



NCHS-HUD Linked Data: Methodology and Analytic Considerations

Data Release Date: June 20, 2019
Documentation Version Date: June 14, 2019

Office of Analysis and Epidemiology
National Center for Health Statistics
Centers for Disease Control and Prevention
datalinkage@cdc.gov

Table of Contents

Introduction	5
Data Sources	6
National Center for Health Statistics Surveys.....	6
HUD Administrative Data.....	6
Linkage of NCHS Survey Data with HUD Administrative Data	8
Linkage Eligibility	8
Child Survey Participants.....	10
Linkage Process.....	10
Linkage Rates	12
Access to NCHS-HUD Linked Data Files.....	12
Description of NCHS-HUD linked files	13
Transaction File	13
Episode Files.....	13
Temporal Alignment File.....	14
Weights File	14
Analytic Considerations and Guidelines	16
General Analytic Guidance	16
Eligibility-adjusted Participant Survey Weights.....	17
Temporal alignment of survey and administrative data.....	18
Identification of Ever and Concurrent HUD-Assisted Participants	18
Analyses of Children in the 1999-2016 NCHS-HUD Linked Data Files	19
Analyses of Rental Assistance Programs.....	19
Unit of Analysis	19
Analytic Considerations for Episode Files	19
MF Housing Program Data: Limitations and Considerations	20
Variable Considerations and Data Anomalies	21
HUD Program Variables.....	21
Transaction Variables	21
Disability Indicator and Count	21
Income Variables	22
Household Payment and Subsidy Variables.....	23
Geocoded Data	24
Variables with Potential Data Quality Concern	24
Appendices	26
Appendix A: Merging Linked NCHS-HUD Files with NCHS Survey Data.....	26
Appendix B: SAS Program to Create Participation Episodes Construction of the Episode Files	28
References	32

Suggested citation: National Center for Health Statistics, Office of Analysis and Epidemiology. NCHS-HUD Linked Data: Methodology and Analytic Considerations. June 2019. Hyattsville, Maryland. Available at the following address: <https://www.cdc.gov/nchs/data-linkage/hud-methods.htm>

Last updated: June 14, 2019

List of Acronyms

CPI, Consumer Price Index

ERB, ethics review board

HCV, Housing Choice Voucher

HIC, health insurance claim

HUD, Department of Housing and Urban Development

IRB, institutional review board

MEC, Mobile Examination Center

MF, Multifamily

MTW, Moving to Work

NCHS, National Center for Health Statistics

NDI, National Death Index

NHANES, National Health and Nutrition Examination Survey

NHIS, National Health Interview Survey

NYSIIS, New York State Identification Intelligence System

PBS8, Project-based Section 8

PH, Public Housing

PHA, Public Housing Agency

PIC, PIH Information Center

PIH, Public & Indian Housing

PII, personally identifiable information

PRAC, project rental assistance contract

RDC, Research Data Center

SSN4, 4-digit social security number

SSN9, 9-digit social security number

TRACS, Tenant Rental Assistance Certification System

TTP, Total Tenant Payment

Introduction

As the nation's principal health statistics agency, the mission of the National Center for Health Statistics (NCHS) is to provide statistical information that can be used to guide actions and policies to improve the health of the American people. As part of its ongoing efforts to fulfill this mission, NCHS conducts several population-based and establishment surveys that provide rich cross-sectional information on risk factors such as smoking, height and weight, health status, and socio-economic circumstances. Although the survey data collected provide information on a wide range of health-related topics, they often lack information on longitudinal outcomes. Through its data linkage program, NCHS has been able to enhance the analytic utility of the survey data by augmenting it with information from vital records and administrative data sources. The linkage of survey information with administrative data and vital records provide a unique opportunity to study factors that influence disability, chronic disease, health care utilization, morbidity, and mortality in specialized populations.

In a collaboration between NCHS and the U.S. Department of Housing and Urban Development (HUD), data for participants from two population-based health surveys (the National Health and Nutrition Examination Survey and National Health Interview Survey) were linked to data from HUD administrative records. These data, collectively referred to as the NCHS-HUD Linked Data, enable researchers to examine the health and well-being of the subset of NCHS survey participants who receive HUD-assisted housing.

This document describes the second and most recent linkage conducted between NCHS survey data and HUD administrative data. A brief overview of the data sources, the methods used for linkage, and descriptions of the resulting linked data files are provided. General data considerations and analytic guidelines for researchers using the restricted-use NCHS-HUD linked data are also included. More information about HUD housing assistance programs can be found in the companion document to these guidelines, "A Primer on HUD Programs and Associated Administrative Data," or on the HUD website.[1] Additional documentation about the variables in the linked data files are available from the NCHS data linkage website.[2]

Detailed information about the previous linkage of NCHS survey data and HUD administrative data has been published elsewhere.[3] The new linked data supersede the previous release and should be used for all new analyses.

Data Sources

National Center for Health Statistics Surveys

Data used in this linkage were from the following population-based NCHS health surveys and years:

- National Health Interview Survey (NHIS), years: 1999-2016;
- National Health and Nutrition Examination Survey (NHANES), cycles: 1999-2016.

A brief description of the NCHS surveys follows.

NHIS is a nationally representative, cross-sectional household interview survey that serves as an important source of information on the health of the civilian, noninstitutionalized population of the United States. It is a multistage sample survey with primary sampling units of counties or clusters of adjacent counties, secondary sampling units of clusters of houses, tertiary sampling units of households, and finally, persons within households. It has been conducted continuously since 1957 and the content of the survey is periodically updated.

Prior to 2007, NHIS traditionally collected full 9-digit Social Security Numbers (SSN) from survey participants. However, in attempt to address respondents' increasing refusal to provide SSN and consent for linkage, in 2007 NHIS began to collect only the last 4 digits of SSN (SSN4) and added an explicit question about linkage for those who refused to provide SSN4. The implications of this procedural change on data linkage activities are discussed later in this report.

For more information on NHIS, visit: <https://www.cdc.gov/nchs/nhis.htm>.

NHANES is a continuous, nationally representative survey consisting of about 5,000 persons from 15 different counties each year. For a variety of reasons, including disclosure issues, the NHANES data are released on public-use data files in two-year increments. The survey includes a standardized physical examination, laboratory tests, and questionnaires that cover various health-related topics. NHANES includes an interview in the household followed by an examination in a mobile examination center (MEC). NHANES is a nationally representative, cross-sectional sample of the U.S. civilian, noninstitutionalized population that is selected using a complex, multistage probability design.

More information on NHANES is available from: <https://www.cdc.gov/nchs/nhanes.htm>.

HUD Administrative Data

Housing providers collect administrative data on families participating in HUD programs and transmit required data to HUD using federal forms. Information collected includes:

- Dates of transactions including enrollment, income recertification, and end of participation

- First and last names of each person living in the housing unit, and personal attributes including sex, race, citizenship status, Social Security Number, date of birth, and relationship to the head of household
- Family characteristics that might qualify the family for tenant selection preferences
- Detailed income and assets information for all household members
- An estimate of the family's anticipated income for the next twelve months and sources of that income
- Geographic information for the newly approved unit

Data for HUD's housing assistance programs are stored in two databases, the Public & Indian Housing (PIH) Information Center (PIC) database and the Tenant Rental Assistance Certification System (TRACS):

- **Public & Indian Housing (PIH) Information Center (PIC)**
The PIC data extract created for the NCHS-HUD data linkage was based on the PIC transactions-level file, originally implemented in May 2001. The term "transaction" refers to any occurrence for which a HUD form is completed (e.g., new admission to a HUD program, annual recertification, end of participation, etc.). The earliest effective date for transactions on this file is 18 months prior to May 2001 (i.e. December 1999), although transactions for a small portion of program participants may exist prior to this date in the transactions-level file. Due to data incompleteness, HUD restricted data from Housing Choice Vouchers (HCV) and Public Housing (PH) PIC to transactions during the dates: December 1, 1999 – December 31, 2016.
- **Tenant Rental Assistance Certification System (TRACS)**
TRACS is a system developed to collect and maintain certified tenant data for processing from owners and management agents of multifamily housing projects. The TRACS data extract created for the NCHS-HUD data linkage was based on TRACS point-in-time quarterly extracts from the TRACS production system. Since December 1997, these data capture only the most recent transaction within the past 18 months at the time of the data extract. For example, the 2012 quarterly extract contains the most recent transaction occurring between June 30, 2011 and December 30, 2012. For the NCHS-HUD data linkage, the December TRACS quarterly extracts for the years 1997-2016 were concatenated. Transactions with the same SSN, effective date, and transaction code were considered duplicates and removed. Similar to the PIC file, the earliest effective date for transactions on the TRACS file is 18 months prior to the date of the extract. While transactions for a small portion of program participants may exist prior to this date in the transactions-level file, HUD restricted data from Multifamily (MF) program TRACS transactions during the dates: June 30, 1996 – December 31, 2016.

