IPUMS Working Papers

A Holistic Approach to Validating Current Population Survey Panel Data

Renae Rodgers†
University of Minnesota

Sarah M. Flood
University of Minnesota

July 2023

Working Paper No. 2023-01
DOI: https://doi.org/10.18128/IPUMS2023-01

†Address correspondence to Renae Rodgers: rodge103@umn.edu. This work was supported by the Integrated Current Population Survey Data for Population Dynamics and Health Research grant R01HD067258, funded by the Eunice Kennedy Shriver National Institute for Child Health and Human Development.

A Holistic Approach to Validating Current Population Survey Panel Data

Renae Rodgers* and Sarah Flood

Institute for Social Research and Data Innovation, University of Minnesota

Abstract:

The Current Population Survey (CPS) is the primary source of labor force data in the United States; the panel component of the CPS provides an excellent opportunity for studying short-run change in areas such as labor force participation and poverty. The Census Bureau's recommended procedures for linking CPS data occasionally yield a collection of matched observations that appears — based on demographic information — to represent different people rather than a single person across CPS files. We describe our methodology for assigning a unique identifier to records that share Census Bureau identifiers and have demographic characteristics that are consistent across observations. The identifier eliminates a currently-recommended step when working with linked CPS data – checking to ensure that records with a shared identifier look like the same individual based on age, sex, and race – and provides researchers a common starting point for analyzing validated CPS panel data. This identifier extends previous efforts at IPUMS CPS to create unique identifiers for leveraging the panel aspect of the CPS and are available from IPUMS CPS.

Keywords: Data integration, linking, panel data, validation, Current Population Survey

1. Introduction

The Current Population Survey (CPS) is the primary source of labor force data in the United States. This household survey is fielded monthly and includes data on employment, wages, occupation, and industry as well as demographics and family structure. Certain months also include supplements on topics such as food security, education, and fertility. IPUMS is a leading provider of CPS data (cps.ipums.org), giving the research community streamlined access to this important data set.

The CPS has a unique but underutilized panel component that can be used to examine short-run changes in labor force participation, health insurance coverage, family structure, and a host of other topics relevant to social science research. Individuals can be observed up to eight times over a 16-month period. Households surveyed in the CPS follow a 4-8-4 rotation pattern where a household is interviewed for four consecutive months, rotates out of the sample for the next eight months, and is then interviewed again for four consecutive months. A household's place in the CPS rotation is called its "month-in-sample" (MISH) and can take on values of one through eight.

While the Census Bureau provides identifiers for the purpose of linking records across CPS files, these identifiers in their original form are difficult to use. Challenges include changing names and number of variables required to uniquely identify households over time (Drew et al., 2014) and duplicate records in data from the late 1970s and early 1980s (Flood et al., 2020). To reduce the technical difficulty of working with linked CPS data, IPUMS CPS currently provides unique identifiers for easily linking CPS Basic Monthly Survey (BMS) files (CPSIDP) (Drew et al., 2014; Flood et al., 2020) and for linking the March BMS respondents to the ASEC file (MARBASECIDP) (Flood & Pacas, 2017; Flood et al., 2020). These identifiers reflect the Census Bureau's guidelines for linking CPS data from 1976 to the present along with some necessary adjustments as described.

The Census Bureau's recommended procedures for linking CPS data (as captured in CPSIDP) matches some records that appear to be different people at different points in the survey based on demographic information. IPUMS CPS documentation for CPSIDP distinguishes between "mechanical" and "plausible" linkages, the former being the results of links using only Census Bureau identifiers and the latter being the subset of the mechanical linkages that have expected age, sex, and race values across all linked observations (Drew et al., 2014; Flood et al., 2020). For example, linking the January and February 2009 BMS files using CPSIDP results in 96,822 linked records; however only 96,049 of those links (96.1%) have the same sex and race values and age values that increase by less than three years across the two observations. Of the 11,528 records of individuals who entered the CPS in January of 2009 and completed all eight interviews, only 10,989 records (64.9%) had the same sex and race values and age values that increased by less than three years across all observations (Drew et al., 2014).

A post-linking validation step is recommended to remove these specious links (Madrian & Lefgren, 2000; Feng, 2001; Drew et al., 2014). However, validating CPS links is more complicated than it first appears and represents a real obstacle to use of the CPS panel component as well as to reproducibility of research utilizing this aspect of the CPS. To address this problem, IPUMS CPS has constructed a new linking key called CPSIDV that only links records across months that are plausible, or "valid", based on the demographic characteristics age, sex, and race. CPSIDV removes technical barriers to working with linked CPS data by implementing a consistent validation methodology across time. CPSIDV eliminates the need for researchers to write their own validation code, which is complicated, error-prone, and may be based on differing definitions of validity. With this linking key, IPUMS CPS has further lowered the barrier to using the panel component of the CPS and improved reproducibility of research by providing researchers with a common starting point for working with linked and validated CPS data. This paper first details the challenges we encountered in producing a single, unique, validated identifier, then describes our methodology for creating such an identifier, and finally presents validation rates for researchers to check their own linkages.

2. Existing validation methods

The need for validating CPS linkages is well documented (Madrian & Lefgren, 2000; Feng, 2001; Drew et al., 2014). While there is no standard approach to validating CPS linkages, most validation strategies recommend using age, sex, and race for this purpose. In this section we describe several existing methods for validating linked CPS data and discuss how our methodology for generating CPSIDV differs from previous approaches.

The earliest methods for validating CPS linkages were intended for use in linking only two time points – specifically the ASEC files in adjacent years. The most popular method for validating linked CPS data was originally promulgated by Madrian and Lefgren (2000). In their method for linking CPS ASEC observations across a one-year period, Madrian and Lefgren employ demographic characteristics to ensure valid matches, requiring sex and race to be identical across both observations and for age to change in expected ways. For a link to be considered valid, they required the difference between age of the second and first observation to be between -1 and 3. In other words, a person may have aged no more than 3 years between observations and may have even aged backwards by one year. While the expected value of age change across the 12-year period that Madiran and Lefgren validate is between zero and two, they extend the accepted range to capture potential measurement error in age (Madrian & Lefgren, 2000). Feng (2001) also uses demographic information in his Bayesian approach to linking ASEC observations. Feng adopts Madrian and Lefgren's criteria for age, sex, and race and requires marital status to be unchanged when considering whether records match across years. Because marital status may legitimately change over the course of an individual's 16-month CPS rotation, requiring this variable to be static over time may lead to the rejection of valid linked observations and MISHs an important demographic transition in the process. For this reason, we do not consider marital status when assigning CPSIDV.

Other validation methods (Drew et al., 2014; Nekarda, 2009) accommodate validation of more than two linked CPS interviews. In their original description of CPSID(P), Drew and colleagues (2014) define a "plausible" link as one whose sex and race does not change over time and whose age either does not change or increases by one or two years. When linking multiple observations across time, the first age value is used as the reference to test all subsequent age values. If any of the subsequent age values do not meet the validation criteria, the link is not considered plausible. Nekarda (2009) takes a less rigid approach to validating CPS. In this methodology records that link successfully on household and person identifiers are given a validity score based on all of their CPS interviews. An observation that does not have the same sex or race as the first observation in the CPS or has an age value that increases by more than two years compared to the first observation in the CPS reduces the linked record's validity score.

Unlike the existing validation approaches described above that can only accommodate two time points or that rely on a fixed reference observation to determine validity, our approach to validating CPS linked data produces a single identifier that researchers can use to both link and validate CPS observations in one step for all CPS basic monthly records from 1976 through the present. By using this identifier, researchers will be able to link only validated CPS observations across months, eliminating the complicated post-linking validation step. While our method still relies on age, sex, and race to validate matches, generating this type of linking key required us to depart from existing methodologies in three ways. First, we are flexible in which observation we use as a basis for comparison when determining a link's validity so as to capture as many valid links as possible. Second, we impose more robust age validation criteria that takes all available age information into account in order to ensure respondents do not age backwards at any point, that age topcodes are appropriately addressed, and that changes in age topcodes over time are adequately bridged. Finally, we impose more robust race validation criteria that bridges changes in available detail about respondents' race over time. In the following section, we describe our method for determining the proper basis for validation, detail our validation criteria, and discuss the technical details of implementing these criteria to produce CPSIDV.

3. Method for creating a unique, validated person-level identifier

There are many different ways to leverage the CPS panel component to create datasets with multiple time points. Researchers can create full panels of all eight CPS interviews, link two observations one year apart, link observations from adjacent months, or link between months that have topical supplement information of interest. CPSIDV is a single, unique identifier that creates links based on household identifiers, person identifiers, age, sex, and race for all types of links made possible by the CPS panel component while maximizing the number of valid matches within the panel. Because of the incorporation of demographic characteristics into identifier creation, we call these links “validated”. CPSIDV is general-purpose and can be used to generate these validated links for any type of possible link in the CPS with any number of time points from adjacent-month links to full-panel links. CPSIDV builds upon the IPUMS CPS mechanical linking key, CPSIDP (Drew et al., 2014; Flood et al., 2020; Flood et al., 2022). Using CPSIDP as our starting point, we assign a new linking key to records that share a CPSIDP value across months and whose sex and race do not change and whose age values change in expected ways over time.

3a. The basis for validation

In the case of only two linked observations, the first observation is the obvious reference point for determining the validity of the second. However, when expanding the scope of linking validation to include more than two time points, a simple comparison of the first observation to all subsequent observations poses two problems. First, this method does not take intervening observations into account, which may lead to invalidating true links by failure to recognize a subset of valid observations or to validating suspicious links, especially when validating age. Second, validation that depends on only the first observation for determining validity does not scale well, as what constitutes the first observation depends on a researcher's chosen first time point and linkage type.

The following example illustrates the limitations of a methodology that relies exclusively on the first observation to determine validity. For example, the respondent with CPSIDP=20180800009001 (shown in Figure 1) is aged 40 from August 2018 to November 2018. However, in August of 2019, this individual is reported to be age 63 and age 64 from September 2019 to November of 2019. If all observations are validated against the first, observations from August to November of 2019 will be considered invalid and will not link to any other observations. However, it appears that the records where CPSIDP=20180800009001 actually represent two different people – one person before the eight-month break and one person after. If a validated linking key were assigned on the basis of the record's first CPS observation alone, observations between August of 2019 and November of 2019 would not link to one another even though, when looking at the 2019 observations only, these observations appear valid compared to one another.

[Figure 1]

Relying only on a comparison to the first observation's demographic characteristics may also validate links that appear suspicious. For example, consider the record first observed in the CPS in October of 2003 whose CPSIDP=20031002261002. Figure 2 shows all eight CPS appearances of CPSIDP=20031002261002. If a researcher were allowing an individual to age up to three years over the course of the survey as Madrian and Lefgren do, and used only the first observation as the basis for validation, both the January 2004 (age=52) and the October 2004 (age=51) observations would be considered valid as compared to the first 2003 October observation (age=49), even though when these observations are viewed in their entirety, the individual appears to have aged backwards between January and October of 2004.

[Figure 2]

Because households appear up to eight times in the CPS survey, what constitutes the first observation is relative and may change according to the linked data sets constructed by individual researchers. Shifting which of a record's eight CPS surveys a researcher considers the reference observation or what type of linkage a researcher constructs may have consequences for whether links within the same CPSIDP value are considered valid. To return to the example in Figure 2, if a researcher chooses October 2003 as their first time point and links to the next year's observation in October 2004, age values do not violate common validation criteria as the individual has aged two years (see below for details of CPSIDV validation criteria), and this link would be considered valid. However, if a researcher makes the same adjacent-year link but uses January 2004 as their first time point, this individual appears to have aged 4 years in the course of 12 months, which violates our age validation rules. As a result, this link would be considered invalid. Comparisons of multiple pairs of observations that share a CPSIDP may thus yield contradictory validation conclusions depending on the pair of observations under consideration. A similar scenario can occur when the same observation is considered the reference observation but the linkage type differs. For example, consider MISH 1 for the record with CPSIDP=20020907222201 (shown in Figure 3) as the reference time point. When linking adjacent months (September to October of 2002), both observations have the same age value. However, when linking adjacent years (September 2002 to September 2003), there is an age gap of 44 years, making this link invalid.

[Figure 3]

Capturing all of the possible combinations that rely on the first observation as the basis for validation in a linking key is neither desirable nor practical. Instead, we take a holistic approach and consider information from all previous CPS responses when validating linkages based on age, sex, and race, and can distinguish between multiple plausible sets of linkages within a set of CPSIDP-linked observations. We take into account the number of months that have passed since the previous CPS interview as well as the number of months that have passed since the last change in age to help determine validity of age rather than a simple comparison of each subsequent age value to the first. When we encounter an observation that does not meet the validation criteria (See the *Validation Criteria* section below for details), we consider this invalid link to be a new person, appearing in the survey for the first time. The validity of subsequent observations is first evaluated against this newly identified individual. If any observations subsequent to the first appearance of this new person are invalid, these observations are checked against previous "persons" within the CPSIDP group for validity (See *Implementation of CPSIDV* section below for details).

This approach is sufficiently robust to avoid validating links that have suspicious variation in age, or any variation in sex or race, but is flexible enough to recognize multiple sets of valid links within a set of CPSIDP-linked observations. In addition, this approach results in a "static" validated identifier – that is, whether a set of observations is valid remains stable regardless of which month a researcher chooses as their reference point. As a result, CPSIDV may not link some sets of observations that appear valid when considered independently of the rest of the CPSIDP observations.

3b. Validation Criteria

In this section, we discuss validation criteria for CPSIDV and describe the approach to bridging changing race codes and age topcodes over time. Sex and race must be constant for links to be considered valid. Age must change in expected ways over time. After comparing two different approaches to age validation – one relaxed and one strict – we implement a relaxed aging validation scheme in CPSIDV in line with previous research (Madrian & Lefgren, 2000; Feng, 2001; Drew et al., 2014). As a result, individuals are allowed to age more quickly than seems logical over the course of their participation in the CPS as long as they do not age backwards or by more than two years during that 16-month period. For a detailed comparison of age validation methods, see Appendix A.

3b1. Sex

For a link to be valid, sex must remain constant. For respondents who complete all eight CPS interviews, validation on sex remains consistently high over time. There are no changes in the way that sex is recorded or coded that requires bridging over time.

3b2. Race

For a link to be valid, race must also remain constant. In instances where a respondent's CPS participation spans a change in race codes, race values from the current month must map to allowed race values in previous months to be considered valid. Race codes have changed four times in the Basic Monthly Survey. These changes occur between December 1988 and January 1989, December 1995 and January 1996, December 2002 and January 2003, and April and May 2012. Table 1 shows how IPUMS maps race codes across all four of these changes.

[Table 1]

Prior to 2003, CPS respondents could choose one, single-race category. Between 1995 and 1996, the "Other" category was eliminated and replaced with American Indian/Alaska Native and Asian/Pacific Islander. Records from 1995 that have an "Other" code may map to any single-race category in later years for validation purposes.

