IPUMS

Working Papers

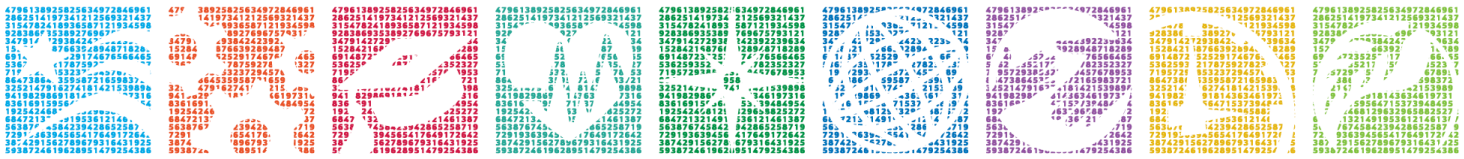
Including the ASEC Oversamples in Linked CPS Data

Renae Rodgers†
University of Minnesota

July 2025

Working Paper No. 2025-01
DOI: https://doi.org/10.18128/IPUMS2025-01

†Address correspondence to Renae Rodgers: rodge103@umn.edu. This work was supported by grant R01HD067258 from the Eunice Kennedy Shriver National Institute for Child Health and Human Development. Comments and suggestions from Sarah Flood and Julia Rivera Drew are gratefully acknowledged.



Including the ASEC Oversamples in Linked CPS Data

Renae Rodgers

Abstract:
The Current Population Survey (CPS) is an important source of labor force data, and its Annual Social and Economic Supplement (ASEC) provides additional information about health insurance coverage, income, and poverty in the United States. The ASEC questionnaire is administered to 1) all respondents to the March CPS Basic Monthly Survey (BMS), 2) an oversample of Hispanic households from 1976 forward, and 3) an oversample of households with children or non-white household members from 2001 forward. The CPS also has a short-term panel component. Creating CPS ASEC panel datasets by linking files from adjacent years allows for the examination of short-run change in health insurance coverage, income, and other topical areas. Many common methods used to link the CPS ASEC files only include March BMS respondents in the ASEC panels; however, many oversample individuals can also be linked. The main challenges to including CPS ASEC oversample records in ASEC panels are the multi-step process required for linking, bridging changing household identifiers over time, and handling duplicated household identifiers and false positive matches that arise due to the recycling of Census Bureau household identifiers prior to 2005. This paper describes our approach for addressing the complexities and challenges of linking ASEC oversamples across adjacent years and the method by which these oversample records were incorporated into the custom IPUMS CPS linking keys CPSID, CPSIDP, and CPSIDV for the period from 1989 to 2024. Properly including these records in linked ASEC data substantially increases the linked sample size. From 1976-2001, linked ASEC sample sizes are between 4 and 9% higher, and from 2002 forward, linked ASEC sample sizes are more than 30% higher than when only the March BMS respondents are included in the panel.

Introduction

The Current Population Survey (CPS), a monthly survey of labor force and demographic information, is the most important source of labor force data in the United States. The CPS has a rotating panel structure, which results in large overlap of samples in adjacent months and adjacent years while avoiding overburdening survey respondents (U.S. Census Bureau, 2019). IPUMS CPS (cps.ipums.org) provides the research community streamlined access to this important data set as well as several custom linking keys that lower the barrier to using the CPS panel component.

The CPS Annual Social and Economic Supplement (ASEC), sometimes referred to as the March Supplement, is the most widely used CPS supplement, including a wealth of additional information about health insurance coverage, income, and poverty, and is an invaluable data source for economic and demographic research in the United States. The ASEC questionnaire is administered to 1) all respondents to the March CPS Basic Monthly Survey (BMS)[1], 2) an oversample of Hispanic households from 1976 forward, and 3) an oversample of households with children or non-white household members from 2001 forward.

The panel component of the CPS provides an excellent opportunity for studying short-run change, yet many methods for linking ASEC files across years exclude the ASEC oversamples (Madrian & Lefgren, 2000; Feng, 2001). The lack of official documentation explaining the relationship between the CPS rotation pattern, the composition of ASEC oversamples, the meaning of the month-in-sample (MIS) value of oversample records in the ASEC file, and the implications of this relationship for the linkability of the ASEC oversample records may explain why popular methods for linking ASEC files fail to accommodate oversample records. Pacas and Rodgers (2023) describe this relationship and demonstrate how to recover linkable oversample records when matching ASEC files across years, resulting in a dramatic increase in linked sample size for linked ASEC data between 2005 and 2020.

This paper describes how IPUMS CPS enables ASEC oversample linkages using IPUMS CPS custom linking keys CPSID, CPSIDP, and CPSIDV from 1989-2024. This effort lowers the barriers to including ASEC oversamples in linked IPUMS CPS datasets and allows researchers to reap the benefits of increased linked sample size. To preview the implications of this work, including ASEC oversample records in ASEC panel datasets increases the linked sample size by between four and nine percent between 1989 and 2001 and by over 30% from 2002 forward.

Several challenges must be overcome to correctly include oversample records in ASEC panels between 1989 and 2024. First, because MIS is not a reliable indicator of an oversample household's place in the CPS rotation pattern, ASEC oversample records cannot be linked using the method applied to the CPS BMS data (Drew et al., 2014; Flood & Pacas, 2017).

[1] While the CPS BMS data has been collected in some form since the 1940s (U.S. Census Bureau, 2006), public BMS data is only available from the Census Bureau back to 1976. CPS ASEC files are publicly available back to 1962. In the years prior to 1976, the CPS ASEC sample is identical to the March BMS sample.

Furthermore, ASEC oversample records are not identified as such in the ASEC file. This difficulty is overcome by implementing a multi-step process for ASEC linking that identifies oversample records by first linking the March BMS to the ASEC file as described by Flood and Pacas (2017). ASEC records that do not match to the March BMS file from the same year are oversample records. Different linking criteria are then applied to oversample records to account for the unreliability of MIS values for these records (Pacas & Rodgers, 2023).

Second, Census Bureau household identifiers change over time. In many instances, these changes can be bridged using the correct combination of linking keys. However, a change in household identifiers between 1995 and 1996 prevents linking ASEC data across these years.

Finally, prior to 2005 the Census Bureau implemented a recycling pattern for household identifiers that poses a particular challenge for ASEC oversample linking. This recycling pattern causes false positive matches across ASEC oversamples that cannot be avoided by using MIS to enforce the expected 4-8-4 rotation pattern. In addition, in ASEC oversamples between 2001 and 2005, this recycling pattern results in households with duplicated Census Bureau identifiers within a single ASEC oversample, posing an additional challenge to linking ASEC data in these years.

Background

The CPS is a monthly survey; in some months, supplemental surveys on topics such as food security and education are also included. The Annual Social and Economic Supplement (ASEC) is the largest of all the CPS supplements and is collected over several months. The CPS rotating panel structure results in a large overlap of samples in adjacent months and adjacent years while avoiding overburdening survey respondents (U.S. Census Bureau, 2019).

The CPS Basic Monthly Survey (BMS) has a rotating panel design wherein each household is first interviewed for four consecutive months, then rotates out of the CPS data collection for eight months, and is then interviewed for another four consecutive months before permanently exiting CPS data collection. A household's place in this 4-8-4 rotation pattern is referred to as the household's "month-in-sample", or MIS. A household's MIS value is the number of times a household[2] has been eligible to participate in the CPS BMS; MIS values of one through four are assigned in the first set of interviews, and interviews after the eight-month break have MIS values of five through eight. This 4-8-4 rotation pattern results in overlap in the CPS samples across months, which researchers can leverage to link observations of members of the same households over time. For example, persons in households in MIS one through four in November of 2022, will be in their second half of the CPS rotation (MIS five through eight) in November of 2023; records with MIS values of 1 in November of 2022 can be linked to those

[2] Note that "household" refers to dwelling in the CPS sample. Addresses are sampled for the CPS and if one or more individuals leaves a physical dwelling between interviews, they are not followed to their new address.

with MIS values of 5 in November of 2023, and so on. See Drew et al. (2014) for in-depth discussion of types of links made possible by the CPS rotation pattern.

The CPS is most well known for the ASEC supplement, which is a primary source of information for national estimates of health insurance coverage and poverty. This supplement, first introduced in 1947 as the Annual Demographic File (ADF), was redubbed the Annual Social and Economic Supplement (ASEC) in 2003 (U.S. Census Bureau, 2006). IPUMS CPS refers to this dataset as the "ASEC" for all years. In addition to poverty, health insurance coverage, and the monthly battery of labor force and demographic questions, the ASEC includes information on income, receipt of noncash benefits, and migration.

[Table 1: Composition of the CPS ASEC files, 1976-present]

From 1947 to 1975, the ASEC sample only included March Basic Monthly respondents who were (and still are) administered the ASEC survey at the time of their March BMS interview (U.S. Census Bureau, 2006). ASEC oversamples have always included respondents identified in other months who would never receive a BMS interview in March due to the 4-8-4 rotation pattern. Table 1 shows the composition of the ASEC file over time, the month and rotation group in which eligible oversample households are identified (their source month and rotation group), as well as interview month for each oversample. Note that most oversample households are administered the ASEC survey in a different month than their source month.

Beginning in 1976, the ASEC sample was expanded to oversample Hispanics, increasing the ASEC sample size by roughly 2,500 households; by 1990, this oversample contains approximately 3,000 households. A household qualifies for the Hispanic oversample if at least one household member identifies as Hispanic (U.S. Census Bureau, 2006). As shown in Table 1, from 1976-2000, Hispanic oversample households were drawn from all MIS groups in the November BMS sample in the prior year (e.g., Hispanic oversample households included in the 1990 ASEC file were interviewed in the 1989 November BMS). In 2001, the Hispanic oversample was expanded to include eligible households from MIS 1 and 5 from the April BMS; this expansion added approximately 2,000 Hispanic oversample households (U.S. Census Bureau, 2006). The 4-8-4 rotation pattern of the CPS BMS ensures that none of the households eligible for the Hispanic oversample would have been interviewed in March.

Another CPS sample expansion occurred in 2001 with the goal of making the CPS suitable for generating reliable state-level estimates of low-income children without health insurance and thus allowing evaluation of the State Children's Health Insurance Program, or SCHIP (U.S. Census Bureau, 2006). This sample expansion occurred in both the BMS sample, where approximately 12,000 households were added across 30 states and the District of Columbia, and the ASEC, which added a SCHIP oversample of approximately 20,000 households. ASEC SCHIP oversample households are those non-Hispanic households that have at least one child aged 18 or younger or a non-White member[3] (U.S. Census Bureau, 2006). The CPS BMS

[3] While it may seem counter intuitive to include households that do not necessarily have children in an oversample with the purported purpose of evaluating the State Children's Health Insurance Program,

expansion was phased in between September of 2000 and July of 2001. However, the 2001 March BMS file available from the Census Bureau appears to be missing the BMS sample expansion records and contains a lower-than-expected number of households relative to BMS household sample sizes in previous and subsequent monthly BMS samples (only 60,015 households). Table 1 shows the composition of the SCHIP oversample from 2001 to the present. When the SCHIP oversample was first introduced to the ASEC, households were drawn from the previous November BMS MIS 1, 5-8 groups as well as April MIS 1 and 5[4]. In 2003, SCHIP households were taken from October MIS 8 instead of November MIS 6. Finally, in 2004, the SCHIP oversample was drawn from MIS 8 groups in the previous August, September, and October BMS samples as well as from MIS 1 and 5 in April. The SCHIP oversample composition introduced in 2004 remains in place at time of writing.

Many oversample households are interviewed more than the maximum of eight times prescribed by the CPS rotation pattern. Prior to 2001, all Hispanic oversample households were eligible to be interviewed ten times - their first ASEC interview occurred during their eight-month break in the BMS rotation pattern and their second ASEC interview occurred after their eighth BMS interview (U.S. Census Bureau, 2002). This pattern of ten possible interviews continues for the November Hispanic oversample households from MIS groups 2-4 and 6-8 beginning in 2001 (U.S. Census Bureau, 2006). When the SCHIP oversample was introduced in 2001, the households identified in November MIS groups 6-8 were administered the ASEC survey at a ninth interview in March following their completion of the eight BMS interviews. Likewise, SCHIP oversample households identified in MIS 8 of August, September, and October are administered the ASEC after they complete the full CPS BMS rotation, resulting in a total of nine possible CPS interviews.

ASEC oversample households that are administered the ASEC survey at the time of one of their regularly scheduled BMS interviews are not administered any topical supplemental questionnaires that may be fielded during that month (U.S. Census Bureau, 2006). For example, Hispanic and SCHIP oversample households that receive the ASEC survey at the time of their April BMS interview do not answer the Child Support Supplement questionnaire that is fielded every other year in April. As a result, ASEC information from oversample households cannot be combined with other CPS topical supplement data from February or April even when the oversample households are interviewed during these months.

Linkable oversample records

A subset of both the Hispanic and the SCHIP oversamples will be observed in adjacent ASEC files and can be linked across years (see Figure 1).

---

these households are in the ASEC oversample data from 2001 onward and that their presence is expected is confirmed by official documentation (U.S. Census Bureau, 2006).

[4] Note that the November MIS 1 and 5 groups are in MIS 4 and 8 in February; these households are given the ASEC supplement during their February BMS interviews.