Linkage of NCHS Survey Data with HUD Administrative Data

NHANES and NHIS data were linked to administrative records for HUD's largest housing assistance program categories, the Housing Choice Vouchers (HCV), Multifamily (MF), and Public Housing (PH) programs. The HUD administrative files contain housing, income, and program participation data for recipients of HCV, MF, and PH programs in all 50 states and the District of Columbia. HUD administrative records for MF program transactions that occurred between June 30, 1996 and December 31, 2016 were included in the linked datasets, and PH and HCV transactions occurring between December 1, 1999 and December 31, 2016 were included in the linked datasets.

For more information on HUD programs, their administration, and the PIC and TRACS data systems, please refer to "A Primer on HUD Programs and Associated Administrative Data."

Linkage Eligibility

Approval for the NCHS-HUD linkage was provided by the NCHS Research Ethics Review Board (ERB), and the linkage was performed only for eligible survey participants. The NCHS ERB, also known as an Institutional Review Board (IRB), is an appointed ethics review committee that is established to protect the rights and welfare of human research subjects. Only a subset of 1999-2016 NHIS and 1999-2016 NHANES participants were eligible for linkage with the HUD administrative data. NCHS-HUD linkage eligibility is distinct from HUD program eligibility; the latter defines whether a person meets eligibility criteria to receive HUD housing assistance through HCVs and related programs, MF programs, or PH programs. More information about HUD eligibility criteria is available from the HUD website (<https://www.huduser.gov/portal/home.html>).

In general, NCHS-HUD linkage eligibility is based on whether the NCHS survey participant provided consent for record linkage activities as well as the necessary personally identifiable information (PII), such as date of birth and SSN, to be uniquely matched to the HUD administrative data. Due to variability of questions across NCHS surveys, changes to PII collection procedures by the surveys over time, and changes in who is asked specific questions, criteria for determining linkage eligibility vary by survey, survey year, and participant characteristics.

Collection of Social Security Numbers in the NHIS and NHANES

For NHIS prior to 2007 and NHANES prior to 2009, a refusal by the survey participant to provide a 9-digit SSN (SSN9) was considered an implicit refusal for data linkage. However, NCHS began to notice an increase in the refusal rate for providing SSN, particularly for NHIS, which reduced the number of survey participants eligible for linkage.[4] In attempt to address declining linkage eligibility rates, NCHS introduced new procedures for obtaining consent for linkage from survey participants. Research was also conducted to assess the accuracy of matching data from NHIS to the National Death Index (NDI) using partial SSN and other PII.[5] The research assessed algorithms using the last four and last six digits of SSN. The results were favorable and provided

sufficient data to support changes in how NHIS collected SSN for linkage.[6] Beginning in 2007, NHIS started requesting only the last four (instead of the full nine) digits of SSN (SSN4) and Medicare Health Insurance Claim (HIC) numbers. In addition, a short introduction before asking for SSN4 was added and participants who declined to provide SSN4 were asked for their explicit permission to link to administrative records without SSN. Also at this time, the NCHS ERB determined that for 2007 NHIS and all subsequent years, only primary respondents (sample adult and sample child) were eligible for data linkage.

For the NCHS-HUD linkage, 1999-2006 NHIS participants were considered eligible for linkage if they:

- Did not refuse to provide SSN9, *and*
- Did not refuse to respond to a question about receiving rental assistance during the interview, *and*
- Provided the required data elements for linkage (SSN9, month of birth, year of birth, and sex).

For 2002-2003 NHIS, SSN9 was asked only for the family respondent and not collected for children in the household. As a result, children were not included in the linkage for these years of NHIS.

Participants in the 2007-2016 NHIS were considered eligible for linkage if they:

- Provided SSN4 *or* an affirmative response to the follow-up question to allow linkage without SSN4, *and*
- Did not refuse to respond to questions about receiving rental assistance during the interview, *and*
- Provided the required data elements for linkage.

The question about receiving rental assistance was asked only of families reporting that they lived in a rental property. Very few families living in a rental property refused to answer the question about receiving rental assistance (less than 1% in each survey year from the 1999–2016 NHIS).

For NHANES, the informed consent procedures changed as well. SSN9 was consistently collected across the survey cycles for 1999-2016. However, beginning with the 2009-2010 NHANES, participants were explicitly asked for consent to be included in data linkage activities during the informed consent process prior to the interview. Only participants who provided an affirmative response to the linkage question were considered linkage-eligible.

For the NCHS-HUD linkage, 1999-2008 NHANES participants were considered eligible for linkage if they:

- Did not refuse to provide SSN9, *and*
- Provided the required data elements for linkage.

Participants in the 2009-2016 NHANES were considered eligible for linkage if they:

- Provided an affirmative response to the linkage consent question, *and*
- Provided the required data elements for linkage (SSN9, month of birth, year of birth, and sex).

Child Survey Participants

NCHS survey participants under age 18 at the time of the survey are considered linkage-eligible if the linkage eligibility criteria described above are met and consent is provided by their parent or guardian. However, the consent provided by the parent or guardian does not apply once the child survey participant becomes a legal adult and there is no opportunity for NCHS to obtain consent to link the child participant's survey data to administrative data based on their adult experiences. As a result, in accordance with NCHS ERB guidance, NCHS only includes administrative data that were generated for program participation, claims, and other events that occurred prior to the participant's 18th birthday on the linked data files provided to researchers.

For example, a 2005 NHIS participant who was 15 years old at the time of interview can only be linked to HUD data for 2007 and earlier years (during which time the child was less than 18 years of age). This participant could not be linked to administrative records with dates after their 18th birthday, in this case dates in 2008 and later years. This practice applies to all analytic projects using linked data, whether they are conducted within the NCHS RDC or in collaboration with NCHS staff in-house.

Linkage Process

HUD provided a data file, referred to as the HUD finder file, containing PII (e.g., SSN, first and last name, sex, date of birth, a unique HUD-created identification [ID], and a sequence number to identify specific HUD program transactions) of individual HCV, MF, or PH program participants. NCHS staff verified that the SSNs on the HUD files were valid before linking to a file of 1999–2016 NHIS and NHANES linkage-eligible survey participants [3].

For 1999-2006 NHIS and 1999-2016 NHANES participants who provided consent for linkage yet did not provide a valid SSN at the time of interview, SSN9 was extracted from their Medicare HIC number, if provided. SSN9 was extracted from the Medicare HIC number only if the survey participant was identified as the primary claimant for Medicare benefits.