Beginning in 2003, respondents could select more than one race category to describe themselves and the number of CPS race categories increased from four to 21. For the purposes of validation, we allow multi-race categories from 2003 to May 2012 to map back to single-race categories in previous years that are a component of their new multi-race designation. For example, consider a record that is categorized as "White-Black" in January of 2003 that also appears in December of 2002. Either "White" or "Black" race values for this record in December of 2002 would yield a valid link to the January 2003 observation. In bridging this code change, records reporting "2 or 3 races" or "4 or 5 races" (codes 20 and 21, respectively) in 2003 and onward may have any of the available single race codes under the 1996-2002 coding scheme.

In May 2012, five more multi-race categories were added to the CPS race variable. The race question in the CPS survey does not change between April and May of 2012. The new categories available starting in May 2012 were possible in April 2012 but not distinguished from other codes, perhaps for privacy reasons. When bridging this code change, records whose May 2012 race category existed prior to May 2012 must have the same code before and after the inclusion of the additional multi-race categories. For example, the category "White-Asian" exists both before and after May of 2012. A record in this race category in April of 2012 must still have this race code in May 2012 forward for the link to be considered valid. Categories that were added in May 2012 must map back to the unspecified multi-race category with the appropriate number of race categories prior to that month. For example, a record categorized as "White-Black-Hawaiian/Pacific Islander" in May of 2012 must be in the "2 or 3 races" category in April of 2012 to be considered a valid link. Similarly, a record categorized as "White-American Indian-Asian-Hawaiian/Pacific Islander" in May of 2012 must be in the "4 or 5 races" category in April of 2012 to be a valid link. Race validation rates are high among respondents to all eight CPS interviews even across changes in race coding. Of those full-panel links that are valid on sex and whose CPS participation does not span a race code change, 99% of all full panel links are valid on race. Those full-panel links whose

CPS participation do span a change in race coding have slightly lower validation rates on race, however even the lowest of these is over 97% (See Table 4 in the *Validation Rates* section).[1]

3b3. Age

There are multiple approaches to using age to validate CPS linkages. Prior approaches have used relatively simplistic age criteria to validate a link. For example, Madrian and Lefgren (2000) allow an increase of up to three years or a decrease of one year from the first observation to the second across a one-year period. Drew et al. (2014) consider a link "plausible" if a linked person's age does not increase by more than two years in age as compared to the first observation and does not decrease. Like Drew et al. (2014), CPSIDV does not allow respondents to age backwards. This choice simplifies the somewhat complex implementation of CPSIDV (see next section) and does not result in much lower validation rates than those reported by Madrian and Lefgren (2000) (See *Validation Rates* section below). We considered two possible age validation schemes, one relaxed and one strict. We ultimately chose to incorporate the relaxed age validation criteria into CPSIDV. The strict approach we considered is described in detail in Appendix A as is a comparison of the strict validation criteria with the relaxed age validation criteria that we ultimately implemented.

Our age validation criteria for CPSIDV are as follows. For a link to be considered valid, age must change in the following expected ways over the course of the CPS rotation:

- 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.

[1] The "Other" category is completely absent from the data in September of 1995 for unknown reasons. Due to the linking discontinuity between 1995 and 1996, the only records that can appear in all eight rotations and end in 1996 begin in this month. As a result, records that ever fall into the RACE category of "Other" in this group will never validate on RACE across the full panel of CPS observations.

- The respondent may age at any point in the 16-month rotation, but they must age by at least one year over the course of the 16-month period, unless they enter the survey with a topcoded age value.
- Beginning in 2004, respondents aged 80 may age 5 years to accommodate topcoding.

Age is topcoded differently across time (See Table 2). These topcode changes must also be bridged when generating CPSIDV. There are four different age topcodes in the years between 1976 and 2020. From 1976-June 1985, ages over 99 are recorded as 99. From July 1985-January 2002, ages over 90 are recorded as 90. From February 2002-March 2004, ages over 80 are recorded as 80. From April 2004 onward, ages from 80 to 84 are recorded as 80 and ages 85 and over are recorded as 85. To bridge these changes in topcodes, respondents with topcoded age values can age backward across these years. Beginning in April 2004, respondents aged 80 may age five years to accommodate the two topcoded age values in this period.[2]

[Table 2]

The following examples illustrate our age validation scheme. Figure 4a shows a record that began the CPS rotation in August of 2018 and appeared all eight times, finishing the rotation in November 2019. The observed ages are valid across all eight CPSIDP links. As such, any link, whether it is a simple month-to-month link or a multi-observation link, will be considered valid.

[Figure 4a]

[2] There are some discrepancies between the original CPS documentation and the data regarding age topcodes. In 2002, the original codebooks indicate that all months should have age values up to 90 (U.S. Census Bureau, 1998), however these values up to age 90 are only present in January of 2002, with age in all other months, including the ASEC, topcoded at 80. Similarly, only the codebook for the 2004 ASEC indicates an age topcode of 85 while those from other months indicate a topcode of 80 from January to April (U.S. Census Bureau, 2003), and of 90 from May on (U.S. Census Bureau, 2004). However, the data are topcoded at 80 from January 2004-March 2004 while the second topcoded value of 85 appears in the 2004 basic monthly data beginning in April. No months in 2004 have age values over 85.

Figure 4b shows another set of records with the same CPSIDP value throughout the CPS rotation record. This CPSIDP value first appears in August 2018, responded to all eight interviews, and finished the rotation in November 2019. However, this record does not validate on age for all eight links. In this case, MISH one through four will link to one another, and MISH five through eight will link to one another, but observations from before and after the eight-month break are invalid relative to one another.

[Figure 4b]

3c. Implementing CPSIDV

In this section we discuss the technical details of creating CPSIDV. In addition, we provide linkage rates using CPSIDV validation criteria for full-panel links, adjacent-year links, and adjacent-month links. For full-panel data, we use linked records that appear in all eight CPS rotations, with their first interview within the given year. For example, the 2019 data contains persons in the CPS who responded to all eight CPS interviews and had their first interview in any month in 2019. For example, a record that entered the CPS in December of 2019 in MISH 1 and exits the CPS after their MISH 8 interview in March of 2021 is included in the 2019 group.

CPSIDV is a holistic and static linking key, meaning that whether a given set of observations constitute valid links is determined based on all appearances of an individual in the CPS, independent of the specific observations being linked. CPSIDP provides the starting point for CPSIDV; a single CPSIDV value will be assigned for records with the same CPSIDP whose age, sex, and race values are identical or change in expected ways (see *Validation Criteria* section above for details). Compared to CPSIDP, CPSIDV has an additional trailing digit which acts as a unique person identifier within the CPSIDP group. In cases where a single CPSIDP value appears to represent multiple separate individuals, the multiple individuals will be distinguished by the additional trailing digit. This approach allows for multiple sets of validated linkages within a set of records with the same CPSIDP.

The first time an individual is observed in the CPS (MISH=1 or the first instance of CPSIDP) a unique CPSIDV is assigned to them. In this case, CPSIDV is equivalent to CPSIDP with a "1" appended. For subsequent instances of CPSIDP, we compare the current age, sex, and race values to the demographic characteristics from the most recent previous appearance of the same CPSIDP value. This previous appearance is usually, but not necessarily the previous month due to the rotation pattern and non-response. This comparison takes into account how much age has increased since entering the survey, how much age has increased since the last interview, and the amount of time elapsed since the last interview. Records with the same CPSIDP, sex, and race values and allowable age values are assigned an existing CPSIDV value. In this way, CPSIDV, like CPSIDP, is backward-looking. Links are validated first on sex, then race, then age. When a link fails to validate on one of these characteristics, the remaining characteristics are not examined for this link. For example, if a link fails to validate on sex, race and age are not checked. If the conditions for a valid link are met, the CPSIDV from the previous appearance is assigned to the record in the current month.

If the comparison to the observation from the most recent previous survey does not meet the validation criteria and the validation digit of the CPSIDV from the previous appearance is one, a new unique CPSIDV value is generated and assigned to the observation in the current month. If the validation criteria are not met and the validation digit of the most recent previous appearance is greater than one, the program performs a validity check against the next previous CPSIDV value for that set of CPSIDP values, and so on, until it either finds a valid match or runs out of CPSIDP instances to check. If no valid match is found, a new CPSIDV is generated by incrementing the validation digit by one. Figure 5 illustrates the assignment of CPSIDV to a record that enters the CPS in July of 2018.

[Figure 5]

When a CPSIDP value first appears the CPS (panel A of Figure 5), we assign a CPSIDV value with a validation digit equal to one and auxiliary variables "years aged since time 1", "months since last observation", and "months since last aged" are set to zero. In panel B of Figure 5, the first two observations have the same CPSIDV, as their age, sex, and race values are the same and only one month has passed since the first observation and these observations appear to belong to the same individual. However, by the third appearance (panel C of Figure 5), age has increased eight years in one month, which violates the validation criteria. As a result, the final digit of CPSIDV is incremented by one and the auxiliary variables are set to zero. The third observation will not link to the first two using CPSIDV, as the third observation appears to be a different individual as compared to the first two. Similarly, the fourth CPS observation (panel D of Figure 5) does not validate against the third, as individuals are not allowed to age backwards. Because the validation digit of the third observation is greater than one, the fourth observation will also be checked against the second (the last observation with the validation digit = 1). Since the fourth observation does not validate against the second observation and the second observation's validation digit is one, we assign a new CPSIDV value.

The age value for the fifth observation (panel E of Figure 5) could validate against either the third or the fourth appearance. However, once a valid match is found moving back in time from the current observation (MISH = 4 in this example), that observation's CPSIDV is assigned and no additional validation attempts are performed. Panel F of Figure 5 shows all appearances in the CPS of the CPSIDP=20180700156401.[3]

---

[3] Though respondents may age at any time during the CPS rotation under our age validation criteria, it is worth pointing out that this age progression may be real even though this individual appears to age two years in the span of 12 months. This is to account for possibility that age increments during the interview week and that the interview occurs before the respondent's birthday in the first year and after in the second year. The CPS is fielded the week that contains the 19th of each month (U.S. Census Bureau, 2006). In the example from Figure 5, suppose

Whereas outcomes of previous approaches to validation depend on which CPS month one chooses as a starting point and which type of link (e.g., adjacent-year or full-panel) is being executed, CPSIDV is static. That is, whether a link or links are considered valid is pre-determined based on all previous instances of a CPSIDP value, regardless of which months of the CPS a researcher chooses to link. This approach fails to make some links that seem plausible. For example, if one was only linking only the MISH 3 and MISH 5 observations in the example above, these links would appear valid, but knowledge of the intervening observations introduces some uncertainty about the validity of that linkage. In this way, CPSIDV is holistic and independent of which months are being linked and what type of link is being made.

3c1. CPSIDV and the ASEC

The Current Population Survey's Annual Social and Economic supplement (ASEC) contains detailed information regarding income, health insurance coverage, and poverty. The ASEC files are composed of all March basic monthly respondents plus a Hispanic oversample, beginning in 1976, and a State Children's Health Insurance Program (SCHIP) oversample beginning in 2001 (U.S. Census Bureau, 2006). As with CPSIDP, CPSIDV can be used to link ASEC to basic monthly (BMS) files. The process of linking CPS March basic monthly and ASEC files using Census Bureau identifiers is described in detail elsewhere (Flood & Pacas, 2017; Flood et al., 2020). As in the case of linking across months, creating CPSIDP for the ASEC does not systematically incorporate demographic characteristics. Validation of the March BMS-ASEC link is straightforward. If sex and race match and age does not differ by more than two years between CPSIDP appearances across these two files, the record's CPSIDV from the March BMS is assigned to the corresponding record in the ASEC file. If any of the three demographic characteristics violate these rules between CPSIDP pairs, the trailing digit of the CPSIDV from the March BMS is

---

the individual was interviewed on the 17th of October 2018 (MIS = 4), turned 58 on the 19th of October 2018, and was then interviewed on the 21st of October 2019, after they had turned 59.

changed to 0 and this value is assigned to the record in the ASEC file. This record in the ASEC file will be unable to link to the March BMS or any other basic monthly file.

The information in the ASEC and March BMS files for records that appear in both files was collected during the same interview (U.S. Census Bureau, 2006). Thus, in theory, all demographic variables, including age, should be identical for the same records across these two files. However, there are some occasions when different Census Bureau editing procedures are applied to the two files. For example, in 2019, the Census Bureau overhauled the ASEC editing procedures to take full advantage of the survey redesign completed in 2014 (Renwick, 2019). This work involved an update to the demographic edit that is applied to all CPS files (R. Rodgers, personal communication, January 24, 2022). However, the new demographic edit was not applied to the Basic Monthly files until 2020. Similar mismatches occur in several earlier years as well, however in almost all years between 1976 and 2021, all ASEC-BMS CPSIDP links match exactly on sex, race, and age. To achieve more realistic validation rates in years where these mismatches occur, we allow age to vary by up to two years across BMS and ASEC files. Table 3 below shows the number of observations that don't match on these characteristics in years where mismatches occur.

There are three years in which a difference in age topcode values between the March BMS and ASEC files needs to be bridged. In 1986 and 1987, the ASEC data have an age topcode of 99 while the March BMS is topcoded at 90. When linking between ASEC and March BMS in these years, any ASEC record with an age value greater than 90 must have an age value of 90 in the March BMS to be considered a valid link. In 2004, the ASEC data has a double topcode where respondents aged 80 to 84 are coded as 80 and those 85 and older are coded as 85. The 2004 March BMS data have a single topcode at 80. When linking between ASEC and March BMS in 2004, an ASEC record with an age of 85 must have an age value of 80 in the March BMS to be considered valid.

Race codes need to be bridged to link the 1988 March BMS and ASEC files. The 1988 ASEC file has codes for "American Indian" and "Asian or Pacific Islander" while the 1988 March BMS file includes these categories in "other". CPSIDP links between the 1988 ASEC and March BMS file that have an "American Indian" or "Asian or Pacific Islander" designation in the former and "other" in the latter are considered valid.

[Table 3]

The Hispanic and SCHIP oversample records in the ASEC files cannot be linked to any Basic Monthly files. Because of this, their CPSIDP value is zero (Flood & Pacas 2017). As CPSIDP links serve as the starting point for CPSIDV, records that cannot be linked using CPSIDP, including the ASEC oversample records, cannot be linked using CPSIDV. Hispanic and SCHIP oversample records have a CPSIDV value of zero.

3c2. Linking Discontinuities and Unlinkable Records

There are several points throughout the CPS where changes in the Census Bureau household identifiers cannot be bridged. In these places, CPSIDP cannot link observations across these changes and, as a result, neither can CPSIDV. These linking discontinuities occur within 1976 and 1977, between 1976, 1977, and 1978, within 1985, between 1985 and 1986, within 1995, and between 1995 and 1996. (Drew et al., 2014; Flood et al., 2020). While there is a change in the household identifiers between 1993 and 1994, sufficient identifying information is retained in 1994 and later that this change can be bridged (Drew et al., 2014). Because CPSIDP is able to bridge this household identifier change, CPSIDV can do so as well. Records that are considered unlinkable due to these linking discontinuities have CPSIDP values of zero; these records also have a CPSIDV value of zero.

