



The Linkage of National Center for Health Statistics Surveys to Medicare Enrollment and Claims Data (1999-2013)

Methodology and Analytic Considerations

Data Release Date: December 14, 2016
Document Version Date: February 7, 2023

Suggested citation:

National Center for Health Statistics, Division of Analysis and Epidemiology. The Linkage of National Center for Health Statistics Surveys to Medicare Enrollment and Claims Data (1999-2013) - Methodology and Analytic Considerations. December 2016. Hyattsville, Maryland.

Available at the following address:

https://www.cdc.gov/nchs/data-linkage/cms/nchs_medicare_linkage_methodology_and_analytic_considerations.pdf

Table of Contents

Introduction	1
Data Sources	2
<i>National Center for Health Statistics</i>	2
<i>Centers for Medicare & Medicaid Services</i>	5
Additional Related Data Sources	10
Linkage of NCHS Surveys with 1999-2013 Medicare Records	10
<i>Linkage Eligibility</i>	10
<i>Child Survey Participants</i>	12
<i>CMS Virtual Research Data Center</i>	12
<i>Linkage Methods</i>	13
<i>Linkage Algorithms using Four-Digit SSN and Partial HIC Number</i>	14
<i>Linkage Algorithms using Nine-Digit SSN and Complete HIC Number</i>	14
<i>Linkage Rates</i>	16
<i>Data Confidentiality</i>	17
Analytic Considerations when using the Linked NCHS-CMS Medicare Files	17
<i>General Notices to Users</i>	17
General Analytic Guidance	18
<i>Merging NCHS-CMS Medicare Linked Data with NCHS survey data</i>	18
<i>Variables to request in RDC applications</i>	18
<i>Sample weights</i>	20
<i>Temporal alignment of survey and administrative data</i>	21
Analysis Using Linked NCHS-CMS Medicare Data	21
<i>NCHS-CMS Medicare Linked Data Feasibility Files</i>	21
<i>Medicare Advantage</i>	22
<i>Importance of the Master Beneficiary Summary File</i>	23
<i>Services not covered (1999-2013)</i>	23
<i>Cost Sharing</i>	24
<i>Gaps and Discrepancies in Coverage Periods</i>	25
<i>On Master Beneficiary Summary File with no claims data</i>	25
<i>Medicare entitlement variables</i>	26
<i>Medicare's Prospective Payment System (PPS)</i>	26
<i>Medicare Provider Analysis and Review File</i>	27

References	28
Acknowledgments	29
Additional Resources	29
Appendix A: Merging Linked NCHS-CMS Medicare Files with NCHS Survey Data	30

**The Linkage of National Center for Health Statistics Surveys to Medicare
Enrollment and Claims Data (1999-2013):
Methodology and Analytic Considerations**

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 policy 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 survey data it collects by augmenting survey information with information from administrative data sources. The linkage of survey information with administrative data provide the unique opportunity to study changes in health status, health care utilization and expenditures in specialized populations, such as the elderly and disabled.

Under an interagency agreement between NCHS and the Centers for Medicare and Medicaid Services (CMS) several NCHS surveys have been linked to Medicare enrollment and claims data. This linkage is the third collaboration between the NCHS and CMS. The two previous linkages were conducted under an interagency agreement among the NCHS, CMS, the Social Security Administration (SSA), and the Office of the Assistant Secretary for Planning and Evaluation (ASPE) of the Department of Health and Human Services (DHHS).

These linkages support various research initiatives of the participating agencies. The resulting linked data files provide the opportunity to examine the administrative data during the year the survey was conducted, in years following the survey, as well as the years prior to the survey for some NCHS survey participants. The NCHS-Medicare linked files, in particular, combine health and socio-demographic information from the

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

surveys with enrollment and claims information from the Medicare program, resulting in unique population-based information that can be used for an array of epidemiological and health services research.

This report describes the most recent linkage conducted between NCHS surveys and CMS administrative records. A brief overview of the data sources, the methods used for linkage, descriptions of the resulting linked data files and analytic guidance are provided. More information about the two previous linkages of NCHS survey and CMS administrative data can found at: http://www.cdc.gov/nchs/data/series/sr_01/sr01_058.pdf (accessed September 28, 2017)

Data Sources

National Center for Health Statistics

The following NCHS surveys were linked to Medicare enrollment and claims data that covered services for the years 1999-2013:

- The 1994-2013 National Health Interview Survey (NHIS)
- The Second Longitudinal Study of Aging (LSOA II)
- The 1999-2012 Continuous National Health and Nutrition Examination Survey (NHANES)
- The Third National Health and Nutrition Examination Survey (NHANES III)
- The NHANES I Epidemiologic Follow-Up Study (NHEFS)
- The 2007 National Home and Hospice Care Survey (NHHCS)
- The 2004 National Nursing Home Survey (NNHS).

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 adjacent counties, secondary sampling units of

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

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. NHIS has been used as the sampling frame for other NCHS surveys focusing on specialized populations, including **LSOA II**. 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, NHIS began, in 2007, to collect only the last 4 digits of SSN and added an explicit question about linkage for those who refused to provide SSN. The implications of this procedural change on data linkage activities are discussed later in this report. NHIS is currently planning a content and structure redesign. For detailed information on the NHIS's contents and methods, refer to the NHIS website, <http://www.cdc.gov/nchs/nhis.htm> (accessed September 28, 2017).