[Figure 1. ASEC oversample MIS groups eligible to link across ASEC files]

Both the Hispanic and SCHIP ASEC oversamples are constructed such that households in these oversamples will never appear in the March BMS sample (to whom the ASEC supplement is always administered at the time of their BMS interview). However, by understanding how the ASEC oversamples are drawn and the 4-8-4 rotation pattern of the CPS BMS, we can link a subset of the oversample records across ASEC files (Pacas & Rodgers, 2023). As shown in Figure 1, Hispanic oversample households from November MIS 1-4 will be in November MIS 5-8 in the following year; when the Hispanic oversample is expanded to include MIS groups from April in 2001, households from April MIS 1 will be in April MIS 5 the following year. Likewise in the SCHIP oversample, households taken from MIS 1 in November and April will be in MIS 5 in these months in the following year. Half of the Hispanic oversample is eligible to link across years (four of eight MIS groups prior to 2001 and five of ten MIS groups from 2001 to present). In the SCHIP oversample, two of seven MIS groups (28.57% of oversample households) are eligible to link across years. A combined total of six of 15 oversample MIS groups (40% of oversample households) are eligible to link across years between 2001 to 2003 and seven of 17 oversample MIS groups (41.2% of oversample households) are eligible to link across years from 2004 to the present.

However, unlike the March BMS portion of the ASEC sample, ASEC oversample records' place in the CPS rotation pattern -- and thus their eligibility to link across years -- is not accurately reflected in their MIS values in the ASEC file. As a result, methods for linking CPS data that rely on MIS to define linkable records (Drew et al., 2014; Feng, 2001; Madrian & Lefgren, 2000) will miss records that are linkable across ASEC oversamples (Pacas & Rodgers, 2023). By enforcing the 4-8-4 rotation pattern when linking March BMS records across ASEC files and ignoring MIS values and relying only on Census identifiers as linking keys for linking oversample records across ASEC files, IPUMS CPS has updated its custom linking keys CPSID, CPSIDP, and CPSIDV to correctly account for the ASEC oversample records when linking adjacent ASEC files. This update allows IPUMS CPS data users to link all possible ASEC records across adjacent years in a single step using a single linking key.

Incorporating ASEC oversample records into IPUMS linking keys

This section describes the method for making ASEC oversample records linkable using IPUMS CPS custom linking keys, CPSID, CPSIDP, and CPSIDV in data from 1989 forward, detailing the challenges encountered and the solutions implemented and, finally, provides linkage rates using the expanded CPSID, CPSIDP, and CPSIDV keys for user reference.

CPSID(P)

CPSID and CPSIDP, collectively referred to henceforth as CPSID(P), are custom linking keys available from IPUMS CPS that lower the barrier to leveraging the Current Population Survey's

rotating panel design (Drew et al., 2014; Flood et al., 2020). CPSID(P) is also available for individuals in the March BMS, who are also in the ASEC, which allows users to link ASEC information for these households to other months of the CPS, including those with topical supplements such as food security (Flood & Pacas, 2017). The procedures described next cover the expansion of CPSID(P) to accommodate links between ASEC oversample records from 1989 forward.

The basic procedure for assigning linkable CPSID(P) values to ASEC oversample records is followed for all ASEC files from 1989 forward. First, CPSID(P) values must be assigned to records in the March Basic Monthly Survey (BMS) file from the same year (See Drew et al., 2014 for detailed description of CPSID(P) generation for BMS files). Second, March BMS records are linked to their counterparts in the ASEC file of the same year. This step generates the IPUMS variables MARBASECIDP and MARBASECIDH that can be used to link March BMS and ASEC records and allows assignment of CPSID(P) values from the March BMS file to the March BMS component of the ASEC file (See Flood & Pacas, 2017 for a detailed description of MARBASECIDP and MARBASECIDH creation). This step also identifies the oversample records within the ASEC file - any records in the ASEC file that do not get matched to the March BMS file must be oversample records.[5] Finally, ASEC oversample records in the current year are matched to ASEC oversample records in the previous year. Records in the current year that link successfully are assigned the CPSID(P) value of their match in the previous year. Records that do not match successfully to the previous year are assigned a new, unique CPSID(P) value.

From 2005 forward, IPUMS follows the method laid out by Pacas and Rodgers (2023) for matching the ASEC oversample records. By this method, once ASEC oversample records have been identified in step two described above, only Census Bureau household and person identifiers are used to link ASEC oversample records across years, and month-in-sample (MIS) values are ignored. As described below, several extra steps are required to address duplicated household identifiers in the ASEC data from 2001 to 2005 and false positive matches that occur prior to 2005 due to the recycling of Census Bureau household identifiers during this period.

Changing household identifiers

The construction and assignment of Census Bureau household identifiers change over time. Changes to original household identifiers can be bridged between the 1993 and 1994 ASEC files and between the 2004 and 2005 ASEC files. However, changes that are meant to preserve confidentiality or that are not explained by official Census Bureau documentation prevent some ASEC-ASEC links between 1994 and 1995 and all ASEC-ASEC links between 1995 and 1996.

In 1994, the corrected H_IDNUM variable used to make the March BMS-ASEC link (Flood & Pacas, 2017) is used to link the 1994 ASEC oversample back to the 1993 ASEC oversample

[5] There are two notable exceptions to this rule: in 1995 where non-response BMS households are not linkable between the March BMS and ASEC files, and in 2001 when BMS sample expansion households are included in the ASEC file but not in the March BMS file.

and to link the 1995 ASEC oversample back to the1994 ASEC oversample. Though the latter years require changing identifiers to be bridged in the BMS data (Drew et al., 2014), the correction of the H_IDNUM error described in Flood and Pacas (2017) is sufficient to successfully link the 1994 ASEC oversample across years.

An oddity in the Census household identifiers of non-response households in 1995 cannot be overcome when linking. Because the ASEC file is composed of all March BMS respondents plus any relevant oversamples, all records in the March BMS sample should also appear in the ASEC file. However, in the 1995 data, none of the non-response households in the March BMS file link to the 1995 ASEC. Within the 1995 ASEC file, many non-response households have duplicated Census identifiers and none of these identifiers appear in the 1995 March BMS file, suggesting that March BMS non-response households are, in fact, present in the 1995 ASEC file, but their household identifiers have been modified such that they cannot be linked across the March BMS and ASEC data. In addition, due to the rotation pattern, roughly half of the households – including non-response households – from the 1995 ASEC file should link back to the 1994 ASEC file. However, none of the non-response households from the 1995 ASEC file appear in the 1994 ASEC file. As a result, no non-response household records can be linked between 1995 and 1994 ASEC files. This inability to separate non-response March BMS households from oversample households and the inability to link any non-response households between the 1995 and 1994 ASEC oversample results in unusually low household linkage rates across these years (see Table 20). The source of this issue is not explained in any official Census Bureau documentation.

Unfortunately, a change in the way Census Bureau identifiers were assigned between 1995 and 1996 cannot be bridged and thus linking any records across these ASEC files is impossible. The change in household identifier assignment may have been implemented to preserve confidentiality in light of a sample reduction that occurred in January of 1996. Due to funding cuts, CPS sample decreased by 6,000 eligible housing units and 38 sampling areas across Illinois, Massachusetts, Michigan, New Jersey, North Carolina, Ohio, and Pennsylvania as well as within Los Angeles and New York City (U.S. Census Bureau, 2002).

The sample redesign that began in May 2004 was also accompanied by a change in Census household identifiers. In the 2004 ASEC file, households are identified by the Census variables H-IDNUM and H-HHNUM; beginning in 2005, the variables H-IDNUM1 and H-IDNUM2 are used to uniquely identify households. H-HHNUM is the last digit of the new H-IDNUM2 identifier and the 2005 oversample is linked back to 2004 using H-IDNUM1 and this H-HHNUM equivalent.

Recycling of Census Bureau household identifiers

Feng (2001) describes a pattern of household identifier recycling in the CPS Basic Monthly data wherein the household identifiers of a household exiting the CPS rotation for good (those in MIS 8) are reassigned to a new household entering the CPS rotation for the first time in the next month (MIS 1). In addition, when this new CPS household begins the eight month break (after MIS 4), their identifier will be passed back to a household that is re-entering the CPS sample after *their* eight month break (MIS 5). The 4-8-4 rotation of the CPS BMS means that identifiers

can safely be re-used in this way without the same identifier being assigned to different households in the same BMS sample. However, this recycling pattern does require that MIS be used to enforce rotation pattern progression when linking records across months to avoid incorrect matches.

Examination of the CPS basic monthly data confirms that this pattern is present from 1989 through May of 2004 and appears to have been ended as part of the CPS sample redesign that took place between April 2004 and July 2005 (U.S. Census Bureau, 2006). Table 2 demonstrates the impact of this identifier recycling scheme for BMS datasets.

[Table 2. The recycling pattern of a Census Bureau household identifier in the CPS Basic Monthly, 1989-2004]

Two major challenges for ASEC oversample linking prior to 2005 arise from this household identifier recycling. First, because the MIS value assigned to ASEC oversample records cannot be used to enforce the CPS rotation pattern, false positive matches may be made across years when using only Census Bureau identifiers in these years. Second, with the introduction of the SCHIP oversample and the addition of two new Hispanic oversample groups in 2001, the recycling pattern results in duplicate household identifiers within ASEC oversamples from 2001-2005. Both of these challenges were largely overcome by recognizing and leveraging alternative patterns in MIS progression in the ASEC panels and creating auxiliary variables in the cross-sectional ASEC data.

False Positive Matches

If the CPS 4-8-4 rotation pattern is not accounted for when linking BMS records across months prior to 2005, the household identifier recycling may result in incorrect matches across years, for example, Household V in year 1 shown in Table 2 above may incorrectly link to Household Y in year 2. Luckily, the potential linking complications presented by this recycling pattern are easily avoided when linking BMS data by using MIS values to enforce the correct movement of households through the CPS rotation pattern across months (Drew et al., 2014; Feng, 2001). CPSID(P) takes care of this enforcement for all BMS links (Drew et al., 2014). However, the MIS values of the ASEC oversample households are not accurate representations of these households' place in the CPS rotation (Flood & Pacas, 2017; Pacas & Rodgers, 2023) and thus using MIS to enforce the correct CPS rotation pattern progression will not avoid false positive matches across files for ASEC oversample records. Fortunately, examination of the linked ASEC data during these periods reveals other patterns of MIS progression that can be used to identify and prevent false positive matches in ASEC panel datasets. These patterns vary with time and oversample type, but are detectable in all linked ASEC data between 1989 and 2003. Unfortunately, no pattern in MIS progression is detectable in ASEC panels that include 2004 ASEC data for either the Hispanic or the SCHIP oversample and, as a result, false positive matches cannot be reliably identified based on identifiers and MIS values in these panels.

[Table 3. An illustration of the impact of household identifier recycling on linking November Hispanic oversample households]

Table 3 illustrates how household identifier recycling can lead to false positive matches across ASEC oversamples. As per the 4-8-4 CPS rotation pattern, MIS 1-4 in November of year 1 are eligible to link to MIS 5-8 in November of year 2. For clarity, only rotation groups MIS 1 and MIS 5 are shown in Table 3. Those households in MIS 1 in November of year 1 who are interviewed again in November of year 2 will correctly link to their MIS 5 interview in November of year 2. However, due to the household identifier recycling pattern, MIS 5 interviews in November of year 1 will incorrectly link to households with their same identifier in MIS 1 in November of year 2 if the 4-8-4 CPS rotation pattern is not enforced using MIS. Because the MIS values of the ASEC oversample records are not representative of the household's place in the rotation pattern, if the households shown in Table 3 are part of the ASEC oversample, Household A will correctly link across years (MIS 1 to MIS 5), but Household D will incorrectly be linked to Household G.

Fortunately, prior to 2002, even though MIS values are not an accurate marker of where a household is in the rotation pattern, MIS values of oversample households linked using only Census Bureau identifiers do have identifiable patterns that make it possible to identify and suppress likely false positive matches. While these patterns vary over time and between the Hispanic and SCHIP oversamples; for each pattern, analysis of age and sex of the linked records supports the identification of false positive matches across ASEC files. As an illustration, Table 4 shows MIS value progressions of households linked between the 1989 and 1990 Hispanic ASEC oversamples using only Census Bureau household and person identifiers.

[Table 4. Month-in-sample values for Hispanic oversample households that link between the 1989 and 1990 ASEC files using only Census identifiers]

In this example, note that there are two distinct groups of MIS value pairs in these linked data - one where MIS values are "on-rotation" and appear to follow the rotation pattern (e.g., MIS=1 in 1989 and MIS=5 in 1990, cells highlighted in red), and those where MIS values are "off-rotation"-- the inverse of the expected MIS rotation (e.g., MIS=5 in 1989 and MIS=1 in 1990). Note also that the households with off-rotation MIS value progressions are much more common than those with on-rotation MIS value progressions. When links from these two groups are validated on age and sex (see Table 7), the vast majority of the links with off-rotation MIS value progressions have the same sex and plausible age values across years, whereas only about half of the links with on-rotation MIS values have the same sex and less than three percent have plausible age values across years.[6] These validation results suggest that the oversample links that have on-rotation MIS value progressions are, in fact, false positive matches; these records are not assigned linking CPSID(P) values.

[6] While many common methods of validating linkages of CPS records across months also use race (Madrian & Lefgren, 2000; Feng 2001; Drew et al., 2014; Rodgers & Flood, 2023; Pacas & Rodgers 2023), I have excluded it here. Because the links being validated here are from the ASEC oversamples only and nearly all individuals that identify as Hispanic are also categorized as white, even false positive links among Hispanic oversample records are likely to be identified as white in the data at both time points.

