Bridging a Survey Redesign Using Multiple Imputation: An Application to the 2014 CPS ASEC

Jonathan Rothbaum*

September 2, 2015

SEHSD Working Paper 2015-15

Abstract

The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) serves as the data source for official income, poverty, and inequality statistics in the United States. In 2014, the CPS ASEC questionnaire was redesigned to improve data quality and to reduce misreporting, item nonresponse, and errors resulting from respondent fatigue. The sample was split into two groups, with nearly 70% receiving the traditional instrument and 30% receiving the redesigned instrument. Due to the relatively small redesign sample, analyses of changes in income and poverty between this and future years may lack sufficient power, especially for subgroups. In this paper, we explore the possibility of using multiple imputation techniques to combine the two subsamples into a single sample that we can use to estimate income and poverty statistics with greater power and smaller standard errors. Multiple imputation is a general approach to analyzing data with missing values. We can treat the traditional sample as if the responses were missing for income sources targeted by the redesign and use multiple imputation to generate plausible responses. We use a flexible semiparametric imputation technique to place individuals into strata along two dimensions: 1) their probability of income recipiency and 2) their expected income conditional on recipiency for each income source. Within each stratum, we randomly select redesign individuals and donate their income recipiency, source, and value information to individuals in the traditional sample as the imputed data. By matching on these two dimensions this approach combines the ideas of propensity score matching (from the probability of recipiency strata) and predictive means matching (from expected income strata). In this paper, we implement this approach, use diagnostics to evaluate the matching models, and analyze the results.

* Social and Economic Housing Statistics Division, U.S. Census Bureau. Corresponding Email: jonathan.l.rothbaum@census.gov

1 Introduction

The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) is among the most widely used surveys conducted by the US Census Bureau. CPS ASEC data is used to calculate measurements of national income and the official poverty rate. Rothbaum (2015) shows that the CPS ASEC suffers from underreporting of certain income types, including property income (especially interest and dividends), retirement income, and income from means-tested government transfer programs. Meyer et al. (2009) also show underreporting of participation in means-tested government programs.

To address this underreporting, the Census Bureau, in consultation with the private sector,[1] implemented a redesign of the survey. As a first step, the redesigned survey instrument was implemented in a nationwide telephone content test of 23,000 households in 2013. Based on favorable results from this test, a more comprehensive assessment was conducted in 2014 using the full survey production environment. In this second test, approximately 30% of the CPS ASEC sample (30,000 housing units) received the redesigned survey instrument, and approximately 70% of the sample (68,000 housing units) received the unchanged traditional instrument (in use since 1994). Assignment into the two groups was random at the household level.[2] A major focus of the redesign was to improve reporting of property income, especially income earned from assets in the form of interest or dividends. In addition, because the nature of retirement savings has shifted from defined benefit to defined contribution plans since 1980,[3] the survey was redesigned to improve reporting of retirement income, which has historically been underreported (Czajka and Denmead, 2008).

The results from the second test were sufficiently favorable, with more reported income in a variety of categories, that the redesigned instrument is being used for the full sample starting with the 2015 CPS ASEC. There were statistically significant increases in income recipiency and aggregate income in a number of categories. Therefore, in order to make apples-to-apples comparisons between the results in 2014 and 2015 and beyond, only 30% of the 2014 sample can

[1] See Czajka and Denmead (2008) and Hicks and Kerwin (2011) for results of that consultation.
[2] For more details about the redesign and the content tests, see Semega and Welniak (2013; 2015).
[3] From 1980 to 2008, the share of private wage and salary workers with defined benefit plans fell from 38% to 20%. The share of private workers with defined contribution plans grew 8% to 31% over the same period (Iams et al. 2009).

be used. This significantly reduces the power of the comparisons that can be made, for example of median income or poverty rates, which is especially relevant for subgroups.[4]

However, while the survey redesign significantly increased recipiency and aggregates for many income types, the majority of income (by dollars) was not affected. For example, earnings comprised 75.9 percent of all income,[5] and there were no statistically significant differences between earnings across the two instruments. Therefore, although we do not observe what respondents to the traditional instrument would have said to the redesigned questions, we do have a considerable amount of information about them that is unaffected by the redesign.

This suggests treating the problem as one of missing data – as if the recipients of the traditional instrument did not respond to the redesigned income questions. To address issues of nonresponse and missing data in surveys, Rubin (1987) developed multiple imputation.[6] Since Rubin's initial work, there has been a tremendous amount of research focused on the theory and application of multiple imputation (see Schafer and Graham, 2002 and Reiter and Raghunathan, 2007 for some examples). Imputation involves modelling responses to replace missing data with plausible values. This is standard practice in survey processing, including the CPS ASEC hot deck imputation procedure (discussed in Section 3). Multiple imputation involves imputing the missing responses repeatedly to incorporate the uncertainty about the correct missing value for each individual in any given analysis.

In this paper, we apply an approach developed by Bondarenko and Raghunathan (2007) to impute these "missing" responses in the traditional sample. This technique was designed for cases where strong parametric assumptions about the distribution of the variable to be imputed and model used may not be satisfied. Individuals are divided into strata along two dimensions: 1) their probability of income recipiency (modelled using logistic regression) and 2) their expected income conditional on recipiency for each income source (modelled using OLS). Within each stratum, redesign respondents are selected at random using an Approximate Bayesian Bootstrap (Rubin and Schenker, 1986; Rubin, 1987) to donate their income recipiency,

[4] The standard error of a mean of a random sample is $\frac{s}{\sqrt{n}}$, where s is the sample standard deviation and n is the number of observations in the sample. Given a reduction of n to 30% of its normal size, the standard error increases by over 80%. This increase is a reasonable approximation for median and poverty rate estimates, as well. Therefore, a change in the poverty rate or mean or median income must be considerably larger to be statistically significant when compared to the 2014 CPS ASEC.

[5] In the redesign sample

[6] In fact, Rubin used income nonresponse in the CPS as one of his primary examples of a problem for multiple imputation to address.

source, and value information to individuals in the traditional sample. The donated information is the imputed data. This approach combines the ideas of propensity score matching (from the probability of recipiency strata) and predictive means matching (from the expected income strata). By matching donors within cells or strata to recipients, this approach is similar to the hot deck procedure used in the normal CPS ASEC processing. That makes it ideal for use in this case as the completed data from the 2014 CPS ASEC can be used to make comparisons with data in subsequent years where all imputation of missing values is done using the hot deck.

Another appealing feature of this approach is the modelling flexibility. Within each income recipiency stratum, the expected income conditional on recipiency is predicted using an OLS regression model. This is advantageous for several reasons. First, the expected income model can vary by stratum. For example, it is possible that the model to predict retirement income differs considerably between those with a low likelihood of recipiency and those with a high likelihood. Second, the analyst does not need to impose his assumptions on the data by preselecting appropriate strata for modelling. Instead, the strata are determined by the relationships in the data between the observed characteristics in the model and the likelihood of income recipiency. Third, because the approach imputes values using an Approximate Bayesian Bootstrap, multiple variables related to a given income category can be imputed simultaneously. This greatly simplifies the modelling process, especially as some variables imputed in the CPS ASEC are unordered categorical variables that can be challenging to model.

We use the technique to create an "income-consistent" full file that uses all of the CPS ASEC sample with imputed income in the affected categories. We call it the income-consistent file as the responses for all individuals are consistent with the questions in the redesign survey instrument.

The paper is organized as follows. In section 2, we describe the CPS ASEC and the survey redesign. In section 3, we discuss the semiparametric imputation methodology. In section 4, we discuss diagnostic results to evaluate the models used. In section 5, we report results relating to income, poverty, and inequality measurement using the imputed data and discuss strategies for selecting a single file to be used for official Census publications. Section 6 concludes.

2 Data and Survey Redesign

The Current Population Survey Annual Social and Economic Supplement (CPS ASEC) is among the most widely used surveys conducted by the US Census Bureau. The CPS ASEC surveys about 100,000 households each year and includes questions on income and health insurance coverage. The data is used to calculate the official poverty rate and measures of national income.

However, research shows that the CPS ASEC suffers from underreporting of income and participation in means-tested government programs. For example, Rothbaum (2015) compares income aggregates in the CPS ASEC to those in the Bureau of Economic Analysis' National Income and Product Account (NIPA) tables. He finds that self-employment, interest, dividend, and retirement and pension income are underreported in the CPS ASEC relative to the NIPA estimates. Meyer et al. (2009) study reported participation in means-tested government transfer programs and finds underreporting in the CPS ASEC.

To address this underreporting, the Census Bureau contracted Westat Inc. and Mathematica, Policy Research in 2011 to evaluate the survey questionnaire and suggest changes to improve it. As a result of this process, a redesigned CPS ASEC survey was developed. The changes, described below, are discussed in greater detail by Semega and Welniak (2015).

2.1 Questionnaire Changes

2.1.1 Remove Family Income Screener

The family income screener for determining which households are asked about low-income sources was removed. Prior to the redesign, only households that reported less than $75,000 in combined family income were asked questions about means-tested transfer programs such as Temporary Assistance to Needy Families (TANF). Semega and Welniak (2015) cite evidence from the American Community Survey (ACS) that some screened households were likely to be recipients of these transfers making it inappropriate to remove them using the income screener.

2.1.2 Dual-Pass Approach

For all income except earnings (wage and self-employment), the questions on income recipiency were separated from the questions on amounts as a part of a "dual-pass" approach. Respondents were asked first about all sources of income received and then later asked about amounts for only the received sources. Prior to the change, if respondents answered "yes" to receiving a source of income, they were immediately asked about the amount (and any other type or source questions). This change was implemented to prevent respondent fatigue from affecting answers to the income recipiency questions. For example, over the course of the survey, respondents may have learned they could avoid value questions by answering "no" to the initial recipiency question.

2.1.3 Tailored Skip Patterns

The order of the income questions is tailored to match those sources most likely to be received by respondents given certain known characteristics. Respondents are separated into three groups of 1) householders 62 and over, 2) lower income households, and 3) all other households. For the 62 and older group, questions on disability and retirement income are prioritized. For the low-income group questions on means-tested government transfers are prioritized. The default group receives questions in the order of the traditional survey instrument.

2.1.4 Income Range Brackets

If an individual responds "don't know" or refuses to provide a specific dollar amount for a given income source for them or a member of their household, new questions on income range are presented. The specific ranges vary by income type. For example, for earnings, the ranges are 1) less than $45,000, 2) $45,000-$60,000, and 3) $60,000 or more. If the respondent chooses the lowest range, a follow-up set of ranges are asked. The range data is not currently used in the data processing and individuals who provide range data have their income imputed (or allocated) using the hot deck procedure.

2.1.5 Changes to Retirement and Asset Income Questions

To better capture retirement income, the survey was redesigned to specifically ask if anyone in the household has a pension and separately if anyone has a retirement account (401(k), 403(b), IRA, or other account designed specifically for retirement savings). The traditional instrument includes one broad question on the receipt of pension and retirement income. The redesigned instrument also asks individuals over 70 years old about required distributions from retirement accounts. To ensure that the distribution is correctly identified as income, a follow up question asks if the required distribution was “rolled over” or reinvested in another account. The traditional ASEC instrument makes no distinction between investment income received in retirement accounts or separately from them. This more detailed set of questions can improve misreporting of income and cue respondents and decrease underreporting.