LSOA II was a prospective study of a nationally representative sample of civilian, non-institutionalized persons 70 years of age and over at the time of their 1994 NHIS interview, which served as the baseline for the study. The LSOA II study design included two follow-up telephone interviews, conducted in 1997-98 and 1999-2000. The LSOA II provides information on changes in disability and functioning, individual health risks and behaviors in the elderly, and use of medical care and services employed for assisted community living. For detailed information on the LSOA II contents and methods, refer to the LSOA II website, <https://www.cdc.gov/nchs/lsoa/lsoa2.htm> (accessed September 28, 2017).

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.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Prior to becoming a continuous survey in 1999, NHANES was conducted periodically, with the last periodic survey, **NHANES III**, conducted between 1988 and 1994. NHANES III was designed to provide national estimates of health and nutritional status of the civilian, non-institutionalized population of the United States aged 2 months and older. Similar to the continuous survey, NHANES III included a standardized physical examination, laboratory tests, and questionnaires that covered various health-related topics.

NHEFS was a national longitudinal study conducted in collaboration with the National Institutes of Health, National Institute on Aging and other agencies of the Public Health Service. The NHEFS cohort included all persons 25–74 years of age who completed a medical examination as part of NHANES I in 1971–75. The NHEFS study design included four follow-up interviews, conducted in 1982–84, 1986, 1987 and 1992, to investigate the relationships between clinical, nutritional, and behavioral factors assessed at baseline, and subsequent morbidity, mortality, and institutionalization.

For detailed information about the Continuous NHANES, NHANES III and NHEFS contents and methods, refer to the NHANES website, <https://www.cdc.gov/nchs/nhanes/index.htm> (accessed September 28, 2017).

NHHCS is one in a series of nationally representative sample surveys of U.S. home health and hospice agencies. It is designed to provide descriptive information on home health and hospice agencies, their staffs, their services, and their patients. NHHCS was first conducted in 1992 and was repeated in 1993, 1994, 1996, 1998, and 2000, and most recently in 2007. Only the most recent year of NHHCS was included in the CMS linkage. For more information on the NHHCS content and methods, refer to the NHHCS website, <http://www.cdc.gov/nchs/nhhcs/index.htm> (accessed September 28, 2017).

NNHS provides information on nursing homes from two perspectives- that of the provider of services and that of the recipient of care. Data for the surveys were obtained through personal interviews with facility administrators and designated staff who used administrative records to answer questions about the facilities, staff, services and programs, and medical records to answer questions about the residents. NNHS was first conducted in

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

1973-1974 and repeated in 1977, 1985, 1995, 1997, 1999, and most recently in 2004. Only the 2004 survey was included in the CMS linkage. For more information on the NNHS content and methods, refer to the NNHS website, <http://www.cdc.gov/nchs/nnhs.htm> (accessed September 28, 2017).

Centers for Medicare & Medicaid Services

Medicare is the primary health insurance program for people age 65 or older, people under age 65 with certain disabilities, and people of all ages with End-Stage Renal Disease (ESRD) - *permanent kidney failure requiring dialysis or a kidney transplant*. Nearly all Medicare beneficiaries receive Part A hospital insurance benefits, which helps cover inpatient hospital care, skilled nursing facility stays (not custodial or long-term care), home health and hospice care. Most beneficiaries also subscribe to Part B medical insurance benefits, which help to cover physician services, outpatient care, durable medical equipment and some home health care¹. In addition, starting in 2006, Medicare prescription drug coverage became available to all Medicare enrollees.

Medicare Enrollment and Summary Data

Master Beneficiary Summary File

For this linkage, Medicare claims information was extracted for the years 1999-2013. The structure of the data files provided by CMS has changed since the previous linkage. Most notably, the Denominator File has been replaced with the **Master Beneficiary Summary File (MBSF)**. The MBSF is an annual file containing demographic and enrollment information about beneficiaries enrolled in Medicare during each calendar year. It does not contain information on all beneficiaries *ever* entitled to Medicare, only those entitled during the calendar year. The MBSF is used to determine beneficiary demographic characteristics, entitlement, and beneficiary participation in a Medicare Advantage (MA)

¹ <https://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html?redirect=/MedicareGenInfo/>

Linked NCHS-CMS Medicare Data (1999-2013)

Linkage Methodology and Analytic Considerations

plan. The MBSF consists of several segments as noted below. Each segment is provided as a separate data file.

The **Base (A/B) segment** includes beneficiary enrollment information, such as date of birth, date of death, sex, race, age, geographic information, monthly entitlement indicators, reasons for entitlement (initial and current), and monthly Medicare Advantage indicators. The **Part D segment** includes variables specific to Medicare Part D Prescription Drug plan enrollment dating back to its inception in 2006. The **Cost & Utilization segment** includes summarized information about the service utilization and Medicare payment amounts by type of claims. The **Chronic Conditions segment** includes a summary of clinical information for Medicare beneficiaries who have been diagnosed with one of the following chronic health conditions or had one of the following events that may indicate the presence of a chronic health condition:

- Acquired Hypothyroidism
- Acute Myocardial Infarction
- Alzheimer's Disease
- Alzheimer's Disease, Related Disorders, or Senile Dementia
- Anemia
- Asthma
- Atrial Fibrillation
- Benign Prostatic Hyperplasia
- Cancer, Colorectal
- Cancer, Endometrial
- Cancer, Female/Male Breast
- Cancer, Lung
- Cancer, Prostate
- Cataract
- Chronic Kidney Disease
- Chronic Obstructive Pulmonary Disease & Bronchiectasis
- Depression