This pattern - off-rotation MIS value progressions are suspected false positives - holds true for all Hispanic oversample links made using Census Bureau identifiers prior to 2002 with four exceptions. In the 1994 and 2000 ASEC Hispanic oversamples, all households have an MIS value of 5. However, as shown in Table 5, panels involving these oversamples have their own pattern which makes it possible to identify probable false matches. When oversample records from these years are linked to the previous year, links that have MIS values of one through four in the previous year are less frequent and validate at similar rates to the false positive on-rotation links in other years during this period. These are false positive matches; these links are not assigned linking CPSID(P) values. When these oversamples are linked to the following year, links with MIS values of five through eight are less frequent and validate at lower rates (see Table 7). These are suspected false positive matches; these records are not assigned linking CPSID(P) values.

[Table 5. Month-in-sample values for households that link across ASEC files, 1994-2000]

In the 2002-2003 ASEC panel, there is an additional group of links made using only Census Bureau identifiers that appears suspicious. As shown in Table 6, in addition to the on-rotation matches, Hispanic oversample households with an MIS value of 4 in 2002 link to records of all MIS values in 2003. However, these links are very few and validation on demographic variables suggest that these Census Bureau identifier links are also false positives; these records are not assigned linking CPSID(P) values. Table 7 shows validation rates of suspected true positives compared to suspected false positive matches.

The SCHIP oversample, included in the ASEC data from 2001 forward, is also susceptible to false positive matches when using only Census Bureau identifiers prior to 2005 due to the household identifier recycling. Between 2001 and 2002 and between 2002 and 2003, MIS values for the SCHIP oversample records also follow patterns that suggest which links are spurious, and validation on age and sex bears out these suspicions. Table 6 shows these suspected false positives highlighted in red. In the 2001-2002 linked ASEC data, linked households whose MIS values follow the CPS rotation pattern (e.g. MIS=1 in 2001 and MIS=5 in 2002) are few and (as shown in Table 7) have low validation rates on age and sex compared to the linked households with off-rotation MIS value progressions. These on-rotation links are suspected false positive matches and these records are not assigned linking CPSID(P) values. In the 2002-2003 ASEC panel created using only Census Bureau identifiers, all SCHIP oversample households with MIS values of 1 or 4 in the 2002 ASEC data are suspected false positives; once again, validation on age and sex supports this suspicion (see Table 7) and these records are not assigned linking CPSID(P) values.

[Table 6. Month-in-sample values for households that link across ASECs, 2001, 2002, 2003]

[Table 7. Validation on age and sex of suspected correct and suspected false positive matches across ASEC oversamples]

Records linked using Census Bureau identifiers from 2003-2004 and 2004-2005 do not show a clear pattern of MIS progressions for either ASEC oversample. Table 8 shows MIS value progressions in the Hispanic and SCHIP linked oversamples from 2003-2004 and 2004-2005. However, as household identifier recycling was still in use during this period, these data linked using only Census Bureau identifiers do likely contain false positive matches. CPSID(P) is designed to provide mechanical links, relying on Census Bureau identifiers and enforcing the 4-8-4 rotation pattern (Drew et al., 2014; Flood et al., 2020). As there is no reliable way to identify false positive matches without resorting to other demographic variables, false positive matches are made by CPSID(P) in the 2003-2004 and 2004-2005 ASEC panels. Users are strongly encouraged to use the IPUMS validated linking key CPSIDV (see below) or perform their own validation of oversample linkages between these years.

[Table 8. Month-in-sample values for households that link across ASECs, 2003, 2004, 2005]

Duplicated household identifiers

The use of household identifier recycling during the period when the SCHIP oversample was introduced and the Hispanic oversample was expanded in 2001 resulted in duplicated household identifiers within ASEC files between 2001 and 2005.[7] The MIS groups from which households are drawn (the source MIS groups) for both the Hispanic and the SCHIP oversamples were chosen because, in the course of the normal CPS rotation, these households are not interviewed in March and do not overlap with one another (that is, households taken from one oversample MIS source group will not appear in any other oversample source group) (U.S. Census Bureau, 2002). However, in the years after the oversample expansion and prior to 2005, the household identifiers of these households *do* appear in other oversample source groups but assigned to different households due to the household identifier recycling scheme. The overall result is that ASEC oversamples from this period contain duplicate household identifiers, both within and across oversamples.

Table 9 shows the source MIS groups for each ASEC oversample between 2001 and 2004[8] along with the MIS groups to which recycled household identifiers move across months; highlighted cells indicate the MIS groups within each month (column) that contribute to the ASEC oversample. Oversample MIS source groups that appear on the same row in different months result in duplicated household identifiers within that ASEC file. For example, in 2001, oversample households from November MIS 4 may have the same household identifiers as different households in April MIS 1 and oversample households from MIS 8 may have the same

[7] Note that while re-designed Census Bureau identifiers H_IDNUM1 and H_IDNUM2 uniquely identify all households in the 2005 ASEC oversample, the keys used to bridge this change in identifiers between 2004 and 2005, H_IDNUM1 and HUHHNUM, do not uniquely identify all ASEC oversample households in 2005.

[8] Note that the composition of the ASEC oversamples is identical between 2004 and 2005 and so the problem of duplicate identifiers when using bridge household identifiers in the 2005 ASEC are reflected in Panel D of Table 9.

household identifiers as different households in April MIS 5.[9] Both of these pairs of MIS groups contributed to the Hispanic and the SCHIP oversamples in 2001, resulting in a potential for duplicated ids both within and across the Hispanic and SCHIP oversamples in the 2001 ASEC file. Each highlighted MIS group on the same row in different months represents an opportunity for duplication. To continue with the 2001 example, household identifiers for Hispanic oversample households from November MIS 8 may be duplicated up to four times in the 2001 ASEC file: once for the Hispanic November MIS 8 group, once for the Hispanic April MIS 5 group, once for the SCHIP November MIS 8 group, and once for the SCHIP April MIS 1 group.

[Table 9. Duplicated Census household identifiers within ASEC oversamples due to household identifier recycling, 2001-2004]

Because CPSID(P) is a mechanical link based only on Census Bureau identifiers (Drew et al., 2014; Flood et al., 2020), these duplicate household identifiers present a challenge. For consistency with the CPSID(P) methodology, demographic characteristics are not used directly as linking keys. The approach for uniquely identifying as many linkable ASEC oversample records as possible is as follows. First, the number of duplicate records that need to be uniquely identified is reduced by retaining only those records whose primary Census household identifier appears in both ASEC oversamples to be linked. This step eliminates duplicated records that could never be linked across the relevant ASEC files, even if they were uniquely identified. Then, three auxiliary variables are created in these years to help account for the recycling pattern and uniquely identify records. These variables are the type of oversample household to which the record belongs, and indicators for whether a household is the first or the last household in the group of households with duplicated identifiers.

Because household identifiers can be duplicated across the ASEC oversamples, the first auxiliary linking variable created identifies the type of oversample household. Records from Hispanic oversample households are assigned a value of 1; SCHIP oversample households who are eligible for this oversample because they have children age 18 or younger are given a value of 2; SCHIP oversample households who are eligible for this oversample because they have a non-white household member[10] are given a value of 3; and SCHIP households that have both a non-white household member and a child age 18 or younger are given a value of 4. This oversample type variable, in combination with Census Bureau household identifiers, uniquely identifies households that have duplicate identifiers across Hispanic and SCHIP oversamples

[9] Because only the subset of the MIS source groups that meet the oversample selection criteria are included in the ASEC file, the recycling pattern will only result in duplicated household identifiers in the ASEC file if the household to which a recycled identifier is assigned also meets oversample selection criteria. For example, the identifier of a household that is included in the 2001 Hispanic or the SCHIP oversample from November MIS 8 will show up again in April MIS 5, but if the household to which that recycled identifier is assigned in April does not meet the selection criteria for the Hispanic or the SCHIP oversample, then no duplication will occur in the ASEC file. Nevertheless, as shown in Table 10, duplicate household identifiers do occur in every ASEC file between 2001 and 2005.

[10] Race categories change over time. Beginning in 2003 when the Census expanded the number of race categories to include multi-racial categories. In these years any race designation other than “white only” is considered non-white for the purposes of creating the oversample type variable.

and some households that have duplicate identifiers within the SCHIP oversample. This accounts for roughly half of the duplicate households in each year between 2001 and 2005; remaining records with duplicated household identifiers are duplicated within the Hispanic oversample and within SCHIP oversample subcategories. Table 10 shows the number of duplicated records using only Census identifiers and with the addition of the oversample type variable to the set of linking keys for the full oversample, when each year is the focal year and the oversample is reduced to households whose identifiers appear in the previous year's oversample, and when each year is the target year and the oversample is reduced to households whose identifiers appear in the following year's oversample .

[Table 10. Records with duplicated identifiers within the ASEC oversample]

Remaining households with duplicated Census Bureau identifiers in the 2001-2005 ASEC files are further de-duplicated by leveraging a household's order in the file. In addition to the Census Bureau household identifiers, the ASEC files contain a variable called H-SEQ, which is a simple 1-n numbering of households within the file. H-SEQ differentiates households even when the Census Bureau identifiers are duplicated within an ASEC file. After accounting for oversample type, most duplicated Census Bureau household identifiers are shared by only two actual households, as per H-SEQ. Table 11 shows the number of person records with a Census Bureau household identifier that is shared across multiple H-SEQ values after accounting for duplicate household identifiers that can be distinguished by oversample type. Records that have a Census Bureau household id that contain more than two unique households as defined by H-SEQ after accounting for oversample type are not considered for linking across ASEC oversamples, as there is no reliable way to distinguish the correct matches for these households that does not directly rely on demographic information. Only 55 potentially linkable records from 2002-2005 are removed from consideration for linking for this reason.

[Table 11. The number of records in Census Households with multiple H-SEQ values after accounting for households uniquely identified by oversample type]

Further examination of the data suggests that within a pair of households with duplicate Census household identifiers, the first household in the set should properly link back to the previous year's oversample and the last duplicate in the set should properly link forward to the next year's oversample. These links are achieved through the second and third auxiliary variables: indicators for whether a household as defined by H-SEQ is the first or last household within the ASEC oversample with the same Census household identifiers. When linking backward, the value of the flag for the last duplicate in the set in the focal year is compared to the value of the flag for the first duplicate in the set in the previous year.

For example, consider the households with the primary Census household identifier 415261329300866 in the 2001 and 2002 ASEC oversamples shown in Table 12.

[Table 12. Records with H_IDNUM = 415261329300866]

Looking at the age and sex values for the individuals in the duplicated households in 2002, it seems clear that the second of the duplicated 2002 households (H-SEQ = 93355) should link back to the 2001 household with the same identifier.

Another example is the households with the primary Census household identifier 316204750300121. Table 13 shows records with this household id in the 2001 and 2002 ASEC oversamples:

[Table 13. Records with H_IDNUM = 316204750300121]

This approach also works when the duplicated household identifier is in the previous year, as in the example shown in Table 14 from 2004 and 2005.

[Table 14. Records with H_IDNUM = 190988401052539]

An extra step for the 2001 ASEC file

As stated above, the only reliable way to identify ASEC oversample records is by elimination - any records in the ASEC file that do not appear in the March BMS file from the same year are ASEC oversample records. However, discrepancies between the March BMS and ASEC file in 2001 require an extra step to identify likely ASEC oversample records in this year and avoid unnecessary duplication of household identifiers when linking oversamples.

As part of the work to make the CPS sample suitable for SCHIP program evaluation, the Basic Monthly CPS sample was also expanded in some states, resulting in an overall sample size increase of approximately 12,000 households each month (U.S. Census Bureau, 2002). This sample expansion was phased in between September of 2000 and July of 2001. However, the 2001 March BMS file available from the Census Bureau appears to be missing this sample expansion, containing only 60,015 households. All of the households in the 2001 March BMS file can be linked to the 2001 ASEC file (Flood & Pacas, 2017), however, as shown in Table 15, the number of records in the 2001 ASEC file that remain unmatched after this merge are considerably higher in 2001 than in subsequent years that include the SCHIP oversample; many of these households do not fit either the Hispanic or SCHIP oversample criteria.

[Table 15. Oversample records as defined by the March BMS-ASEC merge]

Due to the household identifier recycling described above, these unlinked March BMS records mixed in with the actual ASEC oversample records in 2001 result in many more duplicate households based on Census identifiers than in subsequent years where the recycling pattern and the composition of the oversamples also result in household identifier duplication. Furthermore, the inclusion of these BMS households when linking oversample records can lead to incorrect links, even when auxiliary variables described are included in the list of linking keys.

[Table 16. Records with H_IDNUM = 332300081939361]

Persons with Census household identifier 332300081939361, shown in Table 16, illustrate how incorrect links may be made if BMS expansion households are treated as ASEC oversample households in the 2001 data. The first of the duplicated households with this identifier in 2001 (H-SEQ = 3583) is likely a BMS household that is part of the expansion but is missing from the 2001 March BMS file. Even though it happens to have the characteristics of a household that would qualify for the SCHIP ASEC oversample, the household with H-SEQ = 3583 has a very low H-SEQ number[11] compared to the second household in the file with the same Census Bureau identifiers and is in a state that added households as part of the SCHIP sample expansion. This suggests that H-SEQ = 3583 is, in fact, a BMS household. Furthermore, the demographic characteristics of the households shown in Table 16 make it clear that H-SEQ 88692 in 2001 and H-SEQ 93370 are the correct oversample links. However, if the BMS expansion household is treated as an ASEC oversample household, based on the linking rules outlined above, the BMS expansion household will be considered the first duplicate of this Census household in 2001 and incorrectly link to the last duplicate in 2002.