A primarily deterministic or rules-based approach was employed for the linkage. The New York State Identification Intelligence System (NYSIIS) [7], Soundex phonetic algorithms [8], and the SAS COMPGED function [9] were applied for the comparison of name variables to account for typographical errors and spelling variations.[10] Two approaches were used to link NCHS survey data to HUD administrative records. The first approach was used for NCHS survey years when the SSN9 was collected from survey participants (1999–2006 NHIS and all NHANES cycles). This approach linked survey participants to HUD administrative records by matching on valid SSN9,

year of birth, month of birth, and sex. The second approach was designed to accommodate the years of NHIS when SSN4 was collected (2007-2016).

Linkage Algorithm using Four-Digit SSN (SSN4)

For 2007-2016 NHIS, where only 4-digit SSN was collected, a series of rules were implemented to identify successful matches. Survey participants were first required to match exactly on all of the following data elements:

- SSN4
- Month of birth
- Year of birth
- Sex

Survey participants who matched exactly on the elements listed above were also required to meet at least one of the following conditions to be considered a successful match:

- NYSIIS/Soundex/COMPGED value of both non-missing first names and last names agree
- NYSIIS/Soundex/COMPGED value of non-missing first names agree and day of birth match
- NYSIIS/Soundex/COMPGED value of non-missing last names agree and day of birth match
- NYSIIS/Soundex/COMPGED value of non-missing first name agrees with NYSIIS/Soundex/COMPGED value of non-missing last name and NYSIIS/Soundex/COMPGED value of non-missing last name agrees with NYSIIS/Soundex/COMPGED value of non-missing first name

A threshold value of 100, based on the generalized edit distance of the two strings being compared, was used to assess agreement between two names using the SAS COMPGED function [9].

Linkage Algorithms using Nine-Digit SSN

For 1999–2006 NHIS and all NHANES cycles, where a 9-digit SSN was collected the linkage process was completed in two passes:

Pass 1: To be considered a successful match, agreement was required between the NCHS survey participant's record and the HUD administrative record on all of the following items:

- SSN9
- Month of birth
- Year of birth
- Sex

Pass 2: The 4-digit SSN algorithm described above was applied to the SSN9 records that did not match in Pass 1 of the nine-digit SSN9 algorithm.

NCHS participants whose data fields matched the equivalent administrative fields using either of the two approaches were identified as linked records.

To increase the likelihood of finding a match, multiple or alternate NCHS survey submission records were used for each linkage-eligible NCHS survey participant based on variations of names. HUD records could be matched to any or all of the submission records created for a survey participant. For example, the name “Beth” may be a nickname for a formal name like “Elizabeth.” In this situation, a record for “Beth” and a record for “Elizabeth” were created and submitted for linkage.

Once the linkage was performed and the best-matched record was selected, NCHS created an NCHS finder file containing the HUD-created ID and binary indicator flags for survey participants who were linked to the HUD administrative files for HCV, MF, or PH program recipients through 2016. This file was delivered to HUD, and HUD programming staff prepared and delivered to NCHS a single transactions-level file containing data summarized for the person’s household as well as data summarized for the individual HUD housing assistance recipients who were linked to NCHS survey participants.

NCHS programming staff ensured that all linked survey participants identified in the linkage process were returned in the transaction file. NCHS staff edited the HUD Transaction File to resolve discrepancies with linked records so that the final transaction file met ERB guidelines. The final linked transaction file served as the basis for creating the linked data files.

Linkage Rates

The linkage rates for NCHS-HUD linkage are provided in Tables 1 - 2. The tables show for each survey, the total survey sample size, the sample size eligible for linkage, the number of eligible survey participants linked to HUD data and two linkage rates.

The two linkage rates provided in the tables are: a total survey sample linkage rate and an eligible sample linkage rate. The eligible sample for linkage is based upon meeting the linkage eligibility criteria previously described. The two linkage rates for each survey were examined overall by three age groups – less than 18 years, 18 – 39 years, 40-64 years, and 65 years and older, and sex. Age was defined as the survey participant’s age at interview.

Access to NCHS-HUD Linked Data Files

Due to confidentiality requirements, the NCHS-HUD Linked Data Files are considered restricted-use. To ensure confidentiality, all personal identifiers have been removed from the linked NCHS-HUD data files. However, because there remains the small possibility of re-identification, the linked NCHS-HUD data are not available as public-use files.

Researchers who want to obtain the linked NCHS-HUD data must submit a research proposal to the NCHS Research Data Center (RDC) to obtain permission to access the restricted use files -.

Researchers with approved RDC proposals may provide analytic files created from public use survey files to the RDC, which the RDC analyst will merge to the NCHS-HUD Linked Data. Please

refer to [Appendix A: Merging Linked NCHS-HUD Files with NCHS Survey Data](#) for guidance on using and merging the appropriate identification numbers.

Description of NCHS-HUD linked files

The NCHS-HUD linked data are comprised of the Transaction, Episode, Temporal Alignment, and Weights files. These files will be referenced in the remainder of the document. Variables found in each file can be referenced in the codebooks.

The term “transaction” refers to any occurrence for which a HUD form is completed (e.g., new admission to a HUD program, annual recertification, end of participation, etc.). The term “episode” refers to a single continuous period of enrollment in a HUD program based on dates of HUD transactions. The Episode files are constructed from the Transaction file. The begin date of a participant’s first episode is the effective date on their first transaction record. Subsequent episodes for the participant are identified based on the interval between the effective dates on their transaction records.

Transaction File

Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked to HUD administrative data through December 31, 2016.

The transaction file contains a record for each transaction of the linked 1999-2016 NHIS-HUD and 1999-2016 NHANES-HUD participants. As noted previously, transactions for NHIS and NHANES child participants were removed during post-processing if the transaction occurred after their 18th birthday. The transaction file contains detailed member and household attributes that are contained in HUD administrative systems.

Episode Files

There are seven episode files that contain start and end dates for participation episodes in various HUD programs based on the transaction data and assumptions about reasonable intervals between transactions. Most HUD recipients are required to recertify each year, and consequently, a transaction is expected each year. However, some HUD programs (for instance, the Moving to Work (MTW) Demonstration Program) have longer intervals between recertification. The episode files are useful primarily for longitudinal analysis related to the duration and timing of housing assistance episodes, and conditions or outcomes that may have preceded or followed such episodes.

The seven episode files are:

- Episode File – Overall
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with any transaction record in the HUD administrative data.

- Episode File – PH
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with at least one PH program transaction in the HUD administrative data.

- Episode File – HCV
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with at least one HCV program transaction in the HUD administrative data.
- Episode File – MTW PH
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with at least one MTW PH program transaction in the HUD administrative data.
- Episode File – MTW HCV
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with at least one MTW HCV program transaction in the HUD administrative data.
- Episode File – PBS8
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with at least one MF Project-Based Section 8 transaction in the HUD administrative data.
- Episode File – Other MF
Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with at least one Other MF program transaction in the HUD administrative data.

[Appendix B: SAS Program to Create Participation Episodes](#) provides the SAS program code used to create participation episodes.

Temporal Alignment File

Universe: 1999-2016 NHIS and 1999-2016 NHANES participants who were linked with any transaction record in the HUD administrative data.

The temporal alignment file contains variables related to timing of HUD participation relative to the timing of the NHIS or NHANES interview and/or NHANES Mobile Examination Center (MEC) exam, such as: 1) indicator variables for living in HUD-assisted housing on the date of the NCHS interview (or MEC examination, where appropriate), 2) the type of HUD-assisted housing received, and 3) the number of days between the interview and/or examination dates (NHANES MEC participants only) and the previous and/or next transactions. In addition, there are variables that indicate if the survey participant ever participated in the different HUD-assisted housing programs during the entire timespan of the administrative data.

Weights File

Universe: All 1999-2016 NHIS and 1999-2016 NHANES participants.