2.1.6 Other Changes

Several additional changes were made to the CPS ASEC survey. If a respondent was unsure of the income generated from assets, the value of the assets was collected. The questions on disability were clarified to eliminate confusion between disability income from Social Security and Supplemental Security Income (SSI).

2.2 Results of the Redesign

In 2014, the CPS ASEC sample was divided into two groups, with about 30% (30,000 housing units) receiving the redesigned instrument and about 70% (68,000 housing units) receiving the traditional instrument. Within each sample, individual observations were weighted to national population controls, as is standard with the CPS ASEC.

Semega and Welniak (2015) compared income aggregates between the two samples. Table 1 shows a subset of their results for median income, updated to reflect recent edits of the redesign sample file. Household median income was $51,939 in the traditional sample and $53,585 in the redesign, a difference of 3.2%. When decomposed by race, the only statistically significant differences are for whites (and non-Hispanic whites).

Table 2 shows income statistics for total income and various income sources collected in the CPS ASEC. For each source, Semega and Welniak report the number of recipients in the population, the mean income earned by those recipients, and the aggregate value of that income estimated using the traditional and redesign samples separately. For example, for total income the number of income recipients estimated using the traditional sample is 218.7 million compared to 222.0 from the redesigned sample, a statistically significant difference of 1.5%. The estimated difference in mean total income is 2.6% ($41,319 in the traditional vs. 42,394 in the redesign), and the estimated difference in aggregate total income is 4.2% ($9.04 trillion in the traditional vs. 9.41 trillion in the redesign), both statistically significant. At the 90% confidence level, there are a number of income sources that have statistically significant differences in the number of recipients, mean income, or aggregate income. The sources with statistically significant differences in aggregate income include farm self-employment income (-42.1%), public assistance (28.8%), veterans' benefits (-23.1%), disability benefits (36.4%), retirement income (21.9%), interest (113.0%), and dividends (-20.1%).

Mitchell and Renwick (2015) study the effects of the redesign on poverty rates. While they find no statistically significant difference in the overall poverty rate, they do find differences for child and elderly poverty in the redesigned sample. In both cases, they suggest that differences in the sample populations may explain the increase in poverty in the redesigned sample. For child poverty, they show that the redesigned sample has a higher share of children living with female householders than the traditional sample. They also cite the fact that means-tested program recipiency was higher in the redesigned sample as potential evidence that explains subgroup poverty differences.

These potential differences in sample characteristics support the approach taken in this paper. Because we treat the changes in the questionnaire as a problem of missing information, any differences in the samples can be controlled for as a part of the imputation modelling and by combining the samples.

2.3 Selection of Income Sources to be Imputed

Taking these analyses together, the redesign increased aggregate income and increased income recipiency and reporting in a number of income categories. However, some of the differences, especially in income types with no or little change in the questionnaire, may be due to random variation or differences in the samples.

Because of the evidence of sample differences, we focused on those income types which were targeted by the questionnaire redesign. This eliminates farm self-employment, and veterans' benefits. In addition, for each income source where the response could be considered "missing" for the traditional sample due to a question change, there is a tradeoff between imputing the responses using the redesign sample and preserving the information from the responses in the traditional sample.

As a result, we have chosen to focus on income sources that were sufficiently different between the two surveys and were specifically targeted by the questionnaire redesign. This includes three income types: 1) retirement income, 2) interest, and 3) dividends. These three sources had the largest difference in estimated aggregate income of the types affected by the redesign. Figure 1 and Figure 2 show changes in recipiency and aggregate income. For interest income, the number of recipients increased by 37.6 million and aggregate income increased by $206.3 billion. For retirement income, the number of recipients increased by 1.8 million and aggregate income increased by $82.7 billion. For dividend income, the number of recipients decreased by 1.4 million and aggregate income decreased by $29.6 billion.

3 Imputation Methodology

3.1 Hot Deck Imputation

As a part of the standard processing of the CPS ASEC, when an individual does not respond to a particular question, missing values are imputed using a hot deck procedure. In the hot deck, individuals are divided into cells based on the characteristics specified in the hot deck model.[7] Within each cell, individuals without missing information (donors) are randomly selected and their income is assigned to the individuals with missing information (recipients). Donors and recipients in each cell must match on every variable in the hot deck model. If there are no donors in a given recipient's cell, the hot deck model is amended to reduce the number of categories for some variables (for example from nine age groupings to six) and to reduce the number of variables in the model.

The different hot deck models used in the CPS ASEC are called match levels. The 1st match level includes the largest number of variables and categories within each variable. If no matches are found at the 1st level, an attempt match recipients and donors is made using the model at the 2nd match level. This continues until a match level is reached for a given recipient in which at least one donor is present in the same cell. For missing earnings in the longest job, in the 1st match level there are 16 variables in the model and 621 billion possible cells; in the 2nd match level there are 14 variables and 17 billion possible cells; in the 3rd match level there are 11

[7] For example, in the hot deck for earnings, the model include some combination of the following variables: gender, race, age, relationship to householder, education, marital status, presence of children, spousal labor force status, hours and weeks worked, occupation, worker class (private, government, self-employed, etc.), other earnings receipt, type of residence, region, receipt of government transfers, and person status (working-age civilian, armed forces, or child).

variables and 3.8 million possible cells,[8] and by the 6th match level there are 4 variables and 96 possible cells. In the traditional sample for those observations missing earnings from the longest job only, 4.4% matched on the 1st level, 13.0% matched on the 2nd level, 51.5% matched on the 3rd level, and 6.4% matched on the 6th level. The variables and number of categories at each match level are shown in Table 3.

As these numbers make clear, the number of variables that can be included in a hot deck model is clearly limited by the size of the sample. While this is clearly a constraint even in the full CPS ASEC sample of about 200,000 individuals, the constraint is even more binding when imputing income from the redesign sample of about 60,000 individuals. If we were to impute retirement, interest, and dividend income in the traditional sample using the hot deck, we would not be able to incorporate many variables in the model that are potentially correlated with each income type. This would limit the ability of the imputation to accurately match similar individuals as donors and recipients and reduce the quality of the matches.

3.2 Semiparametric Model-Based Imputation

Instead, we implemented a more flexible technique to impute the missing responses to the redesigned questions in the traditional sample for the research file. The approach, developed by Bondarenko and Raghunathan (2007), hereafter BR, matches donors and recipients using regression modelling. Their approach was specifically designed for cases where strong parametric assumptions about the distribution of the variable to be imputed and the functional form of the model may not be satisfied. This is especially a concern for interest income, where nearly 25% of recipients have income values $25 or less.

Another reason the semiparametric approach was chosen in this research is its similarity to the hot deck. As in the hot deck, individuals are matched based on similarities in observable characteristics. In the hot deck, the matching is based on the characteristics directly. In the BR approach, the matching is based on the predicted probability of recipiency and expected income conditional on recipiency, which can both be estimated from observable characteristics. This is advantageous as the imputed data must be comparable to data from subsequent years where all missing data is imputed using the hot deck.

[8] The 1st and 2nd match levels include occupation code to the 4 digit level, whereas the 3rd includes 66 occupation categories. However, even with only 66 occupation categories the 1st and 2nd match levels would have 77.6 billion and 2.1 billion possible cells respectively.

Next, we will describe the BR method, with slight modification for this application. Suppose that the data set has P variables of observable characteristics, X_p, $p = 1,2, \dots, P$ and $X = (X_1, \dots, X_P)$. Suppose that the data set contains Q income types where Y_q, $q = 1,2 \dots, Q$, represents the income value and R_q represents recipiency status ($R_q \in \{0,1\}$). There are two groups in the sample, one for which the income types q are observed (group O) and one for which income types q are unobserved (group M) so that each vector can be partitioned among O and M as $X_p = \left(X_p^O, X_p^M\right)$, $Y_q = \left(Y_q^O, Y_q^M\right)$, and $R_q = \left(R_q^O, R_q^M\right)$. Because missingness is complete for all Y_q^M, we can impute income sequentially without iteration. We therefore define $< q$ as the set of incomes with indices less than q so that $Y_{<q} = \left(Y_1, \dots, Y_{q-1}\right)$ and $R_{<q} = \left(R_1, \dots, R_{q-1}\right)$ and $Y_{<0}$ and $R_{<0}$ are empty sets. We construct two efficient summaries of the income variables through two regression predictions:

1. Probability of recipiency: $\hat{R}_q = \Pr\left(R_q = 1 | Y_{<q}, R_{<q}, X\right)$ estimated using a logistic regression model. It is an efficient summary of R_q that can be used to balance income recipients and non-recipients (Rosenbaum and Rubin, 1983). We stratify $\hat{R}_q$ into K equal size strata, where $k = 1, \dots, K$.
2. Predicted value of income conditional on recipiency within each stratum k: $\hat{Y}_q = E\left(Y_q \middle| R_q = 1, Y_{<q}, R_{<q}, X\right)$ estimated using an OLS regression model on all individuals in stratum k. We then subdivide individuals in stratum k into J equal sized substrata, where $j = 1, \dots, J$. This creates $K \times J$ equal size strata.

Within each stratum k, j there are n individuals with observed income and recipiency and m individuals with missing income and recipiency for income type q. We draw a sample size m from the observed set of n individuals as the imputed values by Approximate Bayesian Bootstrap (ABB). This step is repeated for each stratum k,j and income type q and then sequentially for all $q = 1, \dots, Q$. This entire process is repeated independently to obtain multiple imputations.

There are a number of challenges to implementing BR method in the CPS ASEC. First, many income types do not follow a normal distribution or any simple transformation of a normal distribution. Second, we must select predictors (X) for the modelling of each income variable from a very large set of possible covariates in the CPS ASEC.

As shown in Hokayem, Raghunathan, and Rothbaum (2015), the distribution of income is rarely normally distributed. Simple transformation (such as log) and more flexible ones such as

Tukey's gh distribution (He and Raghunathan, 2006) also can fail to convert the distribution to normal. Therefore, we use an empirical normal transformation proposed by Woodcock and Benedetto (2009) to convert all income values to normal distributions (this includes income and other continuous variables in X as well) prior to imputation.

The most significant challenge to applying BR method to the CPS ASEC was to select the models for each imputed variable. In order to avoid omitted variable bias in the imputation model, we would like to include as many potential predictors as possible. However, if we include too many variables, we run the risk of overfitting the model. The list of potential predictors we use includes all unchanged income information (imputation flag, recipiency, value),[9] spouse/partner earnings, race (separate dummy for each), gender, age (including dummies for each age between 62 and 70), weeks worked last year, hours worked per week, as well as the hot deck categories for relationship to householder, education level, marital status, presence of children, occupation (22 categories), type of residence, Census region, recipiency of means-tested government transfers. We also included a large set of interaction terms in our list of predictors including for major income types (earnings, spouse earnings, etc.), education, weeks and hours worked, race and age, and means-tested transfers. In all, over 1,200 potential predictors and interaction terms can be included in our BR models.[10] The full list of modelling variables and interactions, along with the coding and categories used, are shown in Table 4.

We chose to implement stepwise model selection regressions to prune the list of possible predictors to a more manageable one for each variable. As a part of the logistic and OLS modelling, we use the model selection process to reduce the number of covariates used in predicting $\widehat{R}_q$ and $\widehat{Y}_q$.[11] Another potential concern is that for each stratum k, it is possible that there are a very small number of income recipients. For example, for those with a low probability of receiving retirement income (such as those under 25), there may be few or no

[9] This includes all income variables except the ones being imputed, workers' compensation, public assistance, SSI, and disability income. We chose to impute only retirement, interest, and dividend income as they were the three largest income sources both in terms of recipiency and aggregate income. However, we did not want to include other sources in the imputation model that may also differ between the two survey instruments.