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

- Diabetes
- Glaucoma
- Heart Failure
- Hip / Pelvic Fracture
- Hyperlipidemia
- Hypertension
- Ischemic Heart Disease
- Osteoporosis
- Rheumatoid Arthritis / Osteoarthritis
- Stroke / Transient Ischemic Attack

Medicare Utilization and Claims Data

Medicare Provider Analysis and Review and Part D Prescription Drug Event Files

There are two Medicare utilization files, the **Medicare Provider Analysis and Review (MedPAR)** file that contains final action claim records for inpatient hospitalization and skilled nursing facility (SNF) and the **Part D Prescription Drug Event (PDE)** file that contains the utilization records for beneficiaries enrolled in the Part D Prescription Drug program.

All Medicare Part A short and long stay hospitalization claims and SNF claims for each calendar year are included on the MedPAR file. Each MedPAR record may represent a single claim or multiple claims, depending on the length of stay and the amount of services received during the stay. The file includes dates of service, ICD-9-CM diagnoses, ICD-9-CM procedures, and reimbursement amounts associated with each hospital or SNF stay. Inclusion of hospital stay claim records on the MedPAR file are based on year of discharge. SNF claims are based on year of admission into the facility. There can be multiple claims records per person on the MedPAR file.

Linked NCHS-CMS Medicare Data (1999-2013)

Linkage Methodology and Analytic Considerations

The Part D PDE file contains a summary of prescription drug costs and payment data used by CMS to administer benefits for Medicare Part D enrollees. The PDE file does not contain individual drug claims but are summary extracts submitted to CMS by Medicare Part D prescription drug plan providers. The Part D PDE files contain one record per event. There can be multiple records per person.

Outpatient, Home Health Agency, and Hospice Claims files

The **Outpatient** file contains Part B claims data submitted by institutional outpatient providers, such as hospital outpatient departments, rural health clinics, renal dialysis facilities, comprehensive outpatient rehabilitation facilities, community mental health centers, and ambulatory surgical centers for each calendar year. The **Home Health Agency** file contain claims data submitted by Home Health Agency (HHA) providers and include information on the number of visits, type of visit (skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), the dates of visits, reimbursement amount. The **Hospice** file contain claims data submitted by Hospice providers and include information on the level of hospice care received (e.g., routine home care, inpatient respite care), the dates of service, reimbursement amount, Hospice provider number, and beneficiary demographic information. Each Outpatient, Home Health Care and Hospice record is at the individual claim (or bill) level.

The data for the Outpatient, HHA, and Hospice files were all provided in a similar format. Each of the files are divided into six segments: 1) a base claim segments including demographic information, diagnosis codes, procedures codes, and dates of service; 2) a condition segment, identifying the claim-related condition; 3) an occurrence code segment, identifying a significant claim-related event and date that may affect processing of payment by CMS; 4) a span code segment, identifying a significant claim-related event and time period that may affect payment processing; 5) a value code segment including the billing and reimbursement amounts associated with a claim; and 6) a revenue code segment identifying the cost center or division/unit within a hospital in

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

which a charge is billed. Each segment is available as a separate file but can be combined using a unique claim identification number.

Carrier and Durable Medical Equipment Claims files

The **Carrier** file (formerly known as the Physician/Supplier Part B file) contain claims data submitted by non-institutional providers, such as physicians, physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, and stand-alone ambulatory surgical centers as well as durable medical equipment (DME) claims processed by carriers who also process physician claims. A separate **Durable Medical Equipment** (DME) file containing claims processed by authorized DME suppliers is also provided. DME claim records can contain claims for medical equipment such as oxygen, walkers, and wheelchairs. Information contained in the DME file includes diagnosis codes, description of equipment, dates of service, and reimbursement amount. DME claims on the Carrier file and the DME file are for separate services and are not duplicates.

The Carrier and DME files share similar formats. Each file consists of a base claims segment, containing demographic information and diagnosis codes as well as billing and payment amounts associated with a non-institutionalized claim; and a line items segment; that includes the specific billing and payment amounts for each line item included within the base claim. The base claim and line item segments are available as separate files but can be combined using a unique claim identification number.

Researchers should refer to the data documentation for more information on each file <https://www.cdc.gov/nchs/data-linkage/medicare-restricted.htm> (accessed September 28, 2017). In addition, researchers are encouraged to refer to the Research Data Assistance Center (ResDAC) <http://www.resdac.org/> (accessed September 28, 2017); a CMS contractor providing free assistance to researchers interested in using Medicare data for their research.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Additional Related Data Sources

NCHS has also linked to CMS Medicaid enrollment and claims data. Linkage of the NCHS survey participants with the CMS Medicaid data provides the opportunity to study changes in health status, health care utilization and expenditures in low income families with children, the elderly and disabled U.S. populations. For more information about the linked CMS Medicaid data, please see the data linkage website:

<https://www.cdc.gov/nchs/data-linkage/medicaid.htm> (accessed September 28, 2017).