Children age 14 and under are only present in some of the files between 1976 and 1981. Due to the irregularity with which child records appear in sample during this period, these records are considered unlinkable for CPSIDP (Flood et al., 2020). As these records cannot be linked using CPSIDP, they cannot be linked using CPSIDV either and have a value of zero for this variable. Similarly, children under 14 are present in the ASEC files between 1976 and 1988, but not in the March BMS files from these years. Children in the ASEC files during this period have CPSIDV values of zero.

There is a linking discontinuity between the 1977 ASEC and March BMS files and, as a result, neither CPSIDP nor CPSIDV are available for the 1977 ASEC file. However, the 1977 March BMS file can be matched to and validated against observations in other BMS files and therefore both CPSIDP and CPSIDV have non-zero values in this sample. Due to data quality issues in 1976-1988, some March basic monthly records cannot be linked to the ASEC files in this period using CPSIDP (Flood et al., 2020). As a result, these March BMS records cannot be linked to the ASEC using CPSIDV either. However, because these March BMS records may link to other BMS files using CPSIDP, they have a non-zero CPSIDV value.

4. Validation Rates

The CPS 4-8-4 rotation pattern allows for many different types of linkages. Researchers can create full panels of all eight CPS interviews, link two observations one year apart, link observations from adjacent months, or link between months that have topical supplement information of interest. Full-panel links with all eight responses have the lowest validation rate of any type of link due to the number of responses to validate; adjacent-month links have the highest validation rate of any type of link.

Table 4 shows CPSIDV validation rates on sex, race, and age for full-panel links. Full-panel links that complete the rotation in a given year are included with that year. For example, a record that enters the CPS in January of 2019 and completes the rotation in April of 2020 is included with the 2019 group. Note that due to linking discontinuities before and after June of 1985 and before and after September of 1985, fewer households that enter in 1984 and 1985 can be linked to all eight CPS appearances (Flood et al., 2020). A similar phenomenon occurs for those entering the CPS in 1994 and 1995.

For individuals who enter their CPS rotation in 1980-1993, validation rates slowly decrease over time, yielding linkage rates between 86.96% and 92.45%. With the introduction of Computer Assisted Telephone Interviewing (CATI) in 1994, validation rates increase. For those records beginning their rotation between 1994 and 2001 who complete all eight interviews, validation rates were above 93%. Coincident with the introduction of age perturbation in 2002 (U.S. Census Bureau, 2021), validation rates remain relatively stable. Between 92.17% and 94.46% of mechanical linkages that complete the CPS rotation pattern between 2002 and 2010 are valid linkages. Coincident with a change in age perturbation procedures in 2011 (U.S. Census Bureau, 2021), validation rates for full-panel links rise slightly to between 94.83% and 96.43%.

[Table 4]

Linking observations one year apart is another common CPS linkage type. Table 5 shows adjacent-year links by MISH group for records entering the CPS in 1978-2019. Linkage rates using CPSIDV are higher for adjacent-year links than for full-panel links, but trends over time are similar to those seen in the full-panel links. Adjacent-month links are also popular CPS linkages. Linkage rates using CPSIDV are highest for this type of link. Table 6 shows adjacent-month links for all MISH pairs between 1978 and 2019.

[Table 5]

[Table 6]

CPSIDV is a single, unique and validated identifier that is applicable for all types of CPS links and all available CPS files. This streamlined, general-purpose approach yields similar validation rates to those achieved by popular methods in the literature. Though CPSIDV may fail to make some links that appear to be valid in certain situations, CPSIDV linkage rates are very similar to those achieved using validation methodologies of both Drew et al. (2014) and Madrian and Lefgren (2000). Table 7 illustrates the performance of CPSIDP, Drew and colleagues' validation method, and CPSIDV for links between January and February of 1978, 1996, 2009, and 2018. For each year, there are three columns indicating linkages using CPSIDP only, "Plausible" links as defined by Drew et al. (2014) (sex and race are unchanged and age increases by zero, one, or two years), and CPSIDV. The number of links retained in these adjacent-month links after Drew and colleagues' validation and those linked using CPSIDV is nearly identical in these adjacent-month links in all years and validation rates are within 1 percentage point of one another.

[Table 7]

Table 8 shows the same linkage and validation rates for links between March of 1978 and 1979, 1996 and 1997, 2009 and 2010, and 2018 and 2019 and also includes linkage rates achieved by Madrian and Lefgren's validation criteria for records linked across adjacent years. Linkage rates using CPSIDV are similar to retention rates when using Drew et al. and Madrian and Lefgren's validation methods.

[Table 8]

As Tables 7 and 8 show, CPSIDV achieves slightly fewer validated links for both adjacent-month and adjacent-year linkage types than Drew and colleagues' and Madrian and Lefgren's validation methods. While Madrian and Lefgren's method is only designed to accommodate two time points one year apart and Drew and colleague's method relies on single reference point to determine validity, CPSIDV can be used to make any kind of link made possible by the CPS rotating panel structure. Thus, CPSIDV yields comparable linkage rates to popular methods already in use while offering a flexible, general-purpose solution to linking validation across all available CPS data and all linkage types.

5. Discussion

When CPS data are linked using only Census Bureau identifiers, some linked observations have unexpected age, sex, or race values at one or more time points. These divergent values increase concerns about the validity of linked observations, present a challenge for reproducibility of research results using linked CPS data, and underscore a need for a streamlined holistic approach to validating linked CPS observations. Creating a linked and validated CPS dataset is further complicated by the complex survey rotation pattern and changing age topcodes and race codes over time.

In this paper, we have described the challenges of validating linked CPS data and detailed our methodology for generating a unique identifier that can be used to create validated links across CPS files. Our identifier performs nearly as well as other approaches, incorporates robust age validation that bridges changes in topcodes, and implements an approach for navigating changes in race code detail over time. This new identifier, CPSIDV, successfully matches only those CPS observations who have the same sex and race values and whose age changes in expected ways based on time in the CPS. CPSIDV takes a flexible yet holistic approach to validation, taking previous observations into account when determining a link's validity—starting with the most proximate observations and working backwards in time—and retaining as many valid links as possible. This method results in a "static" linking key which is constant regardless of which CPS sample a researcher chooses as a starting point or which type of link is being executed.

The high linkage rates achieved by CPSIDV are on the order of other common CPS validation strategies (Drew et al., 2014; Madrian & Lefgren, 2000) is an indication of the efficacy of our new approach. CPSIDV eliminates the need for a critical post-linking validation step and provides standardized validation across changing age topcodes and available race categories. Thus, CPSIDV lowers the technical barriers to working with validated CPS panel data. Our innovative approach is flexible enough to capture most valid relationships in the linked CPS data, yielding high validation rates across many CPS linkage types, in a single identifier that is unique but constant regardless of a researcher's chosen first time point or linkage type. CPSIDV thus provides researchers with a common starting point for using linked and validated CPS data, increasing reproducibility of research using CPS panel data.

References

Alexander, J. T., Davern, M., & Stevenson, B. (2010). The Polls-Review: Inaccurate Age and Sex Data in the Census Pums Files: Evidence and Implications. *Public Opinion Quarterly*, *74*(3), 551–569. https://doi.org/10.1093/poq/nfq033

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. M., Rodgers, R., Pacas, J. D., Kristiansen, D., & Klass, B. (2020). *Extending Current Population Survey Linkages: Obstacles and Solutions for Linking Monthly Data from 1976 to 1988*. IPUMS. https://assets.ipums.org/_files/ipums/working_papers/ipums_wp_2020-02.pdf

Flood, S. M., King, M., Rodgers, R., Ruggles, S., Warren, J. R., & Westberry, M. (2022) Integrated Public Use Microdata Series, Current Population Survey: Version 10.0 [dataset]. Minneapolis, MN: IPUMS. https://doi.org/10.18128/D030.V10.0

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

Nekarda, C. J. (2009). *A Longitudinal Analysis of the Current Population Survey* (p. 64).

Ofstedal, M. B., Weir, D., & Chen, K.-T. (Jack). (2011). *Updates to HRS Sample Weights*. Survey Research Center, Institute for Social Research, University of Michigan. https://doi.org/10.7826/ISR-UM.06.585031.001.05.0025.2011

Renwick, T. (2019, September 4). CPS ASEC Redesign and Processing Changes. *The United States Census Bureau*. https://www.census.gov/newsroom/blogs/research-matters/2019/09/cps-asec.html

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

U.S. Census Bureau. (2022, July 8). *CPS Basic Monthly Footnotes*. CPS Basic Monthly Footnotes. https://www.census.gov/programs-surveys/cps/data/datasets/cps-basic-footnotes.html

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

Appendix A: Age Validation Criteria

We considered a stricter age validation scheme than the one described in the body of the paper for the creation of CPSIDV. Here we describe the alternative age validation and compare it to the more relaxed age validation requirements used to generate CPSIDV. Under the strict validation scheme:

- The respondent may age one year over the 16-month time period (two consecutive ages in the data). They must age after four months in the rotation and before thirteen months in the rotation. If they age before their fourth month in the survey, they will be expected to age two years over the course of the panel.
- The respondent may age two years over the course of the 16-month time period (three consecutive ages in the data). The time points at which their age increases by one year must be between 11 and 13 months apart.
- Respondents must age by at least one year between their fourth and thirteenth month in the survey unless they begin the CPS rotation with a topcoded age value. Beginning in 2004, respondents aged 80 may age 5 years to accommodate topcoding.

The difference between the strict age validation scheme described in this appendix and the relaxed age validation scheme described in the body of this paper is the timing and age increments that are considered valid. Under the relaxed validation scheme, a record's age value may increase at any time during the rotation as long as it does increase and does so by two years or less. Under the strict validation scheme, records must age at specific time points and may only do so in one-year increments.

The following examples illustrate the difference between the relaxed and strict age validation schemes. Figure A1 shows an individual who began the CPS rotation in August of 2018 and appeared all eight times, finishing the rotation in November 2019. This individual has valid ages across all eight time points under both relaxed and strict age validation schemes. As CPSIDV values are uniform across all eight appearances of this individual (as defined by CPSIDP), any link, whether it is simple month-to-month link or a multi-observation link, will be considered valid.

[Figure A1]

Figure A2 shows another set of records that entered the CPS rotation in August 2018, responded to all eight interviews, and finished the rotation in November 2019. However, this set of records yields different results under the relaxed and strict age validation schemes. As the individual ages only two years over the 16 months in the CPS panel and never ages backwards, they have a consistent CPSIDV across all appearances when validating age using the relaxed criteria. However, this record has an age value of 72 in the third interview and an age value of 74 in the fifth interview, aging two years in the span of 10 months (recall that there is an eight-month gap between the fourth and fifth interviews). Aging this quickly is not allowed under the strict age validation scheme. When attempting to validate the fifth appearance against previous observations, age is considered invalid and a new CPSIDV value is generated for the fifth appearance. As a result, these observations will not link using the strict age validation criteria across all eight interviews, but will link in all adjacent month links except for November 2018 to August 2019.

[Figure A2]

Comparing age validation schemes

In this section we compare the strict and relaxed age validation schemes for full-panel CPSIDP links, links across adjacent months, and links of the same month in adjacent years between 1980 and 2020. Full-panel links with all eight responses have the lowest validation rate of any type of link due to the number of responses to validate; adjacent-month links have the highest validation rate of any type of link. Differing age criteria are compared for those records that have already been validated on sex and race. Note that due to a break in linkable samples before and after June of 1985 and before and after

September of 1985, no households complete the CPS rotation in 1986 (Flood et al., 2020). Though the strict age validation scheme results in lower validation rates in all years, we find that the strict and relaxed age validation schemes yield similar validation rates of CPSIDP-linked records across all years from 1980-2010. However, the gap between relaxed and strict validation criteria becomes pronounced beginning in 2011 in both full-panel and adjacent-year CPSIDP links. We explore several potential reasons behind this discrepancy.

Full-panel links

Table A1 compares age validation rates using the relaxed and strict criteria for full-panel CPSIDP-linked data and breaks down validation rates under each scheme by number of age values a respondent has across their eight CPS appearances to better assess the impact of the different validation schemes across different types of age progressions. Year groups include all individuals who enter the CPS rotation in any month in the given year and complete all eight CPS interviews. For those that enter the CPS between 1980 and 1993, validation rates for both relaxed and strict criteria slowly decrease over time, with relaxed validation yielding validation rates between 87.27% and 92.45% and the strict validation yielding rates between 83.36% and 90.04%. In this period, there is a greater than two and less than four percentage point differences between the relaxed and strict schemes for individuals completing their rotation in any given year. With the introduction of Computer Assisted Telephone Interviewing (CATI) in 1994, validation rates for both relaxed and strict criteria increase considerably, with both types of validation yielding very similar results. For those individuals entering the rotation between 1994 and 2001, validation rates were above 93% for the relaxed age validation criteria and above 92% for the strict validation criteria, with the strict validation criteria achieving validation rates less than two percentage points lower than the relaxed criteria within a given year.

In 2002, age perturbation is introduced into the CPS data as a privacy measure (U.S. Census Bureau, 2021). Coincident with this introduction, validation rates for both relaxed and strict age validation schemes fall slightly. Between 2002 and 2009, those entering the CPS had between 92.17% and 94.46% validation rates under the relaxed age validation criteria. Those entering between 2002 and 2009 had between 90.68% and 92.52% validation rates under the strict criteria. Relaxed validation rates are never more than 2.11 percentage points greater than the strict validation rates in this period.

In January of 2011, a new age perturbation procedure was implemented, again to preserve privacy (U.S. Census Bureau, 2021). Coincident with this new disclosure avoidance mechanism, validation rates for the strict validation criteria fall sharply, with the lowest point being for those full-panel CPSIDP-linked records that enter the rotation in 2010. For these records, validation rates drop for both the relaxed and validation criteria, with the strict validation criteria yielding the lowest validation rate of all year groups between 1980 and 2018. We suspect these unusually low rates are due to linking records for which age values have been perturbed differently between the 2010 and 2011 observations.

Though validation rates for full-panel CPSIDP-linked records rebound slightly for those who enter the CPS in 2011 and after as compared to 2010, there remains a large gap between the relaxed and strict validation schemes in these groups. As shown in Table A1, this precipitous drop is largely the result of poor validation among those records with three unique age values across all eight appearances, only 44.91%-52.93% of whom are considered valid under the strict aging criteria.