[10] In part, the large number of variables is due to the conversion of categorical variables into separate dummies. For example, there are six education levels so the categorical education variable is converted into six dummy variables, with each interacted with all the other possible interaction terms. This yields a large number of possible predictors from the single education variable.

[11] Note that the variables used to predict $\widehat{R}_q$ and $\widehat{Y}_q$ can differ for the same income type q.

income recipients in a given stratum. In that case, we collapse the cell and use the $\hat{Y}_q$ predictions from the full sample OLS regression for stratum k.[12]

In order to approximate the model selection and parameter uncertainty, for each income type q, prior to running the logistic or OLS regressions, a random sample is taken by ABB. All regressions are run on this ABB sample, but the stratification into groups is done from the original sample.

In summary, the imputation steps to create the income-consistent file are:

1. **Normal transformation** – Transform all income value variables to normal distribution with empirical normal transformation
2. **BR Imputation** – sequentially impute interest, dividends, and retirement income from the redesign (donors/observed) to the traditional sample (recipients/missing). For each income type:
    a. Select a random sample by ABB
    b. Predict probability of income recipiency using logistic regression on the redesign ABB sample with stepwise model selection to choose list of predictors. Only those individuals with non-imputed values of recipiency are included in the regression.
    c. Stratify sample into K equal-sized groups based on probability of income recipiency in the original sample.[13]
    d. Within each stratum k, predict expected income conditional on recipiency using OLS regression on the redesign ABB sample that is within the probability of recipiency bounds of that stratum.
    e. Stratify subsample k into J equal sized substrata based on the expected income of the original sample.
    f. Within each substratum j, select a random sample of m donors from the redesign sample (where m is the number of recipients with missing responses in stratum

[12] For example, in the first stratum of the first implicate for retirement income, the probability of receipt is 0.1%.

[13] The selection of K and J varies by the number of recipients for a given income type. Higher values of each allow finer matching between donors and recipients at the cost of potentially having too few observations for regressions or too small a pool of potential donors to accurately reflect the distribution of potential responses for each recipient. For interest income, $K = 8$ and $J = 16$ for 128 strata, for dividend income, $K = 6$ and $J = 12$ for 72 strata, and for retirement income, $K = 4$ and $J = 8$ for 32 strata. The difference in number of strata are due to the fact that over 50% (18,755) of the non-imputed observations of the redesign sample received interest income compared to 13% (4,830) for dividends and 8% (2,859) for retirement income.

k, j) using ABB. Each donor receives all income, source, and value variables from the recipient.

g. Repeat for each stratum k, j until all missing observations for income type q have been imputed.

3. **Transform to original scale** – return all variables to their original scales.
4. **Repeat the entire process to create ten implicates**

These steps are done *after* processing and allocation of the survey data. This means that hot deck imputed values in the redesign file can be used as part of the imputation process. However, all modelling and prediction is done only on actual responses with allocated values excluded from the modelling step.

4 Diagnostic Results

One way of evaluating the imputation model is to construct an R^2 from the set of regressions on the ABB sample. For the logistic regressions, we use the Tjur-R^2 (Tjur, 2009), which is calculated by comparing the average predicted probability of recipiency for those who did and did not receive income of that type, or

$$R^2_{Tjur} = E\left(\hat{R}_q \middle| R = 1, Y_{<q}, R_{<q}, X\right) - E\left(\hat{R}_q \middle| R = 0, Y_{<q}, R_{<q}, X\right).$$

The Tjur-R^2 is bounded between 0 and 1.

For the OLS regressions, we compute the squared correlation between the transformed income and the predicted income from the strata regressions, shown in Table 5. The average Tjur-R^2 for interest, dividends, and retirement are 0.35, 0.30, and 0.39 respectively.[14] The OLS R^2 values for interest, dividends, and retirement income are 0.12, 0.10, and 0.15 respectively.

The relatively low R^2 are in part due to the fact that predictions are made on ABB samples, not the original one. The regression R^2 are much higher, but they reflect the match between the predictions and the bootstrapped sample, which will by definition be higher than for the original sample, which was not used for the prediction.

Another statistic that can be used to evaluate the value of using the imputation to create the income-consistent file is the estimated rate of missing information, which we denote as γ (Rubin, 1987). Very high values of γ (for example, 0.7) would imply that there is no additional benefit

[14] All R-squared calculations are made comparing only the observations that were not allocated using the hot deck in the redesign file as these observations were the only ones used in the modelling process.

to using the traditional sample with imputed interest, dividend, and retirement income. As the relevance of the missing interest, dividend, and retirement income may differ for different statistics, for each parameter of interest, we can compute a γ. The estimated γ is 0.15 for household median income and 0.08 for poverty. Both of these are low values, which indicates that a considerable amount of information in estimating median income and poverty is contributed by the other variables in the traditional sample. These low γ values also validate the general approach of treating responses to the questions affected by the redesign as missing information to take advantage of the full CPS ASEC sample. It is also encouraging that the rates of missing information are low given the low R^2 in the value regressions used to match donors and recipients.

5 Income and Poverty Statistics

In order to evaluate further the results from the imputation, we calculated median income and poverty statistics that are in the annual Income and Poverty reports.[15] Table 6 shows the median income statistics (Table 1 from the annual Income and Poverty report) comparison between the income-consistent full sample and the traditional and redesign sample. Compared to the redesign sample, the only statistically significant differences are for median income of non-family households with a female householder (3.6% greater) and households headed by individuals without a disability (2.3% greater). At the 90% confidence interval, fewer than 10% of the tested statistics are significantly different. For the comparison with the traditional sample, nearly all of the median income comparisons are statistically significant, including for all households (3.0% greater), for family households (2.6% greater), Whites (2.6% greater), and Asians (7.0% greater). Among the comparisons that are not statistically significantly different are for Blacks, Hispanics, non-citizens, and the Northeast census region.

Table 7 shows the poverty comparison between the redesign sample and the income-consistent file. The headline poverty number for all individuals is not statistically significantly different between the two files. However, there are more statistically significant differences than in the household median income comparisons in Table 6. For example, poverty is statistically significantly lower in the income-consistent file for children (1.1%) and those aged 65 and older (0.7%). This corresponds to the statistically significant differences between the traditional and

[15] The most recent report is available at http://www.census.gov/hhes/www/income/.

redesign sample poverty rates, which suggests that differences in the samples were responsible for these differences in poverty.

Table 8 shows the differences between poverty estimates in the income-consistent file and the traditional sample. Unlike the median income comparison, there are few statistically significant differences in poverty estimates. Poverty is lower in the income consistent-file for blacks (1.0%), naturalized citizens (-0.9%), residents of principal cities (0.5%), and workers (0.2%) and higher for children (0.6%).

To summarize the results, the income-consistent file household median income estimates are more like the redesign file, but the poverty estimates lie between the two files. While the point estimate for poverty of 14.5% is not statistically significantly different from the point estimate for either file, it is much closer to the 14.5% estimate from the traditional file than the 14.8% estimate of the redesign file.

6 Closest Income-Consistent File

All of the diagnostics and results shown up to this point have utilized all 10 implicates and accounted for the uncertainty in the imputation of interest, dividend, and retirement income from the BR methodology. While the microdata of all ten implicates will be released as a research file to the public, more complete processing of the files will only be available for one implicate due to resource constraints. To select the implicate to receive the additional processing, we would like to choose the file that is "closest" to the multiple imputation in some sense. We have chosen to select the file that best matches multiple imputation average for household median income and poverty. These statistics were chosen because the CPS ASEC is the official source of the estimate of US poverty and household median income is another headline statistic that is generated from the CPS ASEC.

To define what is the best match between an individual file and the multiple imputation average, we have chosen to minimize the normalized distance between the estimates from a given file and the average. To do so, we calculated the standard deviation for the estimates of median income and poverty and converted the estimates from each file into a z-score where:

$$z_i = \frac{\rho_i - \mu}{\sigma}$$

where ρ_i is the estimate for file i, μ is the multiple imputation average for that statistic, and σ is the standard deviation in the 10 implicates. We then sum the squared normalized distance from the poverty and median income estimates for each file and selected the file with the lowest squared distance.

Table 9 shows the same information on median income as Table 6, but for the closest income-consistent file. The results are generally comparable between the two tables, in that there are few statistically significant differences between the closest income-consistent file and the redesign file, but there are many statistically significant differences from the traditional file. It is worth noting that some results do differ between the multiple imputation average and the closest single income-consistent file. They are highlighted in the table. These differences could arise for two reasons: 1) the imputation uncertainty is not accounted for in the closest single file, and therefore the standard errors are too narrow or 2) the point estimates differ between the single file and the multiple imputation average. We show the comparable tables on poverty in Table 10 and Table 11. Again, the results largely mirror, with few differences, those for the multiple imputation average.

7 Conclusion

In this paper, we have applied multiple imputation to the problem of a split sample receiving different survey instruments in a bridge year. In this way, we have shown one possible way to use data from all survey respondents even though distinct sets of respondents answered different questions. This idea has an important potential benefit – by making use of all of the data during a bridge year, it potentially lowers the cost in terms of decreased statistical power of survey redesigns and bridges.

To this problem of missing information during a survey bridge year, we have applied specific semiparametric multiple imputation technique proposed by Bondarenko and Raghunathan to the CPS ASEC 2014 redesign. We show that the technique performs reasonably well and analyzed some basic summary statistics to show how this technique affects important economic statistics that are widely reported on from the CPS ASEC.

For the 2014 CPS ASEC, this technique increases the potential sample that can be used to make comparisons to data from subsequent years, which uses the redesigned questionnaire for the entire sample. The larger sample facilitates analyses on subgroups, such as by state, where the redesign sample may lack the statistical power needed for comparisons. By combining the two samples, this technique may also address concerns about differences in sample composition raised in previous research by Mitchell and Renwick (2015).

References

Bondarenko, Irina, and Trivellore E. Raghunathan. 2007. "Multiple Imputations Using Sequential Semi and Nonparametric Regressions." *Proceedings of the Survey Research Methods Section*. 2007.

Czajka, John L. and Gabrielle Denmead. 2008. "Income data for Policy Analysis: A Comparative Assessment of Eight Surveys" Mathematica Reference No.: 6302-601 [http://mathematica-mpr.com/publications/PDFs/incomedata.pdf].

Hokayem, Charles, Trivellore E. Raghunathan, and Jonathan Rothbaum. 2015. "SRMI in the CPS ASEC." *Unpublished Manuscript*.

He, Yulei, and Trivellore E. Raghunathan. 2006. "Tukey's gh distribution for multiple imputation." *The American Statistician* 60 (3).

Hicks, Wendy and Jeffrey Kerwin, 2011. "Cognitive Testing of Potential Changes to the Annual Social and Economic Supplement of the Current Population Survey." Unpublished Westat report to the U.S. Census Bureau. July 25, 2011.

Iams, Howard, Barbara A. Butrica, Karen E. Smith, and Eric J. Toder. 2009 "The Disappearing Defined Benefit Pensions and Its Potential Impact on the Retirement Income of Baby Boomers." *Social Security Bulletin*, (69) 3, 1-28.