In addition, NCHS has linked to a separate set of data files containing information on patients diagnosed with ESRD obtained from the United States Renal Data System (USRDS) <https://www.usrds.org/> (accessed September 28, 2017). The USRDS is a national data system, funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), designed to collect, analyze, and distribute information about ESRD in the United States. The linked ESRD data files can be used by researchers interested in conducting analysis specifically related to patients with ESRD. Although nearly all of the NCHS survey participants who were linked to USRDS records also linked to Medicare records, a small number of the linked USRDS records are not linked to Medicare records (less than 5%). For more information about the data available on the linked ESRD files, please refer to the documentation: <https://www.cdc.gov/nchs/data-linkage/esrd.htm> (accessed September 28, 2017).

Linkage of NCHS Surveys with 1999-2013 Medicare Records²

Linkage Eligibility

The linkage of NCHS survey participants to their Medicare enrollment and claims data was conducted under an interagency agreement between NCHS and CMS. The linkage was performed by NCHS in the CMS Virtual Research Data Center (VRDC) and is not the responsibility of researchers using the data. Approval for the linkage was provided

² Depending on the NCHS survey, Medicare claims data from 1991-1998 may be available for some participants. Please contact the NCHS Data Linkage Team (datalinkage@cdc.gov) for more information.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

by NCHS's Research Ethics Review Board (ERB)³ and the linkage was performed only for eligible NCHS survey participants. Only NCHS survey participants who have provided consent as well as the necessary personally identifiable information (PII), such as date of birth and full or partial SSN or Medicare Health Insurance Claim (HIC) number, are considered linkage-eligible. Linkage-eligibility refers to the potential ability to link data from an NCHS survey participant to 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 NCHS-CMS Medicare linkage eligibility vary by survey and year.

For many of the surveys initiated prior to and during 2007, including 1994-2006 NHIS, LSOA II, 1999-2008 NHANES, NHANES III, NHEFS, 2007 NHHCS, and 2004 NNHS, a refusal by the survey participant to provide a SSN or HIC number was considered an implicit refusal for data linkage. However, NCHS began to notice an increase in the refusal rate for providing SSN and HIC, particularly for NHIS, which reduced the number of survey participants eligible for linkage (1). 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 using partial SSN and other PII (2). 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 and HIC numbers for linkage (3). Beginning in 2007, NHIS started requesting only the last four (instead of the full nine) digits of SSN and HIC numbers. In addition, a short introduction before asking for SSN was added and participants who declined to provide SSN or HIC were asked for their explicit permission to link to administrative records without SSN or HIC. Also, at this time, the NCHS ERB determined that for 2007 NHIS and all subsequent years, only primary respondents

³ The NCHS Research Ethics Review Board (ERB), also known as an Institutional Review Board or IRB, is an administrative body of scientists and non-scientists that is established to protect the rights and welfare of human research subjects.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

(sample adult and sample children) would be eligible for linkage to administrative records.

The informed consent procedures changed for NHANES as well. NHANES continues to collect full nine-digit SSN and HIC numbers. 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.

Child Survey Participants

NCHS survey participants under 18 years of age 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. Since the majority of Medicare beneficiaries are age 65 and older, the ERB guidance pertains to less than 1% of the linked survey participants in the Medicare linkage.

CMS Virtual Research Data Center

The linkage of NCHS survey data with Medicare administrative data was performed by authorized NCHS staff within the CMS VRDC. The CMS VRDC is a virtual research environment that allows approved researchers to access Medicare and Medicaid program data from their own personal workstations. VRDC users are granted direct access to

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

approved CMS data files and are able to conduct analyses for research projects within a secure environment. VRDC users have the ability to download aggregated reports and results to their personal workstations, following disclosure review. More information about the CMS VRDC can be found at <https://www.resdac.org/cms-virtual-research-data-center-vrdc> (accessed October 1, 2019).

Linkage Methods

For the linkage of NCHS surveys with CMS Medicare administrative data, a primarily deterministic or rules-based approach was employed. The New York State Identification Intelligence System (NYSIIS) and Soundex, phonetic algorithms were applied for the comparison of name variables to account for typographical errors and spelling variations. NCHS matched a submission file (noted below) to the Medicare Enrollment Database (EDB) in the CMS VRDC. The submission file contained the following data elements for linkage-eligible NCHS survey participants:

- First name
- Last name
- SSN (9-digit or 4-digit)
- HIC number (complete or partial) - *if available*
- Sex
- Month of birth
- Day of birth
- Year of birth

The file used for linkage did not contain the NCHS survey public-use identification number, nor did it contain any information that could identify the original survey source. The public-use identification number was replaced with an encrypted linkage identification number used by NCHS staff for data linkage projects.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Linkage Algorithms using Four-Digit SSN and Partial HIC Number

For 2007-2013 NHIS, where only 4-digit SSN (SSN4) and partial HIC number (HIC4) were 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 or HIC4
- 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:

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

Linkage Algorithms using Nine-Digit SSN and Complete HIC Number

For all remaining surveys, where a 9-digit SSN (SSN9) and/or complete HIC number (HIC9) were collected the linkage process was completed in multiple passes.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Pass 1: Similar to the approach used in previous linkages of NCHS surveys with Medicare data, to be considered a successful match, agreement was required between the NCHS survey participant's record and the Medicare EDB record on each of the following items:

- SSN9 or HIC9
- Month of birth
- Year of birth
- Sex

For surveys where a 9-digit SSN (SSN9) and/ or complete HIC number (HIC9) were collected, approximately 98% of the matches were picked up in Pass 1.