To minimize incorrect links between 2001 and 2002 and to arrive at a more accurate denominator when calculating linkage rates across these years, the SCHIP oversample records in the 2001 ASEC considered for linking to the 2002 data are restricted to those that are in a state that was not part of the BMS expansion or are in a state that was part of the BMS expansion and have H-SEQ values above 72000.[12]

Limitations of the IPUMS approach

This strategy for dealing with duplicated household identifiers recovers thousands of oversample links across years (see Table 19 below) in spite of difficulty uniquely identifying households within ASEC oversamples between 2001 and 2005 and yields plausible linkage and validation rates across ASEC oversamples in this period. However, this approach does have two notable drawbacks.

First, the use of the oversample type auxiliary variable as a linking key between 2001 and 2005 means that some types of household transitions may result in a missed link. Because the categorization of households into different oversample types requires demographic information about its members, non-response households cannot be categorized as either Hispanic or SCHIP households. Thus, oversample households that transition into or out of non-response across ASEC files will not be captured by this method between 2001 and 2005. Likewise,

---

[11] Census documentation describing how to link ATUS data to 2003 and 2004 CPS ASEC data instruct users to restrict the ASEC files to H-SEQ values below certain values to exclude ASEC oversample records (U.S. Census Bureau, 2014). Though no specific maximum H-SEQ value is given for 2001 in these instructions, they support the conclusion that within a set of duplicated household identifiers, the household with the lowest H-SEQ value is most likely to be a BMS record.

[12] States that had additional households added to their CPS BMS sample as part of the expansion phased in between September 2000 and July 2001 are: Alabama, Alaska, Colorado, Connecticut, Delaware, the District of Columbia, Hawaii, Indiana, Iowa, Kansas, Kentucky, Maine, Maryland, Minnesota, Missouri, Nebraska, Nevada, New Hampshire, North Dakota, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Vermont, Virginia, West Virginia, Wisconsin, Wyoming (U.S. Census Bureau, 2002).

because the SCHIP oversample is further subdivided into its eligibility criteria in this auxiliary variable, households that transition between any of these three categories during this period will not link across years due to mismatch on the oversample type auxiliary variable. For example, if a household has both a child age 18 or under and a non-white household member in its first year in the ASEC oversample and only a non-white household member in its second year in the ASEC oversample, this household will have different values for oversample type across these two years (4 and 3, respectively) and this mismatch will result in no links being made for records in this household across years. Luckily, missed person links across years due to oversample category changes is likely quite small. Examining the linked ASEC oversample data from 2005-2006, 2006-2007, and 2007-2008 shows that roughly 12% of oversample households that link across years using Census identifiers have a change in their oversample category. The vast majority of these are transitions into and out of nonresponse. Of person records that successfully link across oversamples in these years, between 1.5% and 1.7% of person records transition between oversample types (see Table 17).

[Table 17. Oversample type transitions]

The second limitation of this approach is that de-duplicating Census household identifiers within an oversample type as described above does not completely prevent false positive links between 2003 and 2005 caused by household identifier recycling in this period. Consider all of the households with the primary Census household identifier 190988401052539 between 2002 and 2005 shown in Table 18.

[Table 18. Records with H_IDNUM = 190988401052539]

This household identifier is not duplicated in either 2002 or 2003 and, based on age and sex, these records appear to be correct matches between these two oversamples. As seen above, this Census household identifier is duplicated in 2004 and, when linking backward to 2003, the last duplicate household should be retained for linking. However, this link is incorrect as, a) a household can only appear two times in an ASEC oversample and thus, after the household appears in 2002 and 2003, this household will not appear in the ASEC oversample and, b) the demographic composition of the first duplicate household in 2004 makes it clear that these are not the same individuals that appear with this household identifier in 2003. The incorrect match between the household with H-SEQ = 90300 in 2004 and H-SEQ = 80836 in 2003 is due to the household id recycling pattern. However, because there is no clear mechanism to identify false positive matches between 2003 and 2004, this link will be made by CPSID(P). Users are strongly encouraged to perform their own validation or use CPSIDV (see below) when linking ASEC oversamples across these years.

Assigning CPSID(P) values to ASEC oversample records

As with the CPS BMS data, CPSID(P) assignment for ASEC oversample records is backward looking. ASEC oversample records in the focal year are matched to the previous year's ASEC oversample. Records in the focal year that match to the previous year are assigned that record's CPSID(P) value from the previous year. Unlike the BMS records, the linkable ASEC

oversample records are only eligible to link to one other time point[13], and so if no match is found for a record in the previous year, no further samples are searched for a match and a new CPSID(P) value is assigned to the unmatched oversample record in the focal year.

CPSID(P) values for Basic Monthly records have the following components: The year the record first appears in the CPS, followed by the month the record first appears in the CPS, followed by the IPUMS household number (HHNUM), followed by the IPUMS person number within a household (PERNUM) (Drew et al., 2014). CPSID(P) values for the ASEC oversample records are constructed the same way, but all have a value of '13' in the digits that contain the month first appeared. This value ensures that duplicate CPSID(P) values never occur within an ASEC file.

Linkage rates

Including the oversamples in ASEC panel datasets results in an increase in the linked sample size of between 4.5% and 9.01% between 1990 and 2001 and of more than 30% in 2002 and beyond. Table 19 shows the number of March BMS records that link across adjacent ASEC files, the number of linked records when oversamples are included, as well as the percent increase in the linked sample size for each ASEC panel and overall linkage rate for ASEC panels between 1990 and 2024. Making these records linkable using the IPUMS variables CPSID and CPSIDP allows IPUMS CPS users to include these records in a single linking step.

[Table 19. ASEC-ASEC links using CPSID(P), 1989-2024]

Linkage rates differ for Hispanic and SCHIP oversample records. Half of Hispanic oversample records and 28.57% of SCHIP oversample records are eligible to link across years. As shown in Table 20, fewer eligible Hispanic oversample records link across ASEC files than eligible SCHIP records. Between 22.13% and 35.93% of persons in the Hispanic oversample in the focal year ASEC link back to the previous year's ASEC file between 1990 and 2024; 50% of Hispanic oversample records are eligible to link to the previous year's ASEC. By contrast, between 17.34% and 22.80%[14] of SCHIP oversample records link across adjacent ASEC files from 2002 onward when 28.57% of SCHIP oversample records are eligible to link across years.

[Table 20. ASEC oversamples and their linkage rates]

[13] In fact, ASEC oversample records DO appear in other CPS BMS files. However, in ASEC data prior to 2005, Census identifiers are scrambled for oversample records so as to make linking ASEC oversample records back to their BMS interviews impossible. In 2005 and beyond, ASEC oversample records can be linked to their BMS "source" months. However, these linkages are not currently possible using CPSID(P) and CPSIDV. Users can create these linkages themselves using the IPUMS CPS variables HRHHID, HRHHID2, and LINENO; users are cautioned to pay close attention to the relationship between BMS source month and interview month when making these links.

[14] As shown in Table 20, the actual maximum linkage rate of SCHIP oversample records across years is 29.07%, however, this rate includes false positive matches described in the previous section and thus exceeds the fraction of the SCHIP oversample in 2004 that is eligible to link to 2003.

The low portion of the eligible-to-link Hispanic oversample that can actually be matched across ASEC files compared to the SCHIP oversample is likely due to the differing timing of ASEC interviews across the two groups (See Table 1). All households in the SCHIP oversample that are eligible to link across ASEC files are administered the ASEC supplement at the time of their normally scheduled BMS interview in February or April and ASEC data collection does not represent an extra interview. By contrast, many of the Hispanic oversample households that are eligible to link across years must be interviewed a total of 10 times in order to be included in both ASEC oversample files for which they are eligible. Prior to 2001, all Hispanic oversample households were administered the ASEC survey in March during their eight-month break for November MIS groups 1-4 and after their final BMS interview for November MIS groups 5-8 (U.S. Census Bureau, 2002). Beginning in 2001, November MIS groups 1 and 5 and the new April MIS 1 and 5 groups are given the ASEC supplement at the time of their February and April BMS interviews, respectively (U.S. Census Bureau, 2006). However, even after this change, three of five linkable MIS groups from the Hispanic oversample must complete 10 interviews instead of eight in order to be linked across ASEC files. It is possible that this increased interview burden increases attrition among the linkable Hispanic oversample compared to the SCHIP oversample and/or that the extended interview timeline (between 17 and 19 months instead of 16 months to complete all interviews) increases the likelihood that the oversample-eligible households move from the sampled dwelling during this period and are no longer eligible for inclusion in the ASEC oversample.

In addition to by-oversample breakdowns of ASEC oversample linkage rates, Table 20 also highlights several idiosyncrasies of ASEC oversample linking between 1989 and 2024. The unclassified oversample records are non-response households and households in which no member meets the Census Bureau's stated criteria for oversample eligibility and so cannot be classified as belonging to the Hispanic or the SCHIP oversample. As described above, non-response households in the 1995 ASEC file cannot be linked either to the 1995 March BMS file or the 1994 ASEC file. As a result, non-response households that are actually part of the BMS portion of the 1995 ASEC cannot be separated from the 1995 ASEC oversample, leading to the unusually large number of unclassified oversample households in this year that also cannot be linked to the 1994 ASEC file. This unlinkability is a feature of the original Census Bureau household identifiers, though no explanation for this unlinkability is to be found in official documentation.

A similar phenomenon occurs when linking the 2001 ASEC oversample to the 2000 ASEC oversample. Despite almost a quarter of the Hispanic oversample records in the 2001 ASEC linking back to the 2000 file, the overall oversample linkage rate between these years is very low. This seeming discrepancy is due to the introduction of the SCHIP oversample in 2001, which added over 17,000 households that do not appear in the 2000 ASEC oversample.

Finally, recall that the ASEC panels that include the 2004 data likely contain false positive matches. These false positive matches are reflected in elevated linkage rates among the Hispanic oversample records compared to other panels and an impossibly high linkage rate among SCHIP oversample records between 2003 and 2004, exceeding the 28.57% of SCHIP

oversample records that are eligible to link across these files. Though the 2004-2005 ASEC panel also likely includes false positive matches, SCHIP oversample linkage rates are plausible and Hispanic oversample linkage rates are lower than subsequent panels. The lower-than-expected linkage rate among the Hispanic oversample may be due to the CPS sample redesign that occurred between April 2004 and July 2005.

CPSIDV

CPSIDV is a linking key that matches records that share Census identifiers and have consistent demographic characteristics across time. Once linkable CPSIDP values have been assigned to linkable ASEC oversample records, CPSIDV can also be generated for these records. CPSIDV is assigned to ASEC oversample records according to the same validation criteria used to create CPSIDV for the BMS portion of the ASEC (see below). Unlike BMS records linked by CPSIDV, ASEC oversample records can be linked to only one other time point using CPSIDP: an adjacent year's ASEC file. The creation of this IPUMS CPS custom linking key for the CPS BMS and the BMS portion of the ASEC files is described in detail by Rodgers and Flood (2023).

Validation Criteria

The validation criteria for ASEC oversample records linked by CPSIDV are as follows:

Sex

For a link to be valid, sex must remain constant across both ASEC oversample observations.

Race

For a link to be valid, race must also be constant across both ASEC oversample observations. When links span a change in race codes, race values from the current observation must map to allowed race values in the previous year. Race codes changed in the CPS between 1995 and 1996, 2002 and 2003, and 2012 and 2013 ASEC files.

In 1996, the "Other" category was broken into two new categories; linked records with these new values in 1996 must be in the "Other" race category in 1995 to be considered valid on race. When respondents were allowed to report more than one race beginning in 2003, multi-race categories may map back to any single-race category in the 2002 data that is included in the new multi-race category. Beginning in May of 2012, new multi-race categories became available in the public data. Individuals in these categories in the 2013 ASEC oversample must map back to the unspecified multi-race category with the corresponding number of race categories in the 2012 ASEC data. See Rodgers and Flood (2023) for a detailed description of how different race coding schemes are bridged in CPSIDV.

Age

- The respondent's age may increase by up to two years over their participation in the CPS, but may not decrease at any point, unless this decrease is due to a change in topcode values.
- Beginning in 2004, respondents aged 80 may age 5 years to accommodate topcoding.

See Rodgers and Flood (2023) for detailed description of age topcoding changes over time and how these are bridged in CPSIDV.

Implementation

As with CPSIDP, the assignment of CPSIDV is backward looking. When an oversample record links to the previous year's ASEC file using CPSIDP and meets the validation criteria described above, the CPSIDV value from the previous year's ASEC is assigned to this record in the current year's ASEC. When an oversample record links to the previous year's ASEC file using CPSIDP but does **not** meet the validation criteria listed above, this record is assigned to a CPSIDV value that is equal to its CPSIDP value with a "0" appended. These records will not link to any other CPS samples using CPSIDV. ASEC oversample records that do not link to the previous sample are all assigned a new, unique CPSIDV value equal to their CPSIDP value with a "1" appended. These records include those in the current year's oversample that are not eligible to link to the previous year's oversample but are eligible to link to the following year's oversample, records that are eligible to link to the previous year's oversample but do not, and oversample records that are not eligible to link across ASEC oversamples. This method of CPSIDV assignment means that some ASEC oversample records that are unable or ineligible to link across files will appear to have linkable CPSIDV values.

Linkage rates

Validation rates for linked oversample records are extremely high, exceeding 90% in most year-over-year links. Table 21 compares the number of oversample records linked to the previous year's ASEC using CPSIDP and CPSIDV. Validation rates of linked oversamples are noticeably lower in the linked oversamples from 2004-2003 and 2005-2004. These linked samples are those in which false positive matches occur due to Census household identifier recycling and where these matches could not be reliably avoided when creating CPSIDP values. Despite lower validation rates for these linked samples, overall validated linkage rates for these years when both the BMS and oversample components of the ASEC files are linked are plausible compared to validated linkage rates of other linked ASEC files.