Meyer, Bruce D., Wallace K.C. Mok, and James X. Sullivan. 2009. "The under-reporting of transfers in household surveys: its nature and consequences." *National Bureau of Economic Research Working Paper #15181.*

Mitchell, Joshua and Trudi Renwick. 2015. "A Comparison of Official Poverty Estimates in the Redesigned Current Population Survey Annual Social and Economic Supplement." *U.S. Census Bureau SEHSD Working Paper #2014-35*.

Reiter, Jerome P., and Trivellore E. Raghunathan. 2007. "The multiple adaptations of multiple imputation." *Journal of the American Statistical Association* 102(480), 1462-1471.

Rosenbaum, Paul R., and Donald B. Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." *Biometrika* 70 (1), 41-55.

Rothbaum, Jonathan. 2015. "Comparing Income Aggregates: How do the CPS and ACS Match the National Income and Product Accounts, 2007-2012."

Rubin, Donald B., Nathaniel Schenker. 1986. "Multiple Imputation for interval estimation from simple random samples with ignorable nonresponse." *Journal of the American Statistical Association*. 81 (394), 366–374.

Rubin, Donald B. 1987. *Multiple imputation for nonresponse in surveys*. John Wiley and Sons.

Schafer, Joseph L., and John W. Graham. 2002. "Missing data: our view of the state of the art." *Psychological methods* 7(2), 147.

Semega, Jessica and Edward Welniak Jr. 2013. "Evaluating the 2013 CPS ASEC Income Redesign Content Test." Presented at the proceeding of the 2013 FCSM conference. U.S. Census Bureau Income Statistics working paper available online at: http://www.census.gov/hhes/www/income/publications/Evaluating the 2013 CPS ASEC Income Redesign Content Test.pdf.

———. 2015. "The Effects of the Changes to the Current Population Survey Annual Social and Economic Supplement on Estimates of Income" Presented at the Proceedings of the 2015 Allied Social Science Association (ASSA) Research Conference. U.S. Census Bureau Income Statistics working paper available online at: https://www.census.gov/content/dam/Census/library/working-papers/2015/DEMO/ASSA-Income-CPSASEC-Red.pdf

Tjur, Tue. 2009. "Coefficients of determination in logistic regression models—A new proposal: The coefficient of discrimination." *The American Statistician* 63(4), 366-372.

Woodcock, Simon D., and Gary Benedetto. 2009. "Distribution-preserving statistical disclosure limitation." *Computational Statistics & Data Analysis* 53(12), 4228-4242.

Table 1: Comparison of Traditional and Redesign Samples: Median Income

Characteristic	Traditional: Number (Thousands) Estimate	Traditional: Number (Thousands) 90% CI	Traditional: Median Income (Dollars) Estimate	Traditional: Median Income (Dollars) 90% CI	Redesign: Number (Thousands) Estimate	Redesign: Number (Thousands) 90% CI	Redesign: Median Income (Dollars) Estimate	Redesign: Median Income (Dollars) 90% CI	Percentage Change $\left(\frac{R-T}{T}\right)$ Estimate	Percentage Change $\left(\frac{R-T}{T}\right)$ 90% CI
All Households	122,952	723	51,939	455	123,931	942	53,585	1,076	3.2*	2.08
Race[1] and Hispanic Origin of Householder										
White	97,774	605	55,257	699	98,807	756	56,745	850	2.7*	1.81
White, not Hispanic	83,641	544	58,270	1,066	84,432	732	60,329	876	3.5*	2.04
Black	16,108	262	34,598	1,198	16,009	355	35,324	1,410	2.1	5. 13
Asian	5,759	151	67,065	2,830	5,818	215	72,383	5,531	7.9	7.92
Hispanic (Any Race)	15,811	210	40,963	908	16,088	354	39,687	1,953	-3.1	5.00
Earnings of Full-Time Year-Round Workers										
Men with Earnings	60,769	600	50,033	404	61,240	787	50,015	935	-0.4	1.23
Women with Earnings	45,068	510	39,414	596	44,629	659	38,792	1,145	-0.9	3.19

Notes: Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality pr otection, sampling error, nonsampling error, and definitions, see www.census.gov/prod/techdoc/cps/cpsmar14.pdf. * indicates statistically different from zero at the 90-percent confidence level. A 90-percent confidence interval (CI) is a measure of an estimate's variability. The larger the CI in relation to the size of the estimate, the less reliable the estimate. CIs shown in this table are based on standard errors calculated using replicate weights. For more information, see "Standard Errors and Their Use" at <www.census.gov/hhes/www/p60_245sa.pdf>.

[1] Federal surveys now give respondents the option of reporting more than one race. Therefore, two basic ways of defining a rac e group are possible. A group such as Asian may be defined as those who reported Asian and no other race (the race-alone or single-race concept) or as those who reported Asian regardless of whether they also reported another race (the race-alone-or-in-combination concept). This table shows data using the first approach (race alone). The use of the single-race population does not imply that it is the preferred method of presenting or analyzing data. The Census Bureau uses a variety of approaches. Information on people who reported more than one race, such as White and American Indian and Alaska Native or Asian and Black or African American, is available from Census 2010 through American FactFinder. About 2.9 percent of people report ed more than one race in Census 2010. Data for American Indians and Alaska Natives, Native Hawaiians and Other Pacific Islander s, and those reporting two or more races are not shown separately in this table.

Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement.

Table 2: Comparison of Traditional and Redesign Samples: Income Recipiency, Mean Income, and Aggregate Income By Source

	Traditional						Redesign						Percent Change $\left(\frac{R-T}{T}\right)$					
	Number (Thousands)		Mean Income (Dollars)		Aggregate Income (Thousands)		Number (Thousands)		Mean Income (Dollars)		Aggregate Income (Thousands)		Number		Mean Income		Aggregate Income	
Type of Income	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**	**Estimate**	**SE**
Total Income	218,662	311	41,319	279	9,035,004	60,841	222,003	418	42,394	345	9,411,655	79,158	1.5*	0.2	2.6*	1.1	4.2*	1.1
Earnings	158,081	489	44,416	334	7,021,280	58,106	158,655	638	44,999	438	7,139,254	74,929	0.4	0.5	1.3	1.2	1.7	1.3
Wages and Salary	148,752	492	44,931	336	6,683,647	55,005	149,546	684	45,695	457	6,833,462	75,382	0.5	0.5	1.7	1.2	2.2	1.4
Nonfarm Self-Emp	8,702	190	35,145	1,215	305,825	12,033	8,508	298	33,777	1,411	287,376	15,300	-2.2	4.1	-3.9	5.6	-6.0	6.2
Farm Self-Emp	627	59	50,728	6,870	31,808	5,359	601	73	30,662	4,448	18,416	3,278	-4.2	13.7	-39.6*	12.4	-42.1*	14.3
Unemployment	6,818	165	5,841	151	39,825	1,362	6,435	233	5,870	167	37,768	1,806	-5.6	4.1	0.5	4.1	-5.2	5.6
Workers' Comp	1,186	60	9,224	566	10,940	930	952	77	10,156	851	9,671	1,135	-19.7*	7.3	10.1	10.5	-11.6	11.4
Social Security	48,370	332	13,979	55	676,178	5,142	49,055	418	14,052	84	689,325	6,782	1.4	1.0	0.5	0.7	1.9	1.2
SSI	6,053	176	7,782	105	47,104	1,459	6,642	230	7,728	148	51,333	2,021	9.7*	5.0	-0.7	2.3	9.0	5.5
Public Assistance	1,775	81	3,195	149	5,671	340	2,189	124	3,337	160	7,305	508	23.3*	8.8	4.4	6.5	28.8*	10.9
Veterans' Benefits	3,517	127	14,640	424	51,493	2,503	3,296	164	12,021	584	39,619	2,757	-6.3	5.1	-17.9*	4.8	-23.1*	6.1
Survivors' Benefits	3,033	110	12,972	559	39,340	2,260	2,970	156	14,526	995	43,139	3,502	-2.1	6.5	12.0	8.6	9.7	11.0
Disability Benefits	1,771	76	15,543	736	27,524	1,850	3,099	163	12,110	635	37,535	2,654	75.0*	11.8	-22.1*	5.6	36.4*	12.8
Retirement Income	18,871	251	20,034	307	378,054	7,865	20,698	372	22,262	449	460,784	12,668	9.7*	2.4	11.1*	3.0	21.9*	4.2
Interest	86,142	588	2,120	68	182,619	5,963	123,772	887	3,142	107	388,943	13,403	43.7*	1.4	48.2*	6.8	113.0*	9.7
Dividends	33,243	432	4,424	170	147,050	6,225	31,804	568	3,693	211	117,454	6,954	-4.3*	1.9	-16.5*	5.8	-20.1*	5.6

*Statistically different from zero at the 90-percent confidence level.

Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement.

Table 3: CPS Hot Deck Imputation for Missing Earnings from Longest Job

	Match Level					
Match Variable	1	2	3	4	5	6
Sex	2	2	2	2	2	2
Race	3	2	2			
Age	9	6	3	3		
Relationship	7	7	4	4	4	
Years of School Completed	6	5	5	4	4	4
Marital Status	4	4				
Presence of Children	3					
Labor Force Status of Spouse	3					
Weeks Worked	5	5	4	4	4	4
Hours Worked	3	3	3	3	2	
Occupation	528	528	66	66	66	
Class of Worker	5	5	5	3	3	3
Other Earnings	8	8				
Type of Residence	3	2	2			
Region	4	4				
Transfers payments receipt	2	2	2	2		
Number of Cells	620,786,073,600	17,031,168,000	3,801,600	456,192	50,688	96

Table 4: Potential Predictor Variables in Imputation Model

Type	Variable	CPS ASEC Name	Coding/Additional Information	Interaction
Imputation Flags	Earnings Recipiency	I_ERNYN		
	Earnings Value	I_ERNVAL		
	Other Job Wage Recipiency	I_WSYN		
	Other Job Self-Employment Recipiency	I_SEYN		
	Other Job Farm Self-Employment Recipiency	I_FRMYN		
	Other Job Wage Value	I_WSVAL		
	Other Job Self-Employment Value	I_SEVAL		
	Other Job Farm Self-Employment Value	I_FRMVAL		
	Unemployment Compensation Recipiency	I_UCYN		
	Unemployment Compensation Value	I_UCVAL		
	Educational Assistance Recipiency	I_EDYN		
	Educational Assistance Value	I_OEDVAL		
	Child Support Payment Recipiency	I_CSPYN		
	Child Support Payment Value Received	I_CPSVAL		
	Financial Assistance Recipiency	I_FINYN		
	Financial Assistance Value	I_FINVAL		
	Rental Income Recipiency	I_RNTYN		
	Rental Income Value	I_RNTVAL		
	Survivors' Benefits Recipiency	I_SURYN		
	Survivors' Benefits Value (Source 1)	I_SURVA1		
	Survivors' Benefits Value (Source 2)	I_SURVA2		
	Veterans' Benefits Recipiency	I_VETYN		
	Veterans' Benefits Value	I_VETVAL		
	Spouse or Cohabiting Partner Earnings Recipiency	Recode from spouse I_ERNYN		
	Spouse or Cohabiting Partner Earnings Value	Recode from spouse I_ERNVAL		
	Property Value	I_PROPVAL		X

