vary little by source (Table VI.4). The SIPP is striking with almost half of the persons with income from any source having their amounts for that source allocated. But except for asset income and pensions, for which 90 percent and 65 percent of recipients, respectively, have some portion of their income allocated, there is little variation by source in the SIPP allocation rates.

Outside of assets and self-employment income, the CPS also shows little variation by source, although this is undoubtedly influenced by the 8 percent of CPS respondents whose ASEC supplement data are fully allocated. If these persons were removed from the numerator and denominator of the allocation rates, we would see greater variation by source, with SSI and welfare having only a third of the allocation frequency of asset income.

TABLE VI.4

PERCENT OF PERSONS WITH ALLOCATED INCOME BY SOURCE: FIVE SURVEYS

Source of Income	CPS	ACS	SIPP	MEPS	NHIS
	Percent of Persons with Any Allocation Among Persons with Income Source				
Total Income (NHIS family income)	52.2	21.2	80.5	54.3	33.3
Wages and Salaries (NHIS earnings)	32.1	19.5	49.5	44.5	32.9
Self-employment	45.2	18.5	53.2	NA	NA
Asset Income	62.9	19.4	90.3	63.0	NA
Social Security or Railroad Ret.	35.6	18.3	46.9	41.5	NA
SSI	27.9	17.4	47.0	7.9	NA
Welfare	27.6	17.9	50.3	11.7	NA
Pensions	37.2	17.1	65.1	38.4	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

MEPS shows very similar allocation rates—ranging from 39 percent to 43 percent—across wages and salaries, Social Security or Railroad Retirement, and pensions.[56] Yet SSI and welfare have exceedingly low allocation rates—just 8 percent and 12 percent, respectively. These rates are markedly lower than those for any source in any other survey, which prompts us to ask what is different about these items. We have no answer, however.

The NHIS does not collect income by source, but we note that the proportion of persons with allocated family income is the same as the proportion with allocated earnings.

When measured by the percentage of dollars allocated, the patterns of allocation rates by source are reasonably similar across the CPS, SIPP, and MEPS except for the very low allocation rates for SSI (8 percent) and welfare income (14 percent) in MEPS (Table VI.5). About 60 percent of asset income is allocated in all three surveys, whereas the allocation rate for wage and salary income ranges from 29 to 43 percent.

If we look at how much of total income consists of allocated income by source, we see that in every survey, wages and salaries dominate everything else.[57] Allocated wage and salary income is 25 percent of total income in the CPS, 13 percent in the ACS, 21 percent in the SIPP and 36 percent in MEPS (Table VI.6). Allocated self-employment income varies from just 1.4 percent of total income in ACS to 4.2 percent of total income in SIPP. Allocated asset income accounts for only 1 to 3 percent of total income in any of the surveys, and the same is true of

---

[56] What is reported as wage and salary income in MEPS appears to include most of the employment-related income reported by the self-employed, so we do not report allocation rates for self-employment income (see Chapter IV, Section C.2). In addition, nonzero self-employment income in MEPS was allocated only to persons who responded that they received income from self-employment but did not give a dollar amount. That is, persons who did not respond to the recipiency question for self-employment income were allocated no income from this source. This is a departure from how income was allocated to the other sources in MEPS and how income was allocated to self-employment (as well as the other sources) in the other surveys. Given that our estimates of allocation rates are based on nonzero dollars, this allocation strategy would depress an estimated allocation rate for self-employment income.

[57] NHIS does not collect income by source, so these comparisons are based on four surveys.

TABLE VI.5

PERCENT OF INCOME ALLOCATED BY SOURCE: FIVE SURVEYS

Source of Income	CPS	ACS	SIPP	MEPS	NHIS
			Percent of Dollars Allocated		
Total Income (NHIS family income)	34.2	17.6	32.4	42.7	32.4
Wages and Salaries (NHIS earnings)	32.0	17.2	28.9	43.3	31.8
Self-employment	44.7	23.1	39.5	NA	NA
Asset Income	62.6	19.4	60.8	62.0	NA
Social Security or Railroad Ret.	35.5	18.5	31.5	40.7	NA
SSI	28.0	16.7	27.4	7.9	NA
Welfare	29.2	17.9	36.0	13.5	NA
Pensions	35.4	16.2	50.6	40.9	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003
CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC,
and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002,
from the 2002 ACS.

TABLE VI.6

ALLOCATED INCOME BY SOURCE AS A PERCENT OF TOTAL INCOME: FIVE SURVEYS

Source of Income	CPS	ACS	SIPP	MEPS	NHIS
	Allocated Dollars by Source as Percent of Total Income				
Total	34.19	17.63	32.37	42.71	NA
Wages and Salaries	24.85	13.07	20.73	35.81	NA
Self-employment	2.27	1.42	4.23	NA	NA
Asset Income	2.57	0.97	1.46	2.15	NA
Social Security or Railroad Ret.	2.14	1.04	2.03	2.32	NA
SSI	0.11	0.07	0.16	0.05	NA
Welfare	0.03	0.02	0.06	0.01	NA
Pensions	1.36	0.78	3.06	1.61	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Social Security income. Allocated pensions range from 1 to 3 percent of total income while allocated self-employment income is 0.6 to over 4 percent of total income. The contribution of allocated welfare income is measured in hundredths of a percent.

Lastly, we compare the distribution of total income and allocated income by source. Across the four surveys, wages and salaries account for about three-quarters of total income and allocated income, on average, with MEPS higher and SIPP lower (Table VI.7). The possibility that most self-employment income in MEPS may be recorded as wages and salaries has been noted, along with other anomalies (see Chapter IV, section C). Asset income and Social Security income account for 2 to 6 percent of total income but a more consistent 5 to 6 percent of allocated income if we exclude the 7.5 percent share in the CPS. Pensions are 4 to 5 percent of both total income and allocated income in the CPS, ACS, and MEPS, but they account for 6 percent of total income and 9 percent of allocated income in the SIPP. SSI accounts for about half a percent of total income or allocated income across the four surveys while welfare income represents less than 0.2 percent. Clearly, the allocation rate on any source but wage and salary income will have at best a modest impact on total income, and for some sources the potential impact is entirely negligible.

B. METHOD OF ALLOCATION

As explained earlier, we use the term allocation to encompass all methods of filling in missing values besides editing (rarely useful for dollar amounts). Most of the surveys utilize "hot deck" imputation methods to allocate missing values. Hot deck methods involve matching the records with missing values to "donor" records on the basis of a typically large number of characteristics. The missing values are assigned from the donor records. Using other respondents as donors helps to ensure that the allocated values are plausible and have an appropriate

TABLE VI.7

DISTRIBUTION OF TOTAL INCOME AND ALLOCATED INCOME BY SOURCE: FIVE SURVEYS

Source of Income	CPS	ACS	SIPP	MEPS	NHIS
	Percentage Distribution of Total Income				
Total	100.00	100.00	100.00	100.00	NA
Wages and Salaries	77.71	75.91	71.84	82.64	NA
Self-employment	5.07	6.16	10.71	NA	NA
Asset Income	4.10	5.01	2.41	3.47	NA
Social Security or Railroad Ret.	6.02	5.62	6.44	5.70	NA
SSI	0.40	0.43	0.59	0.63	NA
Welfare	0.10	0.13	0.16	0.08	NA
Pensions	3.85	4.82	6.05	3.94	NA
	Percentage Distribution of Allocated Income				
Total	100.00	100.00	100.00	100.00	NA
Wages and Salaries	72.69	74.16	64.03	83.85	NA
Self-employment	6.63	8.08	13.08	NA	NA
Asset Income	7.50	5.53	4.52	5.04	NA
Social Security or Railroad Ret.	6.26	5.89	6.27	5.43	NA
SSI	0.33	0.41	0.50	0.12	NA
Welfare	0.08	0.13	0.18	0.03	NA
Pensions	3.98	4.42	9.46	3.77	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

distribution. Often, multiple variables may be assigned from the same donor to ensure that there is some internal consistency among the allocated values.

A potentially important distinction among allocation methods is whether they make use of "partial" information on the missing amounts. Partial information could include prior wave values (SIPP), detailed income brackets reported in lieu of a dollar amount (MEPS and NHIS), or wage rates and hours worked (used for annual wages in MEPS). Arguably, allocations that make use of partial information are qualitatively different from allocations that rely solely on covariates of the missing items.

The CPS relies heavily on hot deck imputation methods to allocate missing values due to item non-response, and the allocation flags for the income items do not indicate the use of partial information in any form of allocation. The allocation flags in the ACS public use file do not differentiate among types of allocation, so we assume that no ACS allocations use partial information.

In the SIPP, income from earnings may be allocated using a set of procedures specific to labor force items that makes use of prior wave data. Such imputations are designated with a single code (EPPFLAG = 1), so situations in which the respondent provided the equivalent item in a prior wave could not be differentiated from situations where the respondent provided only related items.[58] In examining the frequency with which the survey allocations of income utilize partial information, we will treat these labor force allocations as using partial information. We will do the same with allocations identified as logical edits, allocations from prior wave data, and the ratio adjustments that we made in order to fill in missing months for sample members who had not yet joined the sample or who were not interviewed in those waves. Logical edits are

---

[58] Checking the allocation flags in the prior wave could have been informative, but we were not certain that an amount that was allocated using the labor force procedure in one wave but imputed in the prior wave should unequivocally be regarded as imputed in the current wave.

identified by codes of 3 on the allocation flags, and allocations using prior wave data are identified by codes of 4. The ratio adjustments, as we explained in Chapter III, use data from other waves to allocate amounts to the missing waves. Finally, allocations also include the procedures used to impute what the Census Bureau calls "Type Z non-respondents." These are non-responding persons in responding households. All data for Type Z non-respondents are imputed using hot deck methods; we treat these as allocations *without* partial information.

For MEPS, we consider all income allocations from reported brackets as making use of partial information, and we do the same with allocations of wage and salary income from hourly wage rates and hours worked.[59] Allocations from reported brackets are coded 2 on the allocation flags while allocations of wage and salary income from hourly wage rates and hours worked are identified by a code of 4 on the wage allocation flag. Hot deck imputations that do not use partial information are identified by codes of 5 or 6.

For the NHIS, allocations that were based on detailed bracketed values provided by respondents instead of actual dollar amounts are coded 3 on the allocation flag. We count these allocations as based on partial information.

When we divide allocated dollars of total income into allocations performed with or without partial information, we find that allocations with partial information dominate the allocations for SIPP and MEPS (Table VI.8). In each of these surveys, allocations *without* partial information account for about 7 percent of total income while allocations *with* partial information account for 25 percent of total income in SIPP and 36 percent in MEPS. Allocations with partial information represent only 2 percent of total income in the NHIS while allocations without partial

---

[59] Identifying all allocations from wage rates and hours worked as using partial information is not entirely correct. The allocated amounts are generated from a regression model that uses either reported or imputed wage rates and weeks worked during the calendar year, if such information is available, based on questions asked in the employment section of the interview. In the 2002 data, regression estimates were generated for 3 million weighted persons who have only self-employment recorded in the JOBS file and, therefore, would not have been asked for an hourly rate and usual hours worked.

TABLE VI.8

ALLOCATION OF TOTAL INCOME BY USE OF PARTIAL INFORMATION: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
Amount of Total Income ($billions)	6,468.4	6,346.3	5,766.2	6,257.7	6,115.2
Percent of Dollars Allocated:	34.2	17.6	32.4	42.7	32.4
With Partial Information	0.0	0.0	25.4	35.6	2.2
Without Partial Information	34.2	17.6	6.9	7.1	30.2

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

information account for 30 percent of total income.[60] In both the CPS and ACS, we classified no allocations as using partial information, leaving 34 percent of total income in the CPS and 18 percent in the ACS as allocated without partial information.

Given that we have no direct measure of the quality of the allocated data obtained with the use of partial information, we hesitate to assert that non-response to the income questions is as much less of a problem in the SIPP and MEPS as these results might be read to suggest.

The proportion of total income allocated using partial information does not vary by income quintile in the SIPP whereas it does in MEPS, rising from 25 percent in the lowest quintile to 38 percent in the highest quintile (Table VI.9). In the ACS, as we have noted, there is a modest reduction in the percentage of dollars allocated without partial information as the income quintile increases, and SIPP appears to show the same, but with much lower allocation rates. The lowest quintile in MEPS appears to have a relatively high rate of allocation without partial information to complement its comparatively low rate of allocation *with* partial information. The NHIS does not show a clear relationship between income quintile and allocation with or without partial information.

Differences in the percentage of income allocated by source increase dramatically for SIPP and MEPS when we divide allocations into those generated with or without partial information (Table VI.10). Because of the labor force allocation procedures used for wages and salaries and self-employment income in the SIPP, these two sources emerge with high rates of allocation with partial information and very *low* rates of allocation *without* partial information. The same is true of Social Security income in SIPP, although the reasons are not as obvious, and we do not see the same phenomenon in MEPS. For wages and salaries in MEPS, the combination of bracketed

---

[60] Revisions to the NHIS income questions for the 2007 survey have increased the percentage of respondents providing brackets.

TABLE VI.9

PERCENT OF TOTAL INCOME ALLOCATED WITH OR WITHOUT PARTIAL INFORMATION
BY QUINTILE: FIVE SURVEYS

Family Income Quintile	CPS	ACS	SIPP	MEPS	NHIS
Quintile	Total Income in Billions of Dollars				
Lowest	370.5	368.7	391.4	360.0	356.0
Second	774.1	778.4	750.8	808.4	687.1
Third	1,090.2	1,087.4	1,008.8	1,144.7	1,020.3
Fourth	1,446.8	1,415.8	1,307.2	1,461.8	1,479.1
Highest	2,786.7	2,696.0	2,308.0	2,483.0	2,572.7
Quintile	Percent of Dollars Allocated with Partial Information				
Lowest	0.0	0.0	24.8	24.6	2.3
Second	0.0	0.0	25.5	31.7	2.4
Third	0.0	0.0	25.0	34.6	1.1
Fourth	0.0	0.0	25.3	37.3	1.8
Highest	0.0	0.0	25.8	37.9	2.8
Quintile	Percent of Dollars Allocated without Partial Information				
Lowest	35.1	21.8	8.6	11.5	31.7
Second	33.6	20.1	7.5	7.1	32.2
Third	32.9	18.7	7.2	6.5	28.2
Fourth	32.5	17.2	6.7	6.2	27.1
Highest	35.6	16.1	6.5	7.3	32.0

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003
CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC,
and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002,
from the 2002 ACS.

TABLE VI.10

PERCENT OF INCOME ALLOCATED WITH AND WITHOUT PARTIAL INFORMATION:
FIVE SURVEYS

Source of Income	CPS	ACS	SIPP	MEPS	NHIS
	Percent of Dollars Allocated with Partial Information				
Total Income (NHIS family income)	0.0	0.0	25.4	35.6	2.2
Wages and Salaries (NHIS earnings)	0.0	0.0	24.6	38.9	0.0
Self-employment	0.0	0.0	34.8	NA	NA
Asset Income	0.0	0.0	38.6	35.5	NA
Social Security or Railroad Ret.	0.0	0.0	24.8	13.8	NA
SSI	0.0	0.0	15.3	0.0	NA
Welfare	0.0	0.0	12.4	0.0	NA
Pensions	0.0	0.0	18.7	18.9	NA
	Percent of Dollars Allocated without Partial Information				
Total Income (NHIS family income)	34.2	17.6	6.9	7.1	30.2
Wages and Salaries (NHIS earnings)	32.0	17.2	4.3	4.4	31.8
Self-employment	44.7	23.1	4.7	NA	NA
Asset Income	62.6	19.4	22.2	26.5	NA
Social Security or Railroad Ret.	35.5	18.5	6.7	26.9	NA
SSI	28.0	16.7	12.2	7.9	NA
Welfare	29.2	17.9	23.5	13.5	NA
Pensions	35.4	16.2	31.9	22.0	NA

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: Estimates in the top and bottom panels sum to the estimates reported in Table VI.5.

amounts and predictions based on wage rates and hours worked account for a very high rate of allocation with partial information and very low rate of allocation without partial information. Asset, pension and welfare income have high rates of allocation without partial information in both SIPP and MEPS.

C. ALLOCATION RATES BY INTERVIEW MONTH

With special tabulations of internal ACS data prepared by the Census Bureau, we were able to examine allocation rates in the 2003 survey by interview month. We undertook this assessment in order to determine if non-response to the income questions rose over the course of the calendar year, which might be indicative of respondents having increasing trouble with the concept of income over the past 12 months as they moved away from the prior calendar year. Instead, we found a surprising pattern that challenges long-held notions about the best time of year to ask survey respondents about their annual income.

For total income and for each of seven individual sources of income, the highest allocation rate occurs in or around the month of April, when tax returns are due to be filed (Table VI.11). Generally, the allocation rate in the peak month is one-third to one-half higher than in January, which has the lowest allocation rate for total income, wages and salaries, self-employment income, and interest and dividends and one of the lowest allocation rates for each of the other sources. The existence of a substantial peak around the month of April is true of nearly every income source, including Social Security income. For most of the income sources, allocation rates tend to stabilize in the second half of the year.

While this overall pattern is evident at all levels of relative income, it is weakest among persons below the poverty threshold and grows in strength with rising income (Table VI.12). Among persons in families above 400 percent of poverty, the 24 percent allocation rate in April is 7 percentage points higher than in January and 5 to 6 percentage points higher than in the

TABLE VI.11

PERCENT OF INCOME ALLOCATED BY CALENDAR MONTH AND SOURCE: 2003 ACS

Month	Total Income	Wages & Salaries	Self-Employ-ment	Interest and Dividends	Social Security	SSI	Welfare	Retirement
Jan	18.5	15.2	19.2	14.0	17.0	17.2	15.9	14.6
Feb	19.0	15.6	20.9	**19.5** *	16.8	17.2	17.2	13.2
Mar	**22.8** *	**19.1** *	23.8 *	**22.3** *	**21.3** *	**18.5**	17.0	**19.6** *
Apr	**24.6** *	**20.9** *	24.4 *	**25.1** *	**23.8** *	17.1	**19.5**	**21.5** *
May	**24.2** *	**21.3** *	**27.4** *	19.3 *	**20.9** *	**18.4**	19.1	**17.9** *
Jun	21.9 *	18.6 *	**25.5** *	19.0 *	20.1 *	**17.6**	**25.8** *	16.6 *
Jul	19.6 *	15.9	23.3 *	15.9	17.7	14.6	16.1	14.6
Aug	20.5 *	17.2 *	23.7 *	17.4 *	16.6	16.0	**20.3**	13.1
Sep	19.4	15.8	22.4	14.2	17.5	15.8	11.2	13.9
Oct	19.6 *	16.4 *	20.8	14.1	16.3	14.9	14.3	14.7
Nov	19.7 *	16.3 *	22.7 *	17.3	16.6	15.5	15.1	13.6
Dec	19.8 *	16.6 *	22.2	15.3	16.6	13.9 *	16.1	13.5

Source: U.S. Census Bureau, from the 2003 ACS.

Note: Three highest allocation rates for each source are indicated in bold.

* Statistically significant from January at the .05 level or greater.

TABLE VI.12

PERCENT OF INCOME ALLOCATED BY CALENDAR MONTH AND
POVERTY RELATIVE: 2003 ACS

| | Poverty Relative Based on Family Income | | | | |
Month	Under 100%	100% to < 200%	200% to < 400%	400% or More	Total
Jan	21.8	22.5	20.4	17.5	18.5
Feb	22.1	22.4	20.1	18.1	19.0
Mar	23.0	25.1 *	**23.1** *	**22.4** *	**22.8** *
Apr	**23.6**	**27.8** *	**24.5** *	**24.3** *	**24.6** *
May	**26.2** *	**26.1** *	**25.0** *	**23.7** *	**24.2** *
Jun	22.4	**25.3** *	22.6 *	21.2 *	21.9 *
Jul	**23.4**	24.1	20.9	18.7 *	19.6 *
Aug	23.3	23.3	21.3	19.9 *	20.5 *
Sep	21.3	22.5	21.0	18.6	19.4
Oct	23.4	23.6	20.7	18.7	19.6 *
Nov	24.0	23.8	20.6	18.9 *	19.7 *
Dec	22.9	23.5	21.5	18.8 *	19.8 *

Source: U.S. Census Bureau, from the 2003 ACS.

Note: Three highest allocation rates for each source are indicated in bold.

* Statistically significant from January at the .05 level or greater.

283

months before March and after June. Further, in most months, allocation rates peak among persons between 100 and 200 percent of poverty and then decline as relative income rises. They decline least in the months of March through May.

The same strengthening of the pattern with rising income is evident when respondents are classified by family income quintile (Table VI.13). This holds true even though allocation rates in every month decline with rising income through the fourth quintile. In every quintile April has the highest allocation rate, but the difference between April and the earliest months and second half of the year grows with rising income—at least until the fourth quintile. Allocation rates by month are nearly identical between the fourth and highest quintiles.

These findings suggest that while respondents may be best informed about their annual income when they are putting together their tax returns, they appear to be more sensitive to reporting what they know than at other times of the year. This produces a sharp reduction in response rates between the beginning of the year and the fourth month. The fact that this pattern is weakest among persons least likely to be filing tax returns (those below 100 percent of poverty) and among the sources of income *held* by persons who are least likely to be filing tax returns (SSI and welfare income) is consistent with the inference that the tax season induces sensitivity to reporting income in a federal survey. While strongly suggestive, however, these findings require further study to establish both the stability of the pattern over time and its origins.

D. APPROXIMATION AND ROUNDING

One reason to examine rounding in the context of policy analytic use of income data is that the heaping of incomes at well-spaced values can distort the results of policy simulations involving the use of income thresholds to establish program eligibility. An eligibility threshold that lies near an income amount with excessive heaping will produce dramatically different

TABLE VI.13

PERCENT OF INCOME ALLOCATED BY CALENDAR MONTH AND
FAMILY INCOME QUINTILE: 2003 ACS

| Month | Family Income Quintile | | | | | Total |
	Lowest	Second	Third	Fourth	Highest	
Jan	22.3	22.2	19.7	17.5	17.1	18.5
Feb	23.7	20.9	19.2	18.3	18.0	19.0
Mar	**25.6** *	**24.6** *	**22.9** *	21.7 *	**22.3** *	**22.8** *
Apr	**27.1** *	**26.2** *	**24.7** *	**24.1** *	**24.0** *	**24.6** *
May	**26.2** *	**25.3** *	**24.7** *	**23.6** *	**23.8** *	**24.2** *
Jun	25.5 *	24.0 *	22.2 *	**22.3** *	20.4 *	21.9 *
Jul	24.4 *	21.8	20.5	17.9	18.9	19.6 *
Aug	23.9 *	22.2	21.5 *	19.3 *	19.8 *	20.5 *
Sep	23.2	21.7	20.7	18.2	18.4	19.4
Oct	24.1 *	22.1	20.0	18.3	18.7	19.6 *
Nov	23.8	22.0	20.1	19.3 *	18.6	19.7 *
Dec	23.7	21.9	20.8	19.4 *	18.7	19.8 *

Source: U.S. Census Bureau, from the 2003 ACS.

Note: Three highest allocation rates for each source are indicated in bold.

* Statistically significant from January at the .05 level or greater.

results depending on whether the threshold falls just below or just above that amount. If the former, a simulation will mildly understate the impact of a small change in policy; if the latter, a policy simulation will grossly overstate the impact of the policy change.

Another reason that rounding is a concern when assessing income data for policy work is that a high level of rounding suggests inaccuracy or a lack of precision in the data. This may reduce the analyst's confidence in the data source or the results that it produces.

We examined the extent of rounding in reported incomes below $52,500 for earnings, wages and salaries, Social Security benefits, other retirement income, total personal income, and total family income in the five general population surveys, which allow reported income to be separated from allocated income.[61] Earnings and total family income are the only income amounts collected in the NHIS and, therefore, the only amounts on which all five surveys could be compared. Except in the SIPP, where all annual amounts are built up from monthly values, Social Security benefits reported at the person level will have been collected as a single value. Most respondents reporting wages and salaries, retirement income, and even earnings are likely to have supplied a single value in response to one question even though multiple questions were asked.