Pass 2: For survey participants who were not matched in Pass 1, first name and last name were incorporated into the matching algorithm for a second pass. To be considered a successful match, agreement was required between the NCHS survey participant's record and the Medicare EDB record on SSN9 and at least three of the following items:

- Month of birth
- Year of birth
- Sex
- Soundex or NYSIIS agreement of non-missing first names
Soundex or NYSIIS agreement of non-missing last names

For cases that met only two of the conditions listed for Pass 2, clerical review of potential matched pairs was conducted. This accounted for less than 1% of the overall matched records.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Finally, the 4-digit SSN and partial HIC number algorithm described above was applied to the SSN9 records that did not match in passes 1 or 2 of the nine-digit algorithm. This contributed to <0.5% of the overall matches.

Upon completion of the linkage, a file containing the encrypted NCHS identification number and Medicare beneficiary identification number for successfully matched survey participants was provided to CMS VRDC staff. Medicare enrollment and claims data was extracted for successful matched NCHS survey participants and encrypted data files were shipped to NCHS, where additional quality control checks were performed.

Linkage Rates

The linkage rates for NCHS-CMS Medicare linkage are provided at https://www.cdc.gov/nchs/data/datalinkage/nchs-cms_medicare_linked_data_match_rate_tables.pdf (accessed February 7, 2023). The tables show for each survey, the total survey sample size, the sample size eligible for 1999-2013 Medicare linkage, the number of eligible survey participants linked to Medicare enrollment 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. Medicare has age-based entitlement at age 65. Therefore, the linkage rates for each survey were examined overall and by two age groups – less than 65 years and 65 years and older. Age was defined as the survey participant's age at interview. It should be noted that survey participants who died in the time frame between the survey and the linkage interval were eligible for linkage if they met linkage-eligibility criteria.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Data Confidentiality

The NCHS must provide safeguards for the confidentiality of its survey participants. To ensure confidentiality, all personal identifiers have been removed from the linked NCHS-CMS Medicare data files. However, there remains the small possibility of re-identification and for this reason; the linked NCHS-CMS Medicare data are not available as public-use files. NCHS has provided public-use Feasibility Data Files that include a limited set of variables for researchers to use in determining the feasibility and estimating sample sizes of their proposed research projects. The files can be accessed at <https://www.cdc.gov/nchs/data-linkage/medicare-feasibility.htm> (accessed September 28, 2017).

Researchers who want to obtain the linked NCHS-CMS Medicare data must submit a research proposal to the NCHS Research Data Center (RDC) to obtain permission to access the restricted use files. <https://www.cdc.gov/rdc/index.htm> (accessed September 28, 2017).

Analytic Considerations when using the Linked NCHS-CMS Medicare Files

General Notices to Users

This section summarizes some key analytic issues for users of the linked NCHS survey data and CMS administrative records. It is not an exhaustive list of the analytic issues that researchers may encounter while using the Linked NCHS-CMS Medicare 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). Users of the NCHS-CMS Medicare linked data are encouraged to visit the ResDAC website <http://www.resdac.org/> (accessed September 28, 2017) for more information on Medicare data.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

General Analytic Guidance

Merging NCHS-CMS Medicare Linked Data with NCHS survey data

The NCHS-CMS Medicare Linked Data Files can only be accessed in a RDC. Within the RDC, the NCHS-CMS Medicare Linked 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 identification numbers.

Variables to request in RDC applications

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

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

- Geography— Geography information 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 Medicare data have also been or will soon be linked to death information obtained from the National Death Index (NDI). The linked NDI mortality files provide date and cause of death for each survey participant who

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

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 at <https://www.cdc.gov/nchs/data-linkage/mortality.htm> (accessed September 28, 2017).

- NHANES month and year of examination and interview—The NHANES surveys are released in 2-year cycles. The exact year (and month) of a survey participant's interview and examination is not provided on public-use files. However, many researchers will want to know the time elapsed between a given year (or even month) of the Medicare data and the NHANES interview or examination. The variables that indicate the month and year of NHANES interview or examination must be requested specifically.

It is recommended that researchers request the following variables, available from the public-use NCHS survey files, for inclusion in analytic files:

- Sample weights and design variables—these variables are needed to account for the complex design of the NCHS surveys. The names of the weights and design variables differ depending on which NCHS survey is being used. These can be identified using the documentation for each NCHS survey. As discussed below, NCHS recommends adjusting the sample weights to account for linkage eligibility bias.
- 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 Medicare files. Therefore, when possible, the NCHS data should be used for demographic variables.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Sample weights

The sample weights provided in NCHS population health survey data files adjust for oversampling of specific subgroups and differential nonresponse and are post-stratified to annual population totals for specific population domains to provide nationally representative estimates. The properties of these weights for linked data files with incomplete linkage, due to ineligibility for linkage and non-matches, are unknown. In addition, methods for using the survey weights for some longitudinal analyses require further research. Because this is an important and complex methodological topic, ongoing work is being done at NCHS and elsewhere to examine the use of survey weights for linked data in multiple ways.

One approach is to analyze linked data files using adjusted sample weights. The sample weights available on NCHS population health survey data files can be adjusted for incomplete linkage (nonresponse), using standard weighting domains to reproduce population counts within these domains: sex, age, and race and ethnicity subgroups. These counts are called “control totals” and are estimated from the full survey sample.

