Income Measurement in Federal Surveys

John L. Czajka*
Gabrielle Denmead**

*Mathematica Policy Research
jczajka@mathematica-mpr.com

**Denmead Services & Consulting
gabrielle.denmead@verizon.net

Introduction

Income is a critical variable in policy-related analysis and public program design. Major federal and state programs address the consequences of inadequate resources and the needs resulting from or correlated with low income, and income and poverty status often play key roles in the development of public policy. As a result, regardless of their primary purpose, most major household surveys collect some income data and provide measures of poverty status. Nonetheless, income is difficult to measure well in household surveys. Furthermore, poverty status depends not only on income but on how the family is defined, which differs across surveys. Despite many similarities, there are significant differences among the income and poverty measures in different surveys. More importantly, various federal surveys provide markedly differing estimates of income and poverty even when the same concepts are applied.

A comprehensive and systematic assessment of the income data and their utility for policy-related analyses in eight major surveys was conducted by Mathematica Policy Research, Inc. and its subcontractor, Denmead Services & Consulting, under a contract with the Office of the Assistant Secretary for Planning and Evaluation in the Department of Health and Human Services. The surveys are all federal or federally-sponsored: the Survey of Income and Program Participation (SIPP); the Annual Social and Economic Supplement (ASEC) 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 family core (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 was 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

This paper summarizes key findings on income measurement in the five federal surveys of the general population: the CPS, ACS, SIPP, MEPS, and NHIS. For all surveys, the analysis was restricted to the civilian noninstitutional population, and the income reference period conforms as closely as possible to calendar year 2002.

We begin by discussing differences among surveys, the development of comparable income estimates across surveys, and some limitations of the survey comparisons. Next we compare survey estimates of income, income by quintile, and earned and unearned income aggregates and quintiles. Following that we compare survey estimates of poverty rates, the number and composition of the poor and near poor, and the impact of family definition on measured poverty. Next we present comparative findings on several aspects of data quality, including overall response rates, allocation, internal consistency, and rounding. Finally, we discuss some implications for users of income data from federal surveys.

Development of Comparable Estimates Across Surveys

The assessment's empirical methodology was designed to yield, to the extent feasible, estimates of income with 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 U.S. All analyses in the study use income data for calendar year 2002, except for the rolling reference period in ACS that spans 23 instead of 12 months. Also, since various surveys apply the same term to somewhat different concepts or measures, and may have different names for the same concepts or measures, the terminology used in the study and in this paper has been standardized to Census Bureau and CPS definitions. Since NHIS public use files have no dollar amounts, only brackets, all NHIS tabulations were performed on the internal file, on-site.

Adjustments to Enhance Comparability

Our estimates of income focus on calendar year 2002, but the surveys all differ in the point in time at which they collect income information. We did not attempt to adjust for differences due to differences in survey timing, although the study on which this paper is based explored the impact of survey timing on poverty rates in the light of dynamic changes in family composition, and also on response rates. We did adjust for universe differences that arose from differential treatment of specific subpopulations. 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. Since the ACS in 2002 had not yet been extended to include group quarters, those living in group quarters and included in other surveys could not be included in 2002 ACS tabulations. ACS data for 2006, the first year covering group quarters, include 2.3 million students residing in college or university dormitories, treated as unrelated individuals, and excluded from ACS poverty estimates.

In conformity with CPS income definitions, lump sums and irregular payments were removed where included in survey income—$12.7 billion from SIPP and $65.6 billion from MEPS. Missing components of income that could not be added include educational benefits in SIPP, and tax exempt interest for tax filers in MEPS. 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 other tax-advantaged retirement accounts, contributions to Medical Savings Accounts, and 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.

Official poverty statistics incorporate the definition of a family that is used in the CPS—two or more persons related by blood, marriage, or adoption. The NHIS uses a broader family definition that includes unmarried partners and their relatives, and foster children. Those NHIS families that included unmarried partners or foster children were split into two or more CPS families, and the income of the original family was apportioned among the new families. MEPS uses both the CPS and the broader NHIS family concepts but identifies CPS families on the file; the study and this paper use MEPS income data for the CPS-type families.

Limitations on Comparability

While the goal of these efforts was to make the survey estimates of income as comparable as feasible, there remain a number of differences due to universe definition, survey design, or methodological features for which adjustments were not possible. While the assessment included a number of special studies of the impact of these design and methodological features, these are not discussed in this paper.