The results show the differential impact of few versus many income questions. The NHIS relies on a single, person-level question to collect earnings and a single, family-level question to collect total family income. In this survey, 40 percent of the reported earnings and 36 percent of the reported family incomes below $52,500 are multiples of $5,000, and 23 percent of the earnings and 21 percent of the total family incomes are multiples of $10,000 (Table IV.14). At the opposite extreme, the SIPP, with numerous *monthly* questions, shows very little rounding

---

[61] We selected $52,500 in order that we might examine the frequency of rounding up to levels of $50,000. The ACS public use file has incomes of $50,000 or greater rounded to the nearest $1,000.

TABLE VI.14

REPORTING OF ROUNDED VALUES BY SOURCE OF INCOME BY SURVEY
AMONG POSITIVE DOLLAR AMOUNTS BELOW $52,500

Income Source and Level of Rounding	CPS	ACS	SIPP	MEPS	NHIS
Earnings					
Percent divisible by $5,000	27.8	29.6	1.3	18.6	39.8
Percent divisible by $10,000	15.8	17.4	0.8	9.7	22.9
Percent of income in range	82.1	82.4	88.8	81.8	80.9
Wages and Salaries					
Percent divisible by $5,000	27.2	29.7	0.9	NA	NA
Percent divisible by $10,000	15.4	17.4	0.6	NA	NA
Percent of income in range	82.2	82.7	89.4	NA	NA
Social Security					
Percent divisible by $5,000	0.6	4.3	0.3	6.9	NA
Percent divisible by $10,000	0.4	1.9	0.1	3.6	NA
Percent of income in range	100.0	100.0	100.0	100.0	NA
Retirement Income					
Percent divisible by $5,000	4.5	8.0	1.0	7.4	NA
Percent divisible by $10,000	2.7	4.3	0.5	3.7	NA
Percent of income in range	95.6	95.4	99.0	100.0	NA
Total Personal Income					
Percent divisible by $5,000	13.7	19.7	0.6	9.7	NA
Percent divisible by $10,000	7.8	11.5	0.4	5.1	NA
Percent of income in range	84.6	84.2	90.8	85.5	NA
Total Family Income					
Percent divisible by $5,000	11.0	16.2	0.6	11.1	35.6
Percent divisible by $10,000	6.2	9.5	0.4	6.1	20.9
Percent of income in range	66.9	66.0	77.7	72.0	60.3

Source: Mathematica Policy Research, from tabulations of calendar year 2002 income from the 2003 CPS ASEC supplement, the 2001 SIPP panel, the 2002 Full-year Consolidated MEPS-HC, and the 2003 NHIS, and prior 12 months income, inflation-adjusted to calendar year 2002, from the 2002 ACS.

Note: Allocated amounts are excluded from each source. Family income for the NHIS is based on the NHIS family, which is the level at which family income was reported.

when amounts are aggregated to the annual level. Except for earnings, fewer than 1 percent of the reported amounts are divisible by $5,000.

Among the remaining three surveys, the ACS shows the most rounding, with 30 percent of both earnings and wage and salary income, 20 percent of total personal income, and 16 percent of total family income being divisible by $5,000. The fractions are much lower for Social Security and other retirement income (4 and 8 percent, respectively).

On wages and salaries as well as earnings, the CPS is only marginally better than the ACS, with 27 to 28 percent of the amounts being divisible by $50,000 and 15 to 16 percent being divisible by $10,000. For Social Security benefits, however, the CPS approaches the SIPP, with only 0.6 percent of the reported amounts being multiples of $5,000. On total family income the CPS compares to MEPS, both of which have 11 percent of responses divisible by $5,000. MEPS shows less rounding than the CPS on earnings and total personal income but resembles the ACS on Social Security and retirement income, where 7 percent of the reported responses are divisible by $5,000.

We also examined rounding in the PSID, but because the PSID income variables do not include all of the items reported in Table VI.14 or, for some items, apply to a narrower universe, we present the PSID results separately. Most of the PSID variables in Table VI.15 were constructed (by PSID staff) from the responses to multiple questions, and, as we have seen, this tends to reduce the observed level of rounding.

The most disaggregated variable, the family head's wages and salaries, combines responses to a single wage and salary question over potentially multiple jobs. Since most workers have only one job, however, most of the values of this variable reflect a single response. Not surprisingly, this variable exhibits the highest degree of rounding, with 25 percent of the responses divisible by $5,000 and 14 percent divisible by $10,000. These values compare closely

TABLE VI.15

REPORTING OF ROUNDED VALUES BY SOURCE OF INCOME
AMONG POSITIVE DOLLAR AMOUNTS BELOW $52,500: PSID

Income Source and Level of Rounding	PSID
Combined Earnings of Head and Wife	
Percent divisible by $5,000	18.6
Percent divisible by $10,000	10.7
Percent of income in range	59.4
Family Head's Wages and Salaries	
Percent divisible by $5,000	24.9
Percent divisible by $10,000	13.9
Percent of income in range	73.8
Social Security Income of Family	
Percent divisible by $5,000	2.4
Percent divisible by $10,000	1.0
Percent of income in range	100.0
Combined Transfer Income of Head and Wife	
Percent divisible by $5,000	4.5
Percent divisible by $10,000	2.2
Percent of income in range	97.9
Labor Income of Heads and Wives	
Percent divisible by $5,000	22.8
Percent divisible by $10,000	12.5
Percent of income in range	78.0
Total Family Income	
Percent divisible by $5,000	5.5
Percent divisible by $10,000	3.1
Percent of income in range	58.9

Source: Mathematica Policy Research, from tabulations of calendar year
2002 income from the 2003 PSID.

Note: Allocated amounts are excluded from each source. Family income
is based on the PSID family. Wives include unmarried partners.

to what we found for CPS wages and salaries (27 percent and 15 percent), although the CPS variable includes all workers. It is possible that we would see less rounding in the CPS wages and salaries if we limited them to family reference persons, defined as is done in the PSID, which would be the male in a husband-wife family. We suggest this because it is plausible that there is more rounding in the reporting of wages and salaries for non-principal earners than for the principal earner (who is more likely to be the CPS respondent or spouse of the respondent).

The labor income of heads and wives incorporates additional components of wage and salary income that the PSID collects in separate fields, including overtime pay, bonuses, and tips. Collecting wage and salary income in this way ought to reduce the amount of rounding, even though most families may report only one amount for the head and one for the wife or partner, if present. This wage and salary income is counted separately for family heads and wives/partners. We do see less rounding, but only by a modest amount: 23 percent is divisible by $5,000 and 12.5 percent is divisible by $10,000. Adding income from an unincorporated business and pooling the incomes of heads and wives/partners to create a combined earnings amount for each family reduces the rounding further, down to 19 percent divisible by $5,000 and 11 percent divisible by $10,000. None of the other four surveys shows a marked reduction in rounding between wages and salaries and earnings, although it is possible that it is the pooling of heads' and wives' earnings rather than the addition of self-employment income that accounts for most of the reduction in the PSID.

With 2.4 percent divisible by $5,000 and 1.0 percent divisible by $10,000, family Social Security income in the PSID shows about half as much rounding as personal Social Security income in the ACS but more than the CPS or SIPP. The combined transfer income of the head and wife/partner shows the same level of rounding as retirement income in the CPS, with 4.5 percent divisible by $5,000 and 2.2 percent divisible by $10,000. Transfer income would include

retirement income (other than Social Security) as well as a number of other sources. Lastly, total family income in the PSID exhibits only half the level of rounding as total family income in the CPS and MEPS and one-third as much as total family income in the ACS, with 5.5 percent divisible by $5,000 and 3.1 percent divisible by $10,000.

E. ISSUES IN USING ALLOCATION

The choice of an allocation method or an overall strategy of allocation can have implications for the quality of income data. One example emerged in Chapter IV with our discussion of inconsistencies between family income and the sum of family members' earnings in the NHIS. Another grows out of our study of rounding.

1. Internal Consistency

In surveys that collect detailed sources of income, total personal income and total family income are calculated as sums of the reported or allocated amounts of the individual sources.[62] Amounts can be allocated without a concern that they will be inconsistent with an existing total. In the NHIS, total family income is collected separately from personal earnings and also allocated independently. As we documented in Chapter V, the sum of personal earnings exceeds total family income for an estimated 61.7 million persons or more than one-fifth of the population in the 2003 NHIS. Allocation accounts for 71 percent of the discrepant cases, thus compounding the occasionally inconsistent reports of respondents. Eliminating the inconsistencies will require changes in the allocation strategy as well as editing of responses or the addition of a check to the automated survey instrument.

---

[62] The Census Bureau also applies topcoding to the individual sources before summing them to calculate total personal and total family income. In this way, the totals are assured of being consistent with their components.

2. Replicating Deficiencies in Reported Data

Allocation methods that are based on substituting missing items from other, similar records will tend to replicate any reporting patterns. For example, rounding will be repeated in the allocated values if imputation is done by a hot deck procedure, but it will not be repeated if the imputation procedure is model-based, unless it is explicitly added afterwards. Thus we see in Table VI.16 that the patterns of rounding that were evident for reported values in the previous section are repeated in each of the five general population surveys except the NHIS, which uses model-based imputation. In the NHIS, there is no rounding in the allocated values.

These findings underscore that fact that when choosing an allocation method, data producers need to consider whether it is desirable or undesirable to replicate specific weaknesses in the reported data.

The PSID raises an additional issue with respect to the selection of an allocation method. The PSID, which allocates missing data only for selected items, does not use either hot deck or sophisticated model-based methods but relies on simpler approaches, which seem to produce substantial rounding. When the family head's wage and salary income is allocated, 45 percent of the values are divisible by $5,000, and 44 percent are divisible by $10,000 (data not shown), but the round allocated values are not distributed across the range of allocated values. Instead nearly half of the allocated records are assigned the same value of wages and salaries: $30,000. The only allocated value below $30,000 that is divisible by $5,000 is $15,000. There are no other round values among the allocated amounts below $30,000. Rounded values do appear above $35,000, but they are infrequent. From this distribution of allocated values it would appear that the PSID may employ two different methods of allocation, one of them being a conditional mean imputation of some sort and the other quite possibly a simple regression model. The substantial heaping at a single value suggests that the latter method is much better suited to the PSID application.

292

TABLE VI.16

ALLOCATION OF ROUNDED VALUES BY SOURCE OF INCOME BY SURVEY
AMONG POSITIVE DOLLAR AMOUNTS BELOW $52,500

Income Source and Level of Rounding	CPS	ACS	SIPP	MEPS	NHIS
Earnings					
Percent divisible by $5,000	29.5	19.4	1.4	12.3	0.0
Percent divisible by $10,000	17.1	11.3	0.9	6.8	0.0
Percent of income in range	83.0	86.3	94.1	84.1	78.3
Wages and Salaries					
Percent divisible by $5,000	28.3	19.3	1.0	NA	NA
Percent divisible by $10,000	16.4	11.2	0.6	NA	NA
Percent of income in range	83.2	86.5	95.1	NA	NA
Social Security					
Percent divisible by $5,000	0.6	4.2	0.3	6.0	NA
Percent divisible by $10,000	0.4	1.7	0.1	2.9	NA
Percent of income in range	100.0	100.0	100.0	100.0	NA
Retirement Income					
Percent divisible by $5,000	3.7	6.7	1.1	7.8	NA
Percent divisible by $10,000	2.1	3.5	0.8	3.1	NA
Percent of income in range	96.3	96.3	99.6	100.0	NA
Total Personal Income					
Percent divisible by $5,000	7.4	13.7	0.2	5.0	NA
Percent divisible by $10,000	4.1	7.8	0.1	2.6	NA
Percent of income in range	90.3	88.3	96.7	88.3	NA
Total Family Income					
Percent divisible by $5,000	6.0	11.2	0.2	5.7	0.0
Percent divisible by $10,000	3.3	6.4	0.1	3.1	0.0
Percent of income in range	79.7	78.6	91.3	73.0	62.9

Source: Mathematica Policy Research, from tabulations of the 2003 CPS ASEC
supplement, the 2002 ACS, the 2001 SIPP panel, the 2002 Full-year
Consolidated MEPS-HC, and the 2003 NHIS.

Note: Amounts reported by respondents are excluded from each source. Family
income for the NHIS is based on the NHIS family, which is the level at which
such income was allocated.

VII. SYNTHESIS OF FINDINGS

The purpose of this study was to conduct a comprehensive and systematic assessment of the income data and its utility for policy-related analyses in eight major surveys. To this end we have assembled a detailed descriptive portrait of the eight surveys and conducted an extensive empirical analysis. The empirical analysis included comparisons of the surveys using, to the extent possible, comparable reference periods,[64] universes, income concepts, and family definitions. The empirical portion also included analyses of the impact of various design choices. The assessment focused on three issues:

- The quality and usability of each survey's income and poverty data for policy-related analyses.

- The overall impact of different design and methodological approaches.

- Specific design and processing choices that may be related to the quality and utility of income and poverty data in each survey.

In this synthesis, we pull together findings from both the descriptive and empirical components of the study. We conclude with a brief discussion of next steps.

A. QUALITY AND USABILITY OF INCOME AND POVERTY DATA

As a survey that was designed to support policy analysis over a wide range of topics, SIPP would appear to have a number of advantages over the other general population surveys in measuring income and, especially, its distribution. Consistent with this expectation, SIPP performs much better than the other surveys in identifying program participants and capturing their income. SIPP also captures more income from the bottom of the income distribution than

---

[64] The study uses income data for 2002 (HRS and MCBS income for 2003 were deflated with the CPI-U) that covers a calendar year, except for the rolling reference period in ACS, which spans 23 months.

the other general population surveys, obtaining the most total dollars from the bottom quintile and finding the fewest persons in poverty. Yet this advantage quickly fades as we move up the income ladder or broaden our examination of poverty to subpopulations. Despite finding the fewest poor, SIPP finds *more* persons among the *near-poor* (between 100 and 200 percent of poverty) than any other of the five general population surveys. SIPP also finds more poor *children* than CPS, ACS, or MEPS. Most importantly, the biggest difference among the five surveys with respect to aggregate income is SIPP's capturing just 89 percent as much income as the CPS while NHIS, MEPS, and ACS capture 95 to 98 percent. SIPP fares no better on unearned versus earned income, capturing just 90 percent as much unearned income and 89 percent as much earned income as the CPS.

SIPP's performance raises a number of methodological issues, which are discussed in the next section.

Of the five general population surveys, the CPS remains the most widely used for policy analysis. Yet several limitations of these data are apparent—some of them well known, others not. While the CPS captures the most total income among the five surveys, its greatest advantage is in the top quintile, which is the least relevant for policy analysis. SIPP captures more income from the bottom quintile and finds fewer poor. The ACS captures as much income as CPS from the bottom three quintiles and finds fewer *near-poor*. MEPS collects more income than CPS from the bottom four quintiles, although the MEPS numbers must be qualified because they are not independent of the CPS estimates.

ACS, SIPP, and MEPS all find more persons and a higher percentage of the population with earnings than does the CPS.[65] The higher per capita earnings in the CPS suggest that the shortage

---

[65] While the CPS is the official source of monthly labor force statistics, this status does not extend to annual estimates. The CPS collects much less detailed information on labor force activity in the prior year than in the reference period for the official monthly statistics.

of earners in the CPS may be among lower-income workers, who attract more policy interest than higher-income workers. Overall, the ACS captures more unearned income than the CPS, although the CPS estimates exceed SIPP and MEPS by more than 10 percent. Estimates of persons receiving welfare or Food Stamps, covered by SSI, or enrolled in Medicaid during 2002 are about a third lower in the CPS than SIPP. ACS estimates of welfare or Food Stamp recipients are also markedly higher than the CPS estimates, and both MEPS and NHIS estimates of SSI recipients are higher than the CPS as well. The CPS estimates of persons ever enrolled in Medicaid during 2002 are exceeded by SIPP estimates of Medicaid enrollees in a single month (December). This latter observation ties into a well-known problem with CPS estimates of the uninsured, which represent persons who reported no coverage during the prior calendar year but compare to or are exceeded by SIPP, MEPS, and NHIS estimates of persons uninsured at a point in time.

Overall, ACS income data compare favorably to CPS data in a number of respects and appear to capture more income from selected subpopulations. They have low allocation rates, and the survey itself has a very high overall response rate. In short, the ACS income data look remarkably good given that they are collected in large part through a mailback questionnaire, without the benefit of interviewers, and with a small set of questions administered to a massive sample. In view of the expectations that have been set for ACS as a source of household and family income data at the state and local areas, our findings with respect to this survey should be considered very good news.

Yet there are several important limitations of ACS data for policy analysis. The rolling reference period implies that in a time of significant change in the economy, as we are experiencing currently, estimates of employment and income obtained early and late in the survey year may differ significantly. The suppression of survey month on the public use file

297

limits the analyst's ability to contend with this type of problem as well as other instances of change over the year. The small number of additional variables also restricts the range of policy analyses that can be conducted with ACS data. Counting students where they attend college rather than with their families (where they usually live) will create millions of pseudo-poor. This is not yet evident in our study because college dormitories were not added to the ACS sample frame until a later year. The fact that students will be counted in their parents' homes during the summer months but not the school year is another reason why survey month would be a useful addition to the public use file.

Post-stratification of the MEPS to the CPS poverty distribution leaves us unable to assess how much income MEPS is actually capturing and how it is distributed. With the post-stratification, MEPS has more income than the CPS between the 20th and 80th percentiles of the income distribution but less in the bottom fifth and, especially, the top fifth, but how would it look without the post-stratification? Comparison of MEPS and CPS poverty rates are even less informative, as they are affected directly by the post-stratification.

MEPS users must determine how to work with certain inconsistencies between reported employment and reported income that derive from the collection of these data in separate parts of the survey instrument coupled with a policy of not imposing consistency edits on these items. Users whose analyses require person weights must also determine how to handle a subset of sample members, weighting up to more than six million persons, who have missing data on family members and, because of this, exceedingly high measured poverty rates. Different users will choose to handle these cases in different ways, injecting additional variation into their analytic findings beyond what can be attributed to alternative modeling decisions.

The collection of income data in NHIS has been a low priority for NCHS historically, and restricting the amounts of total family income and personal earnings to an internal file effectively

precludes the use of these income data in time-sensitive analysis. Using a single question to collect total family income (albeit not the family used in official poverty measures), NHIS obtains an aggregate amount that approaches 95 percent of the CPS total and displays a broadly similar distribution but does worst in the bottom quintile, which is the most important from a policy-analytic standpoint. The fact that a significant number of respondents report person-level earnings that sum to more than the reported total family income and that total earnings are even more likely to exceed total family income when one or both were imputed suggests that a different strategy might be more effective. Collecting unearned income for each person, to complement earned income, would yield person-level total income for every person and perhaps a more complete accounting of total family income.

Despite a weighted population that falls short of the CPS by 21 million persons, the PSID captures 4 percent more aggregate income. PSID income exceeds the CPS in every quintile, with the biggest margin, nearly 6 percent, occurring in the top quintile, where the CPS holds the greatest advantage over the other four general population surveys. PSID per capita income, which adjusts for differences in population size, exceeds CPS per capita income by 10 to 14 percent in all five quintiles. PSID also finds a higher percentage of the population with earnings than CPS, SIPP, or ACS. Were it not for the uncertainty regarding the representativeness of the PSID after 40 years, we would see these as evidence of better capture of income in the panel survey. Instead the PSID may simply over-represent higher-income families. While this does not detract from the survey's value for longitudinal analysis, national generalizations from the data are problematic.

Surveys of restricted populations face special challenges in developing representative estimates, owing to the independent selection probabilities of spouse and partners. This was

evident for estimates of aggregate income in both the HRS and MCBS, and it would affect the use of these data to develop cost estimates of legislative proposals.

The income data collected from Medicare beneficiaries in the MCBS are limited to a single dollar amount that includes a spouse's income. For single persons the distribution is consistent with other surveys, but if aggregated, the total income effectively double-counts the incomes of spouses who are also beneficiaries. Limiting the income question to the beneficiary's income would eliminate this double-counting. Asking separately for the incomes of other family members and obtaining family size would enable users to estimate the poverty status of beneficiaries, which is not currently possible for much of the sample. While MCBS data are not released in a public use file, potential users may apply to obtain access to the data for specified uses at their own computing facilities. Whether such uses could encompass time-sensitive policy analysis, as opposed to analyses requiring advance approval, is not clear.

Comparisons of average family income for persons 51 and older in the HRS and the CPS, ACS, and SIPP reveal substantially higher incomes in the HRS. While RAND's construction of family income may play a role in findings for persons living with relatives other than a spouse, we found that average incomes for singles were 22 percent higher than the CPS while average incomes for persons with spouses or partners were 28 percent higher than CPS incomes for persons with spouses. Differences are very consistent across most of the income distribution but grow substantially in the top quintile. These findings would require much more study to determine whether HRS is truly capturing substantially more income than the other surveys or whether there is another explanation.

One general finding on income measurement is that the identification of self-employment income is a particularly weak area, which is reflected in widely varying estimates, with MEPS having both the lowest and highest estimates, depending on whether the estimate is based on

reported income by source or type of employment. Given that self-employed persons may be the focus of policy initiatives related to health insurance and other areas, this is a glaring weakness of income data collection.

A more general area of weakness in survey income data is the comparatively high level of item non-response to income questions. A useful measure of the overall impact of item non-response is the proportion of total income that was allocated. About one-third of total income in the CPS, SIPP, and NHIS was allocated, making the quality of these data dependent on the quality of the allocation methods used to fill in the missing data. The ACS fared markedly better with only 18 percent of total income allocated while 43 percent of total income in MEPS was allocated. Allocation rates show no trend by quintile of family income in the CPS, SIPP and NHIS, but they trend downward in ACS and upward in MEPS. The similarity of allocation patterns in SIPP and NHIS, which ask the most and fewest income questions, respectively, suggests that the level of income detail requested of respondents may have little if any impact on how much income must be "made up" to compensate for non-response. Lastly, SIPP and MEPS are unique among the five surveys in their use of partial information to allocate missing earnings, which dominate total income. SIPP makes extensive use of data collected in prior waves while MEPS predicts earnings from reported wage rates and hours worked or allocates dollar amounts from reported ranges. In both surveys, allocations *without* partial information account for about 7 percent of total income.

We examined the prevalence of rounding in selected income items in the six surveys that differentiated reported and allocated amounts. Significant rounding was evident in reported earnings at the person level in the CPS, ACS, MEPS, NHIS, and PSID. Between 19 and 40 percent of the amounts below $52,500 were multiples of $5,000. Social Security income exhibited substantially less rounding than earnings in every survey. Yet even total family

income, which combines amounts over persons and sources in all but NHIS, had rounded amounts for 11 to 16 percent of families in the CPS, ACS, and MEPS while NHIS had rounded amounts for 36 percent of families. Only SIPP showed no significant degree of rounding on any of the items. All annual amounts in SIPP are sums of monthly values.

B. SURVEY DESIGN AND METHODOLOGY

In the introductory chapter we highlighted the following differences in survey design and methodology as bearing on survey estimates of income: subannual versus retrospective annual income data collection, the breadth and depth of income questions, and strategies for measuring income in the context of a rolling sample. Comparison of survey estimates with an eye to these aspects of survey design raised more questions than it answered. Important questions for follow-up research are suggested by our findings.

SIPP is the only survey that collects income at the monthly level. The annual estimates prepared for this study were built up from monthly amounts. SIPP's approach is clearly effective for program participation, where the SIPP estimates exceed those of other surveys by a wide margin. Given this, why does SIPP end up with 10 percent fewer dollars of total earned income and total unearned income than the CPS—and even 6 percent less total income than NHIS?