[Table 21. ASEC oversample links using CPSIDV, 1989-2024]

As with CPSIDP, including the oversample records in linked ASEC samples using CPSIDV results in a substantial increase in linked sample size. Table 22 compares linkage rates for the BMS portion of the ASEC to the linkage rates that include ASEC oversample records using CPSIDV. In years prior to 2002, when only Hispanic oversample records can be linked across

ASEC files, linked sample size is increased by between 4.37% and 9.13%; this increase is well over 30% in ASEC panels between 2002 to 2024.

[Table 22. ASEC-ASEC links using CPSIDV, 1989-2024]

Discussion

Many methods previously put forward for linking ASEC data across years do not link ASEC oversample records correctly. Including these records in ASEC panels results in a four to nine percent increase in linked sample size between 1976 and 2001 and a more than 30% increase in linked sample size from 2002 forward. However, to achieve this increase using Census identifiers requires multiple steps and overcoming challenges presented by duplicated Census identifiers and false matches in some years.

This paper describes how IPUMS CPS has made ASEC oversample records linkable with its custom linking keys CPSID, CPSIDP, and CPSIDV for data from 1989 to the present. For ASEC data after 2005, IPUMS adopted the method described in Pacas and Rodgers (2023) for correctly linking these records across years; prior to 2005, this method was adapted to account for the Census household identifier recycling pattern. Once linkable ASEC oversample records are assigned a linkable CPSID(P) value using this approach, validation criteria described by Rodgers and Flood (2023) are applied to generate linkable CPSIDV values for linkable oversample records.

By expanding CPSID, CPSIDP, and CPSIDV to make the ASEC oversamples linkable across years, IPUMS CPS has further lowered the barrier to leveraging linked ASEC data for longitudinal analysis. Incorporating linkable ASEC oversample records into CPSID(P) and CPSIDV allows IPUMS CPS users to create linked ASEC datasets that include linkable oversample records and reap the benefits of increased sample size in a single easy step.

References

Drew, J. A., Flood, S., & Warren, J. R. (2014). Making full use of the longitudinal design of the Current Population Survey: Methods for linking records across 16 months. Journal of Economic and Social Measurement, 39(3), 121–144. https://doi.org/10.3233/JEM-140388

Feng, S. (2001). The longitudinal matching of current population surveys: A proposed algorithm. Journal of Economic and Social Measurement, 27(1–2), 71–91. https://doi.org/10.3233/JEM-2003-0197

Flood, S. M., & Pacas, J. D. (2017). Using the Annual Social and Economic Supplement as part of a Current Population Survey panel. Journal of Economic and Social Measurement, 42(3–4), 225–248. https://doi.org/10.3233/JEM-180447

Flood, S., Rodgers, R., Pacas, J., Kristiansen, D., & Klaas, B. (2022). Extending current population survey linkages: Obstacles and solutions for linking monthly data from 1976 to 1988. *Journal of economic and social measurement*, *46*(1), 1-28.

Flood, S., King, M., Rodgers, R., Ruggles, S., Robert Warren, J.R., Backman, D., Chen, A., Cooper, G., Richards, S., Schouweiler, M., & Westberry, M. (2024). IPUMS CPS:Version 12.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V12.0

Madrian, B. C., & Lefgren, L. J. (2000). An approach to longitudinally matching Current Population Survey (CPS) respondents. Journal of Economic and Social Measurement, 26(1), 31–62. https://doi.org/10.3233/JEM-2000-0165

Pacas, J. D., & Rodgers, R. (2023). Research note on linking CPS ASEC files. *Population research and policy review*, 42(3), 50.

Rodgers, R., & Flood, S. (2023). A Holistic Approach to Validating Current Population Survey Panel Data. IPUMS. https://assets.ipums.org/_files/ipums/working_papers/ipums_wp_2023-01.pdf

U.S. Census Bureau & U.S. Bureau of Labor Statistics. (2002). Design and Methodology: Current Population Survey Technical Paper 63RV. U.S. Census Bureau.

U.S. Census Bureau & U.S. Bureau of Labor Statistics. (2006). Design and Methodology: Current Population Survey Technical Paper 66. U.S. Census Bureau.

U.S. Census Bureau & U.S. Bureau of Labor Statistics. (2019). Design and Methodology: Current Population Survey Technical Paper 77. U.S. Census Bureau.

U.S. Census Bureau & U.S. Bureau of Labor Statistics. (2014). America Time Use Survey User's Guide: Understanding ATUS 2003 to 2013. U.S. Census Bureau.

Table 1. Composition of the CPS ASEC files, 1976-present

ASEC Sample Includes:	ASEC interview month	1976-2000	2001-2002	2003	2004-present
March BMS	March	x	x	x	x
Hispanic Oversample					
November [a]					
MIS 1[c]	February[b]	x	x	x	x
MIS 2	March	x	x	x	x
MIS 3	March	x	x	x	x
MIS 4	March	x	x	x	x
MIS 5[d]	February[b]	x	x	x	x
MIS 6	March	x	x	x	x
MIS 7	March	x	x	x	x
MIS 8	March	x	x	x	x
April					
MIS 1	April	--	x	x	x
MIS 5	April	--	x	x	x
SCHIP Oversample					
November					
MIS 1[c]	February	--	x	x	x
MIS 5[d]	February	--	x	x	x
MIS 6	March[e]	--	x	--	--
MIS 7	March[e]	--	x	x	--
MIS 8	March[e]	--	x	x	--
August (MIS 8)	February[e]	--	--	--	x
September (MIS 8)	February[e]	--	--	--	x
October (MIS 8)	April[e]	--	--	x	x
April					
MIS 1	April	--	x	x	x
MIS 5	April	--	x	x	x

Note : This table is based on technical documentation from the U.S. Census Bureau (2006) and Table 2 in Flood & Pacas (2017).

a Hispanic oversample households drawn from the November BMS may be interviewed up to 10 times; in addition to the eight interviews of the CPS BMS rotation, these households are eligible to be interviewed once for the ASEC oversample during their BMS eight-month break and once again for the following year's ASEC oversample after they have completed their full BMS rotation.

b Prior to 2001, all Hispanic oversample households identified in the November sample were administered the ASEC in March. Beginning in 2001 when Hispanic oversample households from these MIS groups are interviewed in February, they are interviewed only eight times instead of ten.

c These households are in MIS 4 in February and are administered the ASEC supplement at the time of their BMS interview.

d These households are in MIS 8 in February and are administered the ASEC supplement at the time of their BMS interview.

e These households are interviewed for a ninth time after completing their BMS rotation.

Figure 1. ASEC oversample MIS groups eligible to link across ASEC files

1989-2000

Hispanic Oversample		SCHIP Oversample	
Year 1	Year 2	Year 1	Year 2
November	*November*	--	--
MIS 1	MIS 1	--	--
MIS 2	MIS 2	--	--
MIS 3	MIS 3	--	--
MIS 4	MIS 4	--	--
MIS 5	MIS 5	--	--
MIS 6	MIS 6	--	--
MIS 7	MIS 7	--	--
MIS 8	MIS 8	--	--

2001-2002

Hispanic Oversample		SCHIP Oversample	
Year 1	Year 2	Year 1	Year 2
November	*November*	*November*	*November*
MIS 1[a]	MIS 1	MIS 1[a]	MIS 1[b]
MIS 2	MIS 2		
MIS 3	MIS 3		
MIS 4	MIS 4		
MIS 5	MIS 5[b]	MIS 5[c]	MIS 5[b]
MIS 6	MIS 6	MIS 6	MIS 6
MIS 7	MIS 7	MIS 7	MIS 7
MIS 8	MIS 8	MIS 8	MIS 8
April	*April*	*April*	*April*
MIS 1	MIS 1	MIS 1	MIS 1
MIS 5	MIS 5	MIS 5	MIS 5

2003

Hispanic Oversample		SCHIP Oversample	
Year 1	Year 2	Year 1	Year 2
November	*November*	*November*	*November*
MIS 1[a]	MIS 1	MIS 1[a]	MIS 1[b]
MIS 2	MIS 2		
MIS 3	MIS 3		
MIS 4	MIS 4		
MIS 5	MIS 5[b]	MIS 5[c]	MIS 5[b]
MIS 6	MIS 6		
MIS 7	MIS 7	MIS 7	MIS 7
MIS 8	MIS 8	MIS 8	MIS 8
		October (MIS 8)	*October (MIS 8)*
April	*April*	*April*	*April*
MIS 1	MIS 1	MIS 1	MIS 1
MIS 5	MIS 5	MIS 5	MIS 5

2004-2024

Hispanic Oversample		SCHIP Oversample	
Year 1	Year 2	Year 1	Year 2
November	*November*	*November*	*November*
MIS 1[a]	MIS 1	MIS 1[a]	MIS 1[b]
MIS 2	MIS 2		
MIS 3	MIS 3		
MIS 4	MIS 4		
MIS 5	MIS 5[b]	MIS 5[c]	MIS 5[b]
MIS 6	MIS 6		
MIS 7	MIS 7		
MIS 8	MIS 8		
		August (MIS 8)	*August (MIS 8)*
		September (MIS 8)	*September (MIS 8)*
		October (MIS 8)	*October (MIS 8)*
April	*April*	*April*	*April*
MIS 1	MIS 1	MIS 1[a]	MIS 1
MIS 5	MIS 5	MIS 5	MIS 5

The linkable portions of the ASEC samples between Year 1 and Year 2 are highlighted and connected with arrows. All MIS source groups that ever comprised the Hispanic and SCHIP oversample are shown; see Table 1 for a breakdown of ASEC oversample composition over time.

a These households are in MIS 4 in February.

b These households are in MIS 8 in February.

Table 2. The recycling pattern of a Census Bureau household identifier in the CPS Basic Monthly, 1989-2004

	Year 1											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Household V	--	--	--	--	5	6	7	8				
Household W	1	2	3	4	--	--	--	--	--	--	--	--
Household X									1	2	3	4
Household Y												
Household Z												

	Year 2											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Household V												
Household W	5	6	7	8								
Household X	--	--	--	--	--	--	--	--	5	6	7	8
Household Y					1	2	3	4	--	--	--	--
Household Z												

	Year 3											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Household V												
Household W												
Household X												
Household Y	--	--	--	--	5	6	7	8				
Household Z	1	2	3	4	--	--	--	--	--	--	--	--

Note : Numbers 1-8 are MIS numbers; "--" indicates the 8-month break in CPS interviews. Households V, W, X, Y, and Z all have the same Census Bureau household identifiers. Note that the recycling pattern can lead to false positive matches between Household V and Household Y between Year 1 and Year 2 and between Household W and Household Z between Year 2 and Year 3 if MIS is not used to enforce the correct rotation pattern. Based on *The recycling pattern of household identifiers* in CPS in Feng (2001).

Table 3. An Illustration of the impact of household identifier recycling on linking November Hispanic oversample households

	Year 1											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Household A											1	2
Household B												
Household C												
Household D											5	6
Household E												
Household F												
Household G												

	Year 2											
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Household A	3	4	--	--	--	--	--	--	--	--	5	6
Household B			5	6	7	8						
Household C							1	2	3	4	--	--
Household D	7	8										
Household E			1	2	3	4	--	--	--	--	--	--
Household F							5	6	7	8		
Household G											1	2

Note : Numbers 1-8 are MIS numbers; "--" indicates the 8-month break in CPS interviews. Households A, B, C, D, E, F, and G all have the same Census Bureau household identifiers. This table illustrates how the household identifier recycling pattern can lead to false positive matches between Hispanic oversample records that are identified in the November BMS and given the ASEC supplement in the following year. Those Hispanic oversample households in MIS 1 in November of year 1 (included in the year 2 ASEC) will correctly link to their MIS 5 interview in November of year 2 (included in the year 3 ASEC); these cells are outlined and highlighted in green. However, MIS 5 interviews in November of year 1 (included in the year 2 ASEC) will incorrectly link to households with their same identifier in MIS 1 in November of year 2 (included in the year 3 ASEC); these cells have shading and are highlighted in red. However, because MIS values of the oversample records in the ASEC files do not indicate the household's true place in the CPS rotation, MIS values cannot be used to enforce correct links. In this example, Household A will correctly link across years (MIS 1 to MIS 5), but Household D will incorrectly be linked to Household G.

Table 4. Month-in-sample values for Hispanic oversample households that link between the 1989 and 1990 ASEC files using only Census identifiers

		1990 MIS							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1989 MIS	**1**	0	0	0	0	102	0	0	0
	2	0	0	0	0	0	98	0	0
	3	0	0	0	0	0	0	102	0
	4	0	0	0	0	0	0	0	70
	5	239	0	0	0	0	0	0	0
	6	0	248	0	0	0	0	0	0
	7	0	0	233	0	0	0	0	0
	8	0	0	0	233	0	0	0	0

Note : This table indicates the number of unique households with each MIS value progression across 1989 and 1990, not the number of person links. MIS value pairs highlighted in red are suspected false positive matches.