First, our adjustments do not compensate for the fact that the surveys represent populations at different times. Family composition is dynamic, and the people living together as a family at the time the data were collected and whose income is summed to determine family income may not have lived together for the entire income reference year. Conversely, other individuals may have lived with the family for some or all of the income reference year but if they are no longer part of the family at the time data are collected, their income is not included in family income. The longer the lag between family composition and the income reference period, the more changes in family composition will occur, with possible bias. 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, on average.[1] 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.

---

[1] CPS ASEC interviews are conducted in February, March, and April in the week that includes the 19th.

Second, our adjustments do not correct for the rolling reference period in ACS that spans 23 months instead of a calendar year. Other surveys represent income during the 12 months of a calendar year, with each month equally represented. In the ACS, respondents are asked to report their income for the past 12 months. For persons completing the survey at the beginning of the year, the past 12 months are January through December of the prior year; for persons completing the survey at the end of the year, the past 12 months are December of the prior year through November of the current year. Thus the 23 months spanned by the rolling reference period are unequally represented. Income data for an ACS "year" are centered on December of the prior year, with half of the income attributable to the seven- month interval from September of the prior year through March of the survey year. The remaining half of income for an ACS "year" is attributable to the 16 months of January through August of the prior year and April through November of the survey year.[2]

Third, we could not compensate for or remove the unique income and poverty-related post-stratification used in generating MEPS weights. All five surveys calculate weights based on selection probabilities adjusted for non-response, then post-stratify to CPS or other Census Bureau control totals by age, race/ethnicity, and sex of the person, family head, or family reference person, and sometimes other information such as household type, or geographic location. MEPS additionally 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.

Fourth, the adjustments do not correct for the differential treatment of college students living away from home while attending school. SIPP, CPS and MEPS include college students living away from home in the parental family. NHIS treats college students living away from home as unrelated individuals, and includes them as such in poverty estimates. The 2002 ACS did not interview in dormitories (group quarters), and also treated any students living in off-campus housing as unrelated individuals. These differences affect the size of the ACS population and the composition of ACS and NHIS families.

Fifth, our adjustments do not compensate for differences in the timing or source of the population controls that were applied or how they were applied. 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, then combined to create a single annual weight on the public use file, with an effective reference date of mid-June 2003. Also, some surveys use Census Bureau population estimates directly as control totals; others use CPS population estimates, or perform their own calculations of control totals based on CPS data. These alternative controls do not agree completely.

Sixth, since the NHIS collects only a family-level income amount and person-level earnings, creating CPS families in the NHIS required assumptions about the allocation of family income other than earnings among the new families when a non-CPS family was split. Lastly, no attempt was made to address differences in how respondents interpreted what they were to include or not include in the income they reported.

Income

In comparing estimates of income across surveys, we look first at total income and income by quintile. We then divide total income into earned and unearned income.

Total Income

As a summary statistic, total or aggregate income is appealing for its simplicity and its use of all the income data in each survey. However, aggregate income 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. Table 1 presents estimates of aggregate income for the whole population and by

---

[2] Income reported in ACS published data and on-line tables that are based on internal files is adjusted across the rolling reference period to the same real dollars, based on the Consumer Price Index (CPI). Income in ACS public use files is not adjusted for inflation, although an average inflation factor is provided.

quintile of family income for the five surveys.[3] Table 1 also shows these amounts as a percentage of the corresponding amounts for the CPS. There is no gold standard for estimates of income, nor do we mean to suggest that the CPS estimates are the best. But the CPS is the official source of household income and poverty statistics for the U.S. and 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, while estimates from the other three surveys lie within two to five percent of the CPS. Since aggregates in the top quintile may be affected by outliers and by differences in survey topcoding practices, we summed the survey aggregates through the four lower quintiles. 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 six percent below the CPS.

When income is examined by quintile, we find that SIPP obtains the largest amount of income for the lowest quintile, at 105.6 percent of the CPS total, but this apparent success erodes noticeably above the lowest quintile. In the second quintile, SIPP collects 97 percent of CPS income, dropping 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.

The similarity of aggregate income across surveys with a vast range in the number and detail of income questions raises a fundamental issue for the collection of income data – how much do question number and detail matter? With a single question NHIS captures 95 percent as much total income as the CPS; with a simple instrument primarily filled out by respondents rather than a trained interviewer, the ACS approximates the CPS more closely than any other survey. SIPP, with far more income questions than the other four surveys, captures 11 percent less total income than the CPS and 6 percent less than the NHIS's single question. It is not possible to determine what impact the post-stratification of MEPS to match CPS poverty rates has on MEPS estimates of aggregate income, but MEPS, with its SIPP-like panel design, might yield SIPP-like income estimates in the absence of this post-stratification, or alternatively, the use of retrospective annual versus monthly income questions may trump the panel design.