As mentioned previously, not all of the 1999-2016 NHIS and 1999-2016 NHANES participants are eligible for linkage. Therefore, NHIS and NHANES sample weights have been adjusted for linkage eligibility. The weights file contains a record for each 1999-2016 NHIS participant and each 1999-2016 NHANES participant. The file contains various sample weights adjusted for

linkage eligibility for survey participants who were eligible for linkage. All participants who were ineligible for linkage are given a missing weight value. Percentages related to linkage eligibility can be found in Tables 1 and 2. Each survey participant may have multiple adjusted weight variables on the weights file based on the number of survey weight variables provided in the public-use survey data.

The weights in the file adjust for the potential bias that may result from differences between the linkage-eligible and linkage-ineligible populations. A model-based calibration approach developed within the SUDAAN software package (Procedure WTADJUST or WTADJX) was used to adjust the statistical weights for linkage eligibility with auxiliary information, specifically age, race and ethnicity, and sex. This is similar to a non-response adjustment. NCHS provides adjusted sample weights in a weights file based on this approach to account for differences in characteristics among linkage-eligible and linkage-ineligible survey participants. Because inferences may depend on the approach used to adjust sample weights—within SUDAAN’s WTADJUST or using a different calibration approach—researchers are advised to seek assistance from a mathematical statistician for guidance on their particular project. Note that other approaches to calculating weights or handling possible linkage-eligibility bias with other statistical software packages can be used. The variable HUD_MATCH_STATUS is provided on the file and indicates whether or not the survey participant was linkage eligible if a researcher wants to create their own adjusted sample weight. Additional information about the eligibility-adjusted participant survey weights is provided in the [Analytic Considerations and Guidelines](#) section later in this document.

Detailed descriptions for the complete list of variables contained in each of the NCHS-HUD linked data files can be found in the data dictionaries available on the NCHS Data Linkage website:

<https://www.cdc.gov/nchs/data-linkage/hud-restricted.htm>

Analytic Considerations and Guidelines

General Notices to Users

This section summarizes some key analytic issues for users of the linked NCHS survey data and HUD administrative records. It is not an exhaustive list of the analytic issues that researchers may encounter while using the Linked NCHS-HUD Data Files. This document will be updated as additional analytic issues are identified and brought to the attention of the NCHS Data Linkage Team (datalinkage@cdc.gov).

General Analytic Guidance

Merging Linked NCHS-HUD Data with NCHS survey data

The Linked NCHS-HUD Data Files can only be accessed in a RDC. Within the RDC, the Linked NCHS-HUD Data Files can be merged with the NCHS restricted (if needed) and public-use survey data files using unique survey person identification numbers. However, the unique survey identifiers are different across surveys and years. Please refer to Appendix A for guidance on using and merging the appropriate survey participant identification numbers.

Variables to request in RDC applications

To create analytic files for use in the RDC, a researcher should provide a file containing the variables from the public-use NCHS survey data to the RDC for merging with the requested restricted variables from NCHS surveys and for use with the HUD file variables. The restricted variables from NCHS surveys and the variables from the HUD files that the researcher will use also need to be specifically requested as part of a researcher's application to the RDC. Staff in the RDC verify the full list of variables (restricted and public-use) and check for potential disclosure risks.

Although the complete list of variables used for specific analyses differs, the following variables from NCHS surveys should be considered for inclusion:

- Geography— Geographical information (e.g., census tract, zip code, county code) is available on the administrative data for linked participants. However, there may be differences in the information available from the survey and administrative data. It is recommended that users who require information on geography, request this information from the NCHS survey.
- Linked mortality data for NCHS surveys—Each of the NCHS surveys that have been linked to the HUD data have also been linked to death information obtained from linkage to the NDI (except for 2015 and 2016 NHIS and 2015-2016 NHANES). The linked NDI mortality files provide date and cause of death for each survey participant who has died. These may be of use to some researchers, but must be specifically requested as part of the researcher's proposal to RDC. More information about the NCHS-NDI linked mortality files can be found on the NCHS Data Linkage website.

- Demographic information about survey participants from the NCHS survey— For variables such as race and ethnicity, NCHS demographic information is self- or family respondent-reported and, thus, may be more accurate than demographic data provided in the HUD files. Therefore, when possible, the NCHS data should be used for demographic variables.

Eligibility-adjusted Participant Survey Weights

The survey weights provided in the NHIS and NHANES data files adjust for oversampling of specific subgroups and differential non-response, and are post-stratified to annual population totals for specific population domains to provide nationally representative estimates. Little is known, however, about the properties of these weights for linked data files with incomplete linkage due to linkage ineligibility. In addition, methods for using the survey weights for some longitudinal analyses are yet to be researched. Because this is an important and complex methodological topic, ongoing work at NCHS and elsewhere is examining the use of survey weights for linked data in multiple ways.

Preliminary guidance is to analyze linked data files using eligibility-adjusted survey weights. Sample weights adjusted for linkage eligibility have been created for the NHIS-HUD and NHANES-HUD linked data files and are available in the Weights data file. The eligibility-adjusted sample weights were derived using a model-based calibration approach implemented with the SUDAAN PROC WTADJUST.[11]

The approach allows auxiliary information to be used to adjust the weights for non-response. This approach and the SUDAAN procedure are provisionally recommended for adjusting survey weights for the linked files. In addition, ‘survwgt’ can be used to poststratify the weights in Stata.[12] As noted above the eligibility-adjusted weights in the Weights data file used age, race and ethnicity, and sex as auxiliary information. Given that inferences may depend on the model used to derive weights, within SUDAAN’s WTADJUST or a different calibration approach, researchers should seek assistance from a mathematical statistician for guidance on creating different eligibility-adjusted weights specific for their particular project.

The choice of which adjusted sample weight to use depends on the analysis and, more specifically, on the variables used in the analyses and the survey years included. Below are important considerations for the two surveys:

NHIS: NCHS did not uniformly ask SSN of all NHIS participants during the 1999-2016 survey period, and this variation affects which survey participants were eligible to be linked and the survey weights that should be used. Since all adult participants of 1999-2006 NHIS were asked for their SSN9, analyses of 1999-2006 NHIS adults should incorporate the person weights (ADJ_PERWT) or the sample adult weights (ADJ_SAWT) (if analytic variables are based on sample adult file). As only adults included in the sample adult module were asked for their SSN4

in 2007-2016 NHIS, analyses of 2007-2016 NHIS adult participants should only incorporate sample adult weights (ADJ_SAWT). SSNs were obtained from all children in the 1999-2001 NHIS, so analyses of children in the 1999-2001 NHIS should incorporate either the person weights (ADJ_PERWT) or the sample child weights (ADJ_SCWT) (if analytic variables are from the sample child file). In the 2004-2016 NHIS only the sample child was asked for their SSN, so analyses of children in the 2004-2016 NHIS should incorporate sample child weights (ADJ_SCWT). As mentioned previously, children in 2002-2003 NHIS were not linked to the HUD data.

NHANES: Analyses should incorporate either the adjusted interview weights (ADJ_INTWT) or, if analytic variables are based on data obtained during the MEC examination, the adjusted MEC examination weights (ADJ_MECWT).

Temporal alignment of survey and administrative data

Each NCHS survey has been linked to multiple years of HUD data. Depending on the survey year, HUD data may be available for survey participants at the time of the survey, as well as before or after the survey period. Several factors may influence the alignment of the survey and administrative data, including: age of the survey participant, program eligibility, and discontinuous program coverage.

Identification of Ever and Concurrent HUD-Assisted Participants

Ever lived in HUD-assisted housing: To identify NCHS participants who live in HUD-assisted housing at any time during the administrative period (i.e., MF program transactions occurring during the dates June 30, 1996 – December 31, 2016, and HCV and PH transactions occurring during the dates December 1, 1999 – December 31, 2016), use the variable EVER_HUD on the Temporal Alignment File. To identify participants who ever lived in HUD-assisted housing through HCV, MF, and PH programs, use the variables EVER_MF, EVER_PH, and EVER_HCV, respectively.