Given that SIPP employs an entirely different approach to collecting income data than any of the other surveys, we cannot conclude from these results that the SIPP approach is flawed; nor can we conclude that the comparatively low estimates of total income are the result of poor implementation. It may be both or neither. Perhaps the SIPP design is more effective among people with erratic income flows and less effective among those with more regular income flows. Alternatively, perhaps the SIPP field staff has focused on getting good data from low-income families with a weaker emphasis on higher-income families. The lower capture of income could also be a function of the dynamic character of the SIPP sample that SIPP

estimation procedures do not properly handle. With their similar panel designs but different approaches to measuring income, SIPP and MEPS could provide useful comparative data on their alternative approaches were it not for the fact that the MEPS data are post-stratified to the distribution of poverty status in the CPS. At the same time, however, we should not dismiss the possibility that asking retrospective questions of a fixed simple—the design element shared by the other four general population surveys—may impart a bias of its own, but this one in an upward direction. That is, SIPP's shortfall may be overstated. The four general population surveys that share the retrospective approach yield surprisingly close estimates of total income despite widely ranging approaches to measurement.

Understanding why SIPP estimates are so much lower than the other surveys is extremely important as the Census Bureau moves forward with a redesign of SIPP that may change many of the features that are unique to SIPP. It is also an exceedingly challenging question from a methodological perspective.

With just a single question asked at the family level, NHIS was able to capture 95 percent as much total income as the CPS. ACS captured 98 percent as much as CPS with seven questions, although these were asked of each person. This suggests that large batteries of questions may not generate much additional total income. Instead, their value lies elsewhere, which may or may not be relevant to the intended use of income data in a given survey. Detailed questions appear to produce less rounding, presumably better accuracy at the family and person level, plus the source detail that may be needed for simulating program eligibility. It is also apparent that the impact of additional questions is not uniform across the income distribution. Compared to the CPS, NHIS misses proportionately more income in the bottom quintile than in quintiles two through four, and one result is a higher estimated poverty rate after differences in family definition are taken into account (see below).

A critical issue for income measurement in a rolling sample is whether a rolling versus fixed reference period for income is to be preferred. Policy applications may favor one over the other, depending on the type of application. For example, a rolling reference period maintains a uniform lag between the income reference period and statuses measured at the time of the interview (such as health insurance coverage or program participation). Equally important, however, is which approach will yield better data. Does the quality of income data for a fixed reference period deteriorate as the interview date moves farther from the reference period? Alternatively, can respondents report income for the past 12 months as accurately as they can for the previous calendar year? Will they fall back on reporting their incomes for the prior calendar year (or show other evidence of diminished quality, such as higher non-response or increased variance)?

Our examination of income reporting by month (ACS) or quarter (NHIS) turned up little evidence that respondents in either survey had difficulty with the income concept in ways that were reflected in reporting patterns over time. Perhaps the low rate of inflation and slow pace of economic change during 2002 and 2003 contributed to our null findings and the findings of a similar assessment conducted with survey data for 2008 would be different. For now, our questions about the choice of reference period remain open questions.

C. SPECIFIC DESIGN AND PROCESSING CHOICES

While we were not able to demonstrate the impact of fundamental survey design features discussed in the preceding section, we were able to simulate or otherwise estimate the impact of a number of other survey design features and aspects of post-survey processing. These elements include:

- Family definition, which determines whose income is aggregated and what poverty threshold is used to determine poverty status

- Contemporaneous versus fixed family composition and income for poverty measurement—that is, whether family composition and income reflect changes in composition over the reference period or whether family composition is measured at a fixed point in time and income collected for the members of this fixed family

- Interview month, which affects recall intervals, family composition, the lag between a fixed family composition and the income reference period, response rates, and the quality of income data

- Choice of imputation methodology, including its impact on the distribution of imputed values and their consistency with reported values

- Application of consistency checks between related items collected at different places in the questionnaire

- Application of inflation adjustments when income reference periods differ

- Adding non-periodic withdrawals from retirement accounts to the income concept

- Post-stratification in general and post-stratification on income in particular

Each of these can affect the quality of the income data ultimately released to users from a survey and how the income and poverty data compare to estimates from other surveys.

Several of the surveys included in this study define the family more broadly than the CPS, including unmarried partners and their relatives as well as foster children. Modifying the family definition in this way reduces the estimated number of persons in poverty. Using both MEPS and NHIS, we found that the broader family concept reduced the estimated number of poor by 2.6 million and the poverty rate by 0.9 percentage points. It also changed to some degree the characteristics of the poor. Agencies that adopt a broader family definition for their surveys and analysts who use such data need to be aware that including unmarried partners and their children in the family reduces the number of poor and changes both their demographic composition and the overall picture of family structure as compared to the official measure of poverty.

In both the CPS and most of the other surveys, poverty is measured by summing the annual incomes of people present in the family at the time of the interview and comparing this total family income to a poverty threshold based on the size of the family and its composition. We

describe this approach as using a fixed family composition. Simulations with SIPP data indicate that this approach yields higher estimates of poverty relative to an alternative approach that defines family composition and family income contemporaneously—that is, based on who lived with the family each month of the year and how much income they received in each month. Compared to fixing family composition in the final month of the income reference year, the contemporaneous approach reduced the estimated poverty rate by nearly half a percentage point. This result is specific to our simulation but illustrative of the general impact of contemporaneous measurement of income and family composition. The PSID makes use of the contemporaneous approach, and SIPP collects the data needed to do so.

Our simulations also addressed the impact of the length of time between the end of the income reference period and the date when family composition is fixed. The longer the lag, the more opportunity for changes in family composition between the survey date and the income reference year. In our simulation, an interview date three months after the end of the income reference year (as the CPS does) added about a third of a percentage point to the poverty rate relative to defining family composition at the end of the reference year (as MEPS does). Lengthening the time interval raised the estimated poverty rate a modest amount, but its bigger impact was on the number of people who were classified differently relative to no lag.

A surprising result emerged from an examination of allocation rates in the ACS by survey month. Intended to show whether data quality deteriorated over the course of the survey year as the income reference period moved farther away from the previous calendar year, these tabulations showed instead that allocation rates (and non-response rates) for the income questions were higher in March, April, May and June than for other months. In other words, respondents were least likely to respond to the income questions in the months that conventional logic suggested were the best months to ask income questions. The association of high non-

response with tax-filing months, and with income levels and income sources usually subject to income taxation, is certainly suggestive but requires further study.

Imputation methods that use respondents as donors will tend to replicate reporting patterns, such as rounding. Allocated income in the CPS, ACS, SIPP, and MEPS shows comparable levels of rounding as reported income. NHIS imputes missing income with a regression model that produces no rounding. PSID includes the imputation of mean values among its allocation methods and shows evidence of very substantial rounding in the allocated values. Clearly, the choice of imputation method has implications for the distribution of imputed values.

The surveys differ in the extent to which they apply consistency checks to related items collected at different points in the survey. Inconsistencies between reported income and reported work activity are notable in MEPS while inconsistencies between the reported receipt of earnings and reported amounts of earnings are observed in NHIS. Inconsistencies such as these present choices to users that will result in different users coming up with different estimates, depending on how they choose to address these inconsistencies. They also provide grounds for critics to question the reliability of any estimates from the survey, even those that may be unaffected by the inconsistencies.

To compensate for the 12 different income reference periods used in an annual ACS, the Census Bureau applies a price adjustment, which converts the reported incomes to constant dollars, using the calendar year in which the survey was conducted as the base. While this achieves a certain uniformity in the income estimates, the approach alters the distribution of income in ways that are inconsistent with actual change over time, as reflected in ACS estimates from consecutive years.

Changes in the way that retirees receive retirement income have been ongoing for decades, yet surveys continue to define and measure retirement income in ways that reflect the earlier

world of defined benefit plans providing regular monthly payments. The CPS income definition used in the study excludes non-periodic or lump sum withdrawals from tax-advantaged retirement accounts, which are likely in the long term to substantially replace monthly pension payments based on defined benefit plans. Two surveys—SIPP and MEPS—request lump-sum payments from a range of sources, but they obtain comparatively little additional income with only marginal impacts on elderly poverty. This suggests that considerable work in this area may be needed to develop significant improvements.

Post-stratification is commonly used to correct survey estimates for differential coverage and response rates by demographic groups. While post-stratification in this manner is widely accepted, one drawback is that if the non-responding units within a demographic group are systematically different from the responding units, post-stratification will not take account of this. Instead, the missing units will be given the same distribution of characteristics as the responding units, in effect. MEPS post-stratifies its person-level sample weights to the distribution of poverty status in the CPS. In forcing the sample to fit the CPS income distribution, this may alter the distribution of other characteristics, which may account for some of the ways in which MEPS departs from other surveys—including substantially more earners and substantially more persons living with spouses or living with no relatives.

D. NEXT STEPS

Our empirical findings using CPS income and family definitions show major differences among the eight surveys, including varying measures of total income, the distribution of income, earnings and earners, number and demographic composition of the poor, poverty rates, program participation, uninsured and low-income uninsured. Additional findings on response rates, allocation and imputation rates and rounding provide information on the quality and reliability of income data. However, standardization cannot adjust for many design features. These include

SIPP's four-month reference period and panel design, ACS's rolling reference period versus NHIS's fixed reference period with a variable recall interval, post-stratification in MEPS, and the contemporaneous poverty measure embedded in PSID. Other survey differences include the identification of relate to unrelated subfamilies, the timing of family composition, and the treatment of students. Simulations were informative about some of these features, but the big differences in design are not amenable to elucidation in this manner.

Lastly, it was not within the scope of this study to make recommendations based on the study findings. However, the study findings provide the groundwork for both a discussion of future directions and work on issues in individual surveys. We hope that we have provided a solid starting place and perhaps the basis for recommendations on survey improvements and future innovations.

APPENDIX A

ANNOTATED BIBLIOGRAPHY

INTRODUCTION

This annotated bibliography was prepared as background to the development of an analysis plan for Assessing the Quality of Income Data across Surveys. In developing the bibliography we surveyed the literature on topics related to methodological issues in measuring income, validation and benchmarking of income data, estimates of accuracy in reported income, and comparisons of income data across household surveys—particularly the eight surveys included in the project. Starting with the references cited in the Department of Health and Human Services working paper (Turek 2005) that provided the conceptual foundation for this project, we extended the list of citations by consulting with the members of the project's Technical Advisory Group and with the Mathematica staff most familiar with the literature on income measurement. This was particularly helpful in expanding our search to encompass the unpublished or "gray" literature. Indeed, many of our final entries are drawn from this literature.

We restricted the scope of our search to the past three decades, but in going back as far as the late 1970s we recognized that changes in survey design, content, and processing may have reduced the relevance of particular types of findings. The earliest reference is from 1977 and the most recent reference is from 2008.

We obtained copies of all potential entries in order to assess their suitability for inclusion and, for those that were selected, to prepare their annotations. Many of the entries identified potential additional references, which we followed up by obtaining copies and reviewing for relevance. Occasionally the same or related findings appeared in more than one venue. To minimize redundancy we sought to include only the most complete or widely accessible version.

The entries that appear in the bibliography were drawn from peer reviewed journal articles, conference proceedings, reports, working papers, and miscellaneous other sources.

The literature that is represented in this bibliography encompasses a range of methodological issues relating to the measurement of income in household surveys. Specific methodological issues include:

- Question wording
- Number of questions
- Question context
- Item and unit nonresponse
- Post survey editing and processing
- Weighting
- Imputation

To assist readers in findings references to these and other topics, an index follows the bibliography.

The purpose of the annotations that accompany the citations is to summarize rather than review. In preparing the annotations we drew from the authors' abstracts and conclusions as a starting point. We supplemented these texts in order to clarify key findings or to expand upon results that were especially germane to this project. The annotations vary in length, which is a function of the relevance of the material that they describe and how easily the main findings could be communicated.

Survey and other acronyms used in the annotations are spelled out the first time that they appear. A list of acronyms used in more than one entry follows this introduction.

Finally, the bibliography includes a number of papers from the Survey of Income and Program Participation (SIPP) working paper series. Most of these papers have no dates, but we understand that the numbering of the papers is sequential, so approximate dates can be inferred from the papers in the series that do have dates.

ACRONYMS

AAPOR – American Association for Public Opinion Research

ACS – American Community Survey

AFDC – Aid to Families with Dependent Children

AGI – Adjusted Gross Income

AHS – Annual Housing Survey

ASEC – Annual Social and Economic Supplement

BEA – Bureau of Economic Analysis

C2SS – Census 2000 Supplementary Survey

CAPI – Computer-assisted personal interviewing

CATI – Computer-assisted telephone interviewing

CE – Consumer Expenditure Survey

CPI – Consumer Price Index

CPS – Current Population Survey

GAO – Government Accountability Office

HRS – Health and Retirement Study

IRA – Individual Retirement Account

IRS – Internal Revenue Service

ISDP – Income Survey Development Program

MCBS – Medicare Current Beneficiary Survey

MEPS – Medical Expenditure Panel Survey

MSA – Metropolitan Statistical Area

NHIS – National Health Interview Survey

NIPA – National Income and Produce Accounts

OASDI – Old-Age, Survivors and Disability Insurance

PSID – Panel Study of Income Dynamics

SIPP – Survey of Income and Program Participation

SMI – Supplementary Medical Insurance (Medicare Part B)

SSA – Social Security Administration

SSI – Supplemental Security Income

SSN – Social Security number

TANF – Temporary Assistance to Needy Families

ANNOTATED BIBLIOGRAPHY

Alternative Measures of Income and Poverty. U.S. Census Bureau, http://www.census.gov/hhes/www/income/incomestate.html#altmeas.

This website contains historical income data tables from the Decennial Census and the March supplement to the Current Population Survey (CPS). The site includes more detailed tables from the renamed CPS Annual Social and Economic Supplement from 1995 forward as well as two reports on the effect of government taxes and transfers on income and poverty. Reports on income inequality and *The Changing Shape of the Nation's Income Distribution, 1947–98* by Arthur F. Jones, Jr. and Daniel H. Weinberg are also accessible form the website.

Atrostic, B.K., and Charlene Kalenkoski. "Item Response Rates, One Indicator of How Well We Measure Income." *Proceedings of the American Statistical Association, Section on Survey Research Methods* [CD-ROM]. Alexandria, VA: American Statistical Association, 2002, pp. 94-99.

The authors of this paper develop a process for defining consistent sets of item nonresponse rates that explicitly account for the survey design. Item response rates are defined in terms of the group eligible for a set of questions and whether group members answered those questions. The paper illustrates the definition with a few examples. The authors find several key points from their calculations of the March 1990 and March 2000 CPS. First, response rates to income items were falling, and the amount of imputed income was increasing. Second, wage and salary income was 102 percent of a benchmark based on the National Income and Product Accounts between 1990 and 1996 (1996 benchmark), interest income was 84 percent of the 1996 benchmark, and dividend income was 60 percent of the 1996 benchmark. Based on their findings, the authors recommended reporting standard income nonresponse rates and continuing research into ways to reduce nonresponse while identifying characteristics of nonrespondents.

Banthin, Jessica S., and Thomas Selden. "Income Measurement in the Medical Expenditure Panel Survey." Agency for Healthcare Research and Quality Working Paper No. 06005, July 2006, http://gold.ahrq.gov.

Using 2002 data, the paper compares the Medical Expenditure Panel Survey (MEPS) and CPS poverty distributions for selected populations of interest. It shows that MEPS income data align relatively closely to CPS estimates. It then compares an experimental poverty status measure based on a single question recently added to the MEPS Round 1 and 3 instrument with the standard poverty status measure based on the more detailed MEPS income questions. An experimental question was added to the MEPS as of 2003, asking respondents to report their total household income within ranges corresponding to five poverty-level status categories. A majority (63.1 percent) of individuals with responses for the single income question provided information that matched the information collected from the detailed questions. However, 26.2 percent underreported family income while 10.7

percent overestimated family income. The paper also finds that, compared to the detailed questions, the single income question overestimates the percent of people in poverty covered by private insurance, underestimates the percent with public coverage, and overestimates the percent uninsured, although these differences are not statistically significant.

Bates, Nancy, and Robert Pedace. "Reported Earnings in the Survey of Income and Program Participation: Building an Instrument to Target Those Likely to Misreport." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2000, pp. 959-964.

The paper analyzes income misreporting propensities by using the 1992 Survey of Income and Program Participation (SIPP) longitudinal file matched to Social Security Summary Earnings Records. Specifically, it focuses on wage and salary and self-employment earnings. The findings suggest that the 1992 SIPP accurately estimated the net number of earnings recipients but underestimated amounts received. The misreporting pattern reveals that respondents at the lowest end of the income distribution tended to overreport earnings while those at the high end were more likely to underreport earnings. The authors fit multinomial models to predict misreporting based on demographic characteristics. Those age 50 and over, males, blacks, Asians, Hispanics, craftspersons, and those with low levels of education were more likely to underreport. Farmers, members of the military, and the self-employed tended to overreport.

Battaglia, Michael P., David C. Hoaglin, David Izrael, Meena Khare, and Ali Mokdad. "Improving Imputation by Using Partial Income Information and Ecological Variables." *Proceedings of the American Statistical Association, Section on Survey Research Methods* [CD-ROM]. Alexandria, VA: American Statistical Association, 2002, pp. 152-157.

This research examines alternative ways of using reported ranges or "partial" income information to impute missing family incomes in the National Immunization Survey, a telephone survey that collects data on children aged 19 to 35 months. Respondents who do not know their total family income for the prior calendar year or refuse to answer the question are asked a cascading sequence of questions designed to assign family income to one of 15 intervals. In the 2000 survey, 27.8 percent of respondents did not answer the initial income question. About half of these completed the follow-up sequence. Of the rest, about 2 in 5 completed part of the sequence, yielding partial information. That is, their incomes could be placed within a broader interval than one of the 15. The authors compared two regression approaches to imputing family income for persons with the most limited partial information or no partial information. The first approach estimated a separate equation for each partial interval, with three additional equations for don't knows, refusals, and those who broke off the interview before the income question. Don't knows and refusals were allotted separate models because refusals reported higher incomes when they responded to the cascading questions. The second approach estimated a single equation over all of these groups. The models were estimated on cases with reported family incomes or, for the don't knows and refusals, cases that completed the cascading questions. Predictor variables included characteristics of the child, mother, family and household as well as ecological characteristics associated with the telephone exchange (such as median

education). In general, the separate regression models provided more accurate imputations than the overall model.

Bavier, Richard. "Reconciliation of Income and Consumption Data in Poverty Measurement." *Journal of Policy Analysis and Management*, vol. 27, no. 1, Winter 2008, pp. 40-62.

Researchers are interested in whether consumption data are superior to income data for poverty measurement. Although the Census Bureau has provided researchers with an experimental series of variables in the CPS that can produce a comprehensive income measure, previous analyses have not fully exploited these variables. The author examines data from the CPS, the Consumer Expenditure Survey (CE), and the SIPP and shows no "huge discrepancy" in federal surveys, as some have suggested, between income and expenditures near the bottom of the distribution. When poverty is measured with a comprehensive income measure that includes the income value of noncash benefits, capital gains and losses, the earned income tax credit, and returns on home equity and subtracts the value of direct taxes, income poverty rates and trends are similar to those of consumption poverty. Arguments that income is measured with more error than consumption at the bottom of the income distribution are shown to derive from inferior income data.

Beebout, Harold. Reporting of Transfer Income on the Survey of Income and Education: Initial Corrections of the Microdata for Underreporting. Mathematica Policy Research, October 14, 1977.

The study attempts to remedy the underreporting of transfer income in the Survey of Income and Education by adjusting for two types of error: (1) fewer individuals reported receipt of an income type than were indicated to have received the income in administrative data; and (2) the number of recipients was acceptable, but they reported too few dollars. In the first case, the study imputed additional recipients by using either a hot deck or simulation technique. In the second case, the study made an upward, proportional adjustment to the class of recipients' benefits to conform to the administrative data. The author attempted to change as little of the original survey data as possible, to edit the data so that the aggregate amount of each income type was approximately equal to the adjusted administrative data, and to preserve major covariances. The procedures were intended to provide a better basis for policy analysis than the unadjusted data but were not intended to satisfy any formal statistical criteria. The results for the total of all work-related transfers indicate that the corrected file has nearly 100 percent of the estimated control income. The total amount of means-tested transfers on the adjusted file, including food stamps, is 99.5 percent of the estimated control total.

Bishaw, Alemayehu, and Sharon Stern. "Evaluation of Poverty Estimates: A Comparison of the American Community Survey and the Current Population Survey." U.S. Census Bureau, June 15, 2006.

At the national level, the CPS Annual Social and Economic Supplement (ASEC) and the American Community Survey (ACS) are relatively consistent in their estimates of poverty. Differences in counting unrelated persons within a household suggest that estimates of poverty may differ, but the data do not show systematic differences between the surveys. For selected characteristics, however, the national estimates of poverty rates differed between the two surveys. The 2003 estimates differed for individuals age 18 to 64 and married-couple families, and the 2002 estimates differed for children under age 18, individuals 65 and older, women, married-couple families, and female-headed households with no husband present. Statistically, the state poverty rates were the same in the ACS and the CPS ASEC for 36 states. The ACS estimates were higher than the CPS ASEC in 12 states and lower in 2 states and the District of Columbia.

Bound, John, and Alan B. Krueger. "The Extent of Measurement Error in Longitudinal Earnings Data: Do Two Wrongs Make a Right?" *Journal of Labor Economics*, vol. 9, no. 1, January 1991, pp. 1-24.

This article reports findings from a study using Social Security earnings data matched to CPS sample records from 1977 and 1978. The analysis is restricted to heads of households who remained at the same address for two years, were successfully matched to their Social Security earnings records, and received earnings from covered employment in both years. The results suggest that the combination of mean reversion and correlated error in reports of wages in consecutive years has a beneficial impact on estimated change in earnings; fully 75 percent of the variation in the change in CPS earnings represents true earnings variation. However, the findings also suggest that the simple models that have been used to characterize measurement error in past studies are not appropriate. The standard assumptions about measurement error as white noise are contradicted by evidence that measurement error is positively autocorrelated and negatively correlated with true earnings.

Bound, John, Charles Brown, Greg J. Duncan, and Willard L. Rodgers. "Evidence on the Validity of Cross-sectional and Longitudinal Labor Market Data." *Journal of Labor Economics*, vol. 12, no. 3, July 1994, pp. 345-368.

This article reports findings from a Panel Study of Income Dynamics (PSID) validation study based on sample members who were employed by a single, large firm. Survey reports from two successive waves of the panel were compared to payroll records. Respondents' reports of annual earnings were fairly accurate, with a very small mean error in the log of earnings but a substantial standard deviation. In addition, errors were negatively correlated with true earnings. This reduces bias when earnings are used as an explanatory variable but adds negative bias when earnings are the dependent variable. Biases were marginally larger for reported changes in earnings. Bias in calculated earnings per hour (annual earnings divided by annual hours worked) were more severe. This was due in part to the error in reported annual hours worked, despite a detailed sequence of questions used to arrive at these hours. However, correlated error was a bigger factor in the magnitude of the bias in hourly earnings.

Bruun, Maria, and Jeffrey Moore. "SIPP 2004 Wave 1 Asset Income Item Nonresponse Results and Nonresponse Follow-up Outcomes." Statistical Research Division, U.S. Census Bureau, October 3, 2005.

The 2004 Wave 1 SIPP questionnaire asked new and expanded follow-up questions in order to combat nonresponse. The questions asked respondents to report their income in a multiple-choice range rather than as a dollar amount. Overall, asset income amount questions had a 40 percent nonresponse rate. Miscellaneous had the lowest (24 percent) nonresponse rate, and stocks and mutual funds had the highest (56 percent). The predominant form of nonresponse to the follow-up question mirrors the form of initial nonresponse. Overall, the nonresponse follow-up questions reduced nonresponse by about half or more. The effectiveness was even greater for those answering "don't know." For these respondents, 75 percent of those that initially said don't know gave a response. Overall, the paper finds that the follow-up questions should greatly improve the quality of the data, suggesting that the benefits of asking the questions far outweighed the extra burden on respondents.