Earned and Unearned Income
Earnings (wages and salaries plus self-employment income) are the dominant income source in the U.S., and account for 82 to 86 percent of total aggregate income. Employment is a major area of policy concern, as a source of self-support and of health insurance coverage. For both these reasons, accuracy in its measurement is key. Table 2 presents estimates of earned income and unearned income, showing these amounts as shares of total income by survey and as percentages of the corresponding amounts for the CPS.

For SIPP, CPS and ACS, earned income accounts for 82.1 to 82.8 percent of total income, with MEPS slightly higher at 84.1 percent and NHIS highest at 86.0 percent. The identical share of earnings in total income for CPS and SIPP is particularly striking, given 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 must be similarly lower in SIPP as compared to CPS, which is confirmed in the bottom panel of the table. The NHIS does not collect data on unearned income, but the difference between total income and total earnings for the NHIS suggests an amount of unearned income that is only 77 percent of the CPS total. This implied shortfall may indicate 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, or may be explained by inconsistencies between earnings and total income among NHIS families, discussed below.

There are much larger differences among survey counts of persons self-employed or working for others and among amounts earned from wages and salaries and from self-employment, than among numbers of earners and total or average earnings. SIPP's higher estimates for self-employment help offset its apparent understatement of wage and salary income; SIPP captures only 82.4 percent as much aggregate wage and salary income as the CPS, but has 188 percent as much self-employment income. The ACS captures four percent less wage and salary income than the CPS but 19 percent more self-

---

[3] 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.

employment income. (MEPS is not discussed due to classification issues that prevent accurate separation of wage and salary and self-employment earnings.)

Table 3 compares survey estimates of earned income by quintile of family income, and Table 4 presents the same comparisons for unearned income. The ACS, SIPP, MEPS, and NHIS all find more earnings in the lowest quintile of family income than does the CPS, but the aggregate amounts are small. The ACS and SIPP find progressively less total earnings relative to the CPS as we move up the quintiles. For MEPS this is true after the first quintile. The NHIS, on the other hand, finds progressively more aggregate earnings relative to the CPS in quintiles two through four. By contrast, there are few consistent patterns for unearned income as one moves up the quintiles. The ACS finds progressively more unearned income than the CPS for higher quintiles, with 23 percent more in the top quintile. SIPP has 99.6 percent as much unearned income as the CPS through the first four quintiles but 35 percent less in the top quintile. MEPS, on the other hand, falls short of the CPS in all quintiles. In the NHIS, unearned income is a residual rather than a reported amount, as already noted, and is erratic. The difference between aggregate total and aggregate earned income in NHIS is as low as 60 percent of the CPS aggregate in one quintile and as high as 88 percent in an adjacent quintile.

Poverty

The income measures most often employed in policy-related analyses concern the lower end of the income distribution, and most frequently are classifications that compare family income to the family-size adjusted income amounts that comprise the set of official poverty thresholds. The poor are those persons with family incomes below these thresholds, and the poverty rate is the percentage they constitute of all persons in the poverty universe of each survey.[4] The composition of the poor and near poor (whom we define as those with family income between 100 and 200 percent of the poverty threshold)[5] and the poverty rates for specific demographic groups such as children or the elderly may also be of great interest. Marked differences across surveys in estimates of the poor and near poor would usually raise questions about income data quality, reliability and utility for policy-related research.

Estimates of the Poor and Near Poor
Table 5 presents total numbers of poor and near poor, and poverty rates, for all ages. The range of estimates—lowest to highest—of the number of poor and the poverty rate is 8.4 million people and 2.9 percentage points. SIPP obtains the lowest poverty rate among the five surveys at 11.8 percent, based on an estimate of 33.2 million poor. The CPS, ACS, and MEPS cluster very close to each other and slightly above SIPP with poverty rates between 12.2 percent and 12.5 percent. As we have noted, the MEPS sample weights 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.[6] 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 for persons in comparable CPS-like families. 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 the other four 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 NHIS by 2.3 million.