Table 4: Potential Predictor Variables in Imputation Model, Continued

Type	Variable	CPS ASEC Name	Coding/Additional Information	Interaction
Earnings Recipiency	Earnings Recipiency	ERN_YN		X
	Other Job Wage Recipiency	WAGEOTR		X
	Other Job Self-Employment Recipiency	SEOTR		X
	Other Job Farm Self-Employment Recipiency	FRMOTR		X
	Unemployment Compensation Recipiency	UC_YN		X
	Veterans' Benefits Recipiency	VET_YN		X
	Survivors' Benefits Recipiency	SUR_YN		
	Rental Income Recipiency	RNT_YN		X
	Educational Assistance Recipiency	ED_YN		X
	Child Support Payment Recipiency	CSP_YN		
	Financial Assistance Recipiency	FIN_YN		X
	Spouse/Partner Present	Recode from A_SPOUSE		X
	Spouse/Partner Earnings Recipiency	Recode from ERN_YN of Spouse		X
	Interest Income Recipiency	INT_YN	Dividends and Retirement Only	
	Dividend Income Recipiency	DIV_YN	Retirement Only	
Earnings Value	Earnings Value	ERN_VAL		X
	Other Job Wage Value	WS_VAL		
	Other Job Self-Employment Value	SE_VAL		
	Other Job Farm Self-Employment Value	FRM_VAL		
	Unemployment Compensation Value	UC_VAL		X
	Veterans' Benefits Value	VET_VAL		X
	Survivors' Benefits Value (Source 1)	SUR_VAL1		
	Survivors' Benefits Value (Source 2)	SUR_VAL2		
	Rental Income Value	RNT_VAL		X
	Educational Assistance Value	ED_VAL		X
	Child Support Payment Value	CSP_VAL		X
	Financial Assistance Value	FIN_VAL		X
	Spouse/Partner Earnings Value	Recode from ERN_VAL of Spouse		X
	Interest Income Value	INT_VAL	Dividends and Retirement Only	
	Dividend Income Value	DIV_VAL	Retirement Only	

Table 4: Potential Predictor Variables in Imputation Model, Continued

Type	Variable	CPS ASEC Name	Coding/Additional Information	Interaction
Other Variables	Race/Ethnicity - White	Recode from PRDTRACE		X
	Race/Ethnicity - Black	Recode from PRDTRACE		X
	Race/Ethnicity - Native American	Recode from PRDTRACE		X
	Race/Ethnicity - Asian	Recode from PRDTRACE		X
	Race/Ethnicity - Pacific Islander	Recode from PRDTRACE		X
	Race/Ethnicity - Hispanic	PEHSPNON		X
	Age	A_AGE	Continuous	X
			Dummies for 62 and older through 70 and older	
	Weeks Worked in Last Year	WKSWORK	Continuous	X
			Dummies for 1, 40, and 50	X (50 Weeks)
	Usual Hours Worked	HRSWK	Continuous	X
			Dummies for 1, 40, and 60	X (40 Hours)
	Property Value (Non-imputed)	HPROP_VAL		X
	Gender	A_SEX		X
	Supplement Weight (Full Sample)	MARSUPWT		
	Relationship to Household Head	A_RRP	Dummy for 1, 2, 3 or 4, 5, 6 or 7 or 8, 9, and 10	
	Education	A_HGA	Dummy for <= 34, 35-38, 39, 40-42, 43, and 44-46	X
	Marital Status	A_MARITL	Dummy for 1-3, 4, 5 or 6, and 7	
	Children in Family	FRELU6 FRELU18	Dummy for FRELU18 = 0, FRELU6 > 0, and FRELU6 = 0 and FRELU18 > 0	
	Occupation	OCCUP	Dummy for 1-950, 1000-3540, 3600-4650, 4700-4960, 5000-5930, 6000-6130, 6200-6940, 7000-7420, 7700-8960, 9000,9750, and 9840	
	City Type of Residence (Urban/CBSA)	GEUR GECBSATY	GEUR = 1 and GECBSATY = 1-2, GEUR=1 and GECBSATY = 3, and GEUR = 2	
	Census Region	GEREG		
	Transfer Payments/ Program Participation	HLORENT HPUBLIC HENGAST HFOODSP UMCAID	Any = 1	X

Table 5: Model Diagnostics – Effective R^2 of Recipiency and Value Regressions

Variable	Implicate	Recipiency	Value
Interest	1	0.33	0.15
	2	0.36	0.12
	3	0.35	0.13
	4	0.36	0.14
	5	0.35	0.15
	6	0.34	0.15
	7	0.34	0.09
	8	0.35	0.14
	9	0.35	0.05
	10	0.35	0.09
Dividends	1	0.32	0.09
	2	0.29	0.08
	3	0.31	0.06
	4	0.32	0.13
	5	0.31	0.12
	6	0.28	0.10
	7	0.31	0.13
	8	0.29	0.03
	9	0.30	0.12
	10	0.30	0.11
Retirement	1	0.41	0.13
	2	0.38	0.17
	3	0.41	0.15
	4	0.37	0.15
	5	0.37	0.18
	6	0.39	0.14
	7	0.43	0.17
	8	0.36	0.18
	9	0.39	0.19
	10	0.40	0.03

Average R^2 across 10 Implicates

Variable	Recipiency	Value
Interest	0.35	0.12
Dividends	0.30	0.10
Retirement	0.39	0.15

The R^2 are calculated by taking the predicted recipiency and values conditional on recipiency from the prediction models used to define the donor/recipient cells and calculating the Tjur R^2 for recipiency and squared correlation for the value.

Table 6: Comparison of Traditional, Redesign, and Multiple Imputation Income-Consistent Files: Household Median Income by Selected Characteristics

	Traditional (T)			Redesign (R)			Income-Consistent (IC)								
	Median income (dollars)			Median income (dollars)			Median income (dollars)			Percentage change in real median income (IC - T)/T			Percentage change in real median income (IC - R)/R		
Characteristic	Number (thousands)	Estimate	90 percent CI	Number (thousands)	Estimate	90 percent CI	Number (thousands)	Estimate	90 percent CI	Estimate		90 percent CI	Estimate		90 percent CI
All Households	122,952	51,939	455	123,931	53,585	1,076	123,229	53,499	687	3.00	*	0.95	-0.16		1.70
Family households	81,192	65,587	643	82,270	66,923	872	81,353	67,301	673	2.61	*	0.86	0.56		1.18
Married-couple families	59,669	76,509	674	59,626	78,897	1,359	59,629	79,275	946	3.61	*	0.93	0.48		1.56
Female householder, no husband present	15,193	35,154	832	16,158	35,412	1,512	15,420	35,896	793	2.11	*	1.62	1.37		3.69
Male householder, no wife present	6,330	50,625	1,503	6,486	52,480	2,730	6,304	52,248	1,528	3.21	*	2.21	-0.44		4.46
Nonfamily households	41,760	31,178	518	41,660	31,480	951	41,877	31,977	539	2.56	*	1.32	1.58		2.51
Female householder	22,266	26,425	795	21,827	26,238	1,019	22,219	27,186	766	2.88	*	2.62	3.61	*	3.49
Male householder	19,494	36,876	937	19,834	39,379	1,674	19,658	38,242	1,094	3.70	*	2.35	-2.89		3.78
White	97,774	55,257	699	98,807	56,745	850	98,052	56,708	606	2.63	*	0.95	-0.06		1.28
White, not Hispanic	83,641	58,270	1,006	84,432	60,329	876	83,892	60,225	657	3.36	*	1.34	-0.17		1.23
Black	16,108	34,598	1,198	16,009	35,324	1,410	16,064	35,429	948	2.40		2.53	0.29		3.59
Asian	5,759	67,065	2,830	5,818	72,383	5,531	5,749	71,743	2,564	6.98	*	3.25	-0.88		6.07
Hispanic (any race)	15,811	40,963	908	16,088	39,687	1,954	15,874	41,341	894	0.92		1.34	4.17		4.54
Under 65 years	94,223	58,448	958	94,862	60,265	771	94,442	60,528	499	3.56	*	1.19	0.44		1.11
15 to 24 years	6,323	34,311	1,808	6,652	33,791	3,156	6,404	34,845	1,610	1.55		3.85	3.12		8.30
25 to 34 years	20,008	52,702	1,489	19,988	52,416	2,098	19,978	53,592	1,366	1.69		2.17	2.24		3.59
35 to 44 years	21,046	64,973	1,620	21,164	67,594	1,976	21,123	66,985	1,198	3.10	*	1.68	-0.90		2.48
45 to 54 years	23,809	67,141	1,265	23,664	70,598	2,114	23,733	70,671	1,214	5.26	*	1.60	0.10		2.57
55 to 64 years	23,036	57,538	1,662	23,395	60,481	1,835	23,205	60,735	1,503	5.56	*	2.78	0.42		2.88
65 years and older	28,729	35,611	722	29,069	37,297	1,283	28,787	36,352	808	2.08		2.46	-2.53		3.22
Native born	105,328	52,779	754	105,900	55,087	940	105,518	54,615	737	3.48	*	1.18	-0.86		1.52
Foreign born	17,624	46,939	1,037	18,031	46,795	1,563	17,712	48,156	1,259	2.59	*	1.91	2.91		3.20
Naturalized citizen	9,491	54,974	2,898	9,489	56,354	3,098	9,476	57,406	1,947	4.42	*	4.41	1.87		5.48
Not a citizen	8,133	40,578	1,113	8,542	40,185	1,944	8,236	41,020	950	1.09		1.77	2.08		4.46
Households with householders aged 18 to 64	94,024	58,492	955	94,699	60,310	742	94,264	60,566	501	3.55	*	1.19	0.42		1.07
With disability	8,794	25,421	1,260	8,614	25,337	1,746	8,778	26,476	1,152	4.15	*	4.15	4.50		6.69
Without disability	84,784	61,979	564	85,549	62,487	1,021	85,018	63,924	892	3.14	*	0.96	2.30	*	1.55
Northeast	22,053	56,775	1,426	22,511	56,868	2,563	22,150	58,121	1,852	2.37		2.46	2.20		4.08
Midwest	27,214	52,082	1,160	27,426	53,426	2,102	27,280	53,207	1,427	2.16	*	2.00	-0.41		3.27
South	46,499	48,128	1,104	46,553	49,854	1,335	46,462	50,136	809	4.17	*	1.61	0.57		2.20
West	27,186	56,181	1,190	27,441	59,525	2,067	27,338	57,775	1,336	2.84	*	1.78	-2.94		3.02
Inside metropolitan statistical areas	103,573	54,042	790	104,128	55,884	810	103,766	55,614	615	2.91	*	1.07	-0.48		1.19
Inside principal cities	41,359	46,778	892	41,360	48,806	1,621	41,290	48,262	1,067	3.17	*	1.50	-1.11		2.77
Outside principal cities	62,213	59,497	1,090	62,768	60,787	937	62,476	60,731	721	2.07	*	1.33	-0.09		1.23
Outside metropolitan statistical areas	19,379	42,881	1,238	19,802	43,601	1,755	19,463	44,925	1,176	4.77	*	2.21	3.03		3.47
Men with earnings	83,555	39,903	718	83,855	40,229	499	83,737	40,054	362	0.38		1.24	-0.44		1.02
Women with earnings	74,545	27,736	599	74,821	27,390	464	74,633	27,533	415	-0.73		1.19	0.52		1.45

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each income-consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates.