Concurrent receipt of HUD assistance: The variables in the Temporal Alignment file can be used to identify concurrent HUD participation (i.e., participants who live in HUD-assisted housing at the time of their NCHS interview or examination, if applicable). Also included on the Temporal Alignment file are variables to identify participants who lived in HUD-assisted housing within a specific number of days from the survey interview or examination (TIME_A_INT, TIME_A_EXM, etc.). These variables cannot be directly accessed by the researcher, but upon request, RDC staff can use them to derive categorical variables for researchers to use in the RDC. For example, to identify participants who were in HUD within 364 days (one year) of their NCHS interview, researchers may request in their RDC proposal that an indicator variable be created that identifies participants who lived in HUD-assisted housing within 364 days of their survey interview. Due to disclosure risks, derived variables will not be provided to researchers requesting episode files.

There may be misclassification due to how the episodes were constructed from transaction data and this may impact the concurrent indicators.

Analyses of Children in the 1999-2016 NCHS-HUD Linked Data Files

As mentioned previously, administrative data for child survey participants generated after their 18th birthday are not available. This limitation impacts the latter two of the following three groups of 1999-2016 NHIS or 1999-2016 NHANES child participants who lived in HUD-assisted housing during the 1996-2016 timeframe:

1. Child survey participants who only lived in HUD-assisted housing as children. There is no impact on this subgroup of children; all transactions are available.
2. Child survey participants who lived in HUD-assisted housing as children and adults. All transactions that occurred prior to the child's 18th birthday are available, but all transactions occurring on or after the child's 18th birthday are not available for release.
3. Child survey participants who lived in HUD-assisted housing only as adults. No transactions would be available for these participants in the NCHS-HUD linked data.

Researchers should keep in mind that for some survey years, adult survey participants may have HUD program participation available for transactions that occurred in the years prior to the interview when the participant was under 18 years of age. Researchers interested in performing analyses of children should take this into consideration.

Analyses of Rental Assistance Programs

Since a small number of HCV housing assistance programs provide homeownership vouchers, these programs are not technically "rental" assistance programs. Researchers using the linked files to specifically examine "rental" assistance programs should exclude transactions from the HCV homeownership program. If researchers wish to broadly examine HUD assistance programs for low-income households, all transactions can be included. Researchers interested in examining only rental assistance programs should indicate this in their RDC proposal. NCHS will remove HCV homeownership vouchers from the requested file. More information about HCV homeownership vouchers is provided in the [Variable Considerations and Data Anomalies](#) section.

Unit of Analysis

When using the NCHS-HUD linked files, the unit of analysis should be the participant, not the household. Survey participants, not households, were linked to HUD administrative data. While one household member's living in HUD-assisted housing directly affects all members in a household, household-level analyses should not be done for several reasons. First, some members of a HUD household who were NCHS survey participants may not have been eligible for linkage; and will not be on the linked file. Second, transactions that occurred on or after the 18th birthday of child survey participants are not included in the linked files. Third, the membership of the HUD household may differ from that of the corresponding NCHS survey household.

Analytic Considerations for Episode Files

If the number of days between two transactions was within the recertification period (12 months for non-MTW recipients, 36 months for MTW recipients), the recipient was assumed to have been receiving assistance during that episode. If the number of days between two

transactions was outside the recertification period, the end date was the previous transaction date.

There are two important considerations when using the episode files. First, transaction type was not taken into account when the episodes were created. The reason for using the number of days between the transactions rather than the type of transaction was that end of participation forms are not always submitted, and requiring that an end of participation transaction define the end of an episode would bias concurrent predictions. As a result, given the way the episodes were defined, it is possible for an “end of transaction” to also appear as the start date of an episode.

It should be noted that researchers can use transaction type and end of participation dates to define their own episodes. However, this is not advisable without program expertise because, as noted above, this requires some assumptions about timing and may lead to misclassification.

The second consideration to keep in mind is that the overall episode file does not always align with the program-specific episode files. This is because the episodes in the overall episode file are created using the same algorithm as each program-specific episode file, which is based on the dates of transactions. The start and end dates are created irrespective of program type, which means that any two effective dates for two different programs may be the start and end date of a single episode. For program-specific analyses, data linkage staff recommend that the program-specific episode files be used in preference to the overall episode file. Episode files cannot be provided in conjunction with some concurrent variables on the temporal alignment file due to disclosure risks. Requests for variables from both episode and temporal alignment files will be subject to review.

MF Housing Program Data: Limitations and Considerations

Although HUD analysts generally do not treat the various MF subprograms as one composite category, a composite MF category was created for the NCHS-HUD linked files in addition to maintaining the MF subprograms. HUD does not recommend that researchers analyze MF subprograms without specialized expertise in these subprograms. HUD provides the following recommendations for analyzing MF programs in the linked data:

- If the research purpose is only to identify low-income individuals receiving HUD rental assistance, then use the pooled variable for MF.
- If the research purpose is to make program-specific policy recommendations related to MF housing, then acquire a comprehensive understanding of the various MF subprogram types and functions. Account for differences among the subprograms in the analysis, especially when inferences are drawn. Depending on the research question, it may be advisable to include only Project-Based Section 8 participants in the analysis.
- In the linked data, the Project Based Section 8 program is the largest MF subprogram and the one most similar to the HCV and PH programs. Depending on the research question, it may be inadvisable to combine this program with the Section 236 or Section 221(d)(3) subprograms; doing so could lead to irrelevant and/or inaccurate results.

- Section 202 and Section 811 MF subprograms serve special populations- elderly households and disabled households. The differences between these populations and those of other HUD programs must be accounted for in the analysis, especially when inferences are drawn.

Variable Considerations and Data Anomalies

HUD Program Variables

The variable `VOUCHER_TYPE`, in the transaction file, can be used to examine specific rental assistance programs within the HCV category. Due to the small number of HCV homeownership vouchers, there is not sufficient data to evaluate the HCV homeownership voucher program, alone. Additionally, NCHS recommends that researchers using the linked files to solely examine "rental" assistance programs should exclude transactions from HCV homeownership program. Researchers interested in examining only rental assistance programs should indicate this in their RDC proposal. NCHS will remove HCV homeownership vouchers from their file.

Individuals may be recipients of more than one HUD program at the same point in time. Cases of dual program participation are rare but nonetheless occur in the linked data and indicate errors in the administrative data. Analysts must consider this potential discrepancy when conducting analyses using the linked data.

Some `PROGRAM_TYPE` variable codes in the transaction file have been re-categorized under the same overall `PROGRAM` variable. Program type codes for 'Indian Housing', 'Certificate', 'Mandatory Conversion', and 'Moderate Rehabilitation' have been recoded to the 'Housing Choice Vouchers' (VO) `PROGRAM_TYPE` category. The program type code for 'Section 811 Project Rental Assistance Demo' has been recoded to the 'Section202 PRAC (Project Rental Assistance Demo)' (H7) `PROGRAM_TYPE` category.

Transaction Variables

Transactions with rare transaction codes were excluded from the NCHS-HUD linked data transaction file. As described previously, the episodes of participation defined in the Episode files do not take into account the transaction type. Researchers interested in creating their own episodes using the type of transaction should understand the recertification process for each HUD program. Recertification rules vary based on program and Public Housing Agency (PHA) participation in the MTW demonstration.

Disability Indicator and Count

The transaction file includes five disability indicator variables (`DISABLED_COHEAD`, `DISABLED_HEAD`, `DISABLED_HOUSEHOLD`, `DISABLED_MBR`, `DISABLED_SPOUSE`) and two disability count variables (`OTHER_ADULT_DISABLED_CNT`, `CHILD_DISABLED_CNT`). Information on disability for HUD recipients is collected on Forms 50058 and 50059. These two HUD forms capture different definitions of disability which are defined according to program. It is important to note that the disability indicators are not related to the impairment variables, which are also on the transaction file.