---

[4] CPS and SIPP exclude unrelated children under 15 from the poverty universe but not the survey universe. MEPS and NHIS exclude unrelated minors (usually under 18) from both survey and poverty universes. ACS excludes unrelated children under 15 from the poverty but not the survey universe, including those that would be in unrelated subfamilies in CPS or SIPP. After group quarters were added to ACS in 2006, students residing in college or university dormitories are excluded from the ACS poverty universe but not the survey universe

[5] There is no standard definition of "near poor". Historically, the term referred to persons between 100 and 125 percent of poverty. We use it for those with low family income, below 200 percent of poverty, but not poor.

[6] 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.

When the low-income population of both poor and near poor is examined, SIPP is higher than all but NHIS in the 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 measures while the ACS is lower. NHIS finds 33.7 percent of the population or 95.5 million persons to be low-income, when a comparable CPS family definition is used. The count of low-income persons in the NHIS is 6.0 million higher than in the SIPP and 9.3 million higher than in the CPS.

Estimates of Poor and Near-Poor Children and Elderly
Table 6 presents numbers of poor and near-poor children under 18, and their poverty rates. SIPP's comparatively high estimates 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. 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. 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.

Table 7 presents numbers of poor and near poor elderly age 65 or over, and their poverty rates. SIPP's comparatively high estimates of low-income persons in the general population do not extend to the elderly. SIPP finds fewer low-income elderly than the CPS, MEPS, or NHIS at 11.6 million compared to 12.9 to 13.6 million, or 34.1 percent compared to 37.6 to 39.7 percent. The ACS finds the fewest low-income elderly at 11.2 million or 33.3 percent, but SIPP finds the fewest poor elderly at 3.0 million and the lowest poverty rate for the elderly at 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.

Impact of the Family Definition
Some surveys utilize family definitions that deviate from the CPS family concept incorporated into the official measure of poverty in the U.S., and surveys also differ in the level of detail with which they capture family relationships. We found that broadening the family concept as compared to the CPS measure produces major changes in family income and poverty rates. Lack of relationship information among persons unrelated to the household head, as is the case for the ACS, also affects family income and poverty rates.

The response unit in the NHIS is the family, and families are defined to include unmarried partners of the same or opposite sex, relatives of 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 their relatives including children, and foster children, from the NHIS family. We assigned each person his or her reported earnings, then apportioned family income over and above individual earnings (if any) among the two or more CPS-type family units into which these NHIS families were split. By comparing the income and poverty estimates that we prepared using the CPS family definition with estimates obtained from the original data, we assessed the impact of using the NHIS rather than the CPS family definition to combine individuals into families for the purposes of estimating family income, numbers of poor, and poverty rates.

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 MEPS sample weights to the CPS poverty distribution, the Agency for Healthcare Research and Quality (or its MEPS survey contractor) identifies 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 earlier, 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 did with

the NHIS. This process did not require us to make any assumptions in order to apportion income, since MEPS collects person-level income data.

Table 8 shows our estimates of the impact of the broader NHIS family definition for both the NHIS and MEPS. The results for the two surveys are remarkably similar. In both surveys we find that the NHIS family definition reduces the number of poor by 2.6 million and reduces the poverty rate by 0.9 percentage points. 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.

While not an issue of family definition, the limited relationship information in the ACS affects ACS estimates of the number of poor and the poverty rate. ACS determines relationship only to the household head and cannot identify unrelated subfamilies. As a result, members of these families, usually a single parent with one or more never-married children, are all treated in the ACS as unrelated individuals regardless of age. While one might expect this treatment to increase the number of poor and the poverty rate, it probably has the opposite effect. Examination of unrelated subfamilies in the CPS showed that almost half of the unrelated individuals that would be created if unrelated subfamilies were not identified were under 15 and would be dropped from the conventional poverty universe, and thus not counted as poor. Given the high poverty rate found for unrelated subfamilies in the CPS, the reduction in poor children under 15 due to eliminating them from the poverty universe more than offsets the increase in poor age 15 or older resulting from treating them as unrelated individuals.

Aspects of Data Quality

The assessment of income data was specifically focused on the quality and utility of income and poverty data for policy-related analysis. A large number of factors affect data quality, including overall survey response rates, item non-response and allocation rates, the use or absence of internal consistency edits, and heaping or rounding. Policy analyses may be particularly sensitive to any upward or downward bias in income data, and to inconsistencies that could significantly change estimates of the poor and near poor.