Canberra Group. *Expert Group on Household Income Statistics: Final Report and Recommendations*. Ottawa, 2001, www.lisproject.org.

The report is a guide for data collectors, data analysts, and other users on how to prepare comparable statistics on income distribution. Within the context of prevailing ideas and best practices, the authors set forth guidelines for understanding the complex nature of income data. The guidelines reflect how economies are organized and how people conduct their lives. Where sufficient consensus exists about best practices, the report makes recommendations in the hope that such recommendations will contribute to the availability of more accurate, complete, and internationally comparable income statistics compiled to common standards. The report includes chapters on the income concept, other conceptual issues, practical considerations, comparing income distributions over time, income dynamics, data presentation, robustness assessment reporting, and issues still to be resolved.

Clark Sandra Luckett, John Iceland, Thomas Palumbo, Kirby Posey, and Mai Weismantle. "Comparing Employment, Income, and Poverty: Census 2000 and the Current Population Survey." Housing and Household Economic Statistics Division, U.S. Census Bureau, September 2003, www.census.gov/hhes/www/laborfor/final2_b8_nov6.pdf.

The report examines the differences between the 2000 Decennial Census and the CPS with regard to employment, income, and poverty numbers as a consequence of different data collection methods. Before 1990, unemployment rates were higher in the CPS than in the census. However, in 2000, unemployment reported in the CPS was 2.1 percentage points lower than the census estimate. The difference occurred across all demographic variables. Median family and household income were both $1,000 to 2,000 higher in the census than in the CPS despite the fact that the CPS asked more questions about income from different sources. The one exception was single male households, for which census income estimates were lower than the CPS estimates. The poverty rate was moderately higher in the census

(12.4 percent) than in the CPS (11.9 percent). The paper did not find a comprehensive explanation of these income, employment, and poverty differences.

Coder, John. Comparisons of Alternative Annual Estimates of Wage and Salary Income from SIPP. Memorandum for Gordon Green, Assistant Division Chief, Population Division, U.S. Census Bureau, March 1988.

The memorandum demonstrates that, with a combination of annual recall reports from the annual round-up module in the SIPP and the annual estimates constructed from subannual amounts, the SIPP wage and salary estimates would exceed the analogous CPS estimates by about 6 percent instead of showing a consistent deficit.

Coder, John, and Lydia Scoon-Rogers. Evaluating the Quality of Income Data Collected in the Annual Supplement to the March Current Population Survey and the Survey of Income and Program Participation. Housing and Household Economic Statistics Division, U.S. Census Bureau, July 1996, www.sipp.census.gov/sipp/ workpapr/wp215.pdf.

The paper extensively covers differences between income estimates of the 1990 March CPS and the 1990 SIPP. It also compares the estimates to benchmark estimates. The paper observes that the SIPP seemed to miss more high-income recipients than did the CPS. The authors offer alternative explanations for this difference, but there is no hard evidence supporting any particular cause. The SIPP's wage and salary estimates are about 5 percent lower than those in the CPS. One explanation is that the SIPP is more conducive to reporting "take-home" pay than is the CPS. The SIPP and CPS definitions of self-employment income are markedly different, making comparisons between the two surveys difficult. The paper also compares the two surveys' estimates of income from Social Security, railroad retirement, unemployment compensation, workers' compensation, Supplemental Security Income, public assistance, veterans' payments, pensions, interest and dividends, rents, royalties, estates and trusts, child support, alimony, and financial assistance as well as "other income." Overall, the comparisons show evidence of deterioration in the SIPP estimates between 1984 and 1990, with the SIPP maintaining an advantage for some sources while falling closer to or below the CPS for others. For other sources, the SIPP estimates remained no better than the CPS estimates. Generally, the SIPP provides more complete estimates of recipients, however.

Cohen S.B., and S.R. Machlin. "Characteristics of Nonrespondents in the MEPS Household Component." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1998, pp. 329-334.

This paper attempts to determine the characteristics of nonrespondents in the MEPS. Using the National Health Interview Survey (NHIS) as the sample frame, the 1996 MEPS sample consisted of about 9,000 reporting units. Several groups were likely to be nonresponders based on the following factors: telephone availability (no telephone number given on the NHIS), size of dwelling unit (single- or two-person), family income of primary reporting unit (higher income), item nonresponse for employment classification (no response),

Metropolitan Statistical Area (MSA) size (large cities), and the dwelling unit–level personal help measure of need (less healthy). In addition, the race, gender, and experience (less experience) of the interviewers had a significant impact on nonresponse. The authors conclude that the MEPS data should be weighted to adjust for these differences in nonresponse.

Cohen S., S. Machlin, and J. Branscome. "Patterns of Survey Attrition and Reluctant Response in the 1996 MEPS." *Health Services & Outcomes Research Methodology*, vol. 1, no. 2, June 2000, pp. 131-148.

This paper examines MEPS sample members who participated cooperatively in the survey, did not respond, or were reluctant to respond. The authors find that reluctant responders in the first round of the survey were much more likely to be nonresponders in the second round. Other characteristics of the round-two nonresponders were membership in a large household, residence in a large metropolitan area, and the presence of elderly members in the household. In addition, reluctant responders were a distinct group whose members shared similar age, MSA residence, and employment characteristics with those who dropped out of the survey; nonetheless, they shared marital status and reporting unit size characteristics with cooperating respondents. The authors find that, in the absence of an effort to convert reluctant respondents, the survey's precision would have dropped, though not substantially (standard errors would have increased by about 6 percent).

Czajka, John L., James Mabli, and Scott Cody. "Sample Loss and Survey Bias in Estimates of Social Security Beneficiaries: A Tale of Two Surveys." Final Report. Washington, DC: Mathematica Policy Research, Inc., February 2008.

This report examines two sources of sample loss that affect the utility of SIPP and CPS data for analysis of Social Security beneficiary populations. One source is survey nonresponse, which includes both initial nonresponse and attrition. The other source is the reluctance of respondents to provide their Social Security numbers, which prevents the Census Bureau from matching their survey records to administrative records. The report documents the growth in sample loss due to nonresponse and nonmatching; provides estimates of match bias and attrition bias; examines discontinuities between consecutive SIPP panels in estimates of beneficiary characteristics as well as poverty rates for the broader population; and examines the comparative strengths of the SIPP and CPS in describing the economic well-being of the population in general and elderly and lower-income persons in particular. Analysis of SIPP full panel and cross-sectional sample records matched to Internal Revenue Service earnings records and Social Security benefit records provides evidence that the Census Bureau's full panel weights are highly effective in compensating for bias due to differential attrition. The authors also found little evidence of match bias in SIPP estimates of a wide range of characteristics when the matched sample was calibrated to the same demographic controls used to weight the SIPP sample. A more limited evaluation of match bias in the CPS focused on retired workers and obtained results very similar to the SIPP findings.

The authors present evidence that discontinuities in SIPP poverty estimates across panels are due in part to a recent tendency for SIPP panels to obtain high estimates of poverty in the first wave, which then decline sharply in the second wave. The authors also present evidence that new entrants who are excluded from a panel over time are a distinctive group that is large enough and potentially unique enough to induce marked shifts in poverty when they are represented in full by a new panel.

Across all age groups but particularly children and the elderly the SIPP has continued to identify more sources of family income than the CPS. With respect to income amounts, however, the SIPP has lost ground to the CPS since the initial SIPP panel. From 1993 on, the most significant losses have occurred in the bottom income quintile, where the SIPP has historically performed best relative to the CPS. In 1993 the SIPP captured 20 percent more aggregate income from this quintile than did the CPS. By 2002, however, the SIPP's advantage had fallen to just 6 percent. These losses were distributed across most income sources. Only for SSI, welfare and pensions did the SIPP maintain or improve its advantage. A comparison of poverty trends in the two surveys raises a number of concerns about the use of either survey for the measurement of trends in economic well-being. These concerns are greatest for estimates of poverty among the elderly.

Davern, Michael, Lynn A. Blewett, Boris Bershadsky, and Noreen Arnold. "Missing the Mark? Examining Imputation Bias in the Current Population Survey's State Income and Health Insurance Coverage Estimates." *Journal of Official Statistics*, vol. 20, no. 3, 2004, pp. 519-549.

This article examines earned income in the 1990 Decennial Census, the Census 2000 Supplemental Survey (C2SS), and the 1998–2000 CPS data to determine the bias at the state level created by the hot deck imputations used in the surveys. It also examines CPS state health insurance coverage rates. For income data, the Census imputes income if any of the income-related questions are missing, whereas the CPS imputes only for the missing question. Through the fitting of a bias model, the results showed little bias at the state level in estimates of income for the 1990 Decennial Census or the C2SS. The CPS income data, however, showed a bias. The CPS health insurance coverage estimates were even more biased because the hot deck procedure did not use geographic region. To correct for this significant bias, the article considers possible approaches to model bias, to change the hot deck procedure in order to capture more between-state variation by adding a geographic proximity preference, or to use a multiple imputation procedure.

Davern, Michael., Holly. Rodin, Timothy J. Beebe, and Kathleen Thiede Call. "The Effect of Income Question Design in Health Surveys on Family Income, Poverty and Eligibility Estimates." *Health Services Research*, vol. 40, no. 5, October 2005, part I, pp. 1534-1552.

The article uses March CPS supplement data and compares omnibus family income estimates (obtained by one overarching income question) to aggregate family income estimates (obtained by asking several income questions about various sources of income). The authors find substantially different income estimates depending on the method used. Only 31 percent of people remained in the same income bracket for both methods. Factors

associated with underreporting were households with three or more family members or those with other sources of income or assistance. One table in the article shows that the omnibus question inflates the amount of poverty by an average of about 1 percentage point. The article concludes that the omnibus household income question is likely biased and that the bias should be recognized when using such question for analysis.

Denmead, Gabrielle, and Joan Turek. "Comparisons of Health Indicators by Income in Three Major Surveys." *Proceedings of the Annual Meeting of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association, 2005, pp. 1532-1538.

The authors compare relationships between income and comparable measures of health status, insurance coverage, and utilization in three surveys: the NHIS, CPS and SIPP. The comparisons use identically defined family income for calendar year 2001. Study findings include differences among the surveys in counts and composition of the low-income population, health status, uninsured, uninsured children, Medicaid coverage, and utilization of inpatient and ambulatory care. The surveys provide different pictures of the needs and target groups for public programs. The NHIS has more poor and low-income than CPS and SIPP despite its broader family definition. The NHIS finds more insurance coverage but less Medicaid coverage on a monthly basis, total and for children, than does SIPP. The CPS finds both less insurance coverage and less Medicaid coverage on an annual basis, total and for children, than does SIPP.

Denmead, Gabrielle, Joan Turek, and Michele Adler. "Annual Income and Working-Age Disability: Estimates from the NHIS and CPS." *Proceedings of the American Statistical Association, Section on Health Policy Statistics* [CD-ROM]. Alexandria, VA: American Statistical Association, 2003, pp. 1203-1208.

This paper develops annual income measure s that can be used in conjunction with disability data and program participation to address health and disability policy issues. The analysis uses data from the NHIS and the CPS from the mid-1990s, when the NHIS collected person-level information on monthly income by source. The authors annualized income reported in the NHIS and conducted validity tests of alternate income measures within a single data base, SIPP. They found that monthly income came closest to total income and poverty rates under Actual annual income, but also had the highest rate of false negatives in determining poverty status. Another difference between the NHIS and the CPS is that the NHIS treats unmarried partners as married, affecting the poverty rate. Overall, the data from the NHIS matched the CPS fairly well. The article goes on to analyze the information on income, disability, and participation.

Doyle, Pat. "The Survey of Income and Program Participation: AAPOR Roundtable: Improving Income Measurement." SIPP Working Paper 241, U.S. Census Bureau, no date.

This summary describes an American Association for Public Opinion Research (AAPOR) Annual Conference roundtable discussion of the findings from the first two field experiments of the SIPP Methods Panel project. Each experiment included a treatment

group, which received the experimental instrument, and a control group, which received the SIPP Wave 1 instrument for the panel in the field at the time. In addition to other changes the second experiment introduced a different approach to collecting earnings. Respondents were allowed more flexibility in choosing the best time period for reporting amounts received (that is hourly, weekly, biweekly, monthly, quarterly, or annually). For unearned income, the experiment introduced screening procedures for effectively targeting need-tested program questions to households potentially eligible for such programs. With regard to assets, the experiment took a three-part approach: (1) determining ownership of Individual Retirement Accounts (IRAs), (2) determining ownership of a set of commonly held asset types, and (3) capturing ownership of the remaining asset types. Overall, the treatment group experienced significantly lower item nonresponse on income amounts than did the control group, especially for asset amounts. For earnings, the treatment group achieved a reduction in item nonresponse of over 40 percent. A comparison of mean income amounts and the proportion of the population with income in the treatment and control groups showed no significant differences.

Doyle, Pat, Betsy Martin, and Jeff Moore. "The Survey of Income and Program Participation (SIPP) Methods Panel: Improving Income Measurement." SIPP Working Paper 234, U.S. Census Bureau, November 13, 2000, http://www.sipp.census.gov/sipp/workpapr /wp234.pdf (an abbreviated version appears in the Proceedings of the American Statistical Association, 2000).

This paper describes experimental research in trying to increase response and accuracy in the 2000 SIPP survey. To test different question designs, the authors randomly assign 1,000 people to the standard SIPP instrument and 1,000 people to the modified instrument. The authors reach several conclusions. Use of nonresponse follow-up improves reporting of income amounts. The high nonresponse to asset income questions is primarily a function of lack of knowledge, suggesting that follow-up questions that request more limited information (such as bracketed values) can improve response rates. For some respondents, a common set of asset types can be used instead of asking about each asset type individually. An income screener can reduce the number of respondents asked about needs-based programs. The seam bias problem remains unresolved, however.

Duncan, Greg J. and Daniel H. Hill. "Assessing the Quality of Household Panel Data: The Case of the Panel Study of Income Dynamics." *Journal of Business and Economic Statistics,* vol. 7, no. 4, October 1989, pp. 441-451.

Evidence from a number of methodological studies is used to assess the overall quality of data from the PSID. Despite substantial cumulative attrition, comparisons with the CPS indicate that after 12 years the PSID sample continued to provide good representation of the nonimmigrant population. The PSID had proportionately fewer low-income families in both 1968 and 1980, but this may reflect the PSID's more complete capture of income. In addition, PSID reports of transfer income appear to compare more favorably with program aggregates than reports from the CPS. The results of a validation study conducted with a subsample of respondents indicate that reports of wages and employment are generally unbiased but contain measurement error that varies from trivial to very large.

Fisher, Patricia J. "Assessing the Effect of Allocated Data on the Estimated Value of Total Household Income in the Survey of Income and Program Participation (SIPP)." SIPP Working Paper 244, U.S. Census Bureau, no date.

This paper examines the individual components of total household income as collected in the SIPP and evaluates the proportion imputed (or allocated) for each component. The author concludes that 28.8 percent of total household monthly income is allocated. Much of the allocation is carried over from previous waves of data collection rather than allocated with hot deck or cold deck imputation or logical imputation.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. "An Analysis of the Impact of Sample Attrition on the Second Generation of Respondents in the Michigan Panel Study of Income Dynamics." *The Journal of Human Resources*, vol. 33, no. 2, Spring 1998a, pp. 300-344.

The authors study the impact of sample attrition on the second generation of respondents in the PSID. They conclude that the intergenerational relationship among earnings, education, and welfare participation of parents and their adult children is stronger for the subsample of children who do not attrite by the end of the panel than for the full sample that includes all children who did not attrite before their mid-20s (but may have attrited afterwards). The differences in intergenerational coefficients are small and seldom statistically different from zero for welfare and earnings. However, the authors do find evidence of attrition bias in estimates for education. They assert that attrition may be random with respect to some outcomes but not others.

Fitzgerald, John, Peter Gottschalk, and Robert Moffitt. "An Analysis of Sample Attrition in Panel Data: The Michigan Panel Study of Income Dynamics." *The Journal of Human Resources*, vol. 33, no. 2, Spring 1998b, pp. 251-299.

The authors study the effect of approximately 50 percent sample loss from cumulative attrition on the unconditional distributions of several socioeconomic variables and on the estimates of several sets of regression coefficients. Their empirical analysis shows that attrition is highly selective and concentrated among individuals with lower socioeconomic status. They also show that attrition is concentrated among those with more unstable and lower earnings. However, cross-sectional comparisons of unconditional moments between the PSID and the CPS show a close correspondence all the way through 1989. The authors conclude that the selection that occurs is moderated by regression-to-the-mean effects from transitory components that fade over time. Therefore, despite a high level of attrition, they find no strong evidence of loss of representativeness.

Garner, T., and L. Blanciforti. "Household Income Reporting: An Analysis of U.S. Consumer Expenditure Survey Data." *Journal of Official Statistics*, vol.10, no. 1, 1994, pp. 69-91.

This paper uses data from the 1987 CE to model income response with socioeconomic factors. The binomial logit model showed significant increases in response associated with age (very young or very old), race (non-black), education (non–college graduate), employment (not self-employed), consumer unit composition (single), and region (West or

South). The expenditure variable was particularly interesting and showed that those reporting higher expenditures were significantly more likely to give complete income information.

Gouskova, Elena and Robert F. Schoeni. "Comparing Estimates of Family Income in the Panel Study of Income Dynamics and the March Current Population Survey, 1968–2005." http://psidonline.isr.umich.edu/Publications/Papers/Report_on_income_quality_v3.pdf, July 2007.

The PSID has experienced substantial cumulative non-response over its 39-year history. Moreover, the PSID has undergone several methodological changes: 1) conversion to computer assisted telephone interviewing (CATI) from paper and pencil telephone interviewing in 1993, 2) suspension of roughly one-half of the low-income sample in 1997, 3) addition in 1997 of a sample of families who immigrated to the US since 1968, 4) switch to biannual interviewing in 1999, and 5) a doubling of the length of the interview between 1995 and 1999. The objective of this study is to reassess the quality of the PSID family income data by comparing estimates of family income between the PSID and the CPS for the survey years 1968 through 2005. Over this period the family income distributions from the two surveys match fairly closely between the 5th and 95th percentiles. Overall, the PSID estimates have been somewhat higher than the CPS estimates, but the trends are quite similar. The two data sets show less agreement at the upper and lower five percentiles of the distribution.

Government Accountability Office. American Community Survey: Key Unresolved Issues. October 2004, GAO-05-82.

In this report, the Government Accountability Office (GAO) considers whether the ACS can provide an adequate replacement for the census long form as the major source of data for small geographic areas. GAO reviews both operational and programmatic aspects of the ACS and identifies a number of issues that the Census Bureau will have to address. One outstanding issue relates to the measurement of income. GAO reports that when the Census Bureau releases ACS data for each new year, it will present only annual estimates adjusted for inflation and will revise all dollar-denominated data for earlier years. Dollar-denominated items include income, housing value, rent, and housing-related expenditures. The Census Bureau also has decided to continue to adjust data collected each month in the ACS to a calendar year basis. It will use the Consumer Price Index (CPI), a national measure of inflation, for the annual and monthly adjustments for all geographic areas. GAO raises serious questions about inflation adjustments. Moreover, GAO finds that the use of a national cost-of-living adjustment does not reflect variations in geographic areas and may not be appropriate when allocating federal funds to states.

Grieger, Lloyd D., Sheldon Danziger, and Robert F. Schoeni. "Estimating and Benchmarking the Trend in Poverty from the Panel Study of Income Dynamics." http://psidonline.isr. umich.edu/Publications/Papers/grieger-danz-schoeni.pdf, November 2007.

This paper guides researchers through the process of calculating the poverty rate from the PSID for each year from 1968 to the present and compares the level and trend in PSID poverty rates to those of the March CPS. The authors explain how to calculate four alternative PSID poverty series, which differ with respect to their income thresholds. Prior to 1973, the trends in the first two PSID poverty rates differ significantly from the CPS series, with the PSID showing greater declines in poverty. The third series, available from 1990 forward, is highly correlated with the CPS series from 1989 through 2002, and the fourth series is highly correlated with the CPS series over the entire period, 1967 to 2002.

Heeringa, S.G., D.H. Hill, and D.A. Howell. "Unfolding Brackets for Reducing Item Non-Response in Economic Surveys." HRS Working Paper 94-029, Institute for Social Research, University of Michigan, June 1995.

This paper describes and analyzes a new survey methodology for reducing item nonresponse on financial measures. A respondent who is unable to provide an exact dollar amount may be able to provide a range, but respondents vary in how precisely they can bound the true value. Giving a respondent a set of fixed brackets is not the most effective way to determine how much the respondent knows. Systematic "unfolding brackets" provide an alternative approach, whereby the respondent is given a series of choices (for example, "Is it more/less than X dollars?") to determine the lower and upper bounds that the respondent is able to provide. Unfolding brackets are applicable in both face-to-face and telephone surveys. The proportion of missing observations for financial variables in national surveys is often in the range of 20 to 25 percent and, in some cases, as high as a one-third. With the unfolding bracket method, the proportion of completely missing data can be cut by two-thirds. Furthermore, with appropriately chosen bracket breakpoints, it is possible to recover a high proportion of the variance of the underlying measure. The authors investigate the effects of bracketing on the empirical validity of survey data. While they find lower empirical validity for data from individuals exposed to brackets early in the survey instrument, this finding appears to result from self-selection rather than from a direct effect of exposure to the methodology.

Hendrick, Mark R., Karen E. King, and Julia L.Bienias. "Research on Characteristics of Survey of Income and Program Participation Nonrespondents Using IRS Data." SIPP Working Paper 211, U.S. Census Bureau, no date.

The paper relies on matching individual 1990 Internal Revenue Service (IRS) data to SIPP data to track the accuracy of SIPP earnings estimates. Differences between the IRS and SIPP definitions of total income necessitated adjustments while cases with IRS income of zero were discarded. The authors use regression models to fit the IRS income and to determine if the relationship between SIPP and IRS earnings differs for respondents and nonrespondents. Married respondents had higher earnings than married nonrespondents while single respondents had lower earnings than single nonrespondents. The authors also find that the relationship between IRS and SIPP earnings data varies by race. Overall, the analysis shows that the SIPP overestimates earnings at low earnings levels and underestimates earnings at high earnings levels. The research also appears to verify a general underreporting of earnings in the SIPP.

Henry, Eric, and Charles Day. "A Comparison of Income Concepts: IRS Statistics of Income, Census Current Population Survey, and BLS Consumer Expenditure Survey." *Proceedings of the Annual Meeting of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association, 2005, pp. 1155-1162.

This paper describes the Adjusted Gross Income (AGI) concept used by the IRS and then explains the most important differences between AGI and the definitions used in the CE and CPS. AGI excludes nontaxable income, which leaves out some sources entirely while discounting other sources. Differences occur in wages and salaries, self-employment income, Social Security, private and government retirement income, interest, dividends, rental and other property income, unemployment and workers' compensation, veterans' benefits, public assistance, Supplemental Security Income, food stamps, regular contributions for support, and other income.

Hess, Jennifer, Jeffrey Moore, Joanne Pascale, Jennifer Rothbag, and Catherine Keeley. "The Effects of Person-level versus Household-level Questionnaire Design on Survey Estimates and Data Quality." *Proceedings of American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2000, pp. 157-162.

This study attempts to identify the best survey method for gaining information about people in a household. The traditional method is a person-level approach whereby the interviewer asks the same questions for every person in the household. A different technique is the household-level approach whereby the interviewer asks questions such as "Does anyone in the household have trouble seeing?" The study was based on two surveys of 908 households. The authors found some limited evidence that the household-level approach increases the risk of underreporting for some summary measures such as asset ownership. However, the reduced risk of underreporting in the person-level survey suggests that the improvement may come at the expense of response reliability. Item nonresponse and behavior coding results did not suggest that either the household- or person-level version was superior. Survey interviewers greatly preferred the household-level survey and thought that it was less burdensome than the traditional person-level survey. The authors suggest that validating data could greatly help determine which survey type is superior. They also suggest that, for some types of information, one might be better than the other and vice-versa.