[Table A1]

Adjacent-year links

Table A2 compares strict and relaxed validation rates for CPSIDP links with two time points, one year apart. In this linking scenario, we see a similar pattern over time to that of the full-panel links. The relaxed age validation criteria always result in a higher validation rate than the strict criteria, validation rates under both schemes improve with the introduction of CATI in 1994. Validation rates are between 93.45% and 96.33% under relaxed validation criteria between 1978 and 1993 and between 88.7% and 94.82% under strict validation criteria in the same period. These rates rise to between 95.59% and 96.75% under relaxed validation criteria and between 94.36% and 95.10% for strict validation criteria for adjacent-year links beginning in 1994. Validation rates under both schemes drop slightly with the introduction of age perturbation in 2002, and the gap between the strict and relaxed age validation criteria diverge noticeably with the new age perturbation methodology in 2011. For adjacent-links beginning in 2001 to 2009, relaxed age validation criteria give validation rates between 91.68% and 94.46%, while the strict validation criteria only give between 89.46% and 92.86% validation rates. For adjacent-year links beginning between 2010 and 2019, validation rates using the relaxed criteria are between 92.61% and 96.30%, while those using the strict criteria fall to between 80.40% and 88.92%, making the magnitude of the discrepancy between our two sets of validation criteria for adjacent-year links similar to the results for full-panel links. For all MISH pairs, the relaxed age validation criteria yield validation rates in the low to mid-90s. However, under the strict age validation criteria, the later in the rotation pattern a link fails, the lower the validation rate (i.e. MISH 4 to MISH 8 links have, on balance, a lower validation rate than MISH1 to MISH 5 links).

[Table A2]

Adjacent-month links

Table A3 shows validation rates for adjacent month-in-sample links within a given year for both strict and relaxed age validation criteria from 1980 to 2020. When examining adjacent-month links, validation rates are high for both strict and relaxed age validation criteria for all month-in-sample pairs in all years between 1980 and 2020. While the relaxed criteria result in slightly higher validation rates, validation rates for the strict criteria are always above 94%. In adjacent-month links before the eight-month break, relaxed age validation criteria yield validation rates less than one percentage point higher than those of strict age validation criteria in all years. After the eight-month break, the difference between the relaxed and strict age validation criteria rise slightly between 1980 and 2010, and again jump noticeably in 2011 through 2020, especially between MISH 5 and 6. This is due to the requirement of the strict age validation criteria that persons must age one year 11 to 13 months after entering the survey. Those who do not age after 13 months in the CPS rotation and whose age values are not topcoded are considered invalid and given a new CPSIDV value in MISH 6. However, even the largest difference between relaxed and strict validation criteria is less than five percentage points. In the adjacent-month links, the large disparity between validation rates for the relaxed and strict criteria in 2002 and in 2011 seen in the full-panel links shrinks considerably in MISH 5 to MISH 6 links and all but disappears in other adjacent-month links.

[Table A3]

Potential factors affecting age validation

We examine several potential factors to understand the sharp drop in validation rates on age beginning in 2011 using the strict validation criteria. These include age perturbation, differing age allocation status across months, and reporting error due to differing household respondents across months. We are unable to identify with certainty the cause of the large discrepancy in validation rates between the strict and relaxed validation criteria in the 2011-2020 period.

Age perturbation

Age perturbation was first introduced to the basic monthly CPS in August of 2002 (U.S. Census Bureau, 2022). Age perturbation procedures were shown to be problematic in the CPS ASEC data from 2004 to 2009 (Alexander et al., 2010), however public use ASEC files for these years were not updated and re-released (Ofstedal et al., 2011). To address this problem, age perturbation procedures were updated for the CPS basic monthly data in January 2011 (U.S. Census Bureau, 2022). Strict validation rates drop sharply for full-panel links that begin the rotation in 2010 (and so span the change in age perturbation methodology) and remain low through the most recent data, while relaxed validation rates remain relatively steady. The most obvious suspect behind this divergence in validation rates between the two schema during this time period is the change in age perturbation methodology; we were unable to find documentation of any other contemporaneous changes to CPS BMS data collection or processing that might cause the observed drop in strict validation rates for full-panel links beginning the rotation in 2010.

Because the details of the age perturbation procedure are not public (for obvious reasons), we are unable to determine empirically whether the 2011 age perturbation update is behind the observed drop in strict validation rates. In email correspondence, Census Bureau staff informed us that steps are taken in the age perturbation procedure to ensure that the perturbation applied to a record in month-in-sample one is held constant throughout its time in the CPS due to the longitudinal nature of the data (R. Rodgers, personal communication, October 19, 2020). Analyses comparing male to female ratios in CPS March Basic Monthly and CPS ASEC data from 2006 to 2019 to published ACS estimates (available upon request) does not reveal any unexpected discrepancies in any age group, indicating that age perturbation procedures are not resulting in unusual estimates in the monthly data. However, we are unable to assess whether the age perturbation procedure is having unintended consequences for individuals' age continuity across months.

We investigated preference for reported over allocated age values across time and multiple household respondents over the CPS interview rotation as other potential sources of the drop in strict validation rates for those full-panel links beginning in 2010. Neither of these scenarios could explain the magnitude of the change in strict validation rates after the introduction of new age perturbation methodology, but we include them below for completeness.

Allocated vs reported age values

Occasionally an individual's age may be allocated in some months and reported in other months; when an individual has a reported age in one month and an allocated age value in another month, the reported age values are retained in the data, even if this violates continuity assumptions about aging in age perturbation (R. Rodgers, personal communication, October 19, 2020). We explored the possibility that individuals who entered the CPS rotation in the year 2010 and completed the full rotation but failed to meet the strict validation criteria were more likely to have both reported and allocated values over the course of their eight appearances in the CPS data. There is no documentation to suggest that respondents are more likely to move between allocated and reported values for age after January 2011. However, given the stark difference between the results of strict and relaxed validation criteria for those in the CPS rotation before and after this time, we wished to verify empirically that this period wasn't associated with unusually high rates of variation in age allocation status across months. If this were the case, favoring reported values for age when available rather than allocating age for consistency with previous months may be an explanation for the drop in validation rates for full-panel links under the strict age validation criteria for those entering the rotation in 2010. However, as Figure A1 shows, the percentage of those full-panel links that do not meet the strict validation criteria and have both allocated and reported ages over the course of their CPS participation decreases for those who enter the rotation in 2010 and after as compared to those who enter the rotation prior to 2010. We can therefore conclude that favoring reported values of age over allocating age for consistency over time is not the cause of the observed decline in strict validation rates.

Potential reporting error

Another potential explanation for the large drop in validation rates using the strict validation criteria beginning in 2011 is an increase in reporting error due to different household members answering interview questions in different months. One person, known as the household respondent, who is aged 15 or older and is knowledgeable about the household is designated to give information about all members of the household (U.S. Census Bureau, 2006). Which member of a household plays this role may change across months and it is possible that not all household members have accurate knowledge of the age of everyone else in the household leading to inconsistent age reports across time. If having multiple household respondents per household over the course of the CPS led to more unreliable demographic information for that household over the course of the CPS rotation, and this was a reason behind the drop in strict validation rates, we would expect those that do not meet those strict validation criteria to be in households with multiple household respondents over the course of the CPS rotation more frequently after 2011 when strict age validation rates are considerably lower. However, we see a decline in the frequency of multi-respondent households for strictly invalid links after 2011 (see Figure A3), suggesting that multi-respondent households are not a factor in the divergence of the relaxed and strict validation criteria validation rates.

[figure A3]

Age validation scheme incorporated into CPSIDV

We are confident that neither the transitions between imputed and reported age values nor a potential increase in reporting error due to multiple household respondents over the course of the CPS rotation can fully explain the sudden drop in validation rates under strict validation criteria beginning in full-panel links finishing the rotation in 2011. It may be that the age perturbation procedure is affecting aging patterns in unexpected ways. However, as details of the age perturbation procedure are not public we are unable to be sure that the low strict validation rates are a result of its application. We are unaware of other contemporaneous changes to data collection or processing that might be causing validation under strict criteria to be significantly lower from 2011 onward.

As we are unable to find a clear explanation for the large discrepancy between the strict and relaxed validation schemes in this period either in any CPS documentation or from the data itself, we have chosen to err on the side of inclusivity. CPSIDV uses the relaxed age validation criteria when assigning validated longitudinal identifiers.

Figure 1. Example of missing potential real links when using MISH=1 as the only basis for validation

CPSIDP	YEAR	MONTH	MISH	SEX	RACE	AGE
20180800009001	2018	August	1	Female	Black	40
20180800009001	2018	September	2	Female	Black	40
20180800009001	2018	October	3	Female	Black	40
20180800009001	2018	November	4	Female	Black	40
eight-month break						
20180800009001	2019	August	5	Female	Black	63
20180800009001	2019	September	6	Female	Black	64
20180800009001	2019	October	7	Female	Black	64
20180800009001	2019	November	8	Female	Black	64

Figure 2. Example of differing validation results depending on the type of linkage

CPSIDP	YEAR	MONTH	MISH	SEX	RACE	AGE
20031002261002	2003	October	1	Female	Asian	49
20031002261002	2003	November	2	Female	Asian	49
20031002261002	2003	December	3	Female	Asian	49
20031002261002	2004	January	4	Female	Asian	52
eight-month break						
20031002261002	2004	October	5	Female	Asian	51
20031002261002	2004	November	6	Female	Asian	58
20031002261002	2004	December	7	Female	Asian	56
20031002261002	2005	January	8	Female	Asian	56

Figure 3. Example of differing validation results depending on the type of linkage

CPSIDP	YEAR	MONTH	MISH	SEX	RACE	AGE
20020907222201	2002	September	1	Female	White	22
20020907222201	2002	October	2	Female	White	22
20020907222201	2002	November	3	Female	White	65
20020907222201	2002	December	4	Female	White	65
eight-month break						
20020907222201	2003	September	5	Female	White	66
20020907222201	2003	October	6	Female	White	67
20020907222201	2003	November	7	Female	White	67
20020907222201	2003	December	8	Female	White	68

Figure 4a. Full panel respondent valid on sex, race, and age

CPSIDP	YEAR	MONTH	MISH	SEX	RACE	AGE	MISH of Reference Observation	MISH of VALID LINKS
20180800770402	2018	August	1	Male	White	26	-	1-8
20180800770402	2018	September	2	Male	White	27	1	1-8
20180800770402	2018	October	3	Male	White	27	2	1-8
20180800770402	2018	November	4	Male	White	27	3	1-8
eight-month break								
20180800770402	2019	August	5	Male	White	27	4	1-8
20180800770402	2019	September	6	Male	White	28	5	1-8
20180800770402	2019	October	7	Male	White	28	6	1-8
20180800770402	2019	November	8	Male	White	28	7	1-8

Figure 4b. Full panel respondent valid on sex, race, but invalid on age

CPSIDP	YEAR	MONTH	MISH	SEX	RACE	AGE	MISH of Reference Observation	MISH of VALID LINKS
20180800009001	2018	August	1	Female	Black	40	-	1-4
20180800009001	2018	September	2	Female	Black	40	1	1-4
20180800009001	2018	October	3	Female	Black	40	2	1-4
20180800009001	2018	November	4	Female	Black	40	3	1-4
eight-month break								
20180800009001	2019	August	5	Female	Black	63	-	5-8
20180800009001	2019	September	6	Female	Black	64	5	5-8
20180800009001	2019	October	7	Female	Black	64	6	5-8
20180800009001	2019	November	8	Female	Black	64	7	5-8

Figure 5. Example of CPSIDV validation sequence for CPSIDP = 20180700156401

	YEAR	MONTH	MISH	AGE	SEX	RACE	years aged since time 1	months since last observation	months since last aged	CPSIDV
Panel A: first CPSIDP record										
	2018	July	1	50	Male	White	0	0	0	201807001564011
Panel B: first two CPSIDP records										
	2018	July	1	50	Male	White	0	0	0	201807001564011
	2018	August	2	50	Male	White	0	1	0	201807001564011
Panel C: first three CPSIDP records										
	2018	July	1	50	Male	White	0	0	0	201807001564011
	2018	August	2	50	Male	White	0	1	0	201807001564011
	2018	September	3	58	Male	White	8	1	1	! DOES NOT VALIDATE!
			3	58	Male	White	0	0	0	201807001564012
Panel D: first four CPSIDP records										
	2018	July	1	50	Male	White	0	0	0	201807001564011
	2018	August	2	50	Male	White	0	1	0	201807001564011
	2018	September	3	58	Male	White	0	0	0	201807001564012
	2018	November	4	57	Male	White	-1	1	1	! DOES NOT VALIDATE!
			4	57	Male	White	0	0	0	201807001564013
Panel E: first five CPSIDP records										
	2018	July	1	50	Male	White	0	0	0	201807001564011
	2018	August	2	50	Male	White	0	1	0	201807001564011
	2018	September	3	58	Male	White	0	0	0	201807001564012
	2018	November	4	57	Male	White	0	0	0	201807001564013
	2019	July	5	58	Male	White	1	9	9	201807001564013
Panel F: all eight CPSIDP records										
	2018	July	1	50	Male	White	0	0	0	201807001564011
	2018	August	2	50	Male	White	0	1	0	201807001564011
	2018	September	3	58	Male	White	0	1	0	201807001564012
	2018	November	4	57	Male	White	0	0	0	201807001564013
	2019	July	5	58	Male	White	1	9	9	201807001564013
	2019	August	6	58	Male	White	1	1	10	201807001564013
	2019	September	7	58	Male	White	1	1	11	201807001564013
	2019	November	8	59	Male	White	2	1	12	201807001564013

Figure A1. Full panel respondent valid under both simple and complex age validation schemes

CPSIDP	YEAR	MONTH	MISH	AGE	SEX	RACE	CPSIDV RELAXED	CPSIDV STRICT
20180800770402	2018	August	1	26	Male	White	201808007704021	201808007704021
20180800770402	2018	September	2	27	Male	White	201808007704021	201808007704021
20180800770402	2018	October	3	27	Male	White	201808007704021	201808007704021
20180800770402	2018	November	4	27	Male	White	201808007704021	201808007704021
				eight-month break				
20180800770402	2019	August	5	27	Male	White	201808007704021	201808007704021
20180800770402	2019	September	6	28	Male	White	201808007704021	201808007704021
20180800770402	2019	October	7	28	Male	White	201808007704021	201808007704021
20180800770402	2019	November	8	28	Male	White	201808007704021	201808007704021

Figure A2. Full panel respondent with different validity under simple and complex age validation schemes

CPSIDP	YEAR	MONTH	MISH	AGE	SEX	RACE	CPSIDV RELAXED	CPSIDV STRICT
20180800835601	2018	August	1	72	Female	White	201808008356011	201808008356011
20180800835601	2018	September	2	72	Female	White	201808008356011	201808008356011
20180800835601	2018	October	3	72	Female	White	201808008356011	201808008356011
20180800835601	2018	November	4	73	Female	White	201808008356011	201808008356011
				eight-month break				
20180800835601	2019	August	5	74	Female	White	201808008356011	201808008356012
20180800835601	2019	September	6	74	Female	White	201808008356011	201808008356012
20180800835601	2019	October	7	74	Female	White	201808008356011	201808008356012
20180800835601	2019	November	8	74	Female	White	201808008356011	201808008356012