Some of the disability variables are derived from other disability variables. For example, several conditions must be met in order to identify a household as disabled. A household is considered to be a HUD disabled household if the head of household, spouse, and co-head are all less than 62 years of age and at least one of them is disabled. This is indicated by a disability indicator in the linked data. Two other disability count variables also are derived from this variable as follows: OTHER_ADULT_DISABLED_CNT is the count of disabled adults in the household other than the head of household, spouse, and co-head. The variable CHILD_DISABLED_CNT is the count of all disabled household members who are under 18 years of age (including foster children).

Income Variables

The Transaction file has summary income variables that provide information about the income amounts and sources for individual household members and the household as a whole. Some income codes are used to establish exclusions or deductions. When potential tenants apply for housing assistance, they must report all sources of income, except income for individuals explicitly excluded (i.e., live-in aides, foster children, and foster adults). Exclusions vary by HUD program.

INCOME_OTHER_MBR is a sum of the following income codes collected from HUD Form 50058: (C) Child Support + (E) Medical Reimbursement + (I) Indian trust/per capita + (N) Other nonwage Sources + (U) Unemployment benefits. Item 73 on Form 50059, entitled "Total Other Income," sums four income codes: (CS) Child Support + (I) Indian Trust + (N) Other Non-Wage source + (U) Unemployment.

INCOME_PENSION_MBR is the sum of the following income codes collected from Form 50058: (P) Pension + (S) SSI + (SS) Social Security. Item 71 on Form 50059, entitled "Total Pension Income," sums the same three income codes utilized on Form 50058: (PE) Pensions + (SI) Supplemental Security Income + (SS) Social Security.

INCOME_WAGE_MBR is the sum of the following variables from Form 50058: (B) Own Business + (F) Federal Wage + (HA) PHA Wage + (M) Military Pay + (W) Other Wage. For Form 50059, INCOME_WAGE_MBR is the sum of the following: (B) Business + (F) Federal Wages + (M) Military Pay + (W) Non-Federal Wage. This calculation is Item 70 "Total Employment Income" on HUD Form 50059.

INCOME_WELFARE_MBR is the sum of the following income codes collected from Form 50058: (G) General Assistance + (IW) Annual Imputed Welfare Income + (T) TANF assistance. Item 72 on Form 50059, entitled "Total Public Assistance Income," sums two income source codes: (G) General Assistance/Welfare + (T) TANF.

The two variables, TOT_A_INCOME and INCOME_TOTAL_HH, provide total household income. TOT_A_INCOME is calculated by PHAs and INCOME_TOTAL_HH is calculated by the HUD data analysts who provided data to NCHS. The household income values from these two variables generally match. If they do not match, one or both of them may be inaccurate, possibly due to

administrative errors. Therefore, HUD recommends that researchers be consistent when choosing a household income variable for use. If a researcher is interested in household income details such as source of income, he/she should use INCOME_TOTAL_HH as opposed to TOT_A_INCOME because INCOME_TOTAL_HH is based on income detail variables for all members in the household provided to NCHS and will therefore be consistent with detailed member level income data.

Note that monetary values in the NCHS-HUD linked data files are not adjusted for inflation. General guidance from HUD's Economic and Market Analysis Division is to use the Consumer Price Index (CPI) [13] when adjusting incomes and rents for comparability across time and geography. Due to fluctuations in the relationship between rent and utilities to gross rent, it is recommended to use 80% of the change in Rent of Primary Residence and 20% of the change in Fuels and Utilities when adjusting gross rent for inflation.

Household Payment and Subsidy Variables

A number of variables in the transaction file pertain to household payment.

CONTRACT_RENT - the monthly contract rent of a housing unit as established by a landlord/owner and a PHA/tenant. This value does not include tenant-paid utilities, but does include the value of any utilities paid by the landlord. The variable UTIL_ALLOW (Utility allowance) is HUD's estimate of monthly tenant-paid utilities.

GROSS_RENT - HUD's estimate of the gross amount paid monthly by tenants and is the sum of CONTRACT_RENT and UTIL_ALLOW.

ASSISTANCE_PAYMENT - the amount of money HUD pays for the rent and utilities of the housing unit. This value generally should not be less than 0, but can be for MTW records, because the amount varies in the MTW data due to the flexibility associated with MTW data reporting. It is recommended to either recode negative values to zero or delete negative values.

Assistance amounts are missing for PH programs because the subsidy is delivered via the operating fund and the capital fund, not to individual households.

Total Household Expenses and Assistance Payments

The TOTAL_HOUSEHOLD_EXPENSES variable in the transaction file gives the total amount paid monthly by a household for expenses (i.e., rent and utilities). The variable ASSISTANCE_PAYMENT in the transaction file represents the amount of money HUD pays for the rent and utilities of the unit. These variables may be inaccurate for participants in MTW programs, but the extent of the inaccuracy is unknown and these calculations are provided by HUD only as an estimate for the researcher. Additionally, these variables were derived from multiple variables that are not available on the linked data files. This value generally should not be less than 0 but can be for MTW records, because the amount varies in the MTW data due to the flexibility associated with MTW data reporting. It is recommended to either recode negative values to zero or delete negative values. Assistance amounts are missing for PH programs because the subsidy is delivered via the operating fund and the capital fund, not to individual

households. Calculations for assistance payments and total household expenses can be found in “A Primer on HUD Programs and Associated Administrative Data,” Appendix D.

TTP (Total Tenant Payment) is the rent determination in PH and is based on anticipated gross annual income less deductions, if any, for an individual or family. The formula used in determining the TTP is the highest of the following, rounded to the nearest dollar: 30% of the monthly adjusted income (equals annual income less deductions allowed by the regulations); 10% of monthly income; Welfare rent, if applicable; or, a \$25 minimum rent or higher amount (up to \$50) set by a PHA.

Geocoded Data

Though the original file received from HUD contained a very detailed level of geographic information (i.e., address), the Transaction file available for release through the RDC does not contain this detailed level of geographic information.

Geocoded data for the linked participant’s residence at the time of their survey interview are available through the RDC. However, it is important to note that although this level of geography is available, NHIS and NHANES samples are only representative at the regional and national level. For this reason, PHAs and private housing providers are not identified in the linked data.

Some NCHS surveys include a measure of urban/rural geographic location, whereas others do not. Please refer to the survey documentation for information about available data. If the survey does not include the urban-rural classification of interest, it can be merged onto the file using state and county identifiers. An urban-rural classification recommended for use with NCHS surveys is the NCHS Urban-Rural Classification Scheme for Counties.[14] When requesting that an urban-rural classification scheme be merged onto the NCHS-HUD linked file, include state and county in the list of restricted variables and request the NCHS Urban-Rural scheme as an additional NCHS data source. State and county identifiers will be removed after the urban-rural codes are merged onto the linked file.

Variables with Potential Data Quality Concern

Administrative data are not designed for research purposes. As a result, it is recommended that researchers use the HUD administrative variables on the linked files with caution because the validity and reliability of some of these variables are not completely known (e.g., not reliably reported, inconsistent with transactions types, not consistent with data on individual household members, etc.). HUD acknowledges that several of these variables may not be accurately nor consistently reported. Variables to approach with particular caution are below. Please refer to the Transaction file Codebook for more details about these variables.

Variable name	Description	HUD Program		
		PH	HCV	MF
HOMELESS_IND	A binary variable to indicate if the household was homeless upon entry into the program	X	X	N/A
PORTABILITY_IND ^[15]	A binary variable to indicate if the household moved from a different PHA	N/A	X	N/A
TOT_HOUSEHOLD_MEMBERS	Total number of household members	X	X	X

Appendices

Appendix A: Merging Linked NCHS-HUD Files with NCHS Survey Data

The restricted-use NCHS-linked files are merged with the public-use NCHS survey data files using unique person identifiers. Therefore, it is important for researchers to include the correct survey person identification number: PUBLICID (for NHIS), or SEQN (for NHANES).