A model-based calibration approach developed within the SUDAAN software package (Procedure WTADJUST or WTADJX) allows auxiliary information to be used to adjust the sample weights for nonresponse. This approach is recommended for adjusting sample weights for the linked files. Because inferences may depend on the approach used to develop weights, within SUDAAN’s WTADJUST or using a different calibration approach, researchers should seek assistance from a statistician for guidance on their particular project. Other approaches or software can be used. NCHS continues to investigate alternate approaches for addressing issues related to missing data, including the use of multiple imputation techniques. More detailed information on adjusting sample weights for linkage eligibility using SUDAAN can be found in Appendix III of *Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services* (http://www.cdc.gov/nchs/data/series/sr_01/sr01_058.pdf). (accessed September 28, 2017).

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Temporal alignment of survey and administrative data

Each NCHS survey has been linked to multiple years of Medicare enrollment and claims data. Depending on the survey year, Medicare 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.

Analysis Using Linked NCHS-CMS Medicare Data

NCHS-CMS Medicare Linked Data Feasibility Files

To maximize the use of the restricted-use linked NCHS-CMS Medicare files, NCHS has provided publicly available Feasibility Files that include a limited set of variables for researchers to use in determining the feasibility and sample sizes of their proposed research projects. The Feasibility Files include:

- (1) A public-use survey participant identification code so that users can merge variables from NCHS public-use survey data to the Feasibility File.
- (2) An indicator (CMS_MEDICARE_MATCH) of whether the NCHS participant was eligible for the matching and whether he/she linked to the Medicare EDB. CMS_MEDICARE_MATCH contains values 1, 2, 3, or 9.
 - a. A “1” indicates that the participant is linked; “2” indicates that the participant is not linked; “3” indicates a child survey participant with partial administrative data available; and 9 indicates that the participant was ineligible for linkage.
 - b. NCHS survey participants were considered ineligible for matching if they are missing key identification data and/or if they refused to provide their SSN or Medicare HIC number at the time of the survey interview (for all surveys) or did not agree to linkage without SSN or HIC number (for 2007-2013 NHIS) or did not provide an affirmative response to the linkage consent question (for 2009-2012 NHANES). Additional ineligibility criteria included refused, missing, or incomplete information on SSN, HIC number, first name, last name and date of birth.

Linked NCHS-CMS Medicare Data (1999-2013)
Linkage Methodology and Analytic Considerations

- c. Ineligible participants should be excluded from all analyses using the linked CMS data.
- (3) Information indicating if a survey participant has a linked data record on any of the Medicare administrative record files for any of the years of Medicare benefit coverage.

Documentation for the NCHS-CMS Medicare Feasibility Files can be found at <https://www.cdc.gov/nchs/data-linkage/medicare-feasibility.htm> (accessed September 28, 2017).

Medicare Advantage

CMS generally does not receive claims data for Medicare beneficiaries who enroll in Medicare Advantage (including private fee-for-service plans paid on a capitation basis). Medicare Advantage plans are also referred to as Medicare Part C and include Health Maintenance Organizations (HMOs), Preferred Provider Organizations (PPOs), Private Fee-for-Service (PFFS) Plans, Special Needs Plans, and Medicare Medical Savings Account Plans. During the time covered by the linked data files, Medicare Advantage enrollment increased from approximately 18% of beneficiaries in 1999 to 28% in 2013. A summary of the percent of NCHS survey participants who were enrolled in a Medicare Advantage plan by year and survey can be found at https://www.cdc.gov/nchs/data/datalinkage/nchs-cms_medicare_linked_data_medicare_advantage_enrollment_tables.pdf (accessed May 24, 2021).

Researchers should consider the percent of participants enrolled in a Medicare Part C program when determining the feasibility and sample sizes of their proposed research projects.

Medicare Advantage enrollment can be identified using the HMO indicators from MBSF – Part A/B Segment. The file includes 12 HMO indicator variables

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

(BENE_HMO_IND_01- BENE_HMO_IND_12), one for each month. During periods of Medicare Advantage enrollment, beneficiaries do not generate claims when using Medicare-covered services, except for selected services. Enrollees in cost-based plans may also generate some claims for inpatient hospital services. Utilization of most Medicare-covered services is unobservable from Medicare claims data during periods of Medicare Advantage enrollment. Therefore, in general, studies based on analysis of claims data should exclude Medicare Advantage enrollees from their beneficiary samples.

For more information on how to create an analytic sample that excludes Medicare beneficiaries enrolled in a Medicare Advantage plan, refer to a document written by ResDAC <https://www.resdac.org/articles/identifying-medicare-managed-care-beneficiaries-master-beneficiary-summary-or-denominator> (accessed October 1, 2019) or contact ResDAC, which provides free consultation for researchers using Medicare files, www.resdac.org (accessed September 28, 2017).

Importance of the Master Beneficiary Summary File

All applications to the NCHS RDC should include a request for the MBSF for the years that the researcher is examining claims data. As noted above, the MBSF contains basic demographic and enrollment information about each beneficiary enrolled in Medicare during each calendar year. This information is needed to construct an analytic data file, particularly to identify Medicare beneficiaries enrolled in a Medicare Advantage plans.