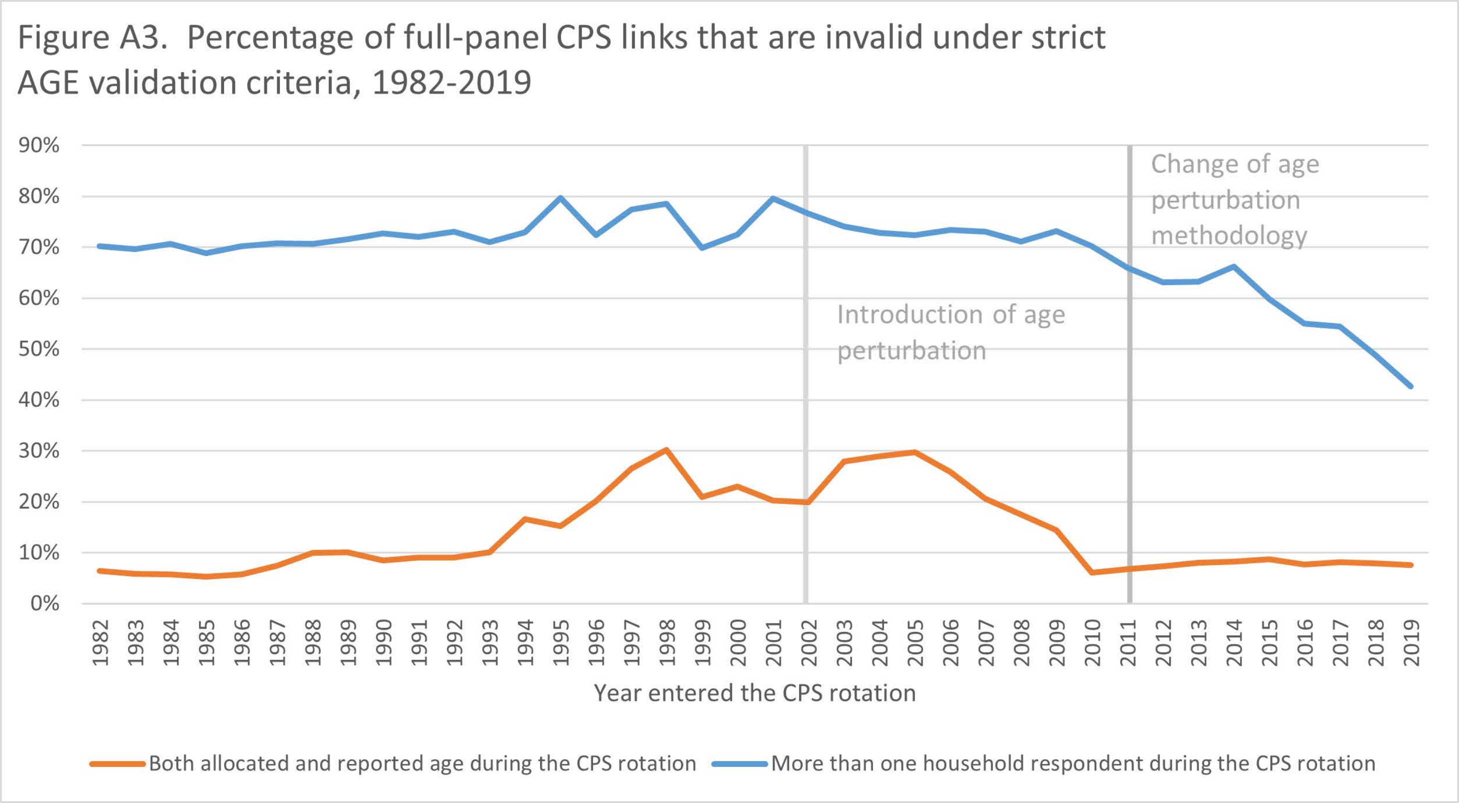
Figure A3. Percentage of full-panel CPS links that are invalid under strict AGE validation criteria, 1982-2019
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Change of age perturbation methodology
Introduction of age perturbation
1982 1983 1984 1985 1986 1987 1988 1989 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003 2004 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018 2019
Year entered the CPS rotation
Both allocated and reported age during the CPS rotation
More than one household respondent during the CPS rotation

Table 1. Bridging changes in race codes over time in linked CPS BMS data

2020-May 2012		May 2012-Apr 2012 bridge		Apr 2012-2003		Jan 2003-Dec 2002 bridge		2002-1996		Jan 1996-Dec 1995 bridge		1989-1995		Jan 1989-Dec 1988 bridge		1988-1976	
code	label	code	label	code	label	code	label	code	label	code	label	code	label	code	label	code	label
1	White only	1	White only	1	White only	1	White	1	White	1	White	1	White	1	White	1	White
										5	Other						
2	Black only	2	Black only	2	Black only	2	Black	2	Black	2	Black	2	Black	2	Black	2	Black
										5	Other						
3	American Indian, Alaskan Native only	3	American Indian, Alaskan Native only	3	American Indian, Alaskan Native only	3	American Indian, Alaskan Native	3	American Indian, Alaskan Native	3	American Indian, Aleut, Eskimo	3	American Indian, Aleut, Eskimo	3	Other	3	Other
										5	Other						
4	Asian only	4	Asian only	4	Asian only	4	Asian/Pacific Islander	4	Asian/Pacific Islander	4	Asian or Pacific Islander	4	Asian or Pacific Islander	3	Other		
										5	Other						
5	Hawaiian/Pacific Islander only	5	Hawaiian/Pacific Islander only	5	Hawaiian/Pacific Islander only	4	Asian/Pacific Islander										
6	White-Black	6	White-Black	6	White-Black	1	White										
						2	Black										
7	White-American Indian	7	White-American Indian	7	White-American Indian	1	White										
						3	American Indian, Alaskan Native										
8	White-Asian	8	White-Asian	8	White-Asian	1	White										
						4	Asian/Pacific Islander										
9	White-Hawaiian/Pacific Islander	9	White-Hawaiian/Pacific Islander	9	White-Hawaiian/Pacific Islander	1	White										
						4	Asian/Pacific Islander										
10	Black-American Indian	10	Black-American Indian	10	Black-American Indian	2	Black										
						3	American Indian, Alaskan Native										
11	Black-Asian	11	Black-Asian	11	Black-Asian	2	Black										
						4	Asian/Pacific Islander										
12	Black-Hawaiian/Pacific Islander	12	Black-Hawaiian/Pacific Islander	12	Black-Hawaiian/Pacific Islander	2	Black										
						4	Asian/Pacific Islander										
13	American Indian-Asian	13	American Indian-Asian	13	American Indian-Asian	3	American Indian, Alaskan Native										
14	American Indian-Hawaiian/Pacific Islander	20	2 or 3 races			4	Asian/Pacific Islander										
15	Asian-Hawaiian/Pacific Islander	14	Asian-Hawaiian/Pacific Islander	14	Asian-Hawaiian/Pacific Islander	4	Asian/Pacific Islander										
16	White-Black-American Indian	15	White-Black-American Indian	15	White-Black-American Indian	1	White										
						2	Black										
						3	American Indian, Alaskan Native										
17	White-Black-Asian	16	White-Black-Asian	16	White-Black-Asian	1	White										
18	White-Black-Hawaiian/Pacific Islander	20	2 or 3 races			2	Black										
						4	Asian/Pacific Islander										
19	White-American Indian-Asian	17	White-American Indian-Asian	17	White-American Indian-Asian	1	White										
20	White-American Indian-Hawaiian/Pacific Islander	20	2 or 3 races			3	American Indian, Alaskan Native										
						4	Asian/Pacific Islander										
21	White-Asian-Hawaiian/Pacific Islander	18	White-Asian-Hawaiian/Pacific Islander	18	White-Asian-Hawaiian/Pacific Islander	1	White										
22	Black-American Indian-Asian	20	2 or 3 races			4	Asian/Pacific Islander										
23	White-Black-American Indian-Asian	19	White-Black-American Indian-Asian	19	White-Black-American Indian-Asian	1	White										
24	White-American Indian-Asian-Hawaiian/Pacific Islander	21	4 or 5 races			2	Black										
						3	American Indian, Alaskan Native										
						4	Asian/Pacific Islander										
25	Other 3 race combinations	20	2 or 3 races	20	2 or 3 races	1	White					5	Other	3	Other		
						2	Black										
						3	American Indian, Alaskan Native										
						4	Asian/Pacific Islander										
26	Other 4 and 5 race combinations	21	4 or 5 races	21	4 or 5 races	1	White										
						2	Black										
						3	American Indian, Alaskan Native										
						4	Asian/Pacific Islander										

Race is bridged backward in time, with more detailed codes being mapped onto less detailed codes. The "bridge" columns in the table show the allowed mappings from race codes in the later period on the left to race codes in the earlier period on the right. If a many-to-one mapping is allowed, multiple categories will be listed in the bridge column. For example, the "May 2012-Apr 2012 bridge" column shows that the code 1=White from May 2012-2020 maps back to 1=White in 2003-April of 2012, and 18="White-Black-Hawaiian/Pacific Islander" in May 2012-2020 maps back to 20="2 or 3 races" in 2003-April 2012. Similarly, the "Jan 2003-Dec 2002 bridge" column shows that 6="White-Black" in 2003-April 2012 maps back to either 1="White" or 2="Black" in 1996-2002.

Table 2. Bridging Age Topcodes in CPS linked data, 1976-2019

Month and year of topcode change	New topcode	Old topcode	Valid ages allowed to bridge topcode change	Months to which the bridge applies
January 1976	99	--	--	--
July 1985	90	99	91-99	N/A
February 2002	80	90	81-90	November 2000-January 2002
April 2004	80, 85	80	80	January 2003-March 2004

Note: Due to a linking discontinuity between July 1985 and previous months, no age topcodes will need to be bridged at this change due to the impossibility of linking backwards from July 1985.

Table 3. March BMS - ASEC CPSIDP links with mismatched demographic characteristics

Year	Total CPSIDP matches	validation fails on… SEX	RACE	AGE	Validation Rate	Allow +/- 1 year difference in AGE	Validation Rate	Allow +/- 2 year difference in AGE	Validation Rate
1982	153017	10	25	44	99.95%	33	99.96%	29	99.96%
1983	153201	42	13	78	99.91%	68	99.92%	58	99.93%
1984	152418	40	13	76	99.92%	57	99.93%	42	99.94%
1985	152466	28	9	102	99.91%	72	99.93%	54	99.94%
1986	147383	39	22	96	99.89%	81	99.90%	65	99.91%
1987	145594	43	16	89	99.90%	74	99.91%	61	99.92%
1988	145041	370	113	775	99.13%	708	99.18%	661	99.21%
~									
1994	140625	117	96	212	99.70%	182	99.72%	161	99.73%
~									
2001	116663	10	213	558	99.33%	530	99.35%	495	99.38%
2002	139660	0	0	128	99.91%	102	99.93%	82	99.94%
2003	141288	0	0	3959	97.20%	2595	98.16%	1286	99.09%
2004	138350	0	154	9	99.88%	9	99.88%	6	99.88%
~									
2009	134650	0	0	6	100.00%	6	100.00%	3	100.00%
~									
2019	118108	53	34	12347	89.47%	416	99.57%	354	99.63%

Note: ASEC oversample records cannot be merged with the March BMS file and so cannot be validated. These records have CPSIDV values of 0 in all years. Years between 1976 and 2019 for which the March BMS - ASEC link has no unmatched BMS records and where all records match across files on SEX, RACE, and AGE are excluded from this table. All records that fail to validate on RACE in the 2004 March BMS - ASEC link have a race value of "White-Asian-Hawaiian/Pacific Islander" in the ASEC file and "2 or 3 races" in the BMS file. We do not make any accomodations for this in assigning CPSIDV values. Links between 1977 March BMS and ASEC files are impossible. Years not shown in this table have identical AGE, SEX, and RACE value for all records that link between the March BMS and ASEC file using CPSIDP.

Table 4. Validation rates by sex, race, and age for individuals who respond to all eight interviews, 1978-2018

Year	Total CPSIDP links	Valid SEX count	Valid SEX %	Valid SEX and RACE count	Valid SEX and RACE %	Valid SEX, RACE, and AGE count	Valid SEX, RACE, and AGE %
1978	101666	99927	98.29%	98862	97.24%	92667	91.15%
1979	106617	105026	98.51%	104152	97.69%	97737	91.67%
1980	116752	115137	98.62%	114223	97.83%	107614	92.17%
1981	116518	114956	98.66%	114186	98.00%	107718	92.45%
1982	143060	139981	97.85%	138866	97.07%	130927	91.52%
1983	142796	138538	97.02%	137455	96.26%	128343	89.88%
1984*	34508	33015	95.67%	32746	94.89%	30227	87.59%
1985*	34887	33434	95.84%	33186	95.12%	30815	88.33%
1986	136854	131244	95.90%	130215	95.15%	120356	87.94%
1987	130640	125150	95.80%	124158	95.04%	113607	86.96%
1988	127921	123045	96.19%	122054	95.41%	111642	87.27%
1989	134770	130043	96.49%	129066	95.77%	118597	88.00%
1990	137811	132908	96.44%	131802	95.64%	121825	88.40%
1991	136394	132128	96.87%	131027	96.07%	121129	88.81%
1992	134345	130152	96.88%	129216	96.18%	118209	87.99%
1993	130292	128024	98.26%	127637	97.96%	119877	92.01%
1994*	20244	20164	99.60%	20097	99.27%	19646	97.05%
1995*	40539	40423	99.71%	40351	99.54%	39367	97.11%
1996	123164	122827	99.73%	122730	99.65%	119886	97.34%
1997	121341	120990	99.71%	120905	99.64%	117869	97.14%
1998	121640	121294	99.72%	121157	99.60%	118061	97.06%
1999	121840	121543	99.76%	121437	99.67%	115958	95.17%
2000	116852	116616	99.80%	116448	99.65%	110611	94.66%
2001	130984	130717	99.80%	129579	98.93%	122231	93.32%
2002	142860	142580	99.80%	139337	97.53%	131679	92.17%
2003	131421	131178	99.82%	130949	99.64%	123432	93.92%
2004	124319	124073	99.80%	123818	99.60%	116719	93.89%
2005	130230	129960	99.79%	129729	99.62%	122011	93.69%
2006	131194	130729	99.65%	130143	99.20%	121997	92.99%
2007	131203	130717	99.63%	130181	99.22%	123599	94.20%
2008	132759	132277	99.64%	131701	99.20%	125398	94.46%
2009	134446	133901	99.59%	133340	99.18%	126613	94.17%
2010	131541	131058	99.63%	130479	99.19%	122957	93.47%
2011	130665	130236	99.67%	129642	99.22%	126002	96.43%
2012	128598	128125	99.63%	127440	99.10%	123682	96.18%
2013	116714	116305	99.65%	115644	99.08%	111895	95.87%
2014	109671	109243	99.61%	108620	99.04%	104693	95.46%
2015	117102	116657	99.62%	115994	99.05%	112077	95.71%
2016	118360	117876	99.59%	117207	99.03%	113319	95.74%
2017	112447	111998	99.60%	111365	99.04%	107701	95.78%
2018	107971	107553	99.61%	106994	99.10%	102385	94.83%

Note: Full-panel links in this table are those that began the CPS rotation in MIS 1 in any month during the given year and complete all eight CPS interviews. Due to a complicated pattern of linking discontinuities between 1976, 1977, and 1978 (Flood et al., 2020), no records that begin the CPS rotation in 1976 or 1977 can be linked to all eight CPS survey responses. These years are excluded from the table. Linking discontinuities in 1985 and 1995 are the cause of low CPSIDP linkage rates for full-panel links entering the rotation in 1984 and 1985 and 1994 and 995 respectively.