For using NHIS data, it also is important to note in the descriptions below that the variable names and locations needed to construct PUBLICID vary by NHIS survey year.

NHIS 1999-2003

The data items 'Survey year' (SRVY_YR), 'Household number' (HHX), 'Family number' (FMX), and 'Person number' (PX) identify a participant within each NHIS*. These data items must be concatenated to obtain the unique personal identifier (PUBLICID) used in the NHIS-HUD linked file.

Public-use

Variable	Location	Length	Description
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial number
FMX	13-14	2	Family number
PX	15-16	2	Person number within Household

SAS example:

```
length publicid $14;  
PUBLICID = trim(left(SRVY_YR||HHX||FMX||PX));
```

Note: The SAS input statements available from the NHIS public-use data website do NOT input all of the variables as character and they must be in character format for the concatenation.

Stata example: (note this will convert the variables to a string variable)

```
egen PUBLICID = concat(SRVY_YR HHX FMX PX)
```

*The data item 'Person number' was called PX in the 1999-2003 NHIS and FPX in the 2004-2012 NHIS. Users may find it necessary to create an FPX variable in the 2003 and earlier datasets (or PX in later datasets).

NHIS 2004

The data items 'Survey year' (SRVY_YR), 'Household number' (HHX), 'Family number' (FMX), and 'Person number' (FPX) identify a participant within NHIS 2004. These data items must be concatenated to obtain the unique personal identifier (PUBLICID) used in the NHIS-HUD linked file.

Variable	Location	Length	Description
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial number
FMX	13-14	2	Family number
FPX	15-16	2	Person number

SAS example:

```
length publicid $14;
PUBLICID = trim(left(SRVY_YR| |HHX| |FMX| |FPX));
```

Stata example: (note this will convert the variables to a string variable)
egen PUBLICID = concat(SRVY_YR HHX FMX FPX)

NHIS 2005 – 2016

The data items ‘Survey year’ (SRVY_YR), ‘Household number’ (HHX), ‘Family number’ (FMX), and ‘Person number’ (FPX) identify a participant within each NHIS. These data items must be concatenated to obtain the unique personal identifier (PUBLICID) used in the NHIS-HUD linked file.

Variable	Location	Length	Description
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial number
FMX	16-17	2	Family number
FPX	18-19	2	Person number

SAS example:

```
length publicid $14;
PUBLICID = trim(left(SRVY_YR| |HHX| |FMX| |FPX));
```

Stata example: (note this will convert the variables to a string variable)
egen PUBLICID = concat(SRVY_YR HHX FMX FPX)

Appendix B: SAS Program to Create Participation Episodes

Construction of the Episode Files

While a transaction is any occurrence for which a HUD form is completed (e.g., new admission to a HUD program, annual recertification, end of participation, etc.), an episode is a single continuous period of enrollment in a HUD program based on dates of HUD transactions. The Episode files are constructed from the Transaction file. The begin date of a participant's first episode is the effective date on their first transaction record. Subsequent episodes for the participant are identified based on the interval between the effective dates on their transaction records. The SAS program (described in below) then cycles through each transaction's effective date and executes one of two actions:

- 1) treat the current transaction as part of the current episode and proceed to the next transaction, or
- 2) treat the current transaction as the start of a new episode, which then forces the previous transaction to be the end date of the previous episode.

The action implemented in the SAS code is determined by the number of days between each transaction as well as the HUD program type. The expected interval between any two transactions for a non-MTW recipient is one year. However, because PHAs are given 60 days leeway to submit reports, 425 days (one year plus 60 days) is used as the standard for determining if there has been a break in assistance. For most MTW PHAs, the expected interval between any two transactions for an MTW recipient is two years. However, MTW PHAs are given the flexibility to conduct recertification as infrequently as every three years. In the NCHS-HUD Linked Data, the estimated interval between any two transactions for the majority of MTW recipients is two years, again with a 60 day leeway; therefore, 790 days (two years and 60 days) is used as the standard.

If the interval between this date and the subsequent transaction's effective date is less than 425 days for non-MTW programs, or 790 days for MTW programs, it is assumed that the two dates are part of the same "episode" of participation. This continues until the interval between two effective dates is greater than 425 days for non-MTW programs, or 790 days for MTW programs. If the interval is greater than these durations, it is assumed that the two dates are from two distinct episodes of enrollment. In this situation, the effective date of the transaction immediately preceding the current transaction becomes the last date of the previous episode and the effective date of the current transaction becomes first date of the subsequent transaction.

The following SAS program was used to generate program participation episodes:

```
*****
*****
***          ***
***PURPOSE: CREATE EPISODE PERIODS FROM HUD TRANSACTION DATA          ***
***                                     ***
```

```

***                                     ***
*****
*****
*@ACTION: INPUT RESTATE TRANSACTION FILE FROM HUD;
DATA NCHS_DATA;
SET INPUT.RESTATE_HUD_TRANSACTIONS_INT;
*@ACTION:RENAME PROGRAMS;
IF PROGRAM EQ 'MTW HCV' THEN PROGRAM='MTW_HCV';
IF PROGRAM EQ 'MTW PH' THEN PROGRAM='MTW_PH';
IF PROGRAM EQ 'Other MF' THEN PROGRAM='OTHER_MF';
RUN;

%MACRO PERIODS (PGRM, PERIOD);
*@ACTION: SORT DATA BY IDS AND EFFECTIVE DATE;
PROC SORT DATA=NCHS_DATA OUT=ALL_DATA;
BY PUBLICID EFFECTIVE_DATE;
*@ACTION: BREAK OUT BY NON MISSING PROGRAM TYPE;
%IF &PGRM GT %THEN %DO;
WHERE PROGRAM EQ "&PGRM";
%END;
RUN;
*@ACTION: CREATE INTERNAL EPISODE FILES BY PROGRAM TYPE;
DATA OUTPUT.EPISODE_DATES_&PGRM._INT (KEEP=PUBLICID SEQN &PGRM._BEG_DATE1 -
&PGRM._BEG_DATE10
&PGRM._END_DATE1 - &PGRM._END_DATE10 SURVEY_NAME);
SET ALL_DATA;
BY PUBLICID;
*@ACTION:CREATE VARIABLES TO HOLD ALL OF THE PERIODS, PLUS BEGINNING AND ENDING
DATES;
RETAIN HOLD_EFFDT PERIOD_1 - PERIOD_66 EPISODE_CNT TRANSACTION_CNT
&PGRM._BEG_DATE1 - &PGRM._BEG_DATE10 &PGRM._END_DATE1 -
&PGRM._END_DATE10;
EFFDT=EFFECTIVE_DATE;
ARRAY PERIODS (66) PERIOD_1 - PERIOD_66;
ARRAY BEGIN_DATE (10) &PGRM._BEG_DATE1 - &PGRM._BEG_DATE10;
ARRAY END_DATE (10) &PGRM._END_DATE1 - &PGRM._END_DATE10;
*@ACTION: LABEL VARIABLES;
LABEL
PUBLICID = "NHIS PUBLIC ID"
SEQN = "NHANES SEQUENCE NUMBER"
SURVEY_NAME = "SURVEY NAME"
&PGRM._BEG_DATE1 = "&PGRM. BEGIN DATE-EPISODE 1"
&PGRM._BEG_DATE2 = "&PGRM. BEGIN DATE-EPISODE 2"
&PGRM._BEG_DATE3 = "&PGRM. BEGIN DATE-EPISODE 3"