Table 7: Comparison of Redesign and Income-Consistent Files: Poverty by Selected Characteristics

Characteristic	Redesign (R)					Income-Consistent Multiple Imputation (IC)					Difference in Poverty (IC-R)/R	
	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
PEOPLE												
Total	313,096	46,269	1,474	14.8	0.5	312,983	45,257	903	14.5	0.3	-1,012	-0.3
Family Status												
In families.	256,070	32,786	1,370	12.8	0.5	255,079	31,668	790	12.4	0.3	-1,118	-0.4
Householder	82,316	9,645	421	11.7	0.5	81,381	9,171	238	11.3	0.3	* -474	* -0.4
Related children under 18	72,246	15,116	723	20.9	1.0	72,454	14,492	421	20.0	0.6	* -624	* -0.9
Related children under 6	23,606	5,590	340	23.7	1.4	23,586	5,339	200	22.6	0.9	-251	-1.0
In unrelated subfamilies.	1,626	776	220	47.7	8.4	1,465	622	102	42.5	5.2	-154	-5.2
Reference person..	661	291	86	44.0	8.2	604	243	39	40.3	4.9	-47	-3.6
Children under 18	844	448	130	53.1	9.3	754	358	65	47.5	5.8	-89	-5.6
Unrelated individual.	55,400	12,707	579	22.9	0.9	56,439	12,967	370	23.0	0.6	260	0.0
Race[3] and Hispanic Origin												
White alone	243,346	31,287	1,073	12.9	0.4	243,144	30,235	682	12.4	0.3	* -1,052	* -0.4
White alone, not Hispanic.	195,118	19,552	815	10.0	0.4	195,288	19,102	574	9.8	0.3	-451	-0.2
Black alone	40,498	10,186	631	25.2	1.6	40,577	10,615	444	26.2	1.1	429	1.0
Asian alone	17,257	2,255	330	13.1	1.9	17,003	1,881	165	11.1	1.0	* -374	* -2.0
Hispanic (of any race).	54,181	13,356	801	24.7	1.5	54,138	12,750	492	23.6	0.9	-606	-1.1
Sex												
Male	153,465	20,294	769	13.2	0.5	153,373	20,101	479	13.1	0.3	-193	-0.1
Female	159,630	25,975	902	16.3	0.6	159,610	25,156	537	15.8	0.3	* -819	* -0.5
Age												
Under 18 years	73,439	15,801	725	21.5	1.0	73,535	15,039	429	20.5	0.6	* -762	* -1.1
18 to 64 years	194,694	25,899	877	13.3	0.5	194,971	26,000	589	13.3	0.3	101	0.0
65 years and over	44,963	4,569	286	10.2	0.6	44,477	4,218	222	9.5	0.5	* -351	* -0.7
Nativity												
Native..	272,423	38,831	1,299	14.3	0.5	272,249	38,064	809	14.0	0.3	-767	-0.3
Foreign born..	40,673	7,438	556	18.3	1.2	40,734	7,193	331	17.7	0.7	-245	-0.6
Naturalized citizen	19,247	2,132	249	11.1	1.3	19,132	2,245	146	11.7	0.7	113	0.7
Not a citizen	21,426	5,306	498	24.8	1.9	21,602	4,948	287	22.9	1.1	-358	* -1.9

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each income-consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates.

Table 7: Comparison of Redesign and Income-Consistent Files: Poverty by Selected Characteristics, continued

Characteristic	Redesign (R)					Income-Consistent Multiple Imputation (IC)					Difference in Poverty (IC-R)/R	
	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
Region												
Northeast	55,529	7,205	700	13.0	1.3	55,481	7,032	424	12.7	0.8	-173	-0.3
Midwest	66,732	9,269	640	13.9	1.0	66,758	8,792	396	13.2	0.6	-477	-0.7
South.	116,956	19,040	968	16.3	0.8	116,959	18,756	627	16.0	0.5	-284	-0.2
West.	73,879	10,754	669	14.6	0.9	73,785	10,676	399	14.5	0.5	-78	-0.1
Residence												
Inside metropolitan statistical areas.	265,301	37,994	1,491	14.3	0.5	265,773	37,604	995	14.1	0.3	-390	-0.2
Inside principal cities.	101,094	18,617	1,140	18.4	1.0	101,874	19,030	781	18.7	0.6	412	0.3
Outside principal cities.	164,207	19,377	1,091	11.8	0.6	163,900	18,574	691	11.3	0.4	-802	-0.5
Outside metropolitan statistical areas	47,795	8,275	891	17.3	1.3	47,210	7,653	633	16.2	0.8	* -622	* -1.1

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each income-consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates.

Table 8: Comparison of Traditional and Income-Consistent Files: Poverty by Selected Characteristics

		Traditional (T)					Income-Consistent Multiple Imputation (IC)				Difference in Poverty (IC-T)/T	
Characteristic	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
PEOPLE												
Total	312,965	45,318	1,014	14.5	0.3	312,983	45,257	903	14.5	0.3	-61	0.0
Family Status												
In families.	254,988	31,530	844	12.4	0.3	255,079	31,668	790	12.4	0.3	138	0.0
Householder	81,217	9,130	247	11.2	0.3	81,381	9,171	238	11.3	0.3	41	0.0
Related children under 18	72,573	14,142	445	19.5	0.6	72,454	14,492	421	20.0	0.6	* 350	* 0.5
Related children under 6	23,585	5,231	225	22.2	1.0	23,586	5,339	200	22.6	0.9	108	0.5
In unrelated subfamilies.	1,413	608	114	43.0	6.3	1,465	622	102	42.5	5.2	15	-0.5
Reference person..	595	246	48	41.3	6.4	604	243	39	40.3	4.9	-3	-1.0
Children under 18	714	340	69	47.7	6.7	754	358	65	47.5	5.8	18	-0.2
Unrelated individual.	56,564	13,181	414	23.3	0.6	56,439	12,967	370	23.0	0.6	-214	-0.3
Race[3] and Hispanic Origin												
White alone	243,085	29,936	816	12.3	0.3	243,144	30,235	682	12.4	0.3	299	0.1
White alone, not Hispanic.	195,167	18,796	722	9.6	0.4	195,288	19,102	574	9.8	0.3	306	0.2
Black alone	40,615	11,041	506	27.2	1.3	40,577	10,615	444	26.2	1.1	* -427	* -1.0
Asian alone	17,063	1,785	176	10.5	1.0	17,003	1,881	165	11.1	1.0	96	0.6
Hispanic (of any race).	54,145	12,744	513	23.5	0.9	54,138	12,750	492	23.6	0.9	6	0.0
Sex												
Male	153,361	20,119	568	13.1	0.4	153,373	20,101	479	13.1	0.3	-19	0.0
Female	159,605	25,199	572	15.8	0.4	159,610	25,156	537	15.8	0.3	-43	0.0
Age												
Under 18 years	73,625	14,659	455	19.9	0.6	73,535	15,039	429	20.5	0.6	* 381	* 0.5
18 to 64 years	194,833	26,429	648	13.6	0.3	194,971	26,000	589	13.3	0.3	-429	-0.2
65 years and over	44,508	4,231	227	9.5	0.5	44,477	4,218	222	9.5	0.5	-13	0.0
Nativity												
Native..	271,968	37,921	943	13.9	0.3	272,249	38,064	809	14.0	0.3	143	0.0
Foreign born..	40,997	7,397	373	18.0	0.8	40,734	7,193	331	17.7	0.7	-204	-0.4
Naturalized citizen	19,147	2,425	173	12.7	0.9	19,132	2,245	146	11.7	0.7	* -180	* -0.9
Not a citizen	21,850	4,972	311	22.8	1.2	21,602	4,948	287	22.9	1.1	-24	0.2

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses w ere eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each income-consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates.

Table 8: Comparison of Traditional and Income-Consistent Files: Poverty by Selected Characteristics, continued

Characteristic	Traditional (T)					Income-Consistent Multiple Imputation (IC)					Difference in Poverty (IC-T)/T	
	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
Region												
Northeast	55,478	7,046	437	12.7	0.8	55,481	7,032	424	12.7	0.8	-13	0.0
Midwest	66,785	8,590	430	12.9	0.7	66,758	8,792	396	13.2	0.6	202	0.3
South.	116,961	18,870	706	16.1	0.6	116,959	18,756	627	16.0	0.5	-113	-0.1
West.	73,742	10,812	434	14.7	0.6	73,785	10,676	399	14.5	0.5	-137	-0.2
Residence												
Inside metropolitan statistical areas.	265,915	37,746	1,007	14.2	0.4	265,773	37,604	995	14.1	0.3	-142	0.0
Inside principal cities.	102,149	19,530	842	19.1	0.7	101,874	19,030	781	18.7	0.6	* -500	* -0.4
Outside principal cities.	163,767	18,217	738	11.1	0.4	163,900	18,574	691	11.3	0.4	358	0.2
Outside metropolitan statistical areas	47,050	7,572	665	16.1	1.0	47,210	7,653	633	16.2	0.8	81	0.1

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for the traditional and redesigned samples, and using replicate weights for each income-consistent implicate and the multiple imputation variance formula from Rubin and Schenker (1986) to combine estimates across implicates for the Income-Consistent estimates.

Table 9: Comparison of Traditional, Redesign, and Closest Income-Consistent Imputation Files: Household Median Income by Selected Characteristics