Table 5. Month-in-sample values for households that link across ASEC files, 1994 and 2000

		1994 MIS							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1993 MIS	**1**	0	0	0	0	109	0	0	0
	2	0	0	0	0	117	0	0	0
	3	0	0	0	0	107	0	0	0
	4	0	0	0	0	101	0	0	0
	5	0	0	0	0	247	0	0	0
	6	0	0	0	0	299	0	0	0
	7	0	0	0	0	278	0	0	0
	8	0	0	0	0	324	0	0	0

		1995 MIS							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1994 MIS	**1**	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0
	5	254	279	276	252	0	0	0	0
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0

		2000 MIS							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1999 MIS	**1**	0	0	0	0	139	0	0	0
	2	0	0	0	0	138	0	0	0
	3	0	0	0	0	132	0	0	0
	4	0	0	0	0	122	0	0	0
	5	0	0	0	0	358	0	0	0
	6	0	0	0	0	354	0	0	0
	7	0	0	0	0	353	0	0	0
	8	0	0	0	0	402	0	0	0

		2001 MIS							
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
2000 MIS	**1**	0	0	0	0	0	0	0	0
	2	0	0	0	0	0	0	0	0
	3	0	0	0	0	0	0	0	0
	4	0	0	0	0	0	0	0	0
	5	311	326	335	318	91	86	88	113
	6	0	0	0	0	0	0	0	0
	7	0	0	0	0	0	0	0	0
	8	0	0	0	0	0	0	0	0

Note : This table indicates the number of unique households with each MIS value progression across the included years, not the number of person links. MIS value pairs highlighted in red are suspected false positive matches.

Table 6. Month-in-sample values for households that link across ASECs, 2001, 2002, 2003

Hispanic Oversample Household Links

2001 MIS \ 2002 MIS	1	2	3	4	5	6	7	8
1	0	0	0	0	97	0	0	0
2	0	0	0	0	0	94	0	0
3	0	0	0	0	0	0	108	0
4	0	0	0	0	0	0	0	195
5	335	0	0	0	0	0	0	0
6	0	312	0	0	0	0	0	0
7	0	0	327	0	0	0	0	0
8	0	0	0	784	0	0	0	0

SHIP Oversample Household Links

2001 MIS \ 2002 MIS	1	2	3	4	5	6	7	8
1	0	0	0	0	476	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	824
5	1675	0	0	0	0	0	0	0
6	0	4	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	2107	0	0	0	0

Hispanic Oversample Household Links

2002 MIS \ 2003 MIS	1	2	3	4	5	6	7	8
1	0	0	0	0	109	0	0	0
2	0	0	0	0	0	92	0	0
3	0	0	0	0	0	0	99	0
4	11	22	17	21	16	20	15	129
5	343	1	0	0	0	0	0	0
6	0	316	0	0	0	0	0	0
7	0	0	334	0	0	0	0	0
8	55	47	53	393	51	55	58	61

SHIP Oversample Household Links

2002 MIS \ 2003 MIS	1	2	3	4	5	6	7	8
1	62	65	90	71	83	75	72	105
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	61	99	100	87	109	96	99	100
5	216	198	211	212	230	202	220	190
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	261	237	258	281	286	222	285	236

Note : This table indicates the number of unique households with each MIS value progression across the included years, not the number of person links. MIS value pairs highlighted in red are suspected false positive matches.

Table 7. Validation on age and sex of suspected correct and suspected false positive matches across ASEC oversamples

	Oversample linkage Validation Rates			
	Suspected Correct Links		Suspected False Positive Links	
	AGE	SEX	AGE	SEX
1990-1989				
Hispanic	80.87%	96.17%	2.34%	59.74%
SCHIP (child under 18)	--	--	--	--
SCHIP (non-white household member)	--	--	--	--
SCHIP (child under 18 AND non-white household member)	--	--	--	--
1994-1993				
Hispanic	96.22%	96.22%	57.99%	57.99%
SCHIP (child under 18)	--	--	--	--
SCHIP (non-white household member)	--	--	--	--
SCHIP (child under 18 AND non-white household member)	--	--	--	--
1995-1994				
Hispanic	94.75%	99.03%	--	--
SCHIP (child under 18)	--	--	--	--
SCHIP (non-white household member)	--	--	--	--
SCHIP (child under 18 AND non-white household member)	--	--	--	--
2000-1999				
Hispanic	99.24%	99.24%	54.30%	54.30%
SCHIP (child under 18)	--	--	--	--
SCHIP (non-white household member)	--	--	--	--
SCHIP (child under 18 AND non-white household member)	--	--	--	--
2001-2000				
Hispanic	96.94%	99.47%	2.63%	52.52%
SCHIP (child under 18)	--	--	--	--
SCHIP (non-white household member)	--	--	--	--
SCHIP (child under 18 AND non-white household member)	--	--	--	--
2002-2001				
Hispanic	93.55%	99.19%	2.81%	54.30%
SCHIP (child under 18)	94.50%	98.69%	4.15%	52.90%
SCHIP (non-white household member)	84.11%	98.30%	2.31%	49.51%
SCHIP (child under 18 AND non-white household member)	91.31%	98.24%	2.64%	57.80%
2003-2002				
Hispanic	92.57%	99.35%	2.41%	51.07%
SCHIP (child under 18)	95.08%	99.10%	4.17%	52.94%
SCHIP (non-white household member)	86.10%	99.04%	2.31%	53.76%
SCHIP (child under 18 AND non-white household member)	94.30%	99.01%	4.83%	57.14%

Note : For sex to be considered valid in matched records, the value must be the same in both the first and the second year; for age to be considered valid in matched records there must either be a) no change in age value across years or b) age in the second year is one year larger than in the first year. Note that these age validation criteria are more conservative than the age validation criteria applied in the IPUMS CPS variable CPSIDV.

Table 8. Month-in-sample values for households that link across ASECs, 2003, 2004, 2005

Hispanic Oversample Household Links

2003 MIS \ 2004 MIS	1	2	3	4	5	6	7	8
1	8	7	18	10	112	5	3	9
2	10	8	10	5	12	96	9	4
3	5	16	3	10	13	7	117	9
4	5	5	4	16	11	8	9	110
5	322	5	5	10	15	11	5	11
6	12	336	5	13	14	6	12	7
7	10	8	315	5	10	7	7	6
8	14	8	7	346	12	5	7	1

SCHIP Oversample Household Links

2003 MIS \ 2004 MIS	1	2	3	4	5	6	7	8
1	59	78	68	70	66	71	64	66
2	58	80	63	70	66	67	85	68
3	68	96	72	73	72	84	84	93
4	64	74	59	78	56	73	84	73
5	80	78	75	67	62	100	69	83
6	98	77	74	60	74	61	85	87
7	80	86	73	55	92	70	63	83
8	96	79	99	59	66	97	54	75

Hispanic Oversample Household Links

2004 MIS \ 2005 MIS	1	2	3	4	5	6	7	8
1	2	14	7	2	7	6	11	5
2	5	10	8	13	6	37	10	14
3	6	12	10	4	10	12	49	6
4	11	5	9	9	4	11	6	39
5	7	10	10	5	6	9	7	5
6	12	309	4	15	5	12	5	11
7	8	8	353	6	3	14	8	9
8	3	5	8	346	7	7	5	9

SCHIP Oversample Household Links

2004 MIS \ 2005 MIS	1	2	3	4	5	6	7	8
1	60	51	57	63	51	65	54	67
2	68	61	66	58	40	39	36	62
3	55	73	73	60	63	54	45	51
4	47	46	53	57	49	61	61	56
5	43	60	76	53	54	68	59	57
6	69	61	45	58	57	47	76	57
7	58	67	48	57	54	61	45	62
8	56	57	78	55	50	65	52	61

Note : This table indicates the number of unique households with each MIS value progression across the included years, not the number of person links.

Table 9. Duplicated Census household identifiers within ASEC oversamples due to household identifier recycling, 2001-2004

Panel A : Composition of the 2001 ASEC oversamples and the MIS values with the same household identifiers as per the recycling pattern

Hispanic

2000					2001			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

SCHIP

2000					2001			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

Panel B : Composition of the 2002 ASEC oversamples and the MIS values with the same household identifiers as per the recycling pattern

Hispanic

2001					2002			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

SCHIP

2001					2002			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

Panel C : Composition of the 2003 ASEC oversamples and the MIS values with the same household identifiers as per the recycling pattern

Hispanic

2002					2003			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

SCHIP

2002					2003			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

Panel D : Composition of the 2004 ASEC oversamples and the MIS values with the same household identifiers as per the recycling pattern

Hispanic

2003					2004			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

SCHIP

2003					2004			
Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Z6	Z7	Z8	A1	A2	A3	A4	Q5	Q6
Y7	Y8	B1	B2	B3	B4	P5	P6	P7
X8	C1	C2	C3	C4	K5	K6	K7	K8
D1	D2	D3	D4	I5	I6	I7	I8	N1
W2	W3	W4	E5	E6	E7	E8	Z1	Z2
V3	V4	F5	F6	F7	F8	M1	M2	M3
U4	G5	G6	G7	G8	L1	L2	L3	L4
H5	H6	H7	H8	J1	J2	J3	J4	O5

Note : this table shows the source MIS groups for each ASEC oversample between 2001 and 2004 along with the MIS groups to which recycled household identifiers move across months. Letters indicate CPS cohorts or rotation groups. Numbers 1-8 are MIS values; highlighting indicates the MIS group in that cell from the month indicated by the column heading contributes to the ASEC oversample. Hispanic and SCHIP oversample source groups are shown seperately. Oversample MIS source groups that appear on the same line in different months result in duplicated household identifiers within the ASEC file from that year. When using the household identifiers required to bridge the change between the 2004 and 2005 ASEC, households also have duplicate identifiers in the 2005 data. Because the composition of the oversamples is the same in 2004 and 2005, the structure of overlapping groups in the 2005 ASEC data is identical to Panel D above.

Table 10. Records with duplicated identifiers within the ASEC oversample

Year	Oversample Records		Full Oversample				Pared down focal oversample				Pared down target oversample			
			Duplicates based on CB Identifiers		*Duplicates based on CB Identifiers + OS type*		*Duplicates based on CB Identifiers*		*Duplicates based on CB Identifiers + OS source*		*Duplicates based on CB Identifiers*		*Duplicates based on CB Identifiers + OS type*	
	H	**P**	**H**	**P**	**H**	**P**	**H**	**P**	**H**	**P**	**H**	**P**	**H**	**P**
1989	2646	7303	0	0	0	0	0	0	0	0	0	0	0	0
1990	3320	9349	0	0	0	0	0	0	0	0	0	0	0	0
1991	3506	10249	0	0	0	0	0	0	0	0	0	0	0	0
1992	3652	10441	0	0	0	0	0	0	0	0	0	0	0	0
1993	3749	10579	0	0	0	0	0	0	0	0	0	0	0	0
1994	3709	10318	6	0	6	0	0	0	0	0	0	0	0	0
1995[a]	18926	11721	4998	0	4998	0	0	0	0	0	0	0	0	0
1996	3666	10293	0	0	0	0	0	0	0	0	0	0	0	0
1997	3832	10865	0	0	0	0	0	0	0	0	0	0	0	0
1998	3957	11110	0	0	0	0	0	0	0	0	0	0	0	0
1999	4086	11548	0	0	0	0	0	0	0	0	0	0	0	0
2000	4295	12516	0	0	0	0	0	0	0	0	0	0	0	0
2001[b]	26937	80687	2740	5467	1216	3362	0	0	0	0	1227	2432	496	1386
2002	26106	77559	2080	4454	1035	2796	1766	3850	878	2410	927	1906	422	1068
2003	26443	75136	4029	7314	1770	4716	2665	4926	1168	3062	1904	3616	910	2474
2004	25806	74891	1734	3360	863	2316	801	1456	375	944	1349	2682	665	1816
2005	25917	74333	721	1273	235	575	271	413	77	145	0	0	0	0
2006	25652	73534	0	0	0	0	0	0	0	0	0	0	0	0
2007	25782	72822	0	0	0	0	0	0	0	0	0	0	0	0
2008	25540	73249	0	0	0	0	0	0	0	0	0	0	0	0
2009	25391	73271	0	0	0	0	0	0	0	0	0	0	0	0
2010	25574	74324	0	0	0	0	0	0	0	0	0	0	0	0
2011	25231	72708	0	0	0	0	0	0	0	0	0	0	0	0
2012	24600	70026	0	0	0	0	0	0	0	0	0	0	0	0
2013	25559	72100	0	0	0	0	0	0	0	0	0	0	0	0
2014	25207	69829	0	0	0	0	0	0	0	0	0	0	0	0
2015	25515	69213	0	0	0	0	0	0	0	0	0	0	0	0
2016	25089	67497	0	0	0	0	0	0	0	0	0	0	0	0
2017	25509	67264	0	0	0	0	0	0	0	0	0	0	0	0
2018	24845	65793	0	0	0	0	0	0	0	0	0	0	0	0
2019	23640	61993	0	0	0	0	0	0	0	0	0	0	0	0
2020	21173	53081	0	0	0	0	0	0	0	0	0	0	0	0
2021	21808	56209	0	0	0	0	0	0	0	0	0	0	0	0
2022	20667	52197	0	0	0	0	0	0	0	0	0	0	0	0
2023	20125	49787	0	0	0	0	0	0	0	0	0	0	0	0
2024	19922	49131	0	0	0	0	0	0	0	0	0	0	0	0

a Paring down the 1995 ASEC to records with household ids that appear in the 1994 ASEC removes all non-response households.

b Duplicates in the full oversample are counted after records suspected to be part of the March BMS sample expansion are eliminated.

Table 11. The number of records in Census Households with multiple H-SEQ values after accounting for households uniquely identified by oversample type

Year	as focal year		as target year	
	2	> 2	2	> 2
2001			2155	
2002	3843		1762	
2003	4982		3944	
2004	1490	35	2827	35
2005	246	20		

Note : After reducing oversamples to records with primary Census household identifiers that appear in the target or focal year and accounting for oversample type, nearly all records in duplicated households as defined by Census Bureau identifiers are in a group with just one other household as defined by H-SEQ in the ASEC files. Records with Census household identifiers that are shared across more than two H-SEQ values, 55 in total, are not considered for linking. The duplicate records counted here for 2001 are after accounting for the BMS expansion records found in the 2001 ASEC file but not the March BMS file.