Table 5. Validation rates of adjacent-year CPSIDP links using CPSIDV, 1978-2019

	MIS 1, Year X to MIS 5, Year X+1		MIS 2, Year X to MIS 6, Year X+1		MIS 3, Year X to MIS 7, Year X+1		MIS 4, Year X to MIS 8, Year X+1	
Year X	N	%	N	%	N	%	N	%
1978	117769	95.17%	121272	95.61%	121586	95.83%	120687	95.30%
1979	122215	95.35%	125297	95.63%	125620	95.79%	124710	95.35%
1980	136208	95.66%	139977	96.15%	140318	96.20%	139106	95.82%
1981	132050	95.74%	136233	96.18%	136563	96.33%	135503	95.84%
1982	164424	95.64%	170334	96.18%	170151	96.19%	168756	95.75%
1983	162022	94.92%	166866	95.43%	166697	95.45%	165629	95.09%
1984*	71045	94.08%	75202	94.25%	78288	94.71%	80387	94.22%
1985*	38791	94.40%	40140	94.91%	40099	94.85%	40011	94.90%
1986	152414	94.30%	157857	94.60%	158012	94.66%	156147	93.90%
1987	147839	94.05%	152127	94.48%	152006	94.37%	149563	93.45%
1988	142998	93.84%	147725	94.17%	147730	94.44%	145844	93.68%
1989	151883	93.88%	156413	94.42%	156669	94.70%	152842	93.91%
1990	155197	94.10%	160468	94.52%	160833	94.63%	159584	94.05%
1991	153463	94.28%	158182	94.74%	158807	94.96%	157440	94.30%
1992	150197	93.75%	156099	94.35%	156443	94.57%	154303	93.75%
1993	142292	94.41%	149191	94.58%	151002	94.42%	151110	93.81%
1994*	55320	95.59%	58696	96.40%	59326	96.63%	59350	96.75%
1995*	43657	95.85%	45951	96.04%	46004	95.88%	46841	96.72%
1996	133354	95.96%	139311	96.36%	140454	96.65%	140737	96.87%
1997	132417	95.93%	138288	96.33%	138867	96.45%	139463	96.70%
1998	133085	95.57%	138961	95.96%	139756	96.18%	140184	96.48%
1999	133259	95.39%	139838	95.64%	140319	95.60%	140380	95.43%
2000	127953	94.73%	134272	95.37%	136428	95.73%	137714	95.97%
2001	142522	93.36%	148777	93.86%	149667	94.00%	150208	93.96%
2002	148941	91.37%	155001	91.70%	156342	91.77%	156386	91.68%
2003	141276	93.13%	149523	93.68%	152067	93.72%	154270	93.58%
2004	134636	92.97%	140565	93.42%	140762	93.63%	139712	93.42%
2005	141825	93.03%	147916	93.41%	149300	93.68%	150424	93.44%
2006	140643	92.27%	146893	92.92%	148055	93.09%	148664	92.99%
2007	141858	93.46%	149072	93.92%	150708	94.11%	151873	93.92%
2008	143638	93.74%	150371	94.19%	151209	94.46%	151146	94.21%
2009	146773	94.00%	152544	94.32%	153903	94.43%	154196	94.13%
2010	142897	92.61%	148359	93.06%	149552	93.18%	150419	93.23%
2011	146551	95.83%	151960	96.08%	152547	96.22%	153159	95.81%
2012	144164	95.62%	150525	96.19%	151498	96.30%	151442	95.79%
2013	132043	95.38%	139956	95.93%	143130	96.12%	146179	95.71%
2014	125187	94.88%	130606	95.73%	129950	95.77%	128160	95.34%
2015	134990	95.30%	141449	95.63%	143118	95.83%	144373	95.38%
2016	136249	95.37%	143444	95.76%	145099	95.82%	145599	95.29%
2017	129985	95.20%	137567	95.81%	139946	95.98%	141376	95.48%
2018	126139	95.24%	133520	95.74%	136130	96.05%	137540	95.44%
2019	115103	94.63%	123261	95.06%	126440	95.36%	128000	94.64%

Note: Due to a complicated pattern of linking discontinuities between 1976, 1977, and 1978 (Flood et al., 2020), linkage rates between 1976 and 1977 and 1977 and 1978 very low. These links are excluded from the table. Linking discontinuities in 1985 and 1995 are the cause of low CPSIDP linkage rates for adjacent-year links in 1984-1985 and 1985-1986 and 1994-1995 and 1995-1996, respectively.

Table 6. Validation rates of adjacent-month CPSIDP links using CPSIDV, 1978-2019

	MIS 1 to MIS 2		MIS 2 to MIS 3		MIS 3 to MIS 4		MIS 5 to MIS 6		MIS 6 to MIS 7		MIS 7 to MIS 8	
Year	N	%	N	%	N	%	N	%	N	%	N	%
1978	139472	97.34%	142722	98.01%	143142	98.14%	142469	97.54%	144402	97.85%	144315	97.84%
1979	142436	97.55%	145648	98.11%	145707	98.20%	140287	97.82%	141492	98.01%	140788	97.62%
1980	167915	97.66%	172310	98.26%	172052	98.34%	166001	97.97%	167634	98.24%	167241	97.92%
1981	156247	97.83%	160097	98.44%	160326	98.52%	157118	98.12%	158719	98.34%	158396	97.99%
1982	192665	97.68%	197349	98.25%	196464	98.32%	193477	98.00%	195384	98.17%	194950	97.90%
1983	190996	97.60%	195338	98.12%	195235	98.23%	190649	97.93%	192606	98.24%	191070	97.81%
1984	184910	96.65%	189953	97.20%	190790	97.39%	190813	97.16%	192465	97.35%	191669	97.11%
1985*	153361	96.71%	157600	97.24%	157062	97.32%	152488	96.89%	155510	97.30%	155449	97.10%
1986	185064	96.83%	188963	97.35%	188525	97.37%	184389	97.07%	185752	97.31%	185590	97.37%
1987	184721	96.59%	189170	97.19%	188602	97.33%	181707	96.91%	183764	97.27%	182390	96.61%
1988	174462	96.28%	178448	96.81%	177655	96.94%	171848	96.66%	173394	96.96%	171285	96.07%
1989	178023	96.53%	182082	97.08%	180965	97.30%	172752	96.79%	174468	97.23%	172958	96.61%
1990	182937	96.63%	188302	97.25%	188374	97.36%	182426	96.84%	185930	97.30%	184543	96.79%
1991	180734	96.70%	185340	97.41%	185603	97.60%	182586	96.99%	186436	97.46%	185630	97.03%
1992	178227	96.42%	182951	97.41%	182789	97.58%	179307	97.01%	182750	97.47%	182207	97.02%
1993	175921	96.55%	181765	97.43%	181922	97.56%	175922	96.75%	179867	97.30%	178567	96.71%
1994	175251	98.04%	182677	99.24%	183135	99.45%	181844	98.87%	185645	99.45%	185913	99.06%
1995*	113490	95.87%	116150	95.96%	116283	96.08%	111989	96.76%	114477	96.95%	115157	97.17%
1996	157622	98.67%	162071	99.24%	162088	99.44%	157865	99.39%	160083	99.57%	160751	99.62%
1997	156752	98.55%	161510	99.25%	162199	99.42%	158251	99.38%	161136	99.53%	162067	99.57%
1998	157180	98.42%	161949	99.20%	162061	99.45%	156354	99.35%	159322	99.53%	160602	99.57%
1999	156635	98.43%	161629	99.06%	162464	99.23%	156506	99.17%	159359	99.37%	160634	99.41%
2000	149474	97.63%	155005	98.28%	156770	98.51%	158983	98.36%	161973	98.56%	163024	98.57%
2001	165789	98.05%	170571	98.80%	171802	99.04%	168983	99.13%	171966	99.31%	174024	99.34%
2002	181137	98.05%	186768	98.68%	187899	98.89%	182255	98.76%	185467	98.95%	186374	98.83%
2003	176058	98.15%	182956	98.92%	185117	99.17%	183059	99.06%	186421	99.30%	187235	99.20%
2004	172015	98.09%	178597	98.90%	179239	99.17%	179722	99.06%	183871	99.27%	184753	99.07%
2005	171635	98.09%	177360	98.86%	178929	99.10%	176153	99.01%	180322	99.21%	182161	99.10%
2006	171171	97.89%	176814	98.85%	178363	99.02%	173665	98.99%	177702	99.23%	179005	98.99%
2007	169684	98.25%	175441	99.05%	177679	99.23%	172667	99.11%	176227	99.34%	177125	99.12%
2008	169296	98.48%	174742	99.13%	176175	99.42%	172148	99.31%	175747	99.48%	177426	99.25%
2009	172817	98.57%	178187	99.21%	179690	99.45%	173763	99.34%	176383	99.54%	177562	99.29%
2010	171467	98.58%	176173	99.26%	177464	99.45%	175683	99.35%	178178	99.55%	179245	99.28%
2011	171146	98.70%	175212	99.34%	175826	99.51%	171488	99.39%	174726	99.59%	176256	99.58%
2012	168877	98.61%	173301	99.33%	174018	99.51%	170911	99.42%	173507	99.58%	173962	99.17%
2013	166291	98.59%	171725	99.30%	173184	99.51%	168714	99.40%	171732	99.59%	172803	99.18%
2014	168091	98.14%	174223	99.22%	174544	99.49%	167075	99.24%	170060	99.55%	171383	99.06%
2015	162376	98.58%	167853	99.20%	169098	99.45%	165271	99.25%	169092	99.53%	170214	99.14%
2016	162318	98.55%	169103	99.24%	170270	99.38%	162524	99.25%	166822	99.48%	168473	99.03%
2017	154577	98.48%	161324	99.18%	163427	99.40%	163798	99.29%	167389	99.51%	168548	98.98%
2018	150770	98.48%	158244	99.18%	160342	99.41%	155732	99.31%	159625	99.51%	161955	99.00%
2019	142832	98.51%	150084	99.18%	152250	99.39%	149973	99.25%	154249	99.49%	155490	99.00%

Note: Due to a complicated pattern of linking discontinuities between 1976, 1977, and 1978 (Flood et al., 2020), adjacent-month linkage rates in 1976 and 1977 very low. These links are excluded from the table. Links between MIS 4 and 5 are excluded, as these interviews are not in adjacent months. Linking discontinuities in 1985 and 1995 are the cause of low adjacent-month CPSIDP linkage rates in these years.

Table 7. Sample size and retention rate for different validation criteria, CPS respondents linked across two consecutive calendar months

	Year X		1978				1996				2009				2018			
	Jan	Feb	Jan	Feb (CPSIDP)	Feb (Drew et al.)	Feb (CPSIDV)	Jan	Feb (CPSIDP)	Feb (Drew et al.)	Feb (CPSIDV)	Jan	Feb (CPSIDP)	Feb (Drew et al.)	Feb (CPSIDV)	Jan	Feb (CPSIDP)	Feb (Drew et al.)	Feb (CPSIDV)
MIS1	Jan$_X$	Feb$_X$	14,079	-	-	-	14,517	-	-	-	16,942	-	-		14,864	-	-	-
MIS2	Dec$_{X-1}$	Jan$_X$	14,137	12,999	12,641	12,641	15,494	13,880	13,668	13,668	16,443	16,365	16,130	16,130	15,352	13,969	13,721	13,721
MIS3	Nov$_{X-1}$	Dec$_{X-1}$	14,514	13,154	12,898	12,897	15,323	14,783	14,618	14,616	17,114	15,863	15,705	15,705	15,560	14,441	14,301	14,300
MIS4	Oct$_{X-1}$	Nov$_{X-1}$	-	13,509	13,214	13,214	-	14,693	14,597	14,597	-	16,397	16,281	16,281	-	14,654	14,554	14,552
MIS5	Jan$_{X-1}$	Feb$_{X-1}$	14,761	-	-	-	14,795	-	-	-	16,791	-	-	-	15,309	-	-	-
MIS6	Dec$_{X-2}$	Jan$_{X-1}$	14,705	13,554	13,247	13,116	14,831	14,057	13,935	13,934	16,537	16,093	15,969	15,969	15,488	14,433	14,346	14,340
MIS7	Nov$_{X-2}$	Dec$_{X-2}$	14,669	13,567	13,314	13,207	15,040	14,212	14,127	14,126	16,923	15,839	15,744	15,742	16,305	14,676	14,605	14,598
MIS8	Oct$_{X-2}$	Nov$_{X-2}$	-	13,556	13,305	13,182	-	14,417	14,348	14,347	-	16,265	16,147	16,144	-	15,513	15,460	15,392
Total			86,865	66,783	65,314	65,075	90,000	86,042	85,293	85,288	100,750	96,822	95,976	95,971	92,878	87,686	86,987	86,903
Retention rate				76.88%	75.19%	74.92%		95.60%	94.77%	94.76%		96.10%	95.26%	95.26%		94.41%	93.66%	93.57%

Note: Table reports the unweighted number and percentage of CPS repondents in January of one year (the shaded box) who responded to the CPS in February of that year. Under "Year X," entries report the month and year in which respondents were in MIS1. Because of the rotation group structure, not all respondents in January are eligible to respond in February. The *CPSIDP* columns contain counts of unvalidated CPS links. Drew et al validation criteria require sex and race to remain constant across the two time points and for age to increase by 0-2 years.