	Traditional (T)			Redesign (R)			Closest Income-Consistent (CIC)								
	Median income (dollars)			Median income (dollars)			Median income (dollars)			Percentage change in real median income (CIC - T)/T			Percentage change in real median income (CIC - R)/R		
Characteristic	Number (thousands)	Estimate	90 percent confidence interval1 (+)	Number (thousands)	Estimate	90 percent confidence interval1 (+)	Number (thousands)	Estimate	90 percent confidence interval1 (+)	Estimate		90 percent confidence interval1 (+)	Estimate		90 percent confidence interval1 (+)
All Households	122,952	51,939	455	123,931	53,585	1,076	123,229	53,516	655	3.04	*	0.85	-0.13		1.63
Family households	81,192	65,587	643	82,270	66,923	872	81,353	67,319	603	2.64	*	0.73	0.59		1.06
Married-couple families	59,669	76,509	674	59,626	78,897	1,359	59,629	79,380	950	3.75	*	0.82	0.61		1.54
Female householder, no husband present	15,193	35,154	832	16,158	35,412	1,512	15,420	36,095	765	2.68	*	1.51	1.93		3.69
Male householder, no wife present	6,330	50,625	1,503	6,486	52,480	2,730	6,304	51,992	1,454	2.70	*	2.06	-0.93		4.34
Nonfamily households	41,760	31,178	518	41,660	31,480	951	41,877	32,131	490	3.06	*	1.29	2.07		2.36
Female householder	22,266	26,425	795	21,827	26,238	1,019	22,219	27,669	896	4.71	*	2.64	5.45	*	3.40
Male householder	19,494	36,876	937	19,834	39,379	1,674	19,658	37,727	962	2.31	*	2.00	-4.19	*	3.46
White	97,774	55,257	699	98,807	56,745	850	98,052	56,764	546	2.73	*	0.88	0.03		1.22
White, not Hispanic	83,641	58,270	1,006	84,432	60,329	876	83,892	60,296	535	3.48	*	1.28	-0.06		1.16
Black	16,108	34,598	1,198	16,009	35,324	1,410	16,064	35,403	867	2.33	*	2.24	0.22		3.45
Asian	5,759	67,065	2,830	5,818	72,383	5,531	5,749	71,140	2,083	6.08	*	2.93	-1.72		6.11
Hispanic (any race)	15,811	40,963	908	16,088	39,687	1,954	15,874	41,236	836	0.67		1.04	3.90		4.45
Under 65 years	94,223	58,448	958	94,862	60,265	771	94,442	60,486	473	3.49	*	1.18	0.37		1.08
15 to 24 years	6,323	34,311	1,808	6,652	33,791	3,156	6,404	34,875	1,502	1.64		3.74	3.21		8.14
25 to 34 years	20,008	52,702	1,489	19,988	52,416	2,098	19,978	53,951	1,415	2.37	*	2.14	2.93		3.57
35 to 44 years	21,046	64,973	1,620	21,164	67,594	1,976	21,123	67,237	1,177	3.48	*	1.72	-0.53		2.44
45 to 54 years	23,809	67,141	1,265	23,664	70,598	2,114	23,733	71,007	1,081	5.76	*	1.50	0.58		2.55
55 to 64 years	23,036	57,538	1,662	23,395	60,481	1,835	23,205	59,487	1,757	3.39	*	2.39	-1.64		2.74
65 years and older	28,729	35,611	722	29,069	37,297	1,283	28,787	36,835	634	3.44	*	2.06	-1.24		2.69
Native born	105,328	52,779	754	105,900	55,087	940	105,518	54,638	665	3.52	*	1.04	-0.81		1.42
Foreign born	17,624	46,939	1,037	18,031	46,795	1,563	17,712	48,096	1,217	2.46	*	1.69	2.78		3.12
Naturalized citizen	9,491	54,974	2,898	9,489	56,354	3,098	9,476	57,613	2,142	4.80	*	4.49	2.23		5.32
Not a citizen	8,133	40,578	1,113	8,542	40,185	1,944	8,236	40,939	885	0.89		1.53	1.88		4.50
Households with householders aged 18 to 64	94,024	58,492	955	94,699	60,310	742	94,264	60,521	474	3.47	*	1.17	0.35		1.04
With disability	8,794	25,421	1,260	8,614	25,337	1,746	8,778	25,873	978	1.78		3.40	2.12		6.15
Without disability	84,784	61,979	564	85,549	62,487	1,021	85,018	63,950	873	3.18	*	0.90	2.34	*	1.49
Northeast	22,053	56,775	1,426	22,511	56,868	2,563	22,150	57,884	1,600	1.95		2.09	1.79		3.88
Midwest	27,214	52,082	1,160	27,426	53,426	2,102	27,280	53,585	1,357	2.89	*	1.72	0.30		3.22
South	46,499	48,128	1,104	46,553	49,854	1,335	46,462	50,222	755	4.35	*	1.51	0.74		2.09
West	27,186	56,181	1,190	27,441	59,525	2,067	27,338	57,310	1,126	2.01	*	1.52	-3.72	*	2.86
Inside metropolitan statistical areas	103,573	54,042	790	104,128	55,884	810	103,766	55,590	568	2.87	*	0.97	-0.52		1.12
Inside principal cities	41,359	46,778	892	41,360	48,806	1,621	41,290	48,126	1,029	2.88	*	1.31	-1.39		2.69
Outside principal cities	62,213	59,497	1,090	62,768	60,787	937	62,476	60,752	676	2.11	*	1.23	-0.06		1.19
Outside metropolitan statistical areas	19,379	42,881	1,238	19,802	43,601	1,755	19,463	45,012	1,131	4.97	*	2.23	3.24		3.35
Men with earnings	83,555	39,903	718	83,855	40,229	499	83,737	40,054	362	0.38		1.24	-0.44		1.02
Women with earnings	74,545	27,736	599	74,821	27,390	464	74,633	27,533	415	-0.73		1.19	0.52		1.45

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for all samples. The IC file is the single closest file (in normalized distance) to the MI median income and poverty averages, without accounting for imputation variance. The cells highlighted in pink are these where the statistical significance differs between this file and the MI average with imputation variance shown in Table 6.

Table 10: Comparison of Redesign and Closest Income-Consistent File: Poverty by Selected Characteristics

	Redesign					Closest Income-Consistent (CIC)					Difference in Poverty (CIC-R)/R	
Characteristic	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
PEOPLE												
Total	313,096	46,269	1,474	14.8	0.5	312,983	45,267	893	14.5	0.3	-1,002	-0.3
Family Status												
In families.	256,070	32,786	1,370	12.8	0.5	255,079	31,792	777	12.5	0.3	-994	-0.3
Householder	82,316	9,645	421	11.7	0.5	81,381	9,238	224	11.4	0.3	* -407	-0.4
Related children under 18	72,246	15,116	723	20.9	1.0	72,454	14,471	417	20.0	0.6	* -645	* -0.9
Related children under 6	23,606	5,590	340	23.7	1.4	23,586	5,318	197	22.5	0.8	-272	-1.1
In unrelated subfamilies.	1,626	776	220	47.7	8.4	1,465	608	101	41.5	5.1	-168	-6.2
Reference person..	661	291	86	44.0	8.2	604	236	37	39.1	4.8	-54	-4.8
Children under 18	844	448	130	53.1	9.3	754	351	64	46.6	5.8	-96	-6.5
Unrelated individual.	55,400	12,707	579	22.9	0.9	56,439	12,867	331	22.8	0.5	160	-0.1
Race[3] and Hispanic Origin												
White alone	243,346	31,287	1,073	12.9	0.4	243,144	30,210	655	12.4	0.3	* -1,077	* -0.4
White alone, not Hispanic.	195,118	19,552	815	10.0	0.4	195,288	19,026	549	9.7	0.3	-526	-0.3
Black alone	40,498	10,186	632	25.2	1.6	40,577	10,696	439	26.4	1.1	510	1.2
Asian alone	17,257	2,255	330	13.1	1.9	17,003	1,884	165	11.1	1.0	* -372	* -2.0
Hispanic (of any race).	54,181	13,356	801	24.7	1.5	54,138	12,760	476	23.6	0.9	-596	-1.1
Sex												
Male	153,465	20,294	769	13.2	0.5	153,373	20,150	467	13.1	0.3	-144	-0.1
Female	159,630	25,975	902	16.3	0.6	159,610	25,117	530	15.7	0.3	* -858	* -0.5
Age												
Under 18 years	73,439	15,801	725	21.5	1.0	73,535	15,009	427	20.4	0.6	* -792	* -1.1
18 to 64 years	194,694	25,899	877	13.3	0.5	194,971	26,208	544	13.4	0.3	309	0.1
65 years and over	44,963	4,569	286	10.2	0.6	44,477	4,050	169	9.1	0.4	* -519	* -1.1
Nativity												
Native..	272,423	38,831	1,299	14.3	0.5	272,249	38,068	808	14.0	0.3	-763	-0.3
Foreign born..	40,673	7,438	556	18.3	1.2	40,734	7,199	319	17.7	0.7	-239	-0.6
Naturalized citizen	19,247	2,132	249	11.1	1.3	19,132	2,217	140	11.6	0.7	85	0.5
Not a citizen	21,426	5,306	498	24.8	1.9	21,602	4,982	279	23.1	1.1	-323	* -1.7

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for all samples. The IC file is the single closest file (in normalized distance) to the MI median income and poverty averages, without accounting for imputation variance. The cells highlighted in pink are these where the statistical significance differs between this file and the MI average with imputation variance shown in Table 7.

Table 10: Comparison of Redesign and Closest Income-Consistent File: Poverty by Selected Characteristics, Continued

Characteristic	Redesign: Total	Redesign: Number in Poverty (Thousands)	Redesign: 90 percent CI	Redesign: Percent in Poverty	Redesign: 90 percent CI	Closest Income-Consistent (CIC): Total	CIC: Number in Poverty (Thousands)	CIC: 90 percent CI	CIC: Percent in Poverty	CIC: 90 percent CI	Difference in Poverty (CIC-R)/R: Number	Difference in Poverty (CIC-R)/R: Percent
Region												
Northeast	55,529	7,205	700	13.0	1.3	55,481	7,021	409	12.7	0.7	-185	-0.3
Midwest	66,732	9,269	641	13.9	1.0	66,758	8,728	390	13.1	0.6	* -542	* -0.8
South.	116,956	19,040	968	16.3	0.8	116,959	18,796	613	16.1	0.5	-245	-0.2
West.	73,879	10,754	670	14.6	0.9	73,785	10,723	379	14.5	0.5	-31	0.0
Residence												
Inside metropolitan statistical areas.	265,301	37,994	1,491	14.3	0.5	265,773	37,611	960	14.2	0.3	-383	-0.2
Inside principal cities.	101,094	18,617	1,140	18.4	1.0	101,874	19,070	783	18.7	0.6	453	0.3
Outside principal cities.	164,207	19,377	1,091	11.8	0.6	163,900	18,540	653	11.3	0.4	-836	-0.5
Outside metropolitan statistical areas	47,795	8,275	891	17.3	1.3	47,210	7,656	621	16.2	0.7	* -618	* -1.1

The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for all samples. The IC file is the single closest file (in normalized distance) to the MI median income and poverty averages, without accounting for imputation variance. The cells highlighted in pink are these where the statistical significance differs between this file and the MI average with imputation variance shown in Table 7.

Table 11: Comparison of Traditional and Closest Income-Consistent File: Poverty by Selected Characteristics

	Traditional (T)					Closest Income-Consistent (CIC)					Difference in Poverty (CIC-T)/T	
Characteristic	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Total	Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
PEOPLE												
Total	312,965	45,318	1,014	14.5	0.3	312,983	45,267	893	14.5	0.3	-51	0.0
Family Status												
In families.	254,988	31,530	845	12.4	0.3	255,079	31,792	777	12.5	0.3	262	0.1
Householder	81,217	9,130	247	11.2	0.3	81,381	9,238	224	11.4	0.3	108	0.1
Related children under 18	72,573	14,142	445	19.5	0.6	72,454	14,471	417	20.0	0.6	* 329	* 0.5
Related children under 6	23,585	5,231	225	22.2	1.0	23,586	5,318	197	22.5	0.8	86	0.4
In unrelated subfamilies.	1,413	608	114	43.0	6.3	1,465	608	101	41.5	5.1	1	-1.5
Reference person..	595	246	48	41.3	6.4	604	236	37	39.1	4.8	-10	-2.2
Children under 18	714	340	69	47.7	6.7	754	351	64	46.6	5.8	11	-1.1
Unrelated individual.	56,564	13,181	414	23.3	0.6	56,439	12,867	331	22.8	0.5	* -314	* -0.5
Race[3] and Hispanic Origin												
White alone	243,085	29,936	816	12.3	0.3	243,144	30,210	655	12.4	0.3	274	0.1
White alone, not Hispanic.	195,167	18,796	722	9.6	0.4	195,288	19,026	549	9.7	0.3	230	0.1
Black alone	40,615	11,041	506	27.2	1.3	40,577	10,696	439	26.4	1.1	* -345	* -0.8
Asian alone	17,063	1,785	176	10.5	1.0	17,003	1,884	165	11.1	1.0	99	0.6
Hispanic (of any race).	54,145	12,744	513	23.5	0.9	54,138	12,760	476	23.6	0.9	16	0.0
Sex												
Male	153,361	20,119	568	13.1	0.4	153,373	20,150	467	13.1	0.3	31	0.0
Female	159,605	25,199	573	15.8	0.4	159,610	25,117	530	15.7	0.3	-82	-0.1
Age												
Under 18 years	73,625	14,659	455	19.9	0.6	73,535	15,009	427	20.4	0.6	* 351	* 0.5
18 to 64 years	194,833	26,429	648	13.6	0.3	194,971	26,208	544	13.4	0.3	-221	-0.1
65 years and over	44,508	4,231	227	9.5	0.5	44,477	4,050	169	9.1	0.4	-181	-0.4
Nativity												
Native..	271,968	37,921	943	13.9	0.3	272,249	38,068	808	14.0	0.3	146	0.0
Foreign born..	40,997	7,397	373	18.0	0.8	40,734	7,199	319	17.7	0.7	-198	-0.4
Naturalized citizen	19,147	2,425	173	12.7	0.9	19,132	2,217	140	11.6	0.7	* -208	* -1.1
Not a citizen	21,850	4,972	311	22.8	1.2	21,602	4,982	279	23.1	1.1	11	0.3