Non-Response and Allocation
Non-response to household surveys is a serious issue, since high rates of refusal or overall survey non-response may lead to non-response bias. For longitudinal surveys, attrition is a lesser issue than initial non-response, because earlier interviews provide information with which to complete or allocate income amounts and to measure and adjust for any attrition bias. But for income data, item non-response is also a serious problem. Income questions generally yield some of the highest rates of missing values,[7] and significant fractions of income data are in fact allocated or created during processing of the raw data.

Table 9 presents initial survey response rates across surveys, and 2002 income data response rates after attrition for the two panel surveys. The highest initial response rate, over 97 percent, is for the only mandatory survey, the ACS. The lowest initial response rate, 70 percent, is for MEPS, which piggy-backs on the NHIS sample and only contacts NHIS respondents. SIPP and the NHIS have initial response rates of 88 and 89 percent; the CPS has a response of 92 percent for the underlying monthly labor force survey, but about 11 percent of persons with income in the CPS are whole imputes who have refused to answer the ASEC supplement. For the two panel surveys, the overall survey response after attrition for the interviews providing income data for 2002 was 72 percent for the SIPP and 65 percent for MEPS.

Table 9 also presents measures of the fraction of total income, in dollars, for which missing values were filled in or allocated, by any method, and further indicates the proportions allocated with and without 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). Allocation methods without

---

[7] For surveys for which income data is not a high priority, this can become a rationale for limiting the length and detail of income questions or relegating them to the end of interview. The latter minimizes impacts on other questions or early break-off of interviews, but may further reduce income question responses rates and adversely affect response quality, as many researchers have noted appears to be the case with CPS health insurance questions.

the use of partial information could be "hot deck"[8] imputations (CPS, ACS, SIPP and MEPS) or regression methods (NHIS and MEPS). Arguably, allocations that make use of partial information are qualitatively different from allocations that rely solely on covariates of the missing items.

The percent of total dollars allocated ranges from 18 percent in the ACS to 43 percent in MEPS, with the CPS, SIPP and the NHIS all having 32 to 34 percent of total dollars allocated. 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. 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. In the NHIS, allocations with partial information represent only 2 percent of total income, and allocations without partial information account for 30 percent of total income.[9] Neither CPS nor ACS had allocations classified as using partial information, leaving 34 percent of total income in the CPS and 18 percent in the ACS as allocated without partial information. However, lacking a direct quality measure for allocated data with 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.

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 questionnaire design, or through consistency edits during processing, and most of the surveys, like the CPS, do such edits or consistency checks. However, there are some exceptions.

NHIS asks total income for families, and earnings (never negative) for persons, but does not determine whether the sum of earnings for persons in the family (family earnings) exceeds the family's income. As shown in Table 10, 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. 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.

The lack of agreement between family earnings and income in the NHIS has dramatic consequences for numbers of poor and the poverty rate. When higher family earnings were substituted for family income, the poverty rate was reduced by 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 improved 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.

Other internal inconsistencies involve earnings and work activity in MEPS and the NHIS. CPS and ACS include consistency checks in data processing to ensure there is income associated with all work activity, and work activity associated with all reported earnings. But in NHIS, 4.3 million persons reported receipt of earnings but have no work activity or earnings in the same year, and another 4.0 million persons reported working, and amounts earned, but no receipt of earnings. Similarly, in MEPS, 6.6 million persons reported earned income for the year but no work activity on the detailed JOBS file of employment, and 2.6 million persons reported details of one or more jobs during the year but no earned income for the same time period. 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.

---

[8] Hot deck methods involve matching records with missing values to "donor" records based on a typically large number of characteristics, then assigning values from the donor records. Using other respondents as donors helps ensure that the allocated values are plausible and have an appropriate distribution. Often, multiple variables may be assigned from the same donor to ensure that there is some internal consistency among the allocated values.

[9] Revisions to the NHIS income questions for the 2007 survey have increased the percentage of respondents providing brackets.

Rounding

We cannot directly assess the accuracy of survey responses to income questions, but round numbers suggest inexact reporting or approximation of actual income amounts. Respondent rounding distorts the distribution of income by creating artificial spikes or heaping at the rounded values, and in fact, rounding is an accepted way to protect confidentiality in public use files.[10] Further, distortions due to rounding in reporting will be exaggerated by hot-deck allocations that carry rounding over from donor records. We tested for rounding by determining the percent of persons with reported, unallocated, income amounts exactly divisible by $5,000 or $10,000 up to $52,500, which tests the frequency of rounding to $50,000[11] or less. The test is most reliable for the NHIS and ACS, where the amounts tested were always responses to single questions. It is least applicable to the SIPP, where all annual amounts are built up from monthly values.