Hurd, Michael D. "Anchoring and Acquiescence Bias in Measuring Assets in Household Surveys." *Journal of Risk and Uncertainty*, vol. 19, 1999, pp. 111-136.

Cognitive psychology has identified and extensively studied several cognitive anomalies that may be important for assessing the economic status of individuals and households. In particular, the use of unfolding brackets to elicit information about income and assets in household surveys can interact with such cognitive anomalies—acquiescence bias and anchoring—to cause bias in the estimates of the distribution of income and assets. This paper uses data from the Health and Retirement Study (HRS) and the Asset and Health Dynamics Study to study the use of brackets to elicit information about income and assets.

The author finds that bracketing can produce bias in population estimates of assets based on matching respondents across two successive March panels for 1992-93 and 1996-97.

Hurd, Michael, F. Thomas Juster, and James P. Smith. "Enhancing the Quality of Data on Income: Recent Innovations from the HRS." *The Journal of Human Resources*, vol. 38, no. 3, Summer 2003, pp. 758-772.

The authors evaluated two survey innovations introduced in the HRS that aimed to improve income measurement. The innovations are (1) the integration of questions for income and wealth and (2) matching the periodicity over which income questions are asked with the typical way such income is received. Both innovations had significant impacts in improving the quality of income reports. For example, the integration of income questions into the asset module produced in HRS an across-wave 63 percent increase in the amount of income derived from financial assets, real estate investments, and farm and business equity. Similarly, asking respondents to answer in terms of a time interval consistent with how they receive income substantially improved the quality of reports on Social Security income based on matching respondents across two successive CPS March panels for 1992-93 and 1996-97.

Huyhn, Minh, Kalman Rupp, and James Sears, Office of Research, Evaluation and Statistics, Social Security Administration. "The Assessment of the Survey of Income and Program (SIPP) Benefit Data Using Longitudinal Administrative Records." SIPP Working Paper 238, U.S. Census Bureau, no date. http://www.sipp.census.gov/ sipp/workpapr/ wp238.pdf

This paper uses administrative records data from the Social Security Administration (SSA) to assess the accuracy of SIPP data concerning Old-Age, Survivors and Disability Insurance (OASDI) and Supplemental Security Income (SSI) benefits. OASDI estimates from the SIPP are consistently and substantially lower than the Monthly Benefit Credited estimates of gross OASDI benefits (6 to 8 percent difference). Using aggregate SSA-SIPP comparisons, both the March 1996 and October 1998 SIPP underestimate aggregate SSI receipt (by 4.5 and 1.8 percent, respectively). The authors also look at the individual-level variation beyond these overall measures of SIPP receipt error. Overall, the accuracy of reporting receipt of "OASDI only" or "neither" is very high. The percent misreporting in the two categories involving SSI receipt is much higher. The SIPP misclassifies a nontrivial fraction of those receiving SSI ("SSI only" and "concurrent" SSI and OASDI) according to SSA records as receiving "OASDI only." Finally, a substantial portion of "SSI only" recipients reports no benefit at all. The authors also examine benefit amounts conditional on receipt. In January 1993, a large plurality (42.5 percent of observations) had OASDI benefit amounts that exceeded the Monthly Benefits Paid by $31 to $40. For each of the other three time points (August 1995, March 1996, and October 1998), less than 2 percent of individuals fell into this category. The difference is likely attributable to a questionnaire change asking respondents to report the total amount each month after any deductions. The authors also find that reporting errors for both SSI and OASDI differ dramatically by imputation status, and they provide evidence that errors may be systematically related to sample attrition and interview status (self, proxy, and refusal). They also provide a brief assessment of the effect

of the lack of Social Security numbers in a nontrivial fraction of cases and find clear evidence of selectivity.

Juster, F. Thomas, and James P. Smith. " Improving the Quality of Economic Data: Lessons from the HRS and AHEAD." *Journal of the American Statistical Association*, vol. 92, no. 440, 1997, pp. 1268-1278.

Juster and Smith provide an overview of "follow-up brackets" as applied to collecting respondent-reported data on assets for (1) the HRS of people age 51 to 61 in order to measure economic transitions in health, work, income, and wealth and (2) the Asset and Health Dynamics Among the Oldest Old Survey of people over age 70 in order to study the relationship between physical and cognitive health in old age, living arrangements, and "asset decumulation." The authors find that when bracketed amounts are given as follow-up to responses of "don't know" or "refuse," the bracketed data are useful for later imputation of the actual amount requested. They also find that respondents who used the bracket amount path early in the survey were more likely to provide estimated dollar amounts (non-bracket) later in the survey. Use of follow-up brackets reduces item nonresponse and provides for more appropriate imputation estimates.

Kalton, Graham, and Michael E. Miller. "The Seam Effect with Social Security Income in the Survey of Income and Program Participation." *Journal of Official Statistics*, vol. 7, 1991, pp. 235–245.

A common finding in SIPP data is that more month-to-month changes in recipiency occur when data are collected in different waves versus the same wave. This phenomenon is called the seam effect. To examine the seam effect further, this paper looks at the January 1984 3.5 percent increase in Social Security payments. One-third of the Social Security recipients in the SIPP did not report an increase in Social Security payments for the period. Using a logistic regression, the authors compare the characteristics of those reporting an increase and those failing to do so. Those most likely to report the change were in rotation group 1, white, self-reporting, and with a January Social Security payment over $413. They had a predicted reporting rate of 75 percent while those with the opposite characteristics had a predicted reporting rate of 26 percent. One explanation for the seam effect is that it a manifestation of the general problem of measuring gross changes in panel surveys. Another explanation is false consistency; that is, people forget that a change has occurred and repeat the same answer as in the past.

Kapteyn A., P. Michaud, J.P. Smith, and A. Van Soest. "Effects of Attrition and Non-Response in the Health and Retirement Study." RAND Working Paper. May 1, 2006.

This study attempts to determine how nonresponse and attrition affect the representativeness over time of members of the HRS sample born between 1931 and 1941. The authors find that most baseline characteristics are not correlated with nonresponse except for race, ethnicity, gender, and age–factors that HRS already weights. The authors advise against using complicated weighting schemes other than the HRS-provided weights. The paper also

finds that those who leave the survey but return later are significantly different from those who leave permanently and those who always complete the survey. Thus, the authors recommend use of the unbalanced sample (which includes those who dropped out and then returned) because returning respondents differ significantly from either of the other two groups; returnees' omission from the sample could compromise representativity. The paper also studies whether there was a difference in those who did not provide their pension summary plan description (SPD) or SSA records. The authors find that many characteristics of respondents are associated with both an SSA and SPD match and that the sample of those providing SSA or SPD information is nonrandom. Use of the weights helps account for nonresponse and attrition, but some differences remain to be addressed.

Kashihara D., and T. Ezzati-Rice. "Characteristics of Survey Attrition in the Household Component of the Medical Expenditure Panel Survey (MEPS)." *Proceedings of the American Statistical Association, Section on Survey Research Methods* [CD-ROM]. Alexandria, VA: American Statistical Association, 2004, pp. 3758-3765.

This study attempts to determine the factors that make a person likely to drop out of the MEPS panel survey. The first analysis looked at Year 1 and those who completed round 1 but then dropped out. The total attrition rate in this case was about 10 percent. The significant variables (5 percent significance rate) were age, race, education, employment status, region, MSA, health insurance status, number of people in the reporting unit, and whether participants were reluctant respondents. The second analysis looked at Year 2 and those who completed rounds two and three but then dropped out. The significant variables were age, marital status, education, region, self-perceived health status, health care expenditures, office-based doctor visits, first respondent, proxy respondent, number of people in the reporting unit, and whether participants were reluctant respondents. Health care expenditures and doctor visits were new variables in the Year 2 analysis.

Kim, Yong-Seong and Frank P. Stafford. "The Quality of PSID Income Data in the 1990's and Beyond." http://psidonline.isr.umich.edu/Guide/Quality/q_inc_data.html, December 2000.

This paper reviews changes to the PSID implemented in the 1990s along with prospective future changes and assesses their actual and potential future impact on the quality of PSID data. Operational changes included conversion to computer assisted telephone interviewing and the introduction of new processing and editing systems. Sample changes included suspension of more than half of the original low-income sample and the introduction of a new sample of immigrants. Based on comparisons between the PSID and CPS the authors conclude that, despite these changes, a number of potential data seams were avoided, and the basic continuity of the income data series has been preserved.

Koenig, Melissa L. "An Assessment of the Current Population Survey and the Survey of Income and Program Participation Data Using Social Security Administrative Data." Federal Committee on Statistical Methodology, 2003 Research Conference papers, pp. 129-137.

This analysis compares survey-reported Social Security and SSI beneficiary information from the CPS and SIPP to the information contained in program administrative records for persons age 65 or older with a Social Security number (SSN) match. Both surveys estimate aggregate Social Security benefits very well for the matched samples. (CPS reported benefits are compared to the gross Social Security benefit while SIPP reported benefits are compared to the net Social Security benefit—that is, excluding the Medicare Part B premium.) The CPS underestimates SSI benefits by 21 percent compared to 8 percent for the SIPP. The SIPP correctly identifies 99 percent of Social Security beneficiaries and 93 percent of SSI beneficiaries. The CPS correctly identifies 95 percent of Social Security beneficiaries but only 69 percent of SSI beneficiaries. However, both surveys incorrectly identify about 40 percent of elderly nonbeneficiaries as Social Security beneficiaries whereas they misclassify less than one percent of SSI nonbeneficiaries. Imputation affects the level of correspondence between the survey and administrative data. For respondents with no Social Security or SSI imputations, substituting the actual benefit amounts for the reported amounts changes poverty status for only 4 percent of persons in the CPS and 2 percent in the SIPP. For those with imputations, poverty status is changed for 10 percent of persons in the CPS and 4 percent in the SIPP.

Kominski, Robert. "Record Use by Respondents." SIPP Working Paper 152, U.S. Census Bureau, 1991. http://www.sipp.census.gov.sipp.workpapr/wp152.pdf

The study seeks to ascertain the basic level of record use by respondents when reporting income. It relies on Senior Field Representatives (SFRs) who performed routine observations of Wave 1 interviews in the 1990 panel of SIPP. The SFRs used an observation form and noted whether respondents used records in reporting certain income sources: wages and salary, assets, and certain public programs. Of persons reporting a wage or salary, 31 percent used some type of record. A similar level of use—28 percent—was reported for assets. About a third of the sample reported receipt of Social Security, but 43 percent of these respondents did not in any way verify such receipt. Of those providing verification, one in three verified that source, but not the amount. Only about a third of all Social Security recipients (35 percent) verified both the source and amount with some type of record. Of respondents reporting Medicare, 78 percent were able to verify enrollment with a record. Two-thirds of those verifying Medicare did so for the source only. With the remaining programs infrequently reported, the authors combined them into one measure. Of those persons reporting in one of these programs, 21 percent verified participation. The analyses also show that the source of the lack of record use is attributable to the interviewer and to respondent characteristics. The fundamental finding is that record use is noticeably low across all elements.

Kominski, Robert. "The SIPP Event History Calendar: Aiding Respondents in the Dating of Longitudinal Events." *Proceedings of American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1990, pp. 553-558.

This paper presents the results of a test of an event history calendar in the SIPP. Designed to reduce seam bias, the calendar was used to collect selected data on employment health insurance coverage, program participation, and pension receipt. The calendar was tested in

one region, comprising the states of Illinois and Indiana, for the duration of the 1989 panel, which was terminated after just three of the planned nine waves. The calendar, displaying all 32 months that were to be covered by the panel, was completed by the interviewer after each interview and presented to the respondent to use as a reference tool during the next interview. Used in this way the calendar served as a form of dependent interviewing by allowing respondents to see their households' responses from prior waves. Some reduction in seam bias was observed for several of the items collected with the aid of the calendar. The calendar also facilitated longitudinal editing and correction of the data. There was no evidence to suggest that the calendar was rejected by either respondents or the field staff.

Lamas, Enrique, Thomas Palumbo, and Judith Eargle. "The Effect of the SIPP Redesign on Employment and Earnings Data." SIPP Working Paper 217, U.S. Census Bureau, no date.

This paper focuses on the difference between the 1993 and 1996 SIPP. The major change was a switch from paper-and-pencil personal interviewing to computer-assisted personal interviewing (CAPI). In addition, the questions about income and employment were grouped together differently. The results show the same percent of persons working all weeks of a month, but a lower percent with no job who are either looking for employment or on layoff. Moreover, CAPI shows higher mean and aggregate earnings, perhaps indicating a reduction in the level of underreporting in the SIPP.

Lamas, Enrique, Jan Tin, and Judith Eargle. "The Effects of Attrition on Income and Poverty Estimates from the Survey of Income and Program Participation (SIPP)." U.S. Census Bureau. Paper presented at the Conference on Attrition in Longitudinal Surveys, May 4, 1994.

Using several models of income and poverty that take attrition into account, the authors examine the effect of attrition from the SIPP on income and poverty correlates. They also use simulations to examine the magnitude of potential attrition bias on poverty estimates. They impute missing information for attritors and calculate poverty estimates for the complete panel. To obtain an estimate of potential attrition bias, they use simulations for attritors to compare poverty estimates for the full panel to those of panel members with complete information. The authors conclude that, although attrition had an effect on income and poverty estimates in the SIPP, the observed differences in the poverty estimates from the SIPP and CPS do not appear to result from either attrition or the other methodological differences between the two surveys. The differences may result from better reporting in the SIPP of income at the lower end of the distribution, especially reporting of means-tested income and other short-term spells of income, but further work in this area is needed.

Liu, Hongji, and Ravi Sharma. Report on Round 30 Income and Assets Imputation for MCBS Community Residents. Memorandum from Westat to Frank Eppig, Centers for Medicare and Medicaid Services, June 12, 2002.

This memorandum reviews the procedures implemented to impute for income and assets in the Medicare Current Beneficiary Survey (MCBS) Round 30 Income and Assets

Supplement. The authors imputed the income and assets dollar amounts by using a hot deck imputation procedure and a predictive mean-matching procedure. The share of responses missing total annual income for 2000 totaled 25.56 percent. To assess the degree to which the imputation preserved the observed relationship among income, assets, and homeownership amounts in Round 30 and the previous round, the authors compute Pearson correlations. The correlation coefficients for 2000 and 1999 income amounts are very similar for observed and completed Round 30 data.

Loomis, Laura, and Jennifer Rothgeb. Final Report on Cognitive Interview Research Results and Revisions to the Welfare Reform Benefits Questions for the March 2000 Income Supplement to the CPS. Survey Methodology #2005-02. Statistical Research Division, U.S. Census Bureau, March 14, 2005.

This report describes the results of cognitive interview research on questions about welfare benefits that were included in both the 1998 and 1999 March Income Supplement of the CPS. The questions represent the CPS's first attempt to measure participation in welfare after a new law passed in 1996 instituted the Temporary Assistance to Needy Families (TANF) program. The report makes recommendations on welfare-reform related questions dealing with receipt of cash assistance, cash diversion assistance, transportation and child care assistance, and participation in work-related training activities. The authors include the final decisions made by the Housing and Household Economic Statistics Division.

Lynn, Peter, Annette Jackle, Stephen P. Jenkins, and Emanuela Sala. "The Effects of Dependent Interviewing on Response to Questions on Income Sources." *Journal of Official Statistics*, vol. 22, no. 3, 2006, pp. 357-384.

The term "dependent interviewing" generally refers to structured interviews whereby the choice and/or wording of questions varies across sample members, depending on information maintained by the survey organization about the sample member. Typically, the information comes from a previous survey, although it may come from administrative data or the sample frame. Using an experimental design, the authors compare two approaches to dependent interviewing to traditional independent interviewing for a module of questions about sources of income. The authors compare the three approaches to questioning in terms of the effect on underreporting of income sources and related bivariate statistics. The study design also permits identification of the characteristics of respondents whose responses are sensitive to interview mode. The authors conclude that underreporting can be significantly greater with independent interviewing than with either form of dependent interviewing, especially for income sources that are relatively common or relatively easy to forget. They also find that dependent interviewing is helpful as a recall aid for respondents below retirement age and for registered disabled persons.

Mack, Stephen, and Rita Petroni. "Overview of SIPP Nonresponse Research." Presented at the Fifth International Workshop on Household Survey Non-Response, Ottawa, Canada, September 26–28, 1994.

In providing an overview of various weighting techniques for the SIPP, the authors find that alternative longitudinal weighting intended to deal with levels of nonresponse and bias provides no strong evidence of reduction of these two problems. The authors use constrained response propensity adjustments for panel nonresponse in an effort to reduce the bias of subsequent waves' nonresponse. The results, however, do not demonstrate any reduction of nonresponse bias from this approach. Finally, the authors build on research suggesting that the use of a Chi-Squared Automatic Interaction Detector algorithm in conjunction with several alternative panel nonresponse adjustments (ranking adjustment, logistic regression, logistic regression/observed, and collapsed cells) offers a possible means of reducing bias in the estimates. Results show, however, that none of seven nonresponse adjustments were better than the others at reducing panel nonresponse bias. Thus, the paper suggests that, while none of the above methods is effective in reducing nonresponse bias between rounds of data collection, the SIPP staff will continue experimenting with different weights in an effort to obtain the highest-quality data.

Marquis, Kent H., and Jeffrey C. Moore. "SIPP Record Check Results: Implications for Measurement Principles and Practice." SIPP Working Paper 126, U.S. Census Bureau, no date. http://sipp.census.gov/sipp/workpapr/wp126.pdf

The SIPP Record Check uses a "full" as opposed to a one-directional design; that is, the evaluation checks both "yes" and "no" reports of program participation and obtains program participation records for eight government transfer programs administered by four states (Florida, New York, Pennsylvania, and Wisconsin) and the federal government. From each agency, the authors obtained identifying information to match records and monthly benefit amounts in order to measure response error. They find that misclassification error percentages for monthly reports of program participation and program participation changes are very low for each program. The net bias in estimates of the mean level of program participation ranges from -3 to -39 percent, indicating that the estimated mean is usually lower than the true mean. They discuss measures that could improve measurement error in the SIPP, such as statistical error correction and control and design changes.

Marquis, Kent H., and S. James Press. Cognitive Design and Bayesian Modeling of a Census Survey of Income Recall, in *Federal Committee on Statistical Methodology, 1999 Research Conference,* pp. 51-64. http://www.fcsm.gov/papers/index.html.

This paper investigates ways of combining Bayesian estimation and cognitive psychology to make estimates of data containing response errors. If respondents can judge the quality of their answers, then the authors' approach may work well. However, the paper shows that asking respondents for a range associated with their income proved burdensome for both respondent and interviewer. Many people had difficulty with the concept of providing a range, even when presented with a practice question. CATI techniques ensured that each respondent's best guess fell in the given range. Still, some respondents' actual values were on the border of their response, and, for the question on interest and dividends, many people did not want to provide a range. Other people appeared not to be motivated to think hard enough to give reasonable answers. Overall, more fine tuning is needed to make the paper's approach useful.

Martini, Alberto. "Research Grant Summaries: Why SIPP and CPS Produce Different Poverty Measures among the Elderly." *Social Security Bulletin*, vol. 60, no. 4, 1997, pp. 50-55.

The purpose of this research is to document the divergence between SIPP and CPS poverty measures, focusing on the elderly and to explain why such divergence arises, with particular focus on the role played by the reporting of various income sources. On average across four years (1987, 1988, 1990, and 1991), the SIPP poverty rates for the elderly are about 27 percent lower than in the CPS (9 versus 12 percent). The author also observes larger SIPP-CPS discrepancies among men than among women and larger discrepancies for married than nonmarried persons and for those living with others versus those living alone. The SIPP not only finds fewer poor people, it also finds that those counted as poor are on average somewhat better off than their CPS counterparts. The average income-to-needs ratio is about 78 percent among the SIPP elderly, whereas it is 71 percent in the CPS. The author notes that the SIPP counts more recipients for all sources of income. However, with the exception of self-employment income and Social Security benefits, average amounts among SIPP recipients are lower than their CPS counterparts. The author concludes that differences in the reporting of Social Security benefits seem to account for at least half of the observed poverty rate differential among the elderly in the SIPP and CPS. The other half of the differential can be explained by a combination of many other factors, of which only some can be precisely identified. Among them, the author notes the role of differences in the treatment of attrition and family composition, the interaction between income sources, and the role of other aspects of income reporting, such as part-year income and small amounts of income.

Mathiowetz, Nancy A., Charlie Brown, and John Bound. "Measurement Error in Surveys of the Low-Income Population," in *Studies of Welfare Populations: Data Collection and Research Issues*, edited by Michele Ver Ploeg, Robert A. Moffitt, and Constance F. Citro. Panel on Data and Methods for Measuring the Effects of Changes in Social Welfare Programs, Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press, 2002.

The authors provide an introduction to sources of measurement error and examine two theoretical frameworks (cognitive and social psychological) for understanding the various sources of error. They review the empirical literature concerning the quality of responses for reports of earnings and transfer income to identify those items most likely to be subject to response error among the welfare population. The paper concludes with suggestions for attempting to reduce the various sources of error through alternative questionnaire and survey designs. Such alternatives include the use of filter questions to determine the complexity of the experience and the use of different follow-up questions for those with simple and complex behavior. For example, the questionnaire might ask the respondent whether the amount of income from a particular income support program varies from month to month, with follow-up questions based on the response. The authors also suggest that simple, single-focus questions are often more effective than complex, compound questions. In addition, they suggest reducing cognitive burden by asking income questions in the form of recognition (Did you receive income from X?) rather than relying on free recall. And, to reduce cognitive burden, the authors suggest requesting earnings for the time period that the respondent is best able to respond. Finally, they suggest that unfolding income brackets may result in less nonresponse to income questions.

McGonagle, Katherine A., and Robert F. Schoeni. "The Panel Study of Income Dynamics: Overview & Summary of Scientific Contributions After Nearly 40 Years." http://psidonline.isr.umich.edu/Publications/Papers/montreal.pdf, January 30, 2006.

The authors describe the history of the PSID design as well as the key features of the design and content of the survey. The PSID sample was originally drawn from two independent samples: an over-sample of approximately 2,000 low income families from the Survey of Economic Opportunity and a national sample of approximately 3,000 households designed by the Survey Research Center, University of Michigan. They describe how the sample changed over the years as children leaving their parents' households were interviewed as their own family units. In addition, in 1990 the PSID added 2,000 Latino households, and while this sample represented a major group of immigrants, it did not cover all immigrants since 1968, especially Asians. Due to this shortcoming and insufficient funding, the Latino sample was dropped after 1995. In 1997 two major changes to the sample were made: 1) a reduction of the core sample and 2) the introduction of a refresher sample of post 1968 immigrant families and their adult children.

The authors conclude with the strengths and weaknesses of the survey. The strengths include consistently high response rates, the longevity of the data collection, a sample that is nationally representative and genealogically-based, content domains that are broad and recurring, and innovative supplements. There are five main weaknesses of the study. First, as a result of the longevity of the panel, cumulative attrition is an issue. Of the 18,192 individuals in the sample in 1968, 5,282 were alive and interviewed in 2001 and the remainder either died, were explicitly dropped from the study in 1997, or attrited. A second limitation is the periodicity of the PSID data collection. Currently, data are collected every other year but for the entire two-year period. The two-year reference period is especially disadvantageous for the collection of income and employment data. Third, until 1997 the PSID did not interview household members other than the family head and wife. This limitation was addressed in 1997 and 2002 with the Child Development Supplements. Moreover, a pilot study was launched in 2005 to interview children who had participated in the supplements and were at least 18 years old but not yet family heads or wives. A fourth weakness is the limited types of data that can be collected by a telephone interview. A fifth weakness is that new immigrants to the U.S. are not continuously represented in the sample. A large number of immigrants have arrived since 1999, and the PSID cannot be used to assess their outcomes.