Table 8. Sample size and retention rate for different validation criteria, CPS respondents linked in March across two consecutive years

	March		March																			
	Year X	Year X+	1978	1979 (CPSIDP)	1979 (Drew et al.)	1979 (Madrian & Lefgren)	1979 (CPSIDV)	1996	1997 (CPSIDP)	1997 (Drew et al.)	1997 (Madrian & Lefgren)	1997 (CPSIDV)	2009	2010 (CPSIDP)	2010 (Drew et al.)	2010 (Madrian & Lefgren)	2010 (CPSIDV)	2018	2019 (CPSIDP)	2019 (Drew et al.)	2019 (Madrian & Lefgren)	2019 (CPSIDV)
MIS1	Mar$_{X}$	Mar$_{X+1}$	14,209	-	-	-	-	14,893	-	-	-	-	17,098	-	-	-	-	14,787	-	-	-	-
MIS2	Feb$_{X}$	Feb$_{X+1}$	14,069	-	-	-	-	15,469	-	-	-	-	16,810	-	-	-	-	15,135	-	-	-	-
MIS3	Jan$_{X}$	Jan$_{X+1}$	14,217	-	-	-	-	14,823	-	-	-	-	17,444	-	-	-	-	15,324	-	-	-	-
MIS4	Dec$_{X-1}$	Dec$_{X}$	14,022	-	-	-	-	15,340	-	-	-	-	16,532	-	-	-	-	15,470	-	-	-	-
MIS5	Mar$_{X-1}$	Mar$_{X}$	-	10,337	9,957	10,003	9,833	-	11,332	10,917	10,956	10,859	-	13,350	12,673	12,740	12,643	-	10,754	10,254	10,296	10,225
MIS6	Feb$_{X-1}$	Feb$_{X}$	-	10,325	9,971	10,017	9,871	-	11,977	11,533	11,569	11,504	-	13,256	12,513	12,582	12,480	-	11,202	10,809	10,841	10,765
MIS7	Jan$_{X-1}$	Jan$_{X}$	-	10,519	10,132	10,190	10,027	-	11,496	11,165	11,192	11,117	-	13,690	13,006	13,066	12,971	-	11,557	11,252	11,274	11,204
MIS8	Dec$_{X-2}$	Dec$_{X-1}$	-	10,396	10,063	10,109	9,894	-	11,867	11,549	11,566	11,503	-	13,190	12,499	12,564	12,394	-	11,820	11,440	11,465	11,358
Total			56,517	41,577	40,123	40,319	39,625	60,525	46,672	45,164	45,283	44,983	67,884	53,486	50,691	50,952	50,488	60,716	45,333	43,755	43,876	43,552
Retention rate				73.57%	70.99%	71.34%	70.11%		77.11%	74.62%	74.82%	74.32%		78.79%	74.67%	75.06%	74.37%		74.66%	72.07%	72.26%	71.73%

Note: Table reports the unweighted number and percentage of CPS repondents in March of one year (the shaded box) who responded to the CPS in March of the next year. Under "Year X," entries report the month and year in which respondents were in MIS1. Because of the rotation group structure, not all respondents in March are eligible to respond the following March. The *CPSIDP* columns contain counts of unvalidated CPS links. Drew et al validation criteria require sex and race to remain constant across the two time points and for age to increase by 0-2 years. Madrian and Lefgren validation criteria require sex and race to remain constant across two time points and for age to increase between -1 and 3 years.

Table A1. Validation rates for full-panel CPSIDP-linked data using simple and complex criteria, 1978-2018

	All					1 unique age					2 unique ages					3 unique ages				
	relaxed		strict		total full-panel	relaxed		strict		total full-panel	relaxed		strict		total full-panel	relaxed		strict		total full-panel
	count	%	count	%	CPSIDP links	count	%	count	%	CPSIDP links	count	%	count	%	CPSIDP links	count	%	count	%	CPSIDP links
1978	92667	91.15%	90151	88.67%	101666	4	0.71%	4	0.71%	564	71489	96.14%	70829	95.25%	74361	21174	82.18%	19318	74.98%	25764
1979	97737	91.67%	94774	88.89%	106617	4	0.78%	4	0.78%	512	75290	96.47%	74485	95.44%	78041	22443	83.16%	20285	75.16%	26989
1980	107614	92.17%	104656	89.64%	116752	3	0.54%	3	0.54%	559	82467	96.70%	81663	95.75%	85284	25144	84.38%	22990	77.15%	29798
1981	107718	92.45%	104911	90.04%	116518	4	0.63%	4	0.63%	632	82843	96.83%	82042	95.90%	85552	24871	84.93%	22865	78.08%	29284
1982	130927	91.52%	127504	89.13%	143060	4	0.50%	4	0.50%	803	100731	96.13%	99753	95.20%	104785	30192	83.40%	27747	76.64%	36202
1983	128343	89.88%	124056	86.88%	142796	5	0.66%	5	0.66%	760	98692	95.63%	97520	94.50%	103198	29646	79.78%	26531	71.40%	37158
1984*	30227	87.59%	29138	84.44%	34508	1	0.38%	1	0.38%	260	23409	95.20%	23084	93.88%	24588	6817	74.77%	6053	66.39%	9117
1985*	30815	88.33%	29712	85.17%	34887	65	18.06%	65	18.06%	360	23943	95.31%	23669	94.22%	25122	6807	76.71%	5978	67.37%	8874
1986	120356	87.94%	116054	84.80%	136854	322	20.50%	322	20.50%	1571	93077	95.15%	91663	93.70%	97824	26957	75.98%	24069	67.84%	35481
1987	113607	86.96%	108899	83.36%	130640	316	17.21%	316	17.21%	1836	88050	94.89%	86520	93.24%	92794	25241	74.24%	22063	64.89%	34000
1988	111642	87.27%	107306	83.88%	127921	275	19.04%	275	19.04%	1444	86375	94.99%	85170	93.67%	90928	24992	74.58%	21861	65.24%	33511
1989	118597	88.00%	114041	84.62%	134770	288	20.17%	288	20.17%	1428	91527	94.95%	90121	93.49%	96396	26782	76.51%	23632	67.51%	35005
1990	121825	88.40%	117446	85.22%	137811	304	22.82%	304	22.82%	1332	94346	94.97%	93028	93.64%	99345	27175	77.12%	24114	68.43%	35239
1991	121129	88.81%	116433	85.37%	136394	329	23.37%	329	23.37%	1408	93015	95.46%	91639	94.04%	97442	27785	77.92%	24465	68.61%	35660
1992	118209	87.99%	113480	84.47%	134345	339	21.83%	339	21.83%	1553	91991	94.91%	90538	93.42%	96920	25879	75.93%	22603	66.32%	34082
1993	119877	92.01%	116516	89.43%	130292	311	31.73%	311	31.73%	980	90697	96.34%	89510	95.08%	94145	28869	85.43%	26695	79.00%	33792
1994*	19646	97.05%	19381	95.74%	20244	50	94.34%	50	94.34%	53	14852	98.53%	14840	98.45%	15074	4744	94.26%	4491	89.23%	5033
1995*	39367	97.11%	38830	95.78%	40539	99	88.39%	99	88.39%	112	29837	98.65%	29820	98.59%	30246	9431	94.15%	8911	88.96%	10017
1996	119886	97.34%	118232	96.00%	123164	352	84.01%	352	84.01%	419	89956	99.08%	89876	98.99%	90794	29578	94.30%	28004	89.28%	31365
1997	117869	97.14%	116524	96.03%	121341	341	83.58%	341	83.58%	408	88359	99.11%	88279	99.02%	89151	29169	93.76%	27904	89.69%	31110
1998	118061	97.06%	116801	96.02%	121640	419	86.93%	419	86.93%	482	88507	99.11%	88423	99.01%	89304	29135	94.02%	27959	90.23%	30987
1999	115958	95.17%	114295	93.81%	121840	247	77.43%	247	77.43%	319	87243	98.68%	87178	98.60%	88413	28468	89.40%	26870	84.38%	31844
2000^	110611	94.66%	109301	93.54%	116852	190	76.00%	190	76.00%	250	83031	98.53%	82963	98.45%	84273	27339	88.32%	26098	84.31%	30955
2001^	122231	93.32%	120515	92.01%	130984	454	50.22%	454	50.22%	904	91550	96.47%	91112	96.00%	94904	30225	88.94%	28949	85.18%	33985
2002	131679	92.17%	129550	90.68%	142860	4186	88.46%	4186	88.46%	4732	96712	94.50%	96098	93.90%	102336	30781	88.96%	29266	84.59%	34599
2003	123432	93.92%	121074	92.13%	131421	2758	84.21%	2758	84.21%	3275	91895	96.17%	91266	95.52%	95550	28779	91.17%	27050	85.69%	31568
2004	116719	93.89%	114904	92.43%	124319	3393	86.34%	3393	86.34%	3930	86025	96.08%	85400	95.38%	89534	27301	91.25%	26111	87.27%	29920
2005	122011	93.69%	119948	92.10%	130230	4039	86.86%	4039	86.86%	4650	89667	95.97%	88966	95.21%	93437	28305	90.83%	26943	86.46%	31161
2006	121997	92.99%	119227	90.88%	131194	4140	86.25%	4140	86.25%	4800	89201	95.50%	88254	94.48%	93406	28656	89.78%	26833	84.07%	31919
2007	123599	94.20%	120931	92.17%	131203	4249	88.14%	4249	88.14%	4821	90400	96.22%	89619	95.39%	93949	28950	91.37%	27063	85.42%	31684
2008	125398	94.46%	122823	92.52%	132759	4310	87.25%	4310	87.25%	4940	91928	96.28%	91110	95.42%	95480	29160	92.19%	27403	86.64%	31630
2009	126613	94.17%	123770	92.06%	134446	4469	90.67%	4469	90.67%	4929	91501	96.53%	90849	95.84%	94794	30643	90.25%	28452	83.79%	33955
2010	122957	93.47%	103975	79.04%	131541	4236	97.18%	4236	97.18%	4359	84620	95.53%	82759	93.42%	88584	34101	90.20%	16980	44.91%	37805
2011	126002	96.43%	107286	82.11%	130665	4662	82.43%	4662	82.43%	5656	91111	98.62%	85638	92.70%	92383	30229	94.19%	16986	52.93%	32092
2012	123682	96.18%	105564	82.09%	128598	4623	82.98%	4623	82.98%	5571	89645	98.56%	84387	92.77%	90959	29414	93.76%	16554	52.77%	31373
2013	111895	95.87%	95262	81.62%	116714	4118	81.71%	4118	81.71%	5040	80877	98.34%	75925	92.32%	82241	26900	93.36%	15219	52.82%	28813
2014	104693	95.46%	89133	81.27%	109671	4009	83.07%	4009	83.07%	4826	75454	98.27%	71008	92.48%	76785	25230	91.81%	14116	51.36%	27482
2015	112077	95.71%	95396	81.46%	117102	4395	81.98%	4395	81.98%	5361	80786	98.31%	75774	92.21%	82176	26896	92.93%	15227	52.61%	28942
2016	113319	95.74%	96595	81.61%	118360	4650	81.66%	4650	81.66%	5694	81651	98.30%	76607	92.23%	83063	27018	93.18%	15338	52.90%	28995
2017	107701	95.78%	91897	81.72%	112447	4291	81.42%	4291	81.42%	5270	77691	98.30%	73103	92.49%	79035	25719	93.31%	14503	52.62%	27563
2018	102385	94.83%	87484	81.03%	107971	4357	82.15%	4357	82.15%	5304	73472	98.04%	68967	92.03%	74939	24556	90.54%	14160	52.21%	27121

Note: Full-panel links in this table are those that began the CPS rotation in MIS 1 in any month during the given year and complete all eight CPS interviews. Due to a complicated pattern of linking discontinuities between 1976, 1977, and 1978 (Flood et al., 2020), no records that begin the CPS rotation in 1976 or 1977 can be linked to all eight CPS survey responses. These years are excluded from the table. Linking discontinuities in 1985 and 1995 are the cause of low CPSIDP linkage rates for full-panel links entering the rotation in 1984 and 1985 and 1994 and 995 respectively. Due to the topcode change in February of 2002, it is possible to have four unique age values and be considered valid. 51 respondents that enter their rotation in 2000 and 2 respondents that enter the rotation in 2001 fall into this category.

Table A2. Comparison of inclusive and strict age validation criteria for year-to-year links, 1978-2019

	MIS 1, Year X to MIS 5, Year X+1				MIS 2, Year X to MIS 6, Year X+1				MIS 3, Year X to MIS 7, Year X+1				MIS 4, Year X to MIS 8, Year X+1			
	relaxed		strict		relaxed		strict		relaxed		strict		relaxed		strict	
Year X	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1978	117769	95.17%	116862	94.44%	121272	95.61%	118585	93.49%	121586	95.83%	117434	92.56%	120687	95.30%	115312	91.05%
1979	122215	95.35%	121116	94.50%	125297	95.63%	122422	93.44%	125620	95.79%	120997	92.26%	124710	95.35%	119283	91.20%
1980	136208	95.66%	134815	94.68%	139977	96.15%	136610	93.84%	140318	96.20%	135263	92.74%	139106	95.82%	133395	91.89%
1981	132050	95.74%	130773	94.82%	136233	96.18%	133146	94.01%	136563	96.33%	131856	93.01%	135503	95.84%	130280	92.14%
1982	164424	95.64%	162849	94.72%	170334	96.18%	166278	93.89%	170151	96.19%	164018	92.72%	168756	95.75%	162026	91.93%
1983	162022	94.92%	160468	94.01%	166866	95.43%	162626	93.00%	166697	95.45%	160357	91.82%	165629	95.09%	158266	90.86%
1984*	71045	94.08%	70184	92.94%	75202	94.25%	73143	91.67%	78288	94.71%	75129	90.88%	80387	94.22%	76853	90.07%
1985*	38791	94.40%	38335	93.29%	40140	94.91%	39276	92.87%	40099	94.85%	38972	92.19%	40011	94.90%	38947	92.37%
1986	152414	94.30%	150870	93.35%	157857	94.60%	152784	91.56%	158012	94.66%	150850	90.37%	156147	93.90%	148451	89.27%
1987	147839	94.05%	146436	93.16%	152127	94.48%	147034	91.32%	152006	94.37%	144622	89.79%	149563	93.45%	141959	88.70%
1988	142998	93.84%	141520	92.87%	147725	94.17%	143126	91.24%	147730	94.44%	140985	90.13%	145844	93.68%	138513	88.97%
1989	151883	93.88%	150228	92.86%	156413	94.42%	151529	91.47%	156669	94.70%	149287	90.24%	152842	93.91%	144784	88.96%
1990	155197	94.10%	153715	93.20%	160468	94.52%	156057	91.92%	160833	94.63%	154074	90.66%	159584	94.05%	151791	89.45%
1991	153463	94.28%	151407	93.01%	158182	94.74%	153275	91.80%	158807	94.96%	151488	90.58%	157440	94.30%	149261	89.40%
1992	150197	93.75%	148317	92.58%	156099	94.35%	151155	91.37%	156443	94.57%	149411	90.32%	154303	93.75%	146932	89.28%
1993	142292	94.41%	140001	92.89%	149191	94.58%	144901	91.86%	151002	94.42%	144802	90.54%	151110	93.81%	144386	89.64%
1994*	55320	95.59%	54791	94.68%	58696	96.40%	57903	95.10%	59326	96.63%	58019	94.50%	59350	96.75%	57889	94.36%
1995*	43657	95.85%	43579	95.68%	45951	96.04%	45742	95.60%	46004	95.88%	45508	94.85%	46841	96.72%	46365	95.74%
1996	133354	95.96%	132440	95.31%	139311	96.36%	138053	95.48%	140454	96.65%	138001	94.96%	140737	96.87%	137464	94.61%
1997	132417	95.93%	131606	95.34%	138288	96.33%	136766	95.27%	138867	96.45%	136363	94.71%	139463	96.70%	136118	94.39%
1998	133085	95.57%	132228	94.95%	138961	95.96%	137598	95.02%	139756	96.18%	137377	94.55%	140184	96.48%	137102	94.36%
1999	133259	95.39%	132260	94.67%	139838	95.64%	138460	94.70%	140319	95.60%	137999	94.02%	140380	95.43%	137170	93.24%
2000	127953	94.73%	127469	94.37%	134272	95.37%	133202	94.61%	136428	95.73%	134008	94.03%	137714	95.97%	134538	93.76%
2001	142522	93.36%	141087	92.42%	148777	93.86%	146400	92.36%	149667	94.00%	145957	91.67%	150208	93.96%	146264	91.49%
2002	148941	91.37%	147169	90.29%	155001	91.70%	152529	90.24%	156342	91.77%	152810	89.70%	156386	91.68%	152599	89.46%
2003	141276	93.13%	139563	92.01%	149523	93.68%	146582	91.83%	152067	93.72%	148010	91.22%	154270	93.58%	149536	90.70%
2004	134636	92.97%	133255	92.02%	140565	93.42%	138136	91.81%	140762	93.63%	137150	91.23%	139712	93.42%	135681	90.72%
2005	141825	93.03%	140093	91.89%	147916	93.41%	145067	91.61%	149300	93.68%	145225	91.12%	150424	93.44%	146152	90.78%
2006	140643	92.27%	138262	90.71%	146893	92.92%	143223	90.60%	148055	93.09%	143247	90.07%	148664	92.99%	143603	89.82%
2007	141858	93.46%	139732	92.06%	149072	93.92%	145772	91.84%	150708	94.11%	146031	91.19%	151873	93.92%	146747	90.75%
2008	143638	93.74%	141542	92.37%	150371	94.19%	147149	92.17%	151209	94.46%	146637	91.60%	151146	94.21%	146224	91.15%
2009	146773	94.00%	144991	92.86%	152544	94.32%	149673	92.54%	153903	94.43%	149853	91.95%	154196	94.13%	149776	91.44%
2010	142897	92.61%	130944	84.86%	148359	93.06%	133502	83.74%	149552	93.18%	133132	82.95%	150419	93.23%	131341	81.40%
2011	146551	95.83%	135977	88.92%	151960	96.08%	132983	84.08%	152547	96.22%	131748	83.11%	153159	95.81%	130501	81.64%
2012	144164	95.62%	133706	88.68%	150525	96.19%	131772	84.21%	151498	96.30%	130838	83.17%	151442	95.79%	129271	81.77%
2013	132043	95.38%	122562	88.53%	139956	95.93%	122496	83.97%	143130	96.12%	123433	82.90%	146179	95.71%	124311	81.39%
2014	125187	94.88%	116166	88.04%	130606	95.73%	114150	83.67%	129950	95.77%	111878	82.45%	128160	95.34%	108783	80.92%
2015	134990	95.30%	125308	88.46%	141449	95.63%	123665	83.61%	143118	95.83%	123223	82.51%	144373	95.38%	122835	81.15%
2016	136249	95.37%	126520	88.56%	143444	95.76%	125562	83.82%	145099	95.82%	125016	82.56%	145599	95.29%	123958	81.12%
2017	129985	95.20%	120770	88.45%	137567	95.81%	120462	83.89%	139946	95.98%	120773	82.83%	141376	95.48%	120602	81.45%
2018	126139	95.24%	117170	88.46%	133520	95.74%	116860	83.79%	136130	96.05%	117362	82.81%	137540	95.44%	116990	81.18%
2019	115103	94.63%	106844	87.84%	123261	95.06%	108384	83.58%	126440	95.36%	109129	82.31%	128000	94.64%	108743	80.40%