Table 12. Records with H_IDNUM = 415261329300866

Panel A: in the 2001 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
415261329300866	1	88628	Maine	1	38	F	SCHIP (<=18)	1	1
415261329300866	1	88628	Maine	2	15	M	SCHIP (<=18)	1	1
415261329300866	1	88628	Maine	3	13	M	SCHIP (<=18)	1	1

Panel B: in the 2002 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
415261329300866	1	80961	Maine	1	26	M	SCHIP (<=18)	1	0
415261329300866	1	80961	Maine	4	2	F	SCHIP (<=18)	1	0
415261329300866	1	80961	Maine	5	0	F	SCHIP (<=18)	1	0
415261329300866	1	80961	Maine	2	28	F	SCHIP (<=18)	1	0
415261329300866	1	80961	Maine	3	9	F	SCHIP (<=18)	1	0
415261329300866	1	93355	Maine	1	39	F	SCHIP (<=18)	0	1
415261329300866	1	93355	Maine	2	16	M	SCHIP (<=18)	0	1
415261329300866	1	93355	Maine	3	14	M	SCHIP (<=18)	0	1

Table 13. Records with H_IDNUM = 316204750300121

Panel A: in the 2001 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
316204750300121	1	88642	Maine	1	26	M	SCHIP (<=18)	1	1
316204750300121	1	88642	Maine	3	0	M	SCHIP (<=18)	1	1
316204750300121	1	88642	Maine	2	25	F	SCHIP (<=18)	1	1

Panel B: in the 2002 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
316204750300121	1	80972	Maine	1	35	F	SCHIP (<=18)	1	0
316204750300121	1	80972	Maine	3	35	M	SCHIP (<=18)	1	0
316204750300121	1	80972	Maine	4	4	F	SCHIP (<=18)	1	0
316204750300121	1	80972	Maine	2	1	M	SCHIP (<=18)	1	0
316204750300121	1	93359	Maine	1	27	M	SCHIP (<=18)	0	1
316204750300121	1	93359	Maine	3	1	M	SCHIP (<=18)	0	1
316204750300121	1	93359	Maine	2	26	F	SCHIP (<=18)	0	1

Table 14. Records with H_IDNUM = 190988401052539

Panel A: in the 2004 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
190988401052539	1	87553	Hawaii	1	57	M	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	87553	Hawaii	2	14	M	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	87553	Hawaii	4	85	F	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	87553	Hawaii	3	65	M	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	90300	Hawaii	1	44	M	SCHIP (<= 18 AND non-white)	0	1
190988401052539	1	90300	Hawaii	2	39	F	SCHIP (<= 18 AND non-white)	0	1
190988401052539	1	90300	Hawaii	4	20	M	SCHIP (<= 18 AND non-white)	0	1
190988401052539	1	90300	Hawaii	3	0	M	SCHIP (<= 18 AND non-white)	0	1

Panel B: in the 2005 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
190988401052539	1	79400	Hawaii	1	58	M	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	79400	Hawaii	2	15	M	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	79400	Hawaii	4	85	F	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	79400	Hawaii	3	67	M	SCHIP (<= 18 AND non-white)	1	1

Table 15. Oversample records as defined by the March BMS-ASEC merge

Year	Total Oversample Records		Hispanic Oversample Records		SCHIP Oversample Records		Unclassified Oversample Records	
	H	P	H	P	H	P	H	P
2001	38001	101,606	5,492	18,766	21,456	74,317	11,053	8,523
2002	26106	77,559	5,256	17,914	17,424	59,645	3,426	0
2003	26443	75,136	5,395	18,263	16,657	56,873	4,391	0
2004	25806	74,891	5,654	18,845	16,513	56,046	3,639	0

Table 16. Records with H_IDNUM = 332300081939361

Panel A: in the 2001 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
332300081939361	1	3583	Maine	1	40	M	SCHIP (<=18)	1	0
332300081939361	1	3583	Maine	2	37	F	SCHIP (<=18)	1	0
332300081939361	1	3583	Maine	3	16	F	SCHIP (<=18)	1	0
332300081939361	1	3583	Maine	4	10	M	SCHIP (<=18)	1	0
332300081939361	1	88692	Maine	1	32	F	SCHIP (<=18)	0	1
332300081939361	1	88692	Maine	2	35	M	SCHIP (<=18)	0	1
332300081939361	1	88692	Maine	3	9	F	SCHIP (<=18)	0	1
332300081939361	1	88692	Maine	4	6	M	SCHIP (<=18)	0	1

Panel B: in the 2002 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
332300081939361	1	93370	Maine	1	33	F	SCHIP (<=18)	1	1
332300081939361	1	93370	Maine	2	36	M	SCHIP (<=18)	1	1
332300081939361	1	93370	Maine	3	10	F	SCHIP (<=18)	1	1
332300081939361	1	93370	Maine	4	7	M	SCHIP (<=18)	1	1

Table 17. Oversample type transitions

Linked Oversamples	Linked households that transition between oversample categories		Linked households that transition between response and non-response		Linked persons that trasition between oversample categories	
	N	%	N	%	N	%
2005-2006	823	12.8	689	10.7	284	1.6
2006-2007	866	13.5	727	11.4	308	1.7
2007-2009	777	11.8	648	9.8	322	1.7

Table 18. Records with H_IDNUM = 190988401052539

Panel A: in the 2002 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
190988401052539	1	86968	Hawaii	1	28	F	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	86968	Hawaii	2	6	F	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	86968	Hawaii	3	4	M	SCHIP (<= 18 AND non-white)	1	1

Panel B: in the 2003 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
190988401052539	1	80836	Hawaii	1	29	F	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	80836	Hawaii	2	7	F	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	80836	Hawaii	3	5	M	SCHIP (<= 18 AND non-white)	1	1

Panel C: in the 2004 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
190988401052539	1	87553	Hawaii	1	57	M	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	87553	Hawaii	2	14	M	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	87553	Hawaii	4	85	F	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	87553	Hawaii	3	65	M	SCHIP (<= 18 AND non-white)	1	0
190988401052539	1	90300	Hawaii	1	44	M	SCHIP (<= 18 AND non-white)	0	1
190988401052539	1	90300	Hawaii	2	39	F	SCHIP (<= 18 AND non-white)	0	1
190988401052539	1	90300	Hawaii	4	20	M	SCHIP (<= 18 AND non-white)	0	1
190988401052539	1	90300	Hawaii	3	0	M	SCHIP (<= 18 AND non-white)	0	1

Panel D: in the 2005 ASEC oversample

Household ID	Household No.	H-SEQ	State	Person Line No.	Age	Sex	Oversample type	first duplicate	last duplicate
190988401052539	1	79400	Hawaii	1	58	M	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	79400	Hawaii	2	15	M	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	79400	Hawaii	4	85	F	SCHIP (<= 18 AND non-white)	1	1
190988401052539	1	79400	Hawaii	3	67	M	SCHIP (<= 18 AND non-white)	1	1

Table 19. ASEC-ASEC links using CPSID(P), 1989-2024

Year	*Total Records*		*Linked to Year - 1, BMS only*		*Linked to Year - 1, BMS + Oversamples*		*% Increase in linked records*		*Total Oversample Linkage Rate*	
	H	P			H	P	H	P	H	P
1990	75269	158079	28534	51598	29487	53919	3.34%	4.50%	39.18%	34.11%
1991	75076	158477	30154	55804	31465	59169	4.35%	6.03%	41.91%	37.34%
1992	74236	155796	29692	54709	30996	58065	4.39%	6.13%	41.75%	37.27%
1993	73878	155197	29638	54124	31061	57705	4.80%	6.62%	42.04%	37.18%
1994	73126	150943	29708	54066	31110	57590	4.72%	6.52%	42.54%	38.15%
1995[a]	72152	149642	21185	47910	22318	51205	5.35%	6.88%	30.93%	34.22%
1996[b]	63339	130476	0	0	0	0	0.00%	0.00%	0.00%	0.00%
1997	64046	131854	26859	46669	28178	50292	4.91%	7.76%	44.00%	38.14%
1998	64659	131617	26933	46771	28258	50318	4.92%	7.58%	43.70%	38.23%
1999	65377	132324	27310	46866	28795	50731	5.44%	8.25%	44.04%	38.34%
2000	64944	133710	27553	47375	29020	51204	5.32%	8.08%	44.68%	38.29%
2001[c]	98015	218269	26895	46060	28485	50212	5.91%	9.01%	29.06%	23.00%
2002	98848	217219	27001	44985	32615	63466	20.79%	41.08%	33.00%	29.22%
2003	99986	216424	32976	54722	38672	73202	17.27%	33.77%	38.68%	33.82%
2004[d]	98979	213241	33374	55199	40836	78250	22.36%	41.76%	41.26%	36.70%
2005[d,e]	98664	210648	29788	47917	35463	65625	19.05%	36.96%	35.94%	31.15%
2006	97352	208562	32261	50904	38701	69152	19.96%	35.85%	39.75%	33.16%
2007	98015	206639	32211	51496	38614	69560	19.88%	35.08%	39.40%	33.66%
2008	97502	206404	32970	52217	39560	71100	19.99%	36.16%	40.57%	34.45%
2009	97066	207921	32542	52117	38866	70397	19.43%	35.07%	40.04%	33.86%
2010	97263	209802	32958	53486	39593	72709	20.13%	35.94%	40.71%	34.66%
2011	96958	204983	32696	52306	39179	70814	19.83%	35.38%	40.41%	34.55%
2012	96659	201398	33062	51587	39290	69493	18.84%	34.71%	40.65%	34.51%
2013	98095	202634	32835	51125	39138	69021	19.20%	35.00%	39.90%	34.06%
2014	97926	199556	33200	50438	39860	69126	20.06%	37.05%	40.70%	34.64%
2015	99461	199024	27064	41728	32768	57321	21.08%	37.37%	32.95%	28.80%
2016	94097	185487	30924	44503	37339	62134	20.74%	39.62%	39.68%	33.50%
2017	95006	185914	29206	41995	35572	59175	21.80%	40.91%	37.44%	31.83%
2018	92139	180084	29040	41217	35469	58638	22.14%	42.27%	38.50%	32.56%
2019	94633	180101	30231	42189	36365	58820	20.29%	39.42%	38.43%	32.66%
2020	91500	157959	32500	42087	38164	56532	17.43%	34.32%	41.71%	35.79%
2021	90759	163543	31834	38022	36656	50068	15.15%	31.68%	40.39%	30.61%
2022	89197	152732	31262	37344	36439	50475	16.56%	35.16%	40.85%	33.05%
2023	88978	146133	31806	35919	36634	48156	15.18%	34.07%	41.17%	32.95%
2024	89473	144265	31828	34575	36678	46525	15.24%	34.56%	40.99%	32.25%

a Non-response households in the 1995 ASEC file cannot be linked either to the 1995 March BMS file or the 1994 ASEC file.

b Due to the sample re-design that took place between the 1995 and 1996 ASEC files, these years of ASEC data cannot be linked to one another.

c Oversample record counts and linkage rates are calculated after accounting for BMS expansion records that are included in the 2001 ASEC but not in the 2001 March BMS. Note also that the total oversample linkage rate in this year is extremely low as the SCHIP oversample is first introduced in 2001 and so these households are not eligible to link back to the 2000 ASEC.

d False positive matches cannot be reliably identified in linking this year to the previous year; due to the household identifier recycling pattern, false positive matches are likely made using CPSID(P). This is especially true of the 2003-2004 link, where the linkage rate for SCHIP oversample records exceeds the eligible-to-link fraction of that oversample.

e Even though false positive matches are likely included in the 2005-2004 ASEC oversample link, overall linkage rates are slightly lower in this linked dataset than in surrounding years. I suspect that this is an artifact of the CPS sample redesign that took place between April 2004 and July 2005, though a specific reason for this low linkage rate is not found in official Census Bureau documentation.