Services not covered (1999-2013)

Although Medicare provides coverage for a wide range of services, there are health care services not covered by Medicare. Examples of services not covered include routine physical exams, long-term care, and some cancer screening procedures. These gaps in coverage mean that there are no claims records for these services or for certain time periods. You may find more information on what is not currently covered by Medicare in

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

the *Medicare and You Handbook* at <https://www.medicare.gov/medicare-and-you/medicare-and-you.html> (accessed September 28, 2017).

In addition, Medicare data contains little information on prescription drugs for years prior to 2006. However, beginning in 2006 prescription drug coverage for Medicare beneficiaries became available through the Medicare Part D program. Prescription drug information paid by Medicare for 2006-2013 is available on the Part D Prescription Drug Event (PDE) File. Prescription drug information for data years 1999-2005 includes:

- Medication given in an inpatient/hospice/SNF setting - although specific medicines dispensed are rarely coded, if at all.
- Chemotherapy administered intravenously (IV), chemotherapy administered orally as a substitute for a medication that could be administered IV, and oral chemotherapeutic agents that break down to a compound comparable to a chemotherapeutic agent administered IV.

Medicare does not pay for chemotherapeutic agents that are administered exclusively in an oral form (e.g., Tamoxifen) and prior to 2006 most outpatient prescription drugs were not covered by Medicare.

Cost Sharing

Medicare beneficiaries often have a number of cost sharing requirements (i.e. deductibles and coinsurance). Although claims are generated for services where beneficiary cost sharing is involved, the Medicare payment amount does not necessarily represent the full cost to the beneficiary for the service. It is not possible to determine whether the beneficiary paid the cost-sharing amount “out-of-pocket” or whether the cost-sharing was paid by a third party, such as Medi-gap. Therefore, the total amount spent may not be captured by relying on the claims data alone.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Gaps and Discrepancies in Coverage Periods

Medicare enrollment and claims data linked to NCHS data are available for the years 1999-2013. Several of the surveys linked to the Medicare data, such as NHEFS (1971-1992), and NHANES III (1988-1994)⁴ have gaps of several years between the end of the study period and the beginning of the Medicare data.

Researchers should be aware that there may be differences in the availability of Medicare data on the linked files depending the survey participant's age and the year the survey was administered. Earlier Medicare data is only available for NCHS surveys that were included in the previous CMS linkage. For example, an NHIS participant who is 80 years old in 1999 at their interview and matched at some point to the 1999-2013 Medicare EDB would only have Medicare data for 1999 to 2013. A similar NHIS participant, who was 80 years old at their interview in 1998 and linked to the 1999-2013 Medicare EDB may have also linked in the previous Medicare linkage. This participant would hypothetically have Medicare data for 1991-1998 (available in a different format). This issue is particularly important for researchers to consider when combining data across survey and Medicare coverage years. Researchers need to determine how to address these discrepancies in coverage periods in their analyses.

On Master Beneficiary Summary File with no claims data

There may be instances where an NCHS survey participant is on the MBSF, but no claims data are available. It is possible to be enrolled in Medicare but not utilize Medicare services during the coverage period. In addition, there may be some record keeping inconsistencies because CMS data are collected for administrative, not research purposes.

⁴ For some surveys, including NHEFS and NHANES III, 1991-1998 linked NCHS-CMS Medicare data are available from a previous linkage. Please contact the NCHS Data Linkage Team (datalinkage@cdc.gov) for more information.

Linked NCHS-CMS Medicare Data (1999-2013)

Linkage Methodology and Analytic Considerations

Medicare entitlement variables

The MBSF includes three variables indicating Medicare entitlement: original reason for entitlement, current reason for entitlement, and Medicare status code.

A beneficiary's *original reason* for Medicare entitlement is found in the variable BENE_ENTLMT_RSN_ORIG. This variable is coded by CMS using information provided by the SSA and/or Railroad Retirement Board. Knowing a beneficiary's original reason for entitlement can be useful for identifying which aged beneficiaries were formerly Medicare disabled, since their cost and utilization profiles tend differ from other aged beneficiaries, especially at ages 65-74. BENE_ENTLMT_RSN_ORIG values include: Old Age and Survivors Insurance (OASI), Disability Insurance Benefits (DIB) and ESRD.

A beneficiary's *current reason* for Medicare entitlement is found in the variable BENE_ENTLMT_RSN_CURR. Possible values include: OASI, DIB and ESRD. This variable is populated from the Medicare EDB. The EDB is a master enrollment file of all people ever entitled to Medicare. Many of the variables on the Master Beneficiary Summary File are extracted from the EDB. The EDB is not available to researchers.

The variable BENE_MDCR_STATUS_CD specifies the most *recent status* of the beneficiary's entitlement to Medicare benefits. Medicare status code is a CMS coded variable that is created from the following variables available on the EDB: Age, original reason for entitlement, current reason for entitlement, and an indicator of ESRD. Possible values include: Aged without ESRD, Aged with ESRD, Disabled without ESRD, Disabled with ESRD, and ESRD only.

Medicare's Prospective Payment System (PPS)

Medicare's PPS refers to a method of reimbursement where the Medicare payment is made based upon a predetermined, fixed amount. Medicare uses a separate PPS for several services, where the particular payment amount is derived based upon the

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

classification system for that particular service (e.g., diagnosis-related groups for inpatient hospital services). Separate PPS's are used for reimbursement to acute inpatient hospitals, home health agencies, hospice, hospital outpatient, inpatient psychiatric facilities, inpatient rehabilitation facilities, long-term care hospitals, and skilled nursing facilities. PPS claims can be identified on the MedPAR file using the PPS claims indicator (PPS_IND_CD). For HHA and Outpatient claims, the variable CLM_PPS_IND_CD can be used.