McGrath, David E. "Comparison of Data Obtained by Telephone versus Face to Face Response in the U.S. Consumer Expenditures Survey." *Proceedings of the Annual Meeting of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association, 2005, pp. 3368-3375.

The CE was designed to collect data by personal visit. However, 42 percent of households report by telephone. The paper examines whether mode of data collection has a significant impact on data quality. White, non–Hispanic, highly educated people are more likely to report by telephone. By modeling expenditure data with a logistic regression model, the paper finds that mode of collection does not affect total expenditures. However, it is true that telephone respondents tend to refuse income questions such that telephone data are

allocated and imputed at significantly higher rates than data disclosed by personal visits. In addition, the paper finds that interviewers rather than respondents have the largest impact on whether the CE is completed by telephone or personal visit.

Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. "The Under-Reporting of Transfers in Household Surveys: Comparisons to Administrative Aggregates." Manuscript, March 7, 2007, bdmeyer@uchicago.edu.

Household surveys often underreport benefit receipt for reasons such as imperfect recall, a desire to reduce interview burden, the stigma of program participation, or the sensitivity of income information. This paper examines survey reports of benefit receipt from unemployment insurance, workers' compensation, Social Security, Supplemental Security Income, food stamps, the earned income tax credit, and Aid to Families with Dependent Children/TANF. The authors analyze data from the CPS ASEC, the PSID, and the SIPP and compare the weighted totals reported by households for these programs with those published by government organizations. The research results show sharp differences across programs and surveys as well as over time. Surveys differ systematically in their ability to capture benefit receipts. The SIPP typically has the highest reporting rate for government transfers, followed by the CPS and PSID. However, unemployment insurance and workers' compensation are reported at a slightly higher rate in the CPS than in the SIPP. These differences are informative as to the relative importance of the various reasons for underreporting. The reporting rates provided by the authors can also be used to adjust estimated program effects on income distribution and estimates of program take-up.

Meyer, Bruce D., and James X. Sullivan. "Measuring the Well-Being of the Poor Using Income and Consumption." *The Journal of Human Resources*, vol. 38, supplement, 2003, pp. 1180-1220.

This article compares income and consumption as measures of the material well-being of the poor. After reviewing the conceptual and pragmatic reasons that favor income or consumption, the authors examine relevant findings from earlier research and present an empirical analysis using income and consumption data from the CE and the PSID. Comparisons of percentile distributions of income, expenditures, and consumption as well as average income and expenditures show that in both surveys, reported expenditures exceed reported income among low-educated single mothers and among all families at the low ends of both distributions. Reported expenditures among families with low reported incomes provide evidence that incomes in this subpopulation are substantially understated. The authors review evidence from other studies that indicate substantial under-reporting of key components of income in the CPS and SIPP. Finally, the authors examine other measures of hardship and material well-being by level of income and consumption among low-educated and all single mothers in the CE and PSID. The findings suggest that reported consumption does a better job than reported income in capturing well-being among disadvantaged families.

Moon, Marilyn, and F. Thomas Juster. "Economic Status Measures in the Health and Retirement Study." *The Journal of Human Resources*, vol. 30, supplement, 1995, pp. S138-S157.

This paper offers a flavor for the major economic status variables in the HRS, provides some preliminary analysis of the quality of the data, and takes a preliminary look at the interrelationships among economic status measures such as income and wealth and other important variables, including health status, pension rights, and health insurance coverage. The authors also compare the first wave of HRS income data with all households headed by a person between the ages of 51 and 61 from the March 1992 CPS. They find strong similarities in the amount and distribution of income in the two data sets. Poverty rates are somewhat lower for the CPS.

Moore, Jeffrey C., and Laura Loomis. "Using Alternative Question Strategies to Reduce Income Nonresponse." *Proceedings of American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2000, pp. 947-952.

This paper describes research that builds on the unfolding brackets approach to asking about income and tests a new form of income range reporting, which the authors label "implicit brackets." The authors conducted the research as part of the Census Bureau's Questionnaire Design Experimental Research Survey, which was a split-sample experiment using a paper-and-pencil instrument in a telephone interview with a random digit dial sample. For the experimental "implicit brackets" treatment, the question format consisted of two parts: (1) whether annual income for 1998 was more or less than $X, where $X was a minimum amount varying by asset type; and (2) if the answer was "more," then the respondent was asked, "How much was it to the nearest $Z?" The second question in effect establishes response brackets of width $Z. The authors evaluate five asset income sources: checking accounts, savings accounts, certificates of deposit, mutual funds, and stocks. For all five asset income sources, the item nonresponse rate for the experimental treatment was lower than for the control. However, all of the improvement came from a reduction in "don't knows" and not from a reduction in refusals. The authors also find that the distribution of income responses did not differ by questionnaire treatment. Finally, they find that the experimental treatment seemed to increase report precision.

Moore, Jeffrey C., Kent H. Marquis, and Karen Bogen. "The SIPP Cognitive Research Evaluation Experiment: Basic Results and Documentation." SIPP Working Paper 212, Statistical Research Division, U.S. Census Bureau, January 11, 1996.

The Census Bureau implemented a test of new procedures designed to reduce measurement error. One procedure asked household respondents to use their personal income records instead of relying on memory. The results indicate that the procedures had no effect on reducing either under- or over-reporting of participation in income programs. However, the new procedures did produce substantial improvement in reporting income amounts.

Moore, Jeffrey C., Linda L. Stinson, and Edward J. Welniak, Jr. "Income Measurement Error in Surveys: A Review." *Journal of Official Statistics*, vol. 16, no. 4, 2000, pp. 331-361.

This paper reviews what is known about income measurement errors. It focuses on response error research by comparing individual survey respondents' reports to external measures of truth obtained by independent record systems. The paper finds that errors in individual surveys include both bias and random error, with substantially varying propensities for these errors across different income types. However, the authors cite several papers indicating that 95 percent of reported wages and salaries is accurate and concluding that over- and underreporting tend to cancel out (although reporting is slightly underestimated). Research on transfer programs indicates a large and consistent negative bias while many sources indicate that assets suffer from severe underreporting. The paper also finds that respondents have trouble understanding income concepts and terms such as "nonwage cash payments" or "total family income." Others have trouble retrieving information and constructing "monthly pay," for example. Some surveys have found that telling respondents to use records increases accuracy but also places further burden on both respondent and interviewer. Overall, the paper concludes that several problems need to be solved in order to improve income measurement.

Moyer, M. Eugene. Counting Persons in Poverty in the Current Population Survey. August 1998, http://aspe.os.gov/rn/rn20.htm.

The Census Bureau estimated that 36.5 million persons were in poverty in 1996. However, if an analyst were to estimate from the CPS the number of persons in families whose income is less than the poverty level, the estimate would be higher. Two reasons explain the difference. First, by definition, unrelated children under age 15 have no income because the CPS does not ask about their income. They tend to live with families that are not poor. The U.S. Department of Health and Human Services estimates that 40 percent of these children are foster children placed with the family, and, while the family is not poor, the children were poor when they were placed with the family and probably will again be poor when they return to their birth parents. Therefore, the Department has always included them in its count of persons in poverty. Second, some families contain subfamilies. If the analyst counts the subfamily as part of the primary family (as the Census Bureau does), the entire family is likely to have income higher than the poverty level, and no one in the family would be counted as being in poverty.

Nelson, Charles. "What Do We Know about Differences between CPS and ACS Income and Poverty Estimates?" Housing and Household Economic Statistics Division, U.S. Census Bureau, August 21, 2006.

The author summarizes methodological and conceptual differences between the CPS ASEC and ACS as well as differences in the timing of estimates and then compares national estimates and measures of sampling and nonsampling error. The methodological differences include mode of data collection, reference period, income question detail, sample size, survey universe, family unit definition, and residence rules. The differences in timing of estimates can be seen at the national level; CPS results released in August 2006 were based on a somewhat more recent time period than the ACS results. The comparisons of national estimates show that the ACS and CPS were similar in 2004 in that both surveys indicated a rise in poverty between 2003 and 2004, with no change in real median household income

over the period. In terms of point estimates, the ACS poverty rate (13.3 percent) in 2005 was higher than the CPS national rate of 12.6 percent. The CPS poverty rate was lower than the ACS rate in five out of six years between 2000 and 2005, and the rates were not statistically different in the sixth year. The relationship between ACS and CPS median household income has not been consistent; two years showed different estimates, and four years had estimates that were not statistically different.

The author continues by comparing measures of sampling and nonsampling error. At the state level, the author finds that the standard errors of the ACS poverty rates are significantly smaller than the comparable CPS single- or three-year poverty rate standard errors. And while the 2004 CPS aggregate total money income estimate of $6.940 trillion was slightly higher than the 2004 ACS aggregate of $6.862 trillion, the author points out three types of income in which the ACS aggregate was higher than the CPS—self-employment income, public assistance, and retirement income. The author speculates that the difference could be attributable to respondent reporting error, differences in the questionnaire, and differences in how the estimates are constructed. The weighted unit response rate for ACS is around 97 percent while the CPS ASEC combined response rate is around 80 percent. Moreover, item nonresponse rates in the CPS ASEC are higher than comparable ACS figures. Therefore, it would appear that differences in imputation methodology between the two surveys should be considered a potential source of differences between the two estimates. Coverage error could also be a source of differences. The ACS coverage rate is 95 percent, and the CPS coverage rate is around 89 percent.

The author concludes with a comparison of state distributions of poverty and income estimates from the CPS and ACS. In 13 states, the 2004–2005 CPS poverty rate was lower than the 2005 ACS rate. The CPS rate was higher than the ACS rate in two states, Maryland and New York. The author concludes from various Chi-squared test results that strong evidence shows that the 2004–2005 CPS and 2005 ACS estimate different geographic distributions of poverty.

Nelson, Charles T., and Patricia Doyle. "Recommendations for Measuring Income and Program Participation in the Post Welfare Reform Era." *Proceedings of the American Statistical Association, Government Statistics and Social Statistics Sections*. Alexandria, VA: American Statistical Association, 1999, pp. 54-63.

Changes to means-tested benefit systems under welfare reform made it necessary for surveys that collect data on program participation and benefit receipt to modify their questions to avoid losing reported benefits. A topical module administered in wave 8 of the 1996 SIPP panel collected data to determine how welfare reform was affecting the way that people maintained program eligibility and received benefits. This paper discusses planned changes to the core content of the SIPP based on early analysis of the wave 8 topical module data and recommends that portions of the wave 8 topical module be added to future SIPP panels to provide a continuous source of information on the changes in forms of benefit receipt brought about by changes in the way that government benefits are delivered.

Olson, Janice A. "Social Security Benefit Reporting in the Survey of Income and Program Participation and in Social Security Administrative Records." SIPP Working Paper 235, U.S. Census Bureau, 2001. http://www.sipp.census.gov/sipp/workpapr/wp235.pdf.

This paper examines the consistency between Social Security benefit amounts reported in the SIPP and provided in SSA administrative records. A particular interest, especially for the elderly, is whether the amounts reported in the SIPP include the amount of Supplementary Medical Insurance (SMI) or the Medicare Part B premium. Only 25 percent of the elderly and 42 percent of the nonelderly reported a Social Security benefit amount in the SIPP that was within $1 of the amount in SSA administrative records. About three-quarters of both groups reported an amount within 10 percent of that in the records. This analysis suggests that beneficiaries *under* age 65 who were retired workers, aged spouses, and aged widows are the best reporters. Roughly half of them reported amounts matching the Monthly Benefit Credited in the SSA data, a result consistent with the idea that those newly on the program are more likely to have accurate recall of the benefit amount they receive. In contrast, only about a quarter of disabled workers and of beneficiaries age 65 and over (regardless of type) reported consistent amounts. In the SIPP, underreporting of Social Security benefit amounts by the amount of the Medicare premium does not appear to be a major problem among elderly or disabled beneficiaries, although disproportionate shares of both groups make such reports. However, possible measurement error, particularly substantial underreporting by those at the low end of reported benefit amounts (and, to a lesser degree, overreporting at the high end), may be a nontrivial problem, especially among the elderly.

Patil, Vrushali, and J. Neil Russell. *Final Report of the 2000 National Health Interview Survey Welfare Pretest*. Centers for Disease Control and Prevention, National Center for Health Statistics, Division of Health Interview Statistics, September 2000.

This report analyzes the test of various versions of welfare reform–based questions. The test was needed to evaluate and revise old questions after the 1996 implementation of welfare reform. The test used a split-ballot questionnaire design to examine the wording of seven questions as well as a split-ballot design for block areas where low-income respondents resided. Given time constraints, the test did not randomly assign questionnaires. The authors use logistic regression to analyze information about the questions and find that different wording would increase understanding of the questions for several items.

Paulin, Geoffrey and David Ferraro. "Imputing Income in the Consumer Expenditure Survey." *Monthly Labor Review*, vol. 117, no. 12, 1994, pp. 23-31.

This article summarizes methods of adjusting for nonresponse bias in the CE. In the early part of the century, account balancing was used to eliminate large gaps between family income and expenditures. More recently, more complex methods have found application. For example, a hot deck method assigns missing values from a donor from the same demographic group but has proven problematic in that the CE sample size is relatively small. Another method is model-based and creates a statistical model to impute missing values. Models can be specified at the member or family level. Research by Paulin and

Ferraro attempts to explore whether income could be modeled on expenditures. Other research by Chand and Alexander uses stochastic methods to impute income separately for each member and each source of income.

Paulin, Geoffrey D., and Elizabeth M. Sweet. "Modeling Income in the U.S. Consumer Expenditure Survey." *Journal of Official Statistics*, vol. 12, no. 4, 1996, pp. 403-419.

Nonresponse to income questions is common in household surveys. This study examines wage and salary income data from the 1988–1990 CE. A large portion (15 percent) of the sampled families are classified as incomplete income reporters, and not all complete income reporters provide a full accounting of all sources of income. The authors explore different procedures designed to yield a model-based imputation strategy for wage and salary income of two-person consumer units. The two-member units represent a link between single-member consumer units (with few inherent difficulties for modeling) and more complex multiple-member consumer units (with several inherent difficulties). Selected variables from each consumer unit are synthesized into a final model that is tested for proper specification. Results of the final model indicate that imputation increases the means of published CE income data.

Pedace, Roberto, and Nancy Bates. "Using Administrative Records to Assess Earnings Reporting Error in the Survey of Income and Program Participation." *Journal of Economic and Social Measurement*, vol. 26, 2000, pp. 173-192.

This paper analyzes income misreporting propensities and magnitudes by using the 1992 SIPP longitudinal file matched to Social Security Summary Earnings Records. Specifically, the authors focus on wage and salary and self-employment earnings. The paper compares SIPP data to SSA records while making the assumption that the SSA data represent the "truth." The findings suggest that the 1992 SIPP accurately estimated the net number of earnings recipients but tended to underestimate amounts received. The mean difference in dollar amounts between the SIPP and SSA records was -$459, although the magnitude and direction of difference varied by income category. An interesting characteristic of SIPP misreporting is that it overreports in the lowest income categories but underreports for those with at least $20,000 in earnings. The authors use a logit model to estimate misreporting. Those age 50 to 64, males, Hispanics, those without a college education, blacks, Asians, and those who are married or divorced all had significantly higher rates of misreporting. Those who work in farming/forestry/fishing, craft, or military operations also had significantly larger reporting errors. The self-employed tended to overreport.

Pleis, John R., and James M. Dahlhamer. "Family Income Nonresponse in the National Health Interview Survey: 1997–2000." *Proceedings of the American Statistical Association, Section on Survey Research Methods* [CD-ROM]. Alexandria, VA: American Statistical Association, 2003, pp. 3309-3316.

The goal of the paper is to analyze different types of nonresponse in NHIS income data. Most studies treat "don't know" and refusals the same way, but the paper finds that different

types of people are more likely to refuse to answer or to answer "don't know." Refusers were likely to have higher incomes, whereas "don't knows" had lower incomes. Education, marital status, and current employment status are variables that seem to indicate whether a respondent is more likely to refuse than say "don't know." The paper also shows that those with a GED were more similar to those without a high school education than those with a diploma. The paper argues that type of nonresponse should be considered when imputing income data, perhaps requiring different follow-up questions depending on type of nonresponse.

Pleis, John R., and James M. Dahlhamer. "Family Income Response Patterns for Varying Levels of Income Detail: An Analysis of the National Health Interview Survey (NHIS)." *Proceedings of the American Statistical Association, Section on Survey Research Methods* [CD-ROM]. Alexandria, VA: American Statistical Association, 2004, pp. 4200-4207.

This paper measures how much detail people were willing to give about their incomes and what characteristics affected their willingness to disclose such information. The categories used in the ordinal regression were no information given, income greater or less than $20,000, income chosen from a list of categories with $5,000 increments, or exact amount given. The variables that appreciably increased the amount of income detail were age (younger), race (multiracial), employment in the previous year (employed), marital status (married), income sources (more), and adults in the family (fewer). The data show that nonresponse bias is likely to affect analyses involving total family income.

Pleis, John R., James M. Dahlhamer, and Peter S. Meyer. "Unfolding the Answers? Income Nonresponse and Income Brackets in the National Health Interview Survey." Proceedings of the American Statistical Association, Section on Survey Research Methods [CD-ROM]. Alexandria, VA: American Statistical Association, 2006, pp. 3540-3547.

Nonresponse to income-related survey questions is problematic and may lead to biased estimates. In the NHIS, respondents are first asked to provide the exact dollar amount of the family's income in the previous calendar year (nonresponse @ 30%). Previously, follow-up questions based on income intervals had had minimal effect on lowering nonresponse. This paper analyzes results of a test that used NHIS screened-out households in April-June of 2006 to pose alternative income questions using unfolding brackets. Respondents were randomly assigned to the existing or alternative method. Alternative methods for asking about the sources of income were used to assess whether item nonresponse for income could be reduced. Instead of asking about each source separately, a flashcard approach was used where families were asked about only the income sources of which they initially indicated receipt. According to results from the 2006 field test, the alternative follow-up income questions (unfolding brackets) performed much better than the follow-up income questions used since the 1997 NHIS. The path completion rate for the alternative income follow-up questions (unfolding brackets) was approximately 47%, while the path completion rate was 12% for the income follow-up questions used since the 1997 NHIS. Based on the favorable results from the 2006 field test, the unfolding bracket follow-up income questions were incorporated into the NHIS beginning in 2007.

Posey, Kirby G., and Edward Welniak. Income in the ACS: Comparisons to the 1990 Census. U.S. Census Bureau, March 25, 1998, http://www.census.gov/acs/www/AdvMeth/ Papers/ACS/Paper16.htm.

This paper compares income estimates in the 1996 ACS to income estimates in the 1990 Decennial Census. The major difference between the two surveys was that the ACS asked about income in the last 12 months while the census asked about income in the previous year. A split-panel study in October to December, 1997, showed significantly less wage and salary income was reported for the last 12 months than for the prior calendar year. In addition, given that the ACS uses 12 possible reference periods (depending on when respondents answered the survey), it must use inflation adjustment factors to account for the various periods. Allocation and imputation schemes used in the ACS were largely the same as those used in the census. The adjusted median income results for the four ACS sites of interest were lower than the census results. The authors attribute this finding to the recession as well as to the use of national CPI factors in local areas. A final analysis also finds that median household income estimates from mail returns and CATI matched the census figures much more closely than did the CAPI interviews (which were lower).

Posey, Kirby G., Edward Welniak, and Charles Nelson. "Income in the American Community Survey: Comparisons to Census 2000." *Proceedings of the Annual Meeting of the American Statistical Association* [CD-ROM]. Alexandria, VA: American Statistical Association, 2003, pp. 352-3359.

The ACS and 2000 Decennial Census both collected information on total income. However, the two surveys use different reference periods. The ACS collects data throughout the year on an ongoing basis and asks for a respondent's income over the "past 12 months." The 2000 Decennial Census collected income for 1999 (the last calendar year). This paper describes a split-panel test conducted over the period October through December 1997 to evaluate the impact of a prior calendar year versus past 12 months reference period. The only statistical differences in median income estimates between the two reference periods occurred in the earnings categories—wages/salary and self-employment. The questionnaire with the "past 12 months" reference period produced slightly higher response rates for every income source. However, only one income source, public assistance, shows a statistically significant difference. The paper also describes the C2SS, which is an ACS program designed to demonstrate the feasibility of collecting long form–type information in a census environment. The C2SS was conducted at the same time as the 2000 Decennial Census but as a separate effort. Median household income estimates were generally lower in the C2SS/ACS than in the 2000 Decennial Census after adjusting the census's 1999 dollar values for inflation. Nationally, median household income was more than 4 percent lower in the C2SS than in the census. Five states reported median household incomes that were more than 8 percent lower in the C2SS than in the census. The C2SS's median household income at the national level matched more closely with CPS estimates. Surprisingly, of the three major Census Bureau household survey-based estimates of median income at the national level, the outlier is the 2000 Decennial Census estimate, not the C2SS or CPS estimate. The authors conclude with possible explanations for the differences in income estimates.

Reichert, W. Jennifer, and John C. Kidelberger. "Reliability of Income Poverty Data from the Current Population Survey Annual Demographic Supplement." *Proceedings of the American Statistical Association, Section on Survey Research Methods.* Alexandria, VA: American Statistical Association, 2000, pp. 151-156.

The CPS Supplement is the source for estimating national poverty, and, for the first time, the Census Bureau used reinterviews to determine if people's responses were consistent with their answers in an earlier interview. The authors use an index of inconsistency to evaluate response consistency by taking the ratio of response variance to total variance for each question. The model assumes that people's responses on each survey are independent of each other. Many of the questions about participation in poverty-related programs were subject to high variability, suggesting that the questions were unreliable and could confuse respondents. Questions relating to household income were fairly reliable.

Rodgers, Willard, Charles Brown, and Greg J. Duncan. "Errors in Survey Reports of Earnings, Hours Worked, and Hourly Wages." *Journal of the American Statistical Association*, vol. 88, no. 424, December 1993, pp. 1208-1218.

Data collected as part of a validation study for the PSID were analyzed to assess the quality of reporting of earnings, hours worked, and hourly wages for hourly employees of a single manufacturing firm. In comparing reported values with the firm's administrative records, the authors found that standard assumptions about measurement error were violated to varying degrees. Signed errors were (negatively) correlated with true values, and errors in reported earnings and hours worked in different periods were generally (positively) correlated with errors in other periods. Reporting errors followed an approximately normal distribution, with departures from normality being due primarily to a small number of outliers. These exerted considerable influence on estimates of relationships between variables. Overall, these results demonstrate the importance of validation studies as a source of realistic assumptions about measurement error.

Roemer, Marc. "Using Administrative Earnings Records to Assess Wage Data Quality in the March Current Population Survey and the Survey of Income and Program Participation." Longitudinal Employer-Household Dynamics Program, Demographic Surveys Division, U.S. Census Bureau, November 19, 2002. www.census.gov/hhes/hhes/income/papers.html.

The March CPS and SIPP produce different aggregates and distributions of annual wages. The former reports an excess of high wages and shortage of low wages; the latter reports the opposite. Exactly matched Detailed Earnings Records from the SSA allow a comparison of March CPS and SIPP wages by using data independent of the surveys. The findings show that the March CPS and SIPP represent a worker's percentile rank better than the dollar amount of wages. In addition, the March CPS accounts for a higher level of "underground" wages than does the SIPP and increasingly so in the 1990s. The March CPS reports a higher level of self-employed income "misclassified" as wages than does the SIPP and increasingly so in the 1990s. These trends explain one-third of the March CPS's 6 percentage point increase in aggregate wages relative to independent estimates from 1993 to 1995. Finally,

the paper delineates March CPS occupations disproportionately likely to be entirely absent from the administrative data or self-employment income misclassified as wages.

Roemer, Marc I. "Assessing the Quality of the March Current Population Survey and the Survey of Income and Program Participation Income Estimates, 1990–1996." Income Surveys Branch, Housing and Household Economic Statistics Division, U.S. Census Bureau, June 16, 2000.