```

```

&PGRM._BEG_DATE4 = "&PGRM. BEGIN DATE-EPISODE 4"
&PGRM._BEG_DATE5 = "&PGRM. BEGIN DATE-EPISODE 5"
&PGRM._BEG_DATE6 = "&PGRM. BEGIN DATE-EPISODE 6"
&PGRM._BEG_DATE7 = "&PGRM. BEGIN DATE-EPISODE 7"
&PGRM._BEG_DATE8 = "&PGRM. BEGIN DATE-EPISODE 8"
&PGRM._BEG_DATE9 = "&PGRM. BEGIN DATE-EPISODE 9"
&PGRM._BEG_DATE10= "&PGRM. BEGIN DATE-EPISODE 10"
&PGRM._END_DATE1= "&PGRM. END DATE-EPISODE 1"
&PGRM._END_DATE2= "&PGRM. END DATE-EPISODE 2"
&PGRM._END_DATE3= "&PGRM. END DATE-EPISODE 3"
&PGRM._END_DATE4= "&PGRM. END DATE-EPISODE 4"
&PGRM._END_DATE5= "&PGRM. END DATE-EPISODE 5"
&PGRM._END_DATE6= "&PGRM. END DATE-EPISODE 6"
&PGRM._END_DATE7= "&PGRM. END DATE-EPISODE 7"
&PGRM._END_DATE8= "&PGRM. END DATE-EPISODE 8"
&PGRM._END_DATE9= "&PGRM. END DATE-EPISODE 9"
&PGRM._END_DATE10= "&PGRM. END DATE-EPISODE 10"
;
*@ACTION: FORMAT THE DATE FIELDS;
FORMAT &PGRM._BEG_DATE1 - &PGRM._BEG_DATE10 &PGRM._END_DATE1 -
&PGRM._END_DATE10 DATE.;
IF FIRST.PUBLICID THEN DO;

    HOLD_EFFDT=EFFDT;
*@ACTION: INITIALIZE FIELDS TO MISSING OR ZERO;
    DO J=1 TO 10;
        BEGIN_DATE (J)=.;
        END_DATE(J)=.;
    END;

    TRANSACTION_CNT=0;
    EPISODE_CNT=1;
    BEGIN_DATE(EPISODE_CNT)=EFFDT;

    DO I = 1 TO 66;
        PERIODS(I)=.;
    END;

END;
*@ACTION: INCREMENT TRANSACTION COUNTER BY ONE;
TRANSACTION_CNT+1;
*@ACTION: CALCULATE PERIODS BETWEEN TRANSACTIONS;
PERIODS(TRANSACTION_CNT)=EFFDT-HOLD_EFFDT;
IF PERIODS(TRANSACTION_CNT) GT &PERIOD THEN DO;

```

```

        END_DATE(EPISEDE_CNT)=HOLD_EFFDT;
        EPISEDE_CNT+1;
        BEGIN_DATE(EPISEDE_CNT)=EFFDT;
END;

HOLD_EFFDT=EFFDT;
*@ACTION: OUTPUT ONE RECORD PER ID;
IF LAST.PUBLICID THEN DO;
        END_DATE(EPISEDE_CNT)=EFFDT;
OUTPUT;
END;
RUN;
*@ACTION: CREATE PUBLIC FROM INTERNAL VERSION;
DATA OUTPUT.EPISEDE_DATES_&PGRM._PUB(DROP=NEW_HUD_ID);
SET OUTPUT.EPISEDE_DATES_&PGRM._INT ;

RUN;
*@ACTION: SHOW CONTESNTS OF INTERNAL AND PUBLIC FILES;
PROC CONTENTS DATA=OUTPUT.EPISEDE_DATES_&PGRM._PUB;
PROC CONTENTS DATA=OUTPUT.EPISEDE_DATES_&PGRM._INT;
RUN;
%MEND PERIODS;
*@ACTION: RUN MACRO FOR ALL PROGRAM TYPE;
%PERIODS (HCV, 425); *NOTE:ONE YEAR PLUS TWO MONTHS;
%PERIODS (PH, 425);
%PERIODS (PBS8, 425);
%PERIODS (OTHER_MF, 425);
%PERIODS (MTW_HCV, 1155); *NOTE:THREE YEARS PLUS TWO MONTHS;
%PERIODS (MTW_PH, 1155);
%PERIODS (, 425);
*@ACTION: TIME STAMP;
        DATA _NULL_;
        LAPTIME=TIME();
        LAPDATE = TODAY();
        PUT "TIME CHECK: " LAPTIME TIMEAMPM14.2 " " LAPDATE YYMMDD10. " **
PROGRAM ENDED" ;
        RUN;

```

References

- ¹ U.S. Department of Housing and Urban Development, Office of Policy Development and Research (PD&R). HUD User. <http://www.huduser.gov>. Accessed November 19, 2018.
- ² National Center for Health Statistics. NCHS Data Linked to HUD Housing Assistance Program Files. <https://www.cdc.gov/nchs/data-linkage/hud.htm>. Accessed November 19, 2018.
- ³ Lloyd PC, Helms VE, Simon AE, et al. Linkage of 1999–2012 National Health Interview Survey and National Health and Nutrition Examination Survey data to U.S. Department of Housing and Urban Development administrative records. National Center for Health Statistics. *Vital Health Stat* 1(60). 2017.
- ⁴ Miller, D.M., R. Gindi, and J.D. Parker, *Trends in record linkage refusal rates: Characteristics of National Health Interview Survey participants who refuse record linkage*. Presented at Joint Statistical Meetings 2011. Miami, FL., July 30–August 4.
- ⁵ Sayer, B. and C.S. Cox. *How Many Digits in a Handshake? National Death Index Matching with Less Than Nine Digits of the Social Security Number in Proceedings of the American Statistical Association Joint Statistical Meetings*. 2003.
- ⁶ Dahlhamer, J.M. and C.S. Cox, *Respondent Consent to Link Survey Data with Administrative Records: Results from a Split-Ballot Field Test with the 2007 National Health Interview Survey*. Paper presented at the 2007 Federal Committee on Statistical Methodology Research Conference, Arlington, VA, 2007.
- ⁷ Hobbs, S. "New York State Identification and Intelligence System (NYSIIS) Phonetic Encoder." Available at <http://www.dropby.com/NYSIIS.html>. Accessed April 30, 2019.
- ⁸ SAS. "SOUNDEX Function." SAS(R) 9.2 Language Reference: Dictionary, Fourth Edition [online]. <http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a000245948.htm>. Accessed April 30, 2019.
- ⁹ SAS. "COMPGED Function." SAS(R) 9.2 Language Reference: Dictionary, Fourth Edition [online]. <http://support.sas.com/documentation/cdl/en/lrdict/64316/HTML/default/viewer.htm#a002206133.htm>. Accessed April 30, 2019.
- ¹⁰ Roesch, A. Matching data using Sounds-Like Operators and SAS compare functions, SAS Global Forum 2012 Paper 122-2012.
- ¹¹ Golden, C., et al., *Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services*. National Center for Health Statistics, 2015. *Vital Health Stat* 1(58).
- ¹² Winter N. Survey sampling weights: adjustment and replicate weight creation. <http://faculty.virginia.edu/nwinter/progs/survwgt.hlp.shtml>. Accessed November 27, 2018.
- ¹³ U.S. Bureau of Labor Statistics. Consumer Price Index. <https://www.bls.gov/cpi/>. Accessed November 27, 2018.
- ¹⁴ U.S. Centers for Disease Control and Prevention. NCHS Urban-Rural Classification Scheme for Counties. https://www.cdc.gov/nchs/data_access/urban_rural.htm. Accessed November 27, 2018.
- ¹⁵ U.S. Department of Housing and Urban Development. Housing Choice Voucher Program Guidebook. <http://www.lb7.uscourts.gov/documents/15c6523.pdf>. Accessed November 19, 2018.