Table 20. ASEC oversamples and their linkage rates

Year	*Cross-sectional Oversample Totals*								*Linked Oversample Totals*								*Oversample Linkage Rates*					
	Total Oversample Records		*Hispanic Oversample Records*		*SCHIP Oversample Records*		*Unclassified Oversample Records*[f]		*Total Oversample linked to Year - 1*		*Hispanic Oversample linked to Year - 1*		*SCHIP Oversample linked to Year - 1*		*Unclassified Oversample linked to Year - 1*		*Total Oversample Linkage Rate*		*Hispanic Oversample Linkage Rate*		*SCHIP Oversample Linkage Rate*	
	H	P	H	P	H	P	H	P	H	P	H	P	H	P	H	P	H	P	H	P	H	P
1990	3320	9349	2746	9349	--	--	574	0	953	2321	811	2321	--	--	142	0	28.70%	24.83%	29.53%	24.83%	--	--
1991	3506	10249	3007	10249	--	--	499	0	1311	3365	1158	3365	--	--	153	0	37.39%	32.83%	38.51%	32.83%	--	--
1992	3652	10441	3066	10441	--	--	586	0	1304	3356	1142	3356	--	--	162	0	35.71%	32.14%	37.25%	32.14%	--	--
1993	3749	10579	3157	10579	--	--	592	0	1423	3581	1233	3581	--	--	190	0	37.96%	33.85%	39.06%	33.85%	--	--
1994	3709	10318	3041	10318	--	--	668	0	1402	3524	1204	3524	--	--	198	0	37.80%	34.15%	39.59%	34.15%	--	--
1995[a]	18926	11721	3323	10945	--	--	15603	776	1133	3295	1133	3295	--	--	0	0	5.99%	28.11%	34.10%	30.11%	--	--
1996[b]	3666	10293	3062	10290	--	--	604	3	0	0	0	0	--	--	0	0	0.00%	0.00%	0.00%	0.00%	--	--
1997	3832	10865	3178	10865	--	--	654	0	1319	3623	1190	3623	--	--	129	0	34.42%	33.35%	37.44%	33.35%	--	--
1998	3957	11110	3278	11110	--	--	679	0	1325	3547	1163	3547	--	--	162	0	33.48%	31.93%	35.48%	31.93%	--	--
1999	4086	11548	3411	11548	--	--	675	0	1485	3865	1300	3865	--	--	185	0	36.34%	33.47%	38.11%	33.47%	--	--
2000	4295	12516	3635	12516	--	--	660	0	1467	3829	1303	3829	--	--	164	0	34.16%	30.59%	35.85%	30.59%	--	--
2001[c]	26937	80687	5492	18766	17873	61745	3572	176	1590	4152	1382	4152	--	--	208	0	5.90%	5.15%	25.16%	22.13%	--	--
2002[c]	26106	77559	5256	17914	17424	59645	3426	0	5614	18481	1758	5796	3786	12685	70	0	21.50%	23.83%	33.45%	32.35%	21.73%	21.27%
2003	26443	75136	5395	18263	16657	56873	4391	0	5696	18480	1767	5690	3771	12790	158	0	21.54%	24.60%	32.75%	31.16%	22.64%	22.49%
2004[d]	25806	74891	5654	18845	16513	56046	3639	0	7462	23051	2222	6759	5081	16292	159	0	28.92%	30.78%	39.30%	35.87%	30.77%	29.07%
2005[d,e]	25917	74333	5681	18980	16355	55353	3881	0	5675	17708	1592	4856	3976	12852	107	0	21.90%	23.82%	28.02%	25.58%	24.31%	23.22%
2006	25652	73534	6031	20005	15918	53529	3703	0	6440	18248	2158	6131	3827	12117	455	0	25.11%	24.82%	35.78%	30.65%	24.04%	22.64%
2007	25782	72822	6105	20272	15698	52550	3979	0	6403	18064	2295	6629	3623	11435	485	0	24.84%	24.81%	37.59%	32.70%	23.08%	21.76%
2008	25540	73249	6258	20610	15804	52639	3478	0	6590	18883	2386	6931	3792	11952	412	0	25.80%	25.78%	38.13%	33.63%	23.99%	22.71%
2009	25391	73271	6343	20861	15720	52410	3328	0	6324	18280	2332	6722	3605	11558	387	0	24.91%	24.95%	36.76%	32.22%	22.93%	22.05%
2010	25574	74324	6497	21889	15708	52435	3369	0	6635	19223	2506	7370	3704	11853	425	0	25.94%	25.86%	38.57%	33.67%	23.58%	22.61%
2011	25231	72708	6519	21813	15247	50829	3465	66	6483	18508	2453	7200	3590	11295	440	13	25.69%	25.46%	37.63%	33.01%	23.55%	22.22%
2012	24600	70026	6391	21280	14698	48683	3511	63	6228	17906	2433	7364	3345	10537	450	5	25.32%	25.57%	38.07%	34.61%	22.76%	21.64%
2013	25559	72100	6950	22708	14932	49324	3677	68	6303	17896	2569	7476	3295	10411	439	9	24.66%	24.82%	36.96%	32.92%	22.07%	21.11%
2014	25207	69829	6957	22401	14484	47376	3766	52	6660	18688	2729	7893	3465	10787	466	8	26.42%	26.76%	39.23%	35.24%	23.92%	22.77%
2015	25515	69213	7016	22591	14362	46560	4137	62	5704	15593	2363	6766	2892	8814	449	13	22.36%	22.53%	33.68%	29.95%	20.14%	18.93%
2016	25089	67497	6909	22186	14061	45252	4119	59	6415	17631	2650	7757	3219	9868	546	6	25.57%	26.12%	38.36%	34.96%	22.89%	21.81%
2017	25509	67264	7205	22772	13834	44458	4470	34	6366	17180	2625	7424	3213	9749	528	7	24.96%	25.54%	36.43%	32.60%	23.23%	21.93%
2018	24845	65793	7044	22268	13535	43458	4266	67	6429	17421	2724	7767	3163	9651	542	3	25.88%	26.48%	38.67%	34.88%	23.37%	22.21%
2019	23640	61993	6896	21628	12552	40314	4192	51	6134	16631	2636	7444	2982	9175	516	12	25.95%	26.83%	38.23%	34.42%	23.76%	22.76%
2020	21173	53081	5839	17895	10980	35134	4354	52	5664	14445	2355	6430	2659	8010	650	5	26.75%	27.21%	40.33%	35.93%	24.22%	22.80%
2021	21808	56209	6232	19351	11680	36826	3896	32	4822	12046	2187	5653	2109	6387	526	6	22.11%	21.43%	35.09%	29.21%	18.06%	17.34%
2022	20667	52197	5946	18268	10833	33905	3888	24	5177	13131	2167	5847	2450	7280	560	4	25.05%	25.16%	36.44%	32.01%	22.62%	21.47%
2023	20125	49787	5858	17861	10252	31905	4015	21	4828	12237	2044	5530	2242	6701	542	6	23.99%	24.58%	34.89%	30.96%	21.87%	21.00%
2024	19922	49131	5992	18043	9860	30945	4070	143	4850	11950	2071	5518	2194	6422	585	10	24.34%	24.32%	34.56%	30.58%	22.25%	20.75%

a Non-response households in the 1995 ASEC file cannot be linked either to the 1995 March BMS file or the 1994 ASEC file. As a result, non-response households that are actually part of the BMS portion of the 1995 ASEC file are classified as part of the oversample, leading to the unusually large number of unclassified oversample households in this year. Additionally, there are 951 individuals in the 1995 March BMS who do not link to the 1995 ASEC file (Flood & Pacas, 2017); the 776 unclassified oversample persons are potentially the ASEC counterparts of some of these BMS records that were not able to link across files. This unlinkability is a feature of the original Census Bureau household identifiers, though no explanation for this unlinkability is to be found in official documentation.

b Due to the sample re-design that took place between the 1995 and 1996 ASEC files, these years of ASEC data cannot be linked to one another.

c Oversample record counts and linkage rates are calculated after accounting for BMS expansion records that are included in the 2001 ASEC but not in the 2001 March BMS. Note also that the total oversample linkage rate in this year is extremely low as the SCHIP oversample is first introduced in 2001 and so these households are not eligible to link back to the 2000 ASEC.

d False positive matches cannot be reliably identified in linking this year to the previous year; due to the household identifier recycling pattern, false positive matches are likely made using CPSID(P). This is especially true of the 2003-2004 link, where the linkage rate for SCHIP oversample records exceeds the eligible-to-link fraction of that oversample.

e Even though false positive matches are likely included in the 2005-2004 ASEC oversample link, linkage rates among the Hispanic oversample are lower in this linked dataset than in surrounding years. I suspect that this is an artifact of the CPS sample redesign that took place between April 2004 and July 2005, though a specific reason for this low linkage rate is not found in official Census Bureau documentation.

f The unclassified oversample records are non-response households and households in which no member meets the Census Bureau's stated criteria for oversample eligibility and so cannot be classified as belonging to the Hispanic or the SCHIP oversample.

Table 21. ASEC oversample links using CPSIDV, 1989-2024

Year	*Total Oversample Records*	*Oversample Records Linked to Year - 1 (CPSIDP)*	*Validated Oversample links to Year - 1 (CPSIDV)*	*Linked Oversample Validation Rate*	*Oversample Linkage Rate (CPSIDP)*	*Validated Oversample Linkage Rate (CPSIDV)*
1990	9349	2321	2117	91.21%	24.83%	22.64%
1991	10249	3365	3025	89.90%	32.83%	29.52%
1992	10441	3356	3014	89.81%	32.14%	28.87%
1993	10579	3581	3284	91.71%	33.85%	31.04%
1994	10318	3524	3197	90.72%	34.15%	30.98%
1995[a]	11721	3295	3187	96.72%	28.11%	27.19%
1996[b]	10293	0	0	0.00%	0.00%	0.00%
1997	10865	3623	3522	97.21%	33.35%	32.42%
1998	11110	3547	3475	97.97%	31.93%	31.28%
1999	11548	3865	3712	96.04%	33.47%	32.14%
2000	12516	3829	3674	95.95%	30.59%	29.35%
2001[c]	75980	4152	4050	97.54%	5.46%	5.33%
2002	77559	18481	17595	95.21%	23.83%	22.69%
2003	75136	18480	17429	94.31%	24.60%	23.20%
2004[d]	74891	23051	18058	78.34%	30.78%	24.11%
2005[d]	74333	17708	15283	86.31%	23.82%	20.56%
2006	73534	18248	17599	96.44%	24.82%	23.93%
2007	72822	18064	17223	95.34%	24.81%	23.65%
2008	73249	18883	18053	95.60%	25.78%	24.65%
2009	73271	18280	17517	95.83%	24.95%	23.91%
2010	74324	19223	18383	95.63%	25.86%	24.73%
2011	72708	18508	17717	95.73%	25.46%	24.37%
2012	70026	17906	17299	96.61%	25.57%	24.70%
2013	72100	17896	17261	96.45%	24.82%	23.94%
2014	69829	18688	18036	96.51%	26.76%	25.83%
2015	69213	15593	14928	95.74%	22.53%	21.57%
2016	67497	17631	16965	96.22%	26.12%	25.13%
2017	67264	17180	16497	96.02%	25.54%	24.53%
2018	65793	17421	16721	95.98%	26.48%	25.41%
2019	61993	16631	15919	95.72%	26.83%	25.68%
2020	53081	14445	13890	96.16%	27.21%	26.17%
2021	56209	12046	11570	96.05%	21.43%	20.58%
2022	52197	13131	12644	96.29%	25.16%	24.22%
2023	49787	12237	11815	96.55%	24.58%	23.73%
2024	49131	11950	11480	96.07%	24.32%	23.37%

a Non-response households in the 1995 ASEC file cannot be linked either to the 1995 March BMS file or the 1994 ASEC file.

b Due to the sample re-design that took place between the 1995 and 1996 ASEC files, these years of ASEC data cannot be linked to one another.

c Oversample record counts and linkage rates are calculated after accounting for BMS expansion records that are included in the 2001 ASEC but not in the 2001 March BMS. Note also that the total oversample linkage rate in this year is extremely low as the SCHIP oversample is first introduced in 2001 and so these households are not eligible to link back to the 2000 ASEC.

d False positive matches cannot be reliably identified in linking this year to the previous year; due to the household identifier recycling pattern, false positive matches are likely made using CPSID(P). Because false positive links are included in the CPSIDP link across these years, validation rates are low compared to years where false links are not made or can be reliably identified and avoided.

e Even when false positive matches are removed through validation on demographic characteristics, the 2005-2004 ASEC oversample records link at slightly lower rates than in surrounding years. I suspect that this is an artifact of the CPS sample redesign that took place between April 2004 and July 2005, though a specific reason for this low linkage rate is not found in official Census Bureau documentation.

Table 22. ASEC-ASEC links using CPSIDV, 1989-2024

Year	*Total Records*	*Linked to Year - 1, BMS only*	*Linked to Year - 1, BMS + Oversamples*	*% Increase in linked records*	*Total Linkage Rate*
1990	158079	48487	50604	4.37%	32.01%
1991	158477	52330	55355	5.78%	34.93%
1992	155796	51587	54601	5.84%	35.05%
1993	155197	50717	54001	6.48%	34.80%
1994	150943	50930	54127	6.28%	35.86%
1995	149642	46139	49326	6.91%	32.96%
1996[a]	130476	0	0	0.00%	0.00%
1997	131854	44980	48502	7.83%	36.78%
1998	131617	45117	48592	7.70%	36.92%
1999	132324	45082	48794	8.23%	36.87%
2000	133710	45579	49253	8.06%	36.84%
2001[b]	218269	44354	48404	9.13%	22.18%
2002	217219	41936	59531	41.96%	27.41%
2003	216424	50248	67677	34.69%	31.27%
2004	213241	51632	69690	34.97%	32.68%
2005[c]	210648	44710	59993	34.18%	28.48%
2006	208562	47631	65230	36.95%	31.28%
2007	206639	47843	65066	36.00%	31.49%
2008	206404	49081	67134	36.78%	32.53%
2009	207921	49063	66580	35.70%	32.02%
2010	209802	50487	68870	36.41%	32.83%
2011	204983	48833	66550	36.28%	32.47%
2012	201398	49619	66918	34.86%	33.23%
2013	202634	49270	66531	35.03%	32.83%
2014	199556	48308	66344	37.34%	33.25%
2015	199024	39928	54856	37.39%	27.56%
2016	185487	42683	59648	39.75%	32.16%
2017	185914	40321	56818	40.91%	30.56%
2018	180084	39601	56322	42.22%	31.28%
2019	180101	40457	56376	39.35%	31.30%
2020	157959	40156	54046	34.59%	34.22%
2021	163543	36737	48307	31.49%	29.54%
2022	152732	35933	48577	35.19%	31.81%
2023	146133	34656	46471	34.09%	31.80%
2024	144265	33374	44854	34.40%	31.09%

a Due to the sample re-design that took place between the 1995 and 1996 ASEC files, these years of ASEC data cannot be linked to one another.

b Note that the linkage rate in this year is lower than surrounding years as the SCHIP oversample is first introduced in 2001 and so these households are not eligible to link back to the 2000 ASEC.

c The 2005-2004 ASEC oversample records link at slightly lower rates than in surrounding years. I suspect that this is an artifact of the CPS sample redesign that took place between April 2004 and July 2005, though a specific reason for this low linkage rate is not found in official Census Bureau documentation.