This paper establishes a methodology for deriving benchmarks from the National Income and Product Accounts (NIPA) and evaluates CPS and SIPP income estimates by comparing them to these benchmarks. It also considers possible misestimates by the two surveys and explains the changes in the relationship between the surveys. Some NIPA figures need adjustment for institutionalized individuals, decedents, those residing overseas, and those in the military without family. Other adjustments address differences in what is considered income, such as lump-sum payments. As for earnings, the March CPS estimate increased from 93 to 96 percent of benchmark. The SIPP earnings estimate decreased from 90 to 88 percent of benchmark. Among general categories of income, only SIPP pensions have improved relative to the March CPS and perhaps just slightly relative to benchmarks. However, the paper concludes that redesigning the SIPP for 1996 does not seem to improve its income estimates. Even though the SIPP identifies more recipients than the March CPS, it has lower income aggregates, posing a challenge to analysts. In addition, analysis of tax returns matched to the March CPS shows the occurrence of both over- and underreporting, suggesting that comparing aggregate data to benchmarks may be a simplistic method of measuring data quality.

Roemer, Marc. "Reconciling March CPS Money Income with the National Income and Product Accounts: An Evaluation of CPS Quality." Paper presented at ASA Joint Statistical Meeting, Baltimore, August 10, 1999.

This paper attempts to create income benchmarks to reconcile differences between the March CPS and NIPA definitions of income. To compare the two, the author adjusts NIPA's universe to include institutionalized individuals, decedents, those residing overseas, and those in the military. In addition, the March CPS includes only cash that people can spend, whereas NIPA includes all economic resources, a difference corrected in the paper's methodology. The article also explains trends in measurement over time for various income measures. The gap between the March CPS and NIPA estimates increased in 1996 as compared with previous years, but the overall completeness of the March CPS improved.

Ruser, John, Adrienne Pilot, and Charles Nelson. Alternative Measures of Household Income: BEA Personal Income, CPS Money Income, and Beyond. Presented to the Federal Economic Statistics Advisory Committee, December 14, 2004. http://www.bea.gov/bea/about/fesac/AlternativemeasuresHHincomeFESAC121404.pdf.

This paper compares personal income and money income and analyzes how they differ. Bureau of Economic Analysis (BEA) personal income is income received from participation

in production, from government and business transfer payments, and from government interest. BEA estimates personal income largely from administrative sources. CPS money income is defined as total pre-tax cash income earned by persons, exclusive of certain lump-sum payments and capital gains. BEA estimates income at $8.678 trillion and CPS estimates $6.446 trillion. Nearly two-thirds of the difference is attributable to differences in income types between the sources and 18 percent to BEA adjustment and underreporting. The Census Bureau has developed alternative measures that better describe economic well-being and reduce the gap between rich and poor. An unresolved issue is whether some types of income (such as pensions) should be counted when accrued or when dispersed. Many proposed alternative measures move toward the theoretical concept of income as the maximum amount that can be consumed while keeping real wealth unchanged.

Schenker, Nathaniel, Trivellore E. Raghunathan, Pei-Lu Chiu, Diane M. Makuc, Guangyu Zhang, and Alan J. Cohen. Multiple Imputation of Missing Income Data in the National Health Interview Survey. *Journal of the American Statistical Association*, vol. 101, no. 475, September 2006a, pp. 924-933.

The NHIS provides a rich source of data for studying relationships between income and health and for monitoring health and health care for persons at different income levels. However, nonresponse rates are high for two key items: total family income in the previous calendar year and personal earnings from employment in the previous calendar year. To handle missing data for family income and personal earnings, the authors perform multiple imputation of these items, along with employment status and the ratio of family income to the federal poverty threshold, for NHIS survey years 1997–2004. This article describes the approach used in the multiple-imputation project and evaluates the methods by analyzing the multiply imputed data. The analyses suggest that imputation corrects for biases that occur in estimates based on data without imputation and that multiple imputation usually results in lower estimated standard errors than analyses of data without imputation.

Schenker, Nathaniel, Trivellore E. Raghunathan, Pei-Lu Chiu, Diane M. Makuc, Guangyu Zhang, and Alan J. Cohen. Multiple Imputation of Family Income and Personal Earnings in the National Health Interview Survey: Methods and Examples, July 30, 2006b, http://www.cdc.gov/nchs/about/major/nhis/2005imputedincome.htm.

The NHIS provides a rich source of data for studying relationships between income and health and for monitoring health and health care for persons at different income levels. However, nonresponse rates are high for two key items: total family income in the previous calendar year and personal earnings from employment in the previous calendar year. To handle the problem of missing data for family income and personal earnings, the authors performed multiple imputation of these items for NHIS survey years 1997–2005 and plan to create multiple imputations for 2006 and beyond as data become available. The multiple imputations used an adaptation of Sequential Regression Multivariate Imputation that handles the hierarchical nature of the data. Examination of observed data on two-category income (less than $20,000 versus $20,000) suggests that multiple imputation corrects for biases that occur in estimates based on data without imputation (that is, based on complete-cases analysis). Further, multiple imputation usually results in lower estimated standard

errors than do analyses of the data without imputation. For each survey year, data sets containing the imputed values, along with related documentation, are available from the NHIS Web site (http://www.cdc.gov/nchs/nhis.htm).

Schwartz, Lisa, and Geoffrey Paulin. "Improving Response Rates to Income Questions." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 2000, pp. 965-970.

Income data in the CE Quarterly Interview Survey had a high missing rate of 17.7 percent in 1997. Brackets, which are categories or ranges offered to respondents who initially refuse to report income, help in eliciting a partial response. This study investigates the usefulness of bracketing techniques for the CE and compared three bracketing methods: (1) conventional bracketing, which presents the respondent with several researcher-determined data ranges; (2) unfolding bracketing, which asks the respondent a series of yes/no questions designed to narrow the respondent's income range; and (3) respondent-generated intervals, which ask the respondent to provide the upper and lower limits on his or her income. Sixty adults participated in mock CE interviews followed by intensive cognitive interviews. The income item nonresponse rate was 18.1 percent but fell to 9.5 percent with the inclusion of brackets. The results indicate that the unfolding technique is the least popular with respondents. Moreover, respondents liked the respondent-generated intervals technique, and respondent-generated intervals tended to be smaller than those generated by researchers.

Scoon-Rogers, Lydia. "Evaluating Respondents' Reporting of Social Security Income in the Survey of Income and Program Participation Using Administrative Data." Federal Committee on Statistical Methodology, 2005 Research Conference paper.

This paper looks at reporting error and the impact of excluding the Medicare Part B (or SMI) premium from reported Social Security benefits in the SIPP. Using Social Security administrative records matched to SIPP records from the 1996 panel, the author finds that adding the SMI deduction to the reported benefit and correcting any additional error reduces the elderly poverty rate by 2.3 percentage points. Unmatched records have a higher poverty rate than matched cases, suggesting that the uncorrected error among these cases could be even greater.

Sears, James, Kalman Rupp, and Melissa L. Koenig. "Exploring Social Security Payment History Matched with the Survey of Income and Program Participation: An Assessment." Federal Committee on Statistical Methodology, 2003 Research Conference papers, pp. 49-57.

All SIPP panels have been matched to Social Security (OASDI) and SSI benefit history records, providing a valuable resource for assessing the quality of SIPP data. This paper examines matched data for the 1996 and 2001 SIPP panels. The match rate for the 1996 panel, 85 percent, is typical of earlier panels, but the match rate for the 2001 panel is only 60 percent. The recent availability of actual payment record data instead of payment eligibility data further enhances the potential of the administrative data to evaluate and improve the

accuracy of the survey data. At all ages the reporting of OASDI benefit receipt is more accurate than the reporting of SSI receipt, but among the elderly about a third of those with neither benefit report receiving OASDI. Nearly a fifth of the elderly with SSI fail to report it. Among those who correctly reported receiving OASDI in the 1996 panel, 53 percent reported benefits that were within $10 of the actual amounts. This compares to 62 percent in the 1996 panel. In addition, large errors grew in frequency, with 22 percent of the 2001 panel versus 16 percent of the 1996 panel misreporting their benefits by $100 or more. The paper concludes that both survey error and the quality of the SSN match need careful consideration. With the sharp decline in the match rate, an important issue is whether to base analysis on survey matches only—as SSA analysts have done previously—or to combine matches and nonmatches.

Short, Kathleen S. "The Relationship between Monthly and Annual Income." Housing and Household Economic Statistics Division, U.S. Census Bureau, October 26, 1990.

This paper examines the relationship between income collected for one month and annual income. At the time, the income reference period for the NHIS was the month before the interview, and the interview collected dollar amounts for several income sources. However, for many analyses, annual income is the preferred measure because it avoids problems with seasonality, covers a sufficiently long period to establish well-being—such as poverty status—and allows for comparability among various surveys. Using the SIPP to simulate the NHIS, predictors of annual income were derived from monthly income, as though monthly income were the only available information. Predicted income was then compared with "actual" annual income to assess the reliability of predicting annual income from a single month's income. First, the author examined a naïve estimator of annual income obtained by multiplying a month's income by 12. The analysis suggests that this simple estimator may be reasonable for monthly income below some given level but that the relationship between annual and monthly income is not linear. For example, very large monthly income typically does not result in very large annual income. The author also fit a regression equation predicting total personal income. The model included person-month income, monthly dummy variables representing the month in which each amount was received, and demographic indicators. The predicted values from the model follow much more closely the pattern of annual income calculated for each person than does the simple estimator. The author also examined the effect of removing outliers on the prediction of annual income and investigated the lower end of the income distribution, concluding that there is some nonlinear relationship between monthly and annual income. The author observed statistically significant differences in income by month, indicating that months matter in measuring income. Examination of the lower end of the income distribution suggests that classification of persons as poor based on monthly income overstates the number of persons in poverty by almost 3 percent. More accurate classification is possible for persons with specific characteristics, such as those not working or those receiving income from government programs.

Smeeding, Timothy M., and Daniel H. Weinberg. "Toward a Uniform Definition of Household Income." *Review of Income and Wealth*, series 47, no. 1, March 2001, pp. 1-24.

This article attempts to provide a unified framework for aggregating income types to create an income definition that enables researchers to make valid comparisons across nation states. An examination of several national household income surveys shows that it is nearly impossible to quantify all elements of any new comprehensive income definition in a way that expedites comparisons. The authors hope that their framework—a combination of national income-based approaches and a microdata perspective—illuminates the differences in current practice and allows researchers to assess the effect of those differences on income distribution measures. The authors also review theoretical approaches to income definition, present recommendations for constructing a new definition of income, and discuss the feasibility of collecting sufficient data to create comparable international measures.

Susin, Scott. Discrepancies between Measured Income in the American Housing Survey (AHS) and the Current Population Survey (CPS). Final Report. U.S. Census Bureau, March 27, 2003. www.census.gov/hhes/income/papers.html.

The CPS measurement of income is more detailed than the AHS, especially with respect to non-wage income. The two surveys also use different recent periods, with the CPS asking about the previous calendar year and the AHS, which is conducted late in the year, asking about income for the previous 12 months. Average household income is 9 percent lower in the AHS than in the CPS. Family earnings are about the same while non-wage income is 32 percent lower because of the failure of many respondents to report non-wage income. The discrepancy has widened over time, especially since 1995. Underreporting increases with the number of adults in the household, indicating that the CPS's practice of asking each person about income makes a difference. The largest sources of underreported income are interest, dividends, Social Security, and pensions. Those with business income in the AHS report 49 percent more earnings than in the CPS, perhaps reflecting self-employment income reported on the wrong line of the survey. Reanalysis of a 1991 AHS experiment indicates that, compared to a paper instrument (also administered by telephone), CATI reduces non-wage income by $308 on average and has a particularly large effect on families receiving business income. Finally, the CPS counts several sources of income not counted by the AHS, including educational and financial assistance, which represents roughly 10 percent of the gap in non-wage income.

Turek, Joan. "Measuring Income on Surveys: Content and Quality, an Overview." Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, August 1, 2005.

This report describes the income data collected in six federal surveys. Three of the surveys are designed and conducted by the Census Bureau: the SIPP, the Annual Social and Demographic Supplement to the CPS (March CPS), and the ACS. Three surveys are designed by the U.S. Department of Health and Human Services: the MEPS sponsored by the Agency for Healthcare Research and Quality, the NHIS sponsored by the National Center for Health Statistics, and the MCBS sponsored by the Centers for Medicare and Medicaid Services. Of these surveys, only the SIPP has as its mandate the collection of income data. The main purpose of the other surveys is to collect employment and/or health information. Currently, the March CPS provides official estimates of income, poverty, and

health insurance status. The paper also presents available findings on the quality and comparability of the collected data.

Turek, Joan. "Poverty Measures from the Current Population Survey (CPS) and the Consumer Expenditure Survey (CE): Why Do They Differ?" Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, Summer 2001, Joan.Turek@hhs.gov.

Official measures of poverty are obtained from CPS data, although poverty measures can also be constructed by using the results of other surveys, such as the CE. Poverty rates estimated from the CE, however are significantly higher than those estimated from the CPS. The author examines the reasons for the differences and calls for caution when using income-based measures obtained from the CE. Differences in income reporting on the two surveys, particularly underreporting of income in the CE, have a dramatic influence on reported poverty rates. The author finds that significant underreporting of income in the CE captured approximately 86 percent of the aggregate income captured by the CPS in 1990 and 83 percent in 1996. In addition, the CE captured less aggregate income for types of income typically received by people who are better off. For example, in 1996, the CE captured about 28 percent of the aggregate property income (interest, dividend, rents, royalties, estates, and trusts) reported in the CPS. As a result, respondent units in the CE were classified as poor when they were not. Therefore, poverty rates from the CE are inflated as a result of income underreporting. Similarly, comparisons of CE units that use income as a classifying variable are also misleading.

Turek, Joan, Gabrielle Denmead, and Brian James. "Poverty Estimates in the ACS and Other Income Surveys: What Is the Impact of Methodology?" Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, Winter 2004, Joan.Turek@hhs.gov.

For users to take full advantage of ACS data, they need to be aware of methodological differences between the ACS and other surveys. The authors examine three features of the ACS that differ from the Annual Social and Economic Supplement to the CPS and the Decennial Census Long Form Survey (Long Form): a rolling sample, a rolling reference period, and CPI adjustments to the (rolling) income data (the ACS uses these adjustments to approximate fixed sampling and references periods). The authors use the 1996 SIPP panel to construct simulated ACS, CPS, and Long Form estimates for 1998. They replicate each survey's sampling, reference period, weighting, and CPI adjustments as accurately as possible. They change each feature in turn to pinpoint each feature's respective contribution to differences in estimates but obviously cannot control for other differences attributable to the number of income questions, recall periods, or family relationship measures. The authors' tests show that the ACS rolling sample for 1997–1998 yields a higher estimate of poverty than the fixed samples and reference periods for the 1998 CPS and Long Form. This finding holds even with CPI adjustments and an adjustment for SIPP panel attrition, which partially offsets the measured differential. Given that the ACS rolling sample is lagged as compared to true calendar year income, the authors' result could reflect an understatement by the ACS of increases in real income over the lag, although use of a two-year CPS average moderates this effect.

U.S. Census Bureau. Guidance on Differences in Income and Poverty Estimates from Different Sources. U.S. Census Bureau, March 12, 2007. http://www.census.gov/hhes/www/income/ newguidance.html..

This Census Bureau Web site offers guidance on income and poverty estimates from different sources and contains a chart of which data source to use for each purpose and geographic level; a fact sheet on the differences between CPS ASEC, and ACS data for income and poverty; and a comparison of household income from ACS 2005 and CPS ASEC 2004–2005 averages. It also contains background information on income and poverty estimates from five Census Bureau national household surveys and programs: (1) the CPS ASEC, (2) ACS, (3) SIPP, (4) 2000 Decennial Census Long Form, and (5) Small Area Income and Poverty Estimates program.

U.S. Census Bureau, Comparability of Current Population Survey Income Data with Other Data. Washington, DC, http://www.census.gov/hhes/www/income/compare1.html, March 9, 2005.

This article compares CPS data with other data sources. The CPS is a cross-sectional survey while the SIPP is a longitudinal survey. Generally, the two surveys define income the same way except for a few types of interest, educational assistance, and lump-sum payments included in the SIPP but not in the CPS. Self-employment income is also defined and measured differently in the two surveys. The BEA produces personal income statistics mainly derived from business and government sources. The aggregates obtained from these sources are more complete than the data collected from household samples. Farm income data published by the Census Bureau are not directly comparable to data published by the U.S. Department of Agriculture. The income data published by the Census Bureau are also not directly comparable to tax return data because of IRS filing and reporting requirements and other factors.

U.S. Census Bureau. "Differences between the Income and Poverty Estimates from the American Community Survey and the Annual Social and Economic Supplement to the Current Population Survey." U.S. Census Bureau, August 19, 2004, www.census.gov/ hhes/income /factsheet081904.html.

This fact sheet outlines the differences between the ACS and the CPS. The ACS tracks cities and eventually even areas as small as census tracts while the CPS tracks only areas as small as states. Sample size for the ACS is about 3 million households while that for the CPS is about 100,000 households. The ACS is mandatory; the CPS is voluntary. The ACS includes the household and group quarters populations, whereas the CPS uses the civilian noninstitutionalized population. The ACS asks about income for the previous 12-month period, but the CPS asks about calendar year income. The ACS uses a series of 8 questions to ask about income, and the CPS asks about more than 50 sources of income. The ACS adjusts income estimates for inflation while the CPS does not.

U.S. Department of Commerce. *SIPP Quality Profile*. SIPP Working Paper Number 230, 3rd Edition. U.S. Census Bureau, 1998.

Wage and salary earnings are the main components of income. The SIPP estimate amounted to 91 percent of the independent NIPA estimate in 1984 and 92 percent in 1990. The CPS estimate was 97 percent of the NIPA benchmark in both years, although the number of earners estimated from the SIPP was higher than from the CPS in both years. The paper speculates that the CPS has an advantage over the SIPP because the latter is conducive to reports of "take-home pay." The SIPP and CPS self-employment estimates fall far short of the NIPA benchmark, but both are far greater than individual tax returns. It is difficult to compare the SIPP and CPS estimates because they use different concepts of the "draw" that people take to meet personal expenses. The SIPP estimates were lower than the CPS estimates. Evaluation studies have consistently shown that estimates of property income are particularly poor. Respondents have difficulty with the definition of terms. Much of the observed difference between the SIPP and CPS estimates results from different methods of collecting opposing data. The SIPP estimates generally exceeded the CPS estimates. The SIPP produced higher estimates of income from Social Security, railroad retirement, and SSI than the CPS and was close to benchmarks. Estimates for AFDC and other public assistance were well short of benchmark, although such income is often misclassified as general welfare. The SIPP estimates were about 84 percent of benchmark for unemployment income and 84 percent of benchmark for veterans' payments. The SIPP was superior to the CPS in estimating pensions in 1984, but apparently not in 1990. The SIPP also exceeded CPS estimates in child support and other sources of income.

U.S. Department of Housing and Urban Development. "American Housing Survey: A Quality Profile." Rameswar P. Chakrabarty, assisted by Georgina Torres. Current Housing Reports H121/95-1. Washington, DC: HUD, July 1996.

The AHS estimates of total income are lower than the independent estimates calculated from NIPA, the SSA, and the Veterans Administration. They are also lower than the CPS for every category other than self-employment income. The CPS is likewise lower than independent estimates but is closer than the AHS. The CPS may be closer to the independent estimates because of differences in income questions and the timing of both the CPS and AHS. The CPS asks more detailed and extensive questions about income sources and amount by source than does the AHS. Moreover, the CPS ASEC is administered during income tax season, when respondents are more aware of non-wage income such as interest, dividends, and the like.

The report also compares AHS and CPS poverty data between 1985 and 1993, noting three procedural differences between the surveys and subsequent impacts on poverty-level reporting. As of 1989, the AHS uses a set of monthly moving poverty thresholds based on 12 sets of poverty thresholds for the 12 months before the interview. The thresholds were intended to align the poverty cutoffs more closely with how income data are collected. However, the result of the procedural change has gone unmeasured. In 1993, the Census Bureau revised the non-wage income section of the AHS questionnaire in order to capture income sources commonly reported in the CPS but not previously specified in the AHS. The percent of households reporting non-wage income increased, but median non-wage income dropped between 1991 and 1993 from $7,400 to $6,212. Moreover, in 1993, the definition of lodgers in the household was expanded to include all persons not related to the

A.40

householder who paid rent or part of the household's housing costs. This question change has led to an increase in the percentage of households reporting rental income.

Vaughan, Denton R. "Reflections on the Income Estimates from the Initial Panel of the Survey of Income and Program Participation." *Studies in Income Distribution*, Social Security Administration, Office of Research and Statistics, SSA Pub. No. 73-11776 (17), May 1993.

This report reviews the quality of SIPP cross-sectional income estimates. The author draws nine principal conclusions. (1) The SIPP has achieved substantial gains over the CPS in the measurement of public and private transfers. (2) SIPP measures of wage and salary earnings are broadly similar to those available from the CPS, but evidence suggests that the SIPP (a) identifies more recipients who do not work full-time year-round and (b) presents a more valid representation of the population of full-time year-round wage and salary recipients. Accordingly, the SIPP improves the representation of the relationship between annual work experience and annual wage and salary earnings. (3) Clear evidence indicates that SIPP estimates of property income receipt are substantially more complete for the principal sources of property income than comparable CPS estimates. (4) The SIPP has materially reduced the impact of item nonresponse. As a result, the percentage of aggregate income attributable to imputation is approximately half that of the CPS. (5) The SIPP's subannual wage and salary amounts appear to be slightly biased relative to CPS estimates. (6) SIPP estimates of unemployment compensation show little if any improvement in completeness over CPS estimates. (7) While public assistance income is more fully reported in the SIPP than in the CPS, AFDC estimates still appear to be subject to misclassification. (8) Income from workers' compensation and associated sources remains underreported. (9) Property income aggregates remain well below independent estimates, especially interest income. The author suggests measures for improving SIPP income estimates.

Vaughan, Denton R. "Errors in Reporting Supplemental Security Income (SSI) Recipiency," in *Reports from the Site Research Test*, edited by Jan. Olson. U.S. Department of Health and Human Services, Office of the Assistant Secretary for Planning and Evaluation, Washington, DC, 1980.

As with AFDC, this study shows that misclassification was the principal reason for underestimating the prevalence of SSI in the site research test sample of known recipients. Recipients most frequently confused SSI with Social Security benefits, with mis-classification most likely to occur when a person received only SSI. Such individuals had higher-than-average payments and were most likely to be over age 65. Given the prevalence of dual recipients among the elderly population and dual recipients' smaller SSI benefits, surveys affected by the misclassification problem will produce biased estimates of the SSI population by age and underestimate benefits to a greater extent than will recipients themselves. Subsequent research shows a tendency for new entrants on the Social Security Disability Insurance (DI) rolls to misclassify their Social Security benefits as SSI because they often begin receiving benefits under SSI while awaiting Social Security DI benefits.

Vaughan, Denton R. "Errors in Reporting Supplemental Security Income Recipiency in a Pilot Household Survey." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, 1978, pp. 288-293.

This paper reports on the pilot test and measurement of SSI conducted in five cities in fall 1977. The author used program records to select a sample of SSI recipients living in four of the five cities. In order to be certain about the exact nature of the recipiency reporting errors in the survey, the author made a case-by-case comparison of the survey and administrative records for SSI sample members tagged as potential nonreporters. The comparison defined "true" SSI nonreporters as sample members included on the household roster but not identified in the survey as SSI recipients. Misclassified cases were defined as sample members not identified as SSI recipients on the questionnaire but reporting income from some other source in the amount of their actual SSI payment. The recipiency reporting error rate was about 13 percent; the nonreporting rate was less than 4 percent. The misclassification rate was slightly less than 10 percent. Therefore, the SSI income amount went completely unreported on the questionnaire in only about a quarter of the apparent nonreporter cases. About three-quarters of the nonreporter cases had SSI reported on the questionnaire as some other type of income. For misclassification errors, somewhat more than 80 percent were reported as one of three forms of Social Security.