Table 11 presents the percent of persons with reported (not allocated) earnings, wages and salaries, Social Security benefits, other retirement income, total person-level income, or total family income amounts exactly divisible by $5,000 or $10,000, across the five surveys. In SIPP, with detailed income questions and monthly data, virtually no one has rounded income amounts. 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. 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. MEPS has less rounding—19 percent of earners report amounts divisible by $5,000, and 10 percent report amounts divisible by $10,000. In contrast to earnings, Social Security and retirement income have little rounding—less than 10 percent of recipients of either source reported amounts divisible by $5,000 in CPS, SIPP, ACS or MEPS, although the Social Security amounts would almost always be a response to a single question in CPS, ACS or MEPS.

Implications

Our findings from this comparison of income estimates across federal surveys have several implications for users of these data. Income and poverty as measured in different surveys are not equivalent, and there are also important differences in the estimated composition and distribution of income among surveys, as well as in the detail available on sources of income. Some surveys have high survey non-response rates, raising the possibility of non-response bias. All but ACS have high income allocation rates, but little is known about the quality of alternative approaches to allocation, and even less about the prevalence of rounding of reported incomes. It would seem that the quality and utility of income data for policy-related analyses varies substantially across surveys, and this is likely to reflect the different levels of importance assigned income data and its low priority for some surveys. Impacts of differences may be greatest for uses that involve detailed simulations of program eligibility, particularly when evaluating reform options that alter eligibility thresholds and benefit formulas. As a result, no one survey is ideal for policy use.

At one extreme, the NHIS obtains family income only for families more broadly defined than in official statistics, has no person-level information with which to compute poverty rates consistent with official statistics or to test alternative filing units and eligibility, and has internal inconsistencies that changed the count of poor by four million persons. In addition, income information on the NHIS public use file is limited to $5,000- and $10,000-wide brackets, and policy users are unlikely to have the time or flexibility to get an analytic plan approved in order to use the internal file on-site, as was done by the assessment study whose data are used in this paper.

While SIPP was specifically designed for policy use and has strong income and program participation information for the lower half of the income distribution, data quality overall has degraded in the last decade due to budget contractions, the files as currently structured are difficult to use, and changing family composition in conjunction with monthly income reporting means that only part-year income information is available in many cases.

---

[10] Limiting the number of significant digits in reported incomes reduces their uniqueness, making them less identifiable. The ACS applies a very well-defined rounding rule.

[11] On the ACS public use file, amounts of $50,000 or greater rounded to the nearest $1,000.

CPS has strong income data, but its program participation and health insurance data are less credible and for decades have required complex modeling efforts for policy use. The basic CPS is a monthly labor force survey to which the ASEC is an annual supplement, and income data are subordinated to the needs of employment and labor force information.

MEPS income data also make a strong showing, but the impact of post-stratifying the file, altering the distribution of reported incomes to match CPS poverty rates, is not understood. In addition, MEPS has the highest initial survey non-response rate, thus the highest risk of non-response bias, and has internal inconsistencies that are also not well understood.

The ACS is a massive survey with exceptionally good response rates and surprisingly good income data, but has been so narrowly focused on demographics that the needs of policy users have yet to be acknowledged. The ACS has major definitional and methodological differences from the other surveys and from the Decennial Long Form that it has replaced, but these differences have not been fully explored, since most such work requires use of internal Census Bureau data files.[12] The impact of these differences on income measures, especially on the income distribution and poverty rates for subpopulations, is also not fully understood.

---

[12] The Census Bureau produced, pro bono, an extensive set of tabulations on the internal ACS files with which the study could examine some aspects of the unique ACS rolling sample, rolling reference period and the within-year inflation adjustments.

TABLE 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.

TABLE 2

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.

TABLE 3

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 4

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.

TABLE 5

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.

TABLE 6

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 7

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.

TABLE 8

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 9

OVERALL RESPONSE RATES AND ALLOCATION OF TOTAL INCOME: FIVE SURVEYS

Estimate	CPS	ACS	SIPP	MEPS	NHIS
Initial Overall Response Rate	92.2%	97.3%	87.7%	70.0%	89.2%
2002 Response After Attrition	NA	NA	72.5%	64.7%	NA
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 detailed descriptive analysis chapter, and 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 10

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).

TABLE 11

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.