Note: Due to a complicated pattern of linking discontinuities between 1976, 1977, and 1978 (Flood et al., 2020), linkage rates between 1976 and 1977 and 1977 and 1978 very low. These links are excluded from the table. Linking discontinuities in 1985 and 1995 are the cause of low CPSIDP linkage rates for adjacent-year links in 1984-1985 and 1985-1986 and 1994-1995 and 1995-1996, respectively.

Table A3. Comparing AGE validation techniques for adjacent month links by year, 1978-2019

	MIS 1 to MIS 2				MIS 2 to MIS 3				MIS 3 to MIS 4				MIS 5 to MIS 6				MIS 6 to MIS 7				MIS 7 to MIS 8			
	relaxed		strict		relaxed		strict		relaxed		strict		relaxed		strict		relaxed		strict		relaxed		strict	
Year	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
1978	139472	97.34%	139221	97.17%	142722	98.01%	142518	97.87%	143142	98.14%	142936	98.00%	142469	97.54%	139702	95.65%	144402	97.85%	139619	94.61%	144315	97.84%	137574	93.27%
1979	142436	97.55%	142173	97.37%	145648	98.11%	145451	97.98%	145707	98.20%	145433	98.01%	140287	97.82%	138518	96.59%	141492	98.01%	140083	97.04%	140788	97.62%	139597	96.80%
1980	167915	97.66%	167606	97.48%	172310	98.26%	172036	98.10%	172052	98.34%	171800	98.19%	166001	97.97%	164153	96.88%	167634	98.24%	166021	97.30%	167241	97.92%	165736	97.04%
1981	156247	97.83%	155966	97.66%	160097	98.44%	159881	98.31%	160326	98.52%	160101	98.38%	157118	98.12%	155287	96.97%	158719	98.34%	157166	97.37%	158396	97.99%	157197	97.25%
1982	192665	97.68%	192324	97.50%	197349	98.25%	197038	98.09%	196464	98.32%	196176	98.17%	193477	98.00%	191547	97.03%	195384	98.17%	193885	97.42%	194950	97.90%	193823	97.33%
1983	190996	97.60%	190670	97.43%	195338	98.12%	195057	97.97%	195235	98.23%	194916	98.06%	190649	97.93%	188274	96.71%	192606	98.24%	190719	97.27%	191070	97.81%	189741	97.13%
1984	184910	96.65%	184536	96.45%	189953	97.20%	189595	97.02%	190790	97.39%	190425	97.20%	190813	97.16%	188129	95.79%	192465	97.35%	190314	96.26%	191669	97.11%	189861	96.19%
1985*	153361	96.71%	153044	96.51%	157600	97.24%	157355	97.09%	157062	97.32%	156791	97.16%	152488	96.89%	150917	95.89%	155510	97.30%	154319	96.56%	155449	97.10%	154635	96.60%
1986	185064	96.83%	184675	96.63%	188963	97.35%	188633	97.18%	188525	97.37%	188152	97.18%	184389	97.07%	183442	96.57%	185752	97.31%	184959	96.90%	185590	97.37%	184693	96.90%
1987	184721	96.59%	184281	96.36%	189170	97.19%	188809	97.00%	188602	97.33%	188215	97.13%	181707	96.91%	178348	95.12%	183764	97.27%	181685	96.17%	182390	96.61%	181226	96.00%
1988	174462	96.28%	174068	96.06%	178448	96.81%	178072	96.61%	177655	96.94%	177256	96.72%	171848	96.66%	168092	94.55%	173394	96.96%	171167	95.72%	171285	96.07%	170195	95.46%
1989	178023	96.53%	177665	96.34%	182082	97.08%	181723	96.89%	180965	97.30%	180593	97.10%	172752	96.79%	169612	95.03%	174468	97.23%	172256	95.99%	172958	96.61%	171730	95.92%
1990	182937	96.63%	182549	96.43%	188302	97.25%	187951	97.07%	188374	97.36%	188030	97.18%	182426	96.84%	179196	95.13%	185930	97.30%	183469	96.01%	184543	96.79%	182926	95.95%
1991	180734	96.70%	180345	96.49%	185340	97.41%	184978	97.22%	185603	97.60%	185259	97.42%	182586	96.99%	179522	95.36%	186436	97.46%	183986	96.18%	185630	97.03%	183957	96.16%
1992	178227	96.42%	177799	96.19%	182951	97.41%	182623	97.24%	182789	97.58%	182472	97.41%	179307	97.01%	176236	95.34%	182750	97.47%	180383	96.21%	182207	97.02%	180596	96.16%
1993	175921	96.55%	175542	96.34%	181765	97.43%	181422	97.25%	181922	97.56%	181546	97.36%	175922	96.75%	172574	94.91%	179867	97.30%	177438	95.98%	178567	96.71%	177443	96.10%
1994	175251	98.04%	174982	97.89%	182677	99.24%	182537	99.17%	183135	99.45%	182988	99.37%	181844	98.87%	179989	97.86%	185645	99.45%	184234	98.70%	185913	99.06%	185195	98.68%
1995*	113490	95.87%	113360	95.76%	116150	95.96%	116046	95.87%	116283	96.08%	116178	96.00%	111989	96.76%	111833	96.63%	114477	96.95%	114169	96.69%	115157	97.17%	114905	96.96%
1996	157622	98.67%	157445	98.56%	162071	99.24%	161934	99.16%	162088	99.44%	161995	99.38%	157865	99.39%	157701	99.29%	160083	99.57%	159785	99.38%	160751	99.62%	160441	99.43%
1997	156752	98.55%	156551	98.42%	161510	99.25%	161391	99.18%	162199	99.42%	162075	99.35%	158251	99.38%	157968	99.20%	161136	99.53%	160230	98.97%	162067	99.57%	161150	99.01%
1998	157180	98.42%	156961	98.29%	161949	99.20%	161800	99.11%	162061	99.45%	161932	99.37%	156354	99.35%	156038	99.15%	159322	99.53%	158536	99.04%	160602	99.57%	159802	99.07%
1999	156635	98.43%	156413	98.29%	161629	99.06%	161465	98.96%	162464	99.23%	162311	99.14%	156506	99.17%	156146	98.94%	159359	99.37%	158535	98.85%	160634	99.41%	159817	98.90%
2000	149474	97.63%	149087	97.37%	155005	98.28%	154669	98.07%	156770	98.51%	156456	98.31%	158983	98.36%	158460	98.04%	161973	98.56%	160833	97.86%	163024	98.57%	161877	97.87%
2001	165789	98.05%	165556	97.91%	170571	98.80%	170354	98.67%	171802	99.04%	171644	98.95%	168983	99.13%	168605	98.91%	171966	99.31%	171059	98.78%	174024	99.34%	173053	98.78%
2002	181137	98.05%	180743	97.84%	186768	98.68%	186476	98.52%	187899	98.89%	187643	98.75%	182255	98.76%	181272	98.23%	185467	98.95%	184332	98.35%	186374	98.83%	185403	98.31%
2003	176058	98.15%	175796	98.00%	182956	98.92%	182742	98.80%	185117	99.17%	184942	99.07%	183059	99.06%	182314	98.65%	186421	99.30%	185520	98.82%	187235	99.20%	186623	98.88%
2004	172015	98.09%	171723	97.92%	178597	98.90%	178390	98.79%	179239	99.17%	179063	99.07%	179722	99.06%	178904	98.61%	183871	99.27%	182868	98.73%	184753	99.07%	184119	98.73%
2005	171635	98.09%	171347	97.92%	177360	98.86%	177176	98.76%	178929	99.10%	178728	98.98%	176153	99.01%	175329	98.54%	180322	99.21%	179340	98.67%	182161	99.10%	181592	98.79%
2006	171171	97.89%	170894	97.73%	176814	98.85%	176561	98.71%	178363	99.02%	178171	98.92%	173665	98.99%	172713	98.45%	177702	99.23%	176653	98.64%	179005	98.99%	178561	98.74%
2007	169684	98.25%	169491	98.14%	175441	99.05%	175285	98.96%	177679	99.23%	177566	99.17%	172667	99.11%	171664	98.53%	176227	99.34%	175325	98.83%	177125	99.12%	176714	98.89%
2008	169296	98.48%	169109	98.37%	174742	99.13%	174620	99.07%	176175	99.42%	176093	99.38%	172148	99.31%	171356	98.86%	175747	99.48%	174639	98.86%	177426	99.25%	176806	98.91%
2009	172817	98.57%	172644	98.48%	178187	99.21%	178052	99.13%	179690	99.45%	179581	99.39%	173763	99.34%	172907	98.85%	176383	99.54%	175420	98.99%	177562	99.29%	177120	99.05%
2010	171467	98.58%	171302	98.48%	176173	99.26%	176063	99.19%	177464	99.45%	177381	99.40%	175683	99.35%	174837	98.87%	178178	99.55%	177252	99.03%	179245	99.28%	178789	99.03%
2011	171146	98.70%	170983	98.60%	175212	99.34%	175148	99.31%	175826	99.51%	175774	99.48%	171488	99.39%	167947	97.34%	174726	99.59%	171926	97.99%	176256	99.58%	172254	97.32%
2012	168877	98.61%	168725	98.52%	173301	99.33%	173221	99.28%	174018	99.51%	173951	99.47%	170911	99.42%	162980	94.80%	173507	99.58%	171498	98.43%	173962	99.17%	172040	98.07%
2013	166291	98.59%	166131	98.50%	171725	99.30%	171643	99.26%	173184	99.51%	173101	99.46%	168714	99.40%	161107	94.92%	171732	99.59%	169779	98.46%	172803	99.18%	171118	98.21%
2014	168091	98.14%	167935	98.04%	174223	99.22%	174121	99.16%	174544	99.49%	174452	99.43%	167075	99.24%	159891	94.97%	170060	99.55%	168169	98.44%	171383	99.06%	169736	98.10%
2015	162376	98.58%	162228	98.49%	167853	99.20%	167763	99.15%	169098	99.45%	169046	99.42%	165271	99.25%	158711	95.31%	169092	99.53%	167292	98.47%	170214	99.14%	168704	98.26%
2016	162318	98.55%	162152	98.45%	169103	99.24%	169026	99.19%	170270	99.38%	170192	99.34%	162524	99.25%	155153	94.75%	166822	99.48%	164815	98.28%	168473	99.03%	166937	98.13%
2017	154577	98.48%	154422	98.39%	161324	99.18%	161243	99.13%	163427	99.40%	163365	99.36%	163798	99.29%	156271	94.73%	167389	99.51%	165415	98.34%	168548	98.98%	167035	98.09%
2018	150770	98.48%	150628	98.38%	158244	99.18%	158157	99.12%	160342	99.41%	160273	99.37%	155732	99.31%	148575	94.74%	159625	99.51%	157711	98.31%	161955	99.00%	160485	98.10%
2019	142832	98.51%	142702	98.42%	150084	99.18%	149989	99.12%	152250	99.39%	152195	99.36%	149973	99.25%	143051	94.67%	154249	99.49%	152327	98.25%	155490	99.00%	154058	98.09%

Note: Due to a complicated pattern of linking discontinuities between 1976, 1977, and 1978 (Flood et al., 2020), adjacent-month linkage rates in 1976 and 1977 very low. These links are excluded from the table. Links between MIS 4 and 5 are excluded, as these interviews are not in adjacent months. Linking discontinuities in 1985 and 1995 are the cause of low adjacent-month CPSIDP linkage rates in these years.