Vaughan, Denton R. (with K. Goudreau and H. Oberheu). "An Assessment of the Quality of Survey Reports of Income from the Aid to Families with Dependent Children (AFDC) Program." *Journal of Business and Economic Statistics*, April 1984, pp. 179-186.

This multistate record check study establishes the importance of part-period recipients in the phenomenon of nonreporting of means-tested transfers with substantial turnover. It also deals with the impact of misclassification on reported aggregates for an important means-tested source. It shows substantial state-to-state variation in misclassifcation rates and demonstrates that the study—either directly or when misclassification was taken into account--captures approximately 90 percent of aggregate AFDC benefits received by the test sample.

Vaughan, Denton R. (with Bruce Klein). "Validating Recipiency Reporting of AFDC Mothers in a Pilot Household Survey," in *Reports from the Site Research Test*, edited by J. Olson. Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, 1980.

The paper shows the importance of misclassification errors affecting recipiency reports of AFDC in the site research tests of the SIPP development program. Using a set of matched cases of known AFDC cases drawn from program records, the author demonstrates that misclassification was more important than nonreports of recipiency in the overall underestimates of the number of AFDC cases.

Vaughan, Denton R. (with C. Whiteman and C. Lininger). "The Quality of Income and Program Data in the 1979 ISDP Research Panel: Some Preliminary Findings." *The Review of Public Data Use*, June 1984.

This paper establishes a pattern of SIPP measurement characteristics in comparison with the March CPS that was evident from the Income Survey Development Program (ISDP) Research Panel, which preceded the SIPP program. The ISDP Research Panel obtained almost universally greater recipiency estimates, especially for property income, and much lower item nonresponse rates for income recipiency and, to a lesser extent, amounts. At the source level, ISDP aggregates tended to exceed CPS aggregates, with the exception of wages, salaries, and interest income.

Vaughan, Denton R., Charles A. Lininger, Robert E. Klein. "Differentiating Veterans' Pensions and Compensation in the 1979 ISDP Panel." *Proceedings of the American Statistical Association, Section on Survey Research Methods*. Alexandria, VA: American Statistical Association, Washington, DC, 1983, pp. 191-196.

This study explores possible ways to differentiate veterans' pensions from compensation. It relies on attributes of each payment type (for example, disability ratings, death of a spouse while recipient was in the service, or death from a service-connected cause) to determine recipients' awareness that they are recipients of a pension or compensation payment. Later items added to the CPS and SIPP help directly identify pension recipients by asking whether the Department of Veterans Affairs required an individual to respond to an income questionnaire.

Waldo, Dan. Income and Asset Measurement in the Medicare Current Beneficiary Survey. Centers for Medicare and Medicaid Services, August 25, 2005.

This study uses multiple surveys to study the measurement of income for the MCBS and finds people reluctant to answer questions with respect to specific sources of income. As a result, the author imputes 30 percent of the total income data by using a hot deck imputation. The response was higher for those items prominent in lower incomes such as Social Security, SSI, and pensions but worse for items such as bonds, dividends, and interest. Overall, the paper finds that lower-income households report income more accurately when income comes from only a few sources. In addition, data from tax returns are more reliable for higher-income households but may not be accurate for lower-income households in that the latter may be exempt from filing. The MCBS staff now uses a hybrid income measure that features both survey and tax return data sets. The staff assigns each respondent a new income listed as a multiple of the poverty threshold based on the hybrid income measure.

Weinberg, Daniel H. "Income Data Quality Issues in the CPS." *Monthly Labor Review,* June 2006, vol. 129, no. 6, pp. 38-45.

This paper focuses on CPS ASEC questionnaire design, data collection, and data processing and suggests areas for improvement and issues for future research. CPS collects or imputes nearly all of the Canberra Group's components of income recommendation. In the actual

collection, response rates to the CPS are usually about 92 or 93 percent. Hot deck imputation is used for missing data. Some have said that the CPS underreports income, especially government transfers, property income, and self-employment. Another difficulty with the CPS lies in how to value non-cash income, especially employer-provided benefits such as health care. The article identifies the valuation of non-cash income as an area in need of more work and suggests collecting more information on other income sources such as fringe benefits and interhousehold transfers, reducing item nonresponse, and developing additional probes for income sources with notable misreporting.

Weinberg, Daniel H., Charles T. Nelson, Marc I. Roemer, and Edward J. Welniak, Jr. *The American Economic Review*, vol. 89, no. 2. Papers and Proceedings of the One Hundred Eleventh Annual Meeting of the American Economic Association, May 1999, pp. 18-22.

The U.S. Census Bureau has been computing income statistics annually since 1947. Until 1980, the Census Bureau gradually increased the number of income questions in the CPS from 2 to 11. Then, in 1980, the survey underwent a major overhaul and started to ask respondents about over 50 sources of income. The Census Bureau reports on 17 definitions of income based on various combinations of money income, benefits, and so forth. Wages, Social Security, SSI, veterans' payments, and pensions are all reliably estimated. Property income and unemployment have improved greatly as well. However, the remaining income sources such as military retirement, rents, and royalties have seen declines in accuracy of reporting as compared to benchmarks.

Wheaton, Laura L. "CPS Underreporting Across Time." Final Deliverable. Memorandum to Joan Turek and Reuben Snipper. The Urban Institute, March 5, 2007.

The work presented here updates and expands upon the research reported in Wheaton and Giannarelli (2000), which found substantial underreporting of transfer program income in the CPS. Recipients and benefits identified in the CPS and SIPP are compared to targets developed from administrative data for each of several transfer programs. The extent to which the CPS recipient and benefit data are allocated is also examined. The analysis of CPS data covers calendar years 1993 through 2004. The analysis of SIPP data covers calendar years 1997-1998 and 2001-2002. The memorandum concludes by showing the effects of TRIM3's correction for underreporting on poverty estimates for 2004.

Wheaton, Laura, and Linda Giannarelli. "Under-reporting of Means-Tested Income in the March CPS." *Proceedings of the American Statistical Association, Section on Social Statistics*. Alexandria, VA: American Statistical Association, 2000, pp. 236-241.

This paper examines the underreporting of means-tested transfer benefits in the CPS and finds a large decline in the portion of AFDC or TANF benefits captured by the CPS from 1993 to 1998. Possible reasons for the decline include confusion or stigma. Food stamp reporting has remained about constant. The amount of SSI reporting has fluctuated, though for no known reason. The Census Bureau uses allocations and imputations to attempt to account for underreporting. Another way to correct for underreporting is through

microsimulation, which steps through the CPS one household at a time, performing the same steps that a caseworker would perform in determining program eligibility and benefits for household members. The simulation captures 90 percent of AFDC/TANF benefit dollars and 94 percent of food stamp and SSI benefit dollars. Correction for underreporting of AFDC/TANF and SSI has a substantial effect on the estimated number of persons removed from poverty through means-tested cash transfers and an even greater effect on the estimated extent to which these programs reduce the poverty gap. The reduction in the poverty gap from these programs appears 36 percent higher after correction for underreporting. The reduction in the poverty gap from food stamp benefits appears 57 percent higher after correction for underreporting.

INDEX

Benchmark estimates

Atrostic and Kalenkoski 2002

Coder and Scoon-Rogers 1996

Meyer and Sullivan 2003

Meyer, Mok, and Sullivan 2007

Roemer 2000

Roemer 1999

U.S. Department of Commerce 1998

Vaughan 1993

Weinberg et al. 1999

Wheaton 2007

Wheaton and Giannarelli 2000

Comparisons of income estimates

Banthin and Selden 2006

Clark et al. 2003

Czajka, Mabli, and Cody 2008

Denmead and Turek 2005

Denmead, Turek and Adler 2003

Gouskova and Schoeni 2007

Martini 1997

Meyer, Mok, and Sullivan 2007

Nelson 2006

Posey, Welniak, and Nelson 2003

Roemer 2002

Roemer 2000

Susin 2003

U.S. Department of Commerce 1998

U.S. Department of Housing and Urban Development 1996

Comparisons of survey designs

Nelson 2006

Turek 2005

Turek, Denmead, and James 2004

U.S. Census Bureau 2007

U.S. Census Bureau 2004

Consumer Expenditure Survey

Bavier 2008

Ferraro and Paulin 1994

Garner and Blanciforti 1994

Henry and Day 2005

McGrath 2005

Meyer and Sullivan 2003

Paulin and Ferraro 1994

Paulin and Sweet 1996

Schwartz and Paulin 2000

Turek 2001

Correction for underreporting

Beebout 1977

Wheaton 2007

Wheaton and Giannarelli 2000

Current Population Survey

Alternative Measures of Income and Poverty [website]

Atrostic and Kalenkoski 2002

Banthin and Selden 2006

Bavier 2008

Bishaw and Stern 2006

Bound and Krueger 1991

Clark et al. 2003

Coder and Scoon-Rogers 1996

Czajka, Mabli, and Cody 2008

Davern et al. 2005

Davern et al. 2004

Denmead and Turek 2005

Denmead , Turek, and Adler 2003

Gouskova and Schoeni 2007

Henry and Day 2005

Hurd, Juster, and Smith 2003

Koenig 2003

Lamas, Tin, and Eargle 1994

Loomis and Rothgeb 2005

Martini 1997

Meyer and Sullivan 2003

Meyer, Mok, and Sullivan 2007

Moon and Juster 1995

Moyer 1998

Nelson 2006

Reichert and Kidelberger 2001

Roemer 2002

Roemer 2000

Roemer 1999

Ruser, Pilot, and Nelson 2004

Susin 2003

Turek 2005

Turek 2001

Turek, Denmead, and James 2004

U.S. Census Bureau 2005

U.S. Department of Commerce 1998

U.S. Department of Housing and Urban Development 1996

Vaughan 1993

Weinberg 2006

Weinberg et al. 1999

Welniak 1986

Wheaton 2007

Wheaton and Giannarelli 2000

Decennial Census

Alternative Measures of Income and Poverty [website]

Clark et al. 2003

Davern et al. 2004

Posey and Welniak 1998

Posey, Welniak, and Nelson 2003

Turek, Denmead, and James 2004

Event History Calendar

Kominski 1990

Health and Retirement Study

Heeringa, Hill, and Howell 1995

Hurd 1999

Hurd, Juster, and Smith 2003

Juster and Smith 1997

Kapteyn et al. 2006

Moon and Juster 1995

Historical data

Alternative Measures of Income and Poverty [website]

Imputation

Battaglia et al. 2002

Czajka, Mabli, and Cody 2008

Davern et al. 2004

Fisher (no date)

Juster and Smith 1997

Liu and Sharma 2002

Paulin and Ferraro 1994

Paulin and Sweet 1996

Schenker et al. 2006a

Schenker et al. 2006b

Income concepts

Bavier 2008

Canberra Group, The 2001

Hendrick et al. (no date)

Henry and Day 2005

Meyer and Sullivan 2003

Ruser, Pilot, and Nelson 2004

Smeeding and Weinberg 2001

U.S. Census Bureau 2005

Income Survey Development Program (see also Survey of Income and Program Participation)

Vaughan (with Whiteman and Lininger) 1984

Vaughan, Lininger, and Klein 1983

Item nonresponse

Atrostic and Kalenkoski 2002

Bruun and Moore 2005

Garner and Blanciforti 1994

Heeringa, Hill, and Howell 1995

Juster and Smith 1997

McGrath 2005

Moore and Loomis 2000

Nelson 2006

Paulin and Sweet 1996

Pleis and Dahlhamer 2004

Pleis and Dahlhamer 2003

Pleis, Dahlhamer, and Meyer 2006

Schwartz and Paulin 2000

Waldo 2005

Measurement of income

Banthin and Selden 2006

Coder 1988

Davern et al. 2005

Doyle (no date)

Doyle, Martin, and Moore 2000

Heeringa, Hill, and Howell 1995

Hess et al. 2000

Hurd 1999

Hurd, Juster, and Smith 2003

Juster and Smith 1997

Kominski 1991

Lamas, Palumbo, and Eargle (no date)

Loomis and Rothgeb 2005

Lynn et al. 2006

Marquis and Press 1999

Mathiowetz, Brown, and Bound 2002

Moore and Loomis 2000

Nelson and Doyle 1999

Patil and Russell 2000

Schwartz and Paulin 2000

Measurement error

Bound et al. 1994

Bound and Krueger 1991

Mathiowetz, Brown, and Bound 2002

Meyer and Sullivan 2003

Moore, Marquis, and Bogen 1996

Moore, Stinson, and Welniak 2000

Reichert and Kidelberger 2001

Medical Expenditure Panel Survey

Banthin and Selden 2006

Cohen and Machlin 1998

Cohen et al. 2000

Kashihara and Ezzati-Rice 2004

Turek 2005

Medicare Current Beneficiary Survey

Liu and Sharma 2002

Turek 2005

Waldo 2005

National Health Interview Survey

Denmead and Turek 2005

Denmead, Turek, and Adler 2003

Patil and Russell 2000

Meyer and Sullivan 2003

Meyer, Mok, and Sullivan 2007

Rodgers, Brown, and Duncan 1993

Poverty measurement

Bavier 2008

Grieger, Danziger, and Schoeni 2007

Moyer 1998

Nelson 2006

Turek 2001

Turek, Denmead, and James 2004

Reference period

Hurd, Juster, and Smith 2003

Posey and Welniak 1998

Posey, Welniak, and Nelson 2003

Short 1990

Susin 2003

Turek, Denmead, and James 2004

Reporting accuracy, based on matched administrative records

Bates and Pedace 2000

Bound et al. 1994

Bound and Krueger 1991

Duncan and Hill

Hendrick et al. (no date)

Huyhn, Rupp, and Sears 2001

A.56

Koenig 2003

Marquis and Moore (no date)

Olson 2001

Pedace and Bates 2000

Rodgers, Brown, and Duncan 1993

Roemer 2002

Scoon-Rogers 2005

Sears, Rupp, and Koenig 2003

Vaughan 1980

Vaughan 1978

Vaughan (with Goudreau and Oberheu) 1984

Vaughan (with Klein) 1980

Survey of Income and Education

Beebout 1977

Survey of Income and Program Participation (see also Income Survey Development Program)

Bates and Pedace 2000

Bavier 2008

Bruun and Moore 2005

Coder 1988

Coder and Scoon-Rogers 1996

Czajka, Mabli, and Cody 2008

Denmead and Turek 2005

Doyle (no date)

Doyle, Martin, and Moore 2000

Fisher (no date)

Hendrick et al. (no date)

Huyhn, Rupp, and Sears 2001

Kalton and Miller 1991

Koenig 2003

Kominski 1991

Kominski 1990

Lamas, Palumbo, and Eargle (no date)

Lamas, Tin, and Eargle 1994

Mack and Petroni 1994

Marquis and Moore (no date)

Martini 1997

Meyer, Mok, and Sullivan 2007

Moore, Marquis, and Bogen 1996

Nelson and Doyle 1999

Olson 2001

Pedace and Bates 2000

Roemer 2002

Roemer 2000

Scoon-Rogers 2005

Sears, Rupp, and Koenig 2003

Turek 2005

U.S. Census Bureau 2005

U.S. Department of Commerce 1998

Vaughan 1993

Wheaton 2007

APPENDIX B

QUESTIONNAIRES, DATA DICTIONARIES AND DOCUMENTATION

QUESTIONNAIRES, DATA DICTIONARIES AND DOCUMENTATION

This Appendix provides internet addresses for detailed documentation and other information for each survey used in the study—the 2001 SIPP panel, 2002 files for ACS and MEPS, 2003 files for CPS (ASEC), NHIS, MCBS and PSID, and 2004 files for HRS.

The Appendix first lists main web sites, then internet addresses for questionnaires, data dictionaries, file- and year-specific documentation, information on survey design including sampling and weights, and last, other documentation. However, some information on MCBS, as well as any information on internal NHIS files, is available only after a research plan has been approved. NHIS information listed below pertains to the public use files that have income brackets, not income amounts.

MAIN WEB SITES

SIPP	<http://www.census.gov/sipp/>
CPS	<http://www.census.gov/cps/>
ACS	<http://www.census.gov/acs/www/>
MEPS	<http://www.meps.ahrq.gov/mepsweb/>
NHIS	<http://www.cdc.gov/nchs/nhis.htm>
MCBS	<http://www.cms.hhs.gov/MCBS/>
HRS	<http://hrsonline.isr.umich.edu/> and <http://www.rand.org/labor/aging/dataprod/#randhrs
PSID	<http://psidonline.isr.umich.edu/>

QUESTIONNAIRES

SIPP	In documentation for each interview core component and each interview topical module component at <http://www.census.gov/apsd/techdoc/sipp/sipp.html>. Questionnaire sections are also at <http://www.census.gov/sipp/core_content/2001/quests/wave1.html> and <http://www.census.gov/sipp/core_content/2001/quests/wave2.html> for the first and subsequent core interviews and at <http://www.census.gov/sipp/top_mod/2001/top_mod_sched.html> for topical modules.
CPS	In documentation at <http://www.census.gov/apsd/techdoc/cps/cpsmar03.pdf>

ACS	<http://www.census.gov/acs/www/Downloads/SQuest.pdf>
MEPS	The 42 questionnaire sections for 2002 can be accessed at <http://www.meps.ahrq.gov/mepsweb/survey_comp/survey.jsp#Questionnaires>. The questionnaire for the income data on the 2002 Full Year file (asked in 2003) is at <http://www.meps.ahrq.gov/mepsweb/survey_comp/hc_survey/2002/IN65.pdf>
NHIS	<ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2003/qhoushld.pdf> and <ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Survey_Questionnaires/NHIS/2003/qfamilyx.pdf>
MCBS	The 32 sections of the baseline and core community questionnaires for 2003 are at <http://www.cms.hhs.gov/MCBS/Ques/list.asp#TopOfPage> after entering "2003" as Published Year
HRS	The 21 questionnaires for 2004 can be accessed at <http://hrsonline.isr.umich.edu/meta/sho_meta.php?hfyle=qnaires>
PSID	<ftp://ftp.isr.umich.edu/pub/src/psid/questionnaires/q2003.pdf>

DATA DICTIONARIES

SIPP	In documentation for each interview core component and each interview topical module component at <http://www.census.gov/apsd/techdoc/sipp/sipp.html>. Also at <http://www.census.gov/sipp/dictionaries/2001_l01puw1d.txt> for core questions including income and at <http://www.census.gov/sipp/diction.html> for topical modules.
CPS	In documentation at <http://www.census.gov/apsd/techdoc/cps/cpsmar03.pdf> and at <ftp://www.bls.census.gov/pub/cps/march/cpsmar03dd.txt>
ACS	<http://www.census.gov/acs/www/Downloads/DataDict.pdf>
MEPS	<http://www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h70/h70cb.pdf>
NHIS	<ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2003/Househld.pdf> <ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2003/Familyxx.pdf> and <ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2003/Personsx.pdf>
MCBS	The 23 data dictionaries for the 2003 Cost and Use files are at <http://www.cms.hhs.gov/MCBS/CDFS/list.asp#TopOfPage> after entering "2003" as Published Year and "Cost and Use" as Type
HRS	The 37 data dictionaries for the 2004 HRS files and the data dictionary for 2004 imputations are at <http://hrsonline.isr.umich.edu/meta/2004/core/codebook/h04_00.html> and <http://hrsonline.isr.umich.edu/meta/2004/impute/codebook/h2004icb.txt> Data dictionary for the RAND person-based flat file and concordance of HRS variables at <http://www.rand.org/labor/aging/dataprod/randhrsg.pdf>
PSID	Provided as part of data download of all or selected variables from year-specific family and individual files and cross-year individual file, and requires registration first at <https://simba.isr.umich.edu/U/Login.aspx> then selection of file type(s), year(s) and variables at <https://simba.isr.umich.edu/> or use of a topical index by file and year at <https://simba.isr.umich.edu/VS/i.aspx> or use of a codebook search for key words by file and year at <https://simba.isr.umich.edu/VS/s.aspx>. All methods allow the option of downloading only a customized data dictionary without downloading data

DOCUMENTATION

SIPP	SIPP Users Guide 3rd Edition has not been updated but is mostly applicable to the 2001 panel, at <http://www.census.gov/sipp/usrguide/sipp2001.pdf>
	Documentation for each interview core component and each interview topical module component of the 2001 panel is at <http://www.census.gov/apsd/techdoc/sipp/sipp.html>.
	Documentation includes questionnaires and data dictionaries as well. Additional information is available on-line at <http://www.census.gov/sipp/technical.html>.
CPS	<http://www.census.gov/apsd/techdoc/cps/cpsmar03.pdf>. Documentation includes questionnaire and data dictionary as well.
ACS	Code lists and top codes but no separate documentation for public use files. See <http://www.census.gov/acs/www/Products/PUMS/codelist2002.html> and <http://www.census.gov/acs/www/Products/PUMS/C2SS/minmaxval2.htm>
MEPS	<http://www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h70/h70doc.pdf>
NHIS	<ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/Dataset_Documentation/NHIS/2003/Srvydesc.pdf>
MCBS	Provided with data
HRS	<http://hrsonline.isr.umich.edu/meta/2004/core/desc/h04dd.pdf> and RAND documentation at <http://www.rand.org/labor/aging/dataprod/randhrsg.pdf>
PSID	<http://psidonline.isr.umich.edu/data/Documentation/Fam/2003/doc.txt> and see also <http://psidonline.isr.umich.edu/Guide/Overview.html> and <http://psidonline.isr.umich.edu/Guide/FAQ.aspx>

DESIGN, SAMPLING AND WEIGHTS

SIPP	Information on variance estimation and weights is available on-line at <http://www.census.gov/sipp/sam_and_wt.html>. General design information is contained in the SIPP Users Guide 3rd Edition, at <http://www.census.gov/sipp/usrguide/sipp2001.pdf> but has not been updated to the 2001 panel
CPS	Technical Paper 66: Design and Methodology at <http://www.census.gov/prod/2006pubs/tp-6.pdf>
ACS	Technical Paper 67: Design and Methodology at <http://www.census.gov/acs/www/Downloads/tp67.pdf>
MEPS	Methodology Report: Sample Design of the Medical Expenditure Panel Survey Household Component, 1998-2007 at <http://www.meps.ahrq.gov/mepsweb/data_files/publications/mr22/mr22.pdf>
NHIS	Data Evaluation and Methods Research No. 130 at <http://www.cdc.gov/nchs/data/series/sr_02/sr02_130.pdf>
MCBS	Provided with data
HRS	<http://hrsonline.isr.umich.edu/intro/sho_uinfo.php?hfyle=sample_new_v3&xtyp=2> and <http://hrsonline.isr.umich.edu/meta/tracker/desc/wghtdoc.pdf>

PSID	For original sample see <http://psidonline.isr.umich.edu/Publications/Papers/tsp/1996-03_Notes_on_the_SEO_C_Brown.pdf>. For current weights see <http://psidonline.isr.umich.edu/Data/weights/Long-weights-doc.pdf> and <http://psidonline.isr.umich.edu/Data/weights/psidweights.pdf>

OTHER DOCUMENTATION

ACS	Accuracy statement and glossary at <http://www.census.gov/acs/www/Downloads/2002/AccuracyPUMS.pdf>, and <http://www.census.gov/acs/www/Downloads/2002/usedata/Subject_Definitions.pdf>
MEPS	The 2002 JOBS file data dictionary and documentation are at <http://www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h63/h63cb.pdf> and <http://www.meps.ahrq.gov/mepsweb/data_stats/download_data/pufs/h63/h63doc.pdf>
NHIS	Description of imputations at <http://www.cdc.gov/nchs/data/series/sr_02/sr02_130.pdf>
PSID	Interviewer instructions at <http://psidonline.isr.umich.edu/data/Documentation/Fam/2003/QxQs.pdf> and Family Identification and Mapping System (FIMS) at <http://simba.isr.umich.edu/FIMS/>