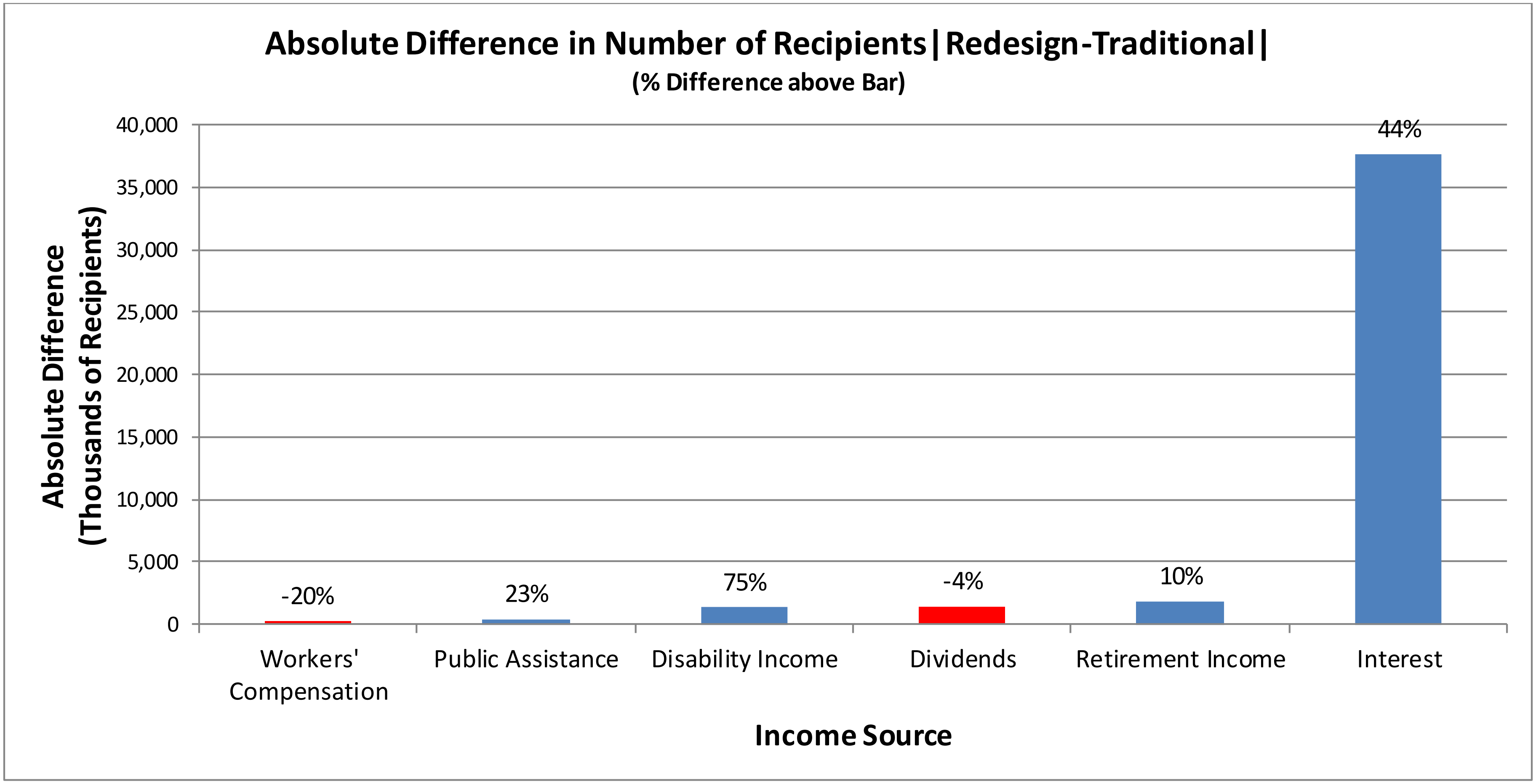
The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for all samples. The IC file is the single closest file (in normalized distance) to the MI median income and poverty averages, without accounting for imputation variance. The cells highlighted in pink are these where the statistical significance differs between this file and the MI average with imputation variance shown in Table 8.

Table 11: Comparison of Traditional and Closest Income-Consistent File: Poverty by Selected Characteristics, Continued

Characteristic	Total	Traditional (T)				Total	Closest Income-Consistent (CIC)				Difference in Poverty (CIC-T)/T	
		Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI		Number in Poverty (Thousands)	90 percent CI	Percent in Poverty	90 percent CI	Number	Percent
Region												
Northeast	55,478	7,046	437	12.7	0.8	55,481	7,021	409	12.7	0.7	-25	0.0
Midwest	66,785	8,590	430	12.9	0.7	66,758	8,728	390	13.1	0.6	137	0.2
South.	116,961	18,870	706	16.1	0.6	116,959	18,796	613	16.1	0.5	-74	-0.1
West.	73,742	10,812	434	14.7	0.6	73,785	10,723	379	14.5	0.5	-90	-0.1
Residence												
Inside metropolitan statistical areas.	265,915	37,746	1,007	14.2	0.4	265,773	37,611	960	14.2	0.3	-136	0.0
Inside principal cities.	102,149	19,530	842	19.1	0.7	101,874	19,070	783	18.7	0.6	* -459	* -0.4
Outside principal cities.	163,767	18,217	738	11.1	0.4	163,900	18,540	653	11.3	0.4	324	0.2
Outside metropolitan statistical areas	47,050	7,572	665	16.1	1.0	47,210	7,656	621	16.2	0.7	84	0.1

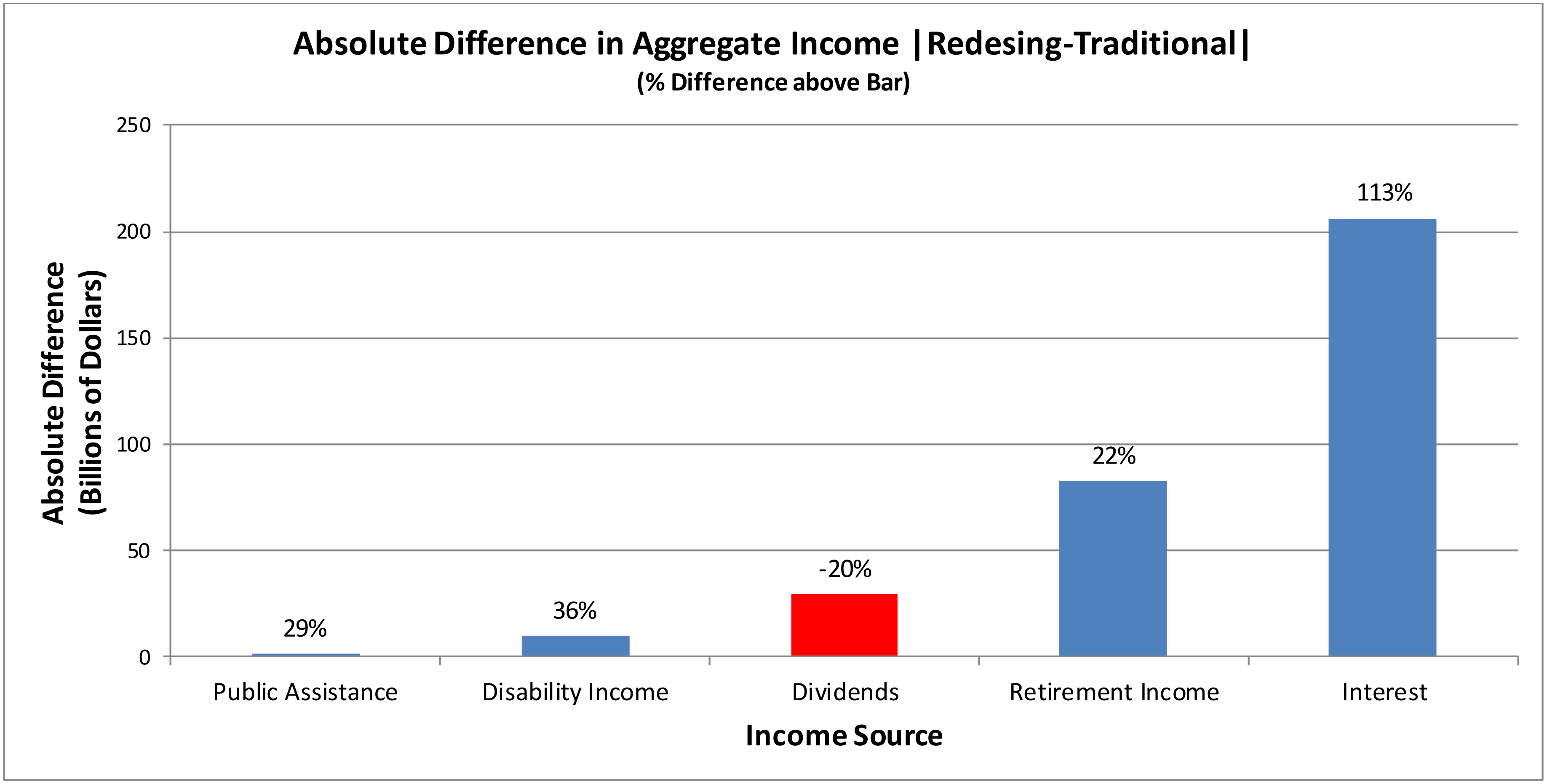
The 2014 CPS ASEC included redesigned questions for income and health insurance coverage. All of the approximately 98,000 addresses were eligible to receive the redesigned set of health insurance coverage questions. The redesigned income questions were implemented to a subsample of these 98,000 addresses using a probability split panel design. Approximately 68,000 addresses were eligible to receive a set of income questions similar to those used in the 2013 CPS ASEC and the remaining 30,000 addresses were eligible to receive the redesigned income questions. * if statistically significant difference at the 90% confidence level. Income in 2013 dollars. Households and people as of March of the following year. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see ftp://ftp2.census.gov/programs-surveys/cps/techdocs/cpsmar13.pdf. Standard errors calculated using replicate weights for all samples. The IC file is the single closest file (in normalized distance) to the MI median income and poverty averages, without accounting for imputation variance. The cells highlighted in pink are these where the statistical significance differs between this file and the MI average with imputation variance shown in Table 8.

Figure 1: Income Recipiency Differences between the Traditional and Redesign Samples



Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement.

Figure 2: Aggregate Income Differences between the Traditional and Redesign Samples



Source: U.S. Census Bureau, Current Population Survey, 2014 Annual Social and Economic Supplement