For more information on the PPS, please visit <https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/ProsperMedicareFeeSvcPmtGen/index.html> (accessed September 28, 2017).

Medicare Provider Analysis and Review File

The MedPAR file includes all hospitalizations that had a discharge date during the calendar year and all SNF stays with an admission date during the calendar year. Hospital stays starting in one calendar year and continuing past the end of the calendar year are not provided on the MedPAR file until the year of discharge. To determine if a record is for a long- or short-stay hospitalization, use the short stay/long/stay/SNF indicator variable SS_LS_SNF_IND_CD.

Each MedPAR record represents a stay in an inpatient "acute stay" or "long stay" hospital. An inpatient stay record summarizes all services rendered to a beneficiary from the time of admission to a facility through discharge. Each MedPAR record may represent one claim or multiple claims, depending on the length of a beneficiary's stay and the amount of inpatient services used throughout the stay.

The following fields on MedPAR files are not used for payment purposes and should be used with caution:

Source of inpatient admission (SRC_IP_ADMSN_CD)

Group health organization payment code (GHO_PD_CD)

In addition, mortality information is provided on the MBSF and MedPAR file. However, if the outcome of interest is mortality, researchers should use the NCHS Linked Mortality Files (<https://www.cdc.gov/nchs/data-linkage/mortality.htm>) (accessed September 28, 2017) for the most recent mortality information.

Linked NCHS-CMS Medicare Data (1999-2013)
Linkage Methodology and Analytic Considerations

References

1. 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.
2. 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.
3. 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.

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Acknowledgments

Information about the Medicare enrollment and claims files was compiled from the following sources:

Centers for Medicare & Medicaid Services (CMS)
<http://www.cms.gov/> (accessed September 28, 2017))

Medicare and You Handbook
<https://www.medicare.gov/medicare-and-you/medicare-and-you.html> (accessed September 28, 2017)

Research Data Assistance Center (ResDAC)
<http://www.resdac.org/> (accessed September 28, 2017)

American Association of Retired Persons (AARP) Medicare Resource Center
<http://www.aarp.org/health/medicare-insurance> (accessed September 28, 2017)

Additional Resources

NHANES–CMS Linked Data Tutorial
<https://www.cdc.gov/nchs/tutorials/NHANES-CMS/index.htm> (accessed September 28, 2017)

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

Appendix A: Merging Linked NCHS-CMS Medicare Files with NCHS Survey Data

The data provided on the 1994-2013 NHIS, NHEFS, 1999-2012 NHANES, NHANES III, LSOA II, 2004 NNHS, and the 2007 NHHCS linked CMS Medicare files can be merged with the NCHS restricted and public use survey data files using the unique survey specific Public Identification number (PUBLICID/SEQN/RESNUM/PATNUM).

Note: At this time the linked CMS Medicare data files are only available for research use through the NCHS restricted access data center (RDC). Approved RDC researchers may choose to provide their own analytic files created from public use survey files to the RDC. Therefore, it is important for researchers to include survey specific Public Identification number on any analytic files sent to the RDC. The RDC will merge data (using PUBLICID, SEQN, RESNUM or PATNUM) from the linked CMS Medicare files to the analyst's file. The merged file will be held at the RDC and made available for analysis.

Information on how to identify and/or construct the NCHS survey specific PUBLICID, SEQN, RESNUM or PATNUM is provided below.

I. National Health Interview Survey (NHIS)

1994 NHIS

<u>Variable</u>	<u>Public-use Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
QUARTER	5	1	Calendar quarter of interview
PSUNUMR	6-8	3	Random recode of PSU
WEEKCEN	9-10	2	Week of interview within quarter
SEGNUM	11-12	2	Segment number
HHNUM	13-14	2	Household number within quarter
PNUM	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID =  
trim(left(YEAR||QUARTER||PSUNUMR||WEEKCEN||SEGNUM||HHNUM||PNUM));
```

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

```
egen PUBLICID = concat(YEAR QUARTER PSUNUMR WEEKCEN SEGNUM  
HHNUM PNUM)
```

Linked NCHS-CMS Medicare Data (1999-2013)
Linkage Methodology and Analytic Considerations

1995-1996 NHIS

<u>Variable</u>	Public-use <u>Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
HHID	5-14	10	Household ID number
PNUM	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID = trim(left(YEAR||HHID||PNUM));
```

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

```
egen PUBLICID = concat(YEAR HHID PNUM)
```

1997-2003 NHIS

<u>Variable</u>	Public-use <u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household number
FMX	13-14	2	Family number
PX	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

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

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

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

*The person identifier was called PX in the 1997-2003 NHIS and FPX in the 2004 (and later) NHIS; users may find it necessary to create an FPX variable in the 2003 and earlier datasets (or PX in later datasets).

Linked NCHS-CMS Medicare Data (1999-2013) Linkage Methodology and Analytic Considerations

2004 NHIS

<u>Variable</u>	Public-use		<u>Description</u>
	<u>Location</u>	<u>Length</u>	
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household number
FMX	13-14	2	Family number
FPX	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

`length publicid $14;`

`PUBLICID = trim(left(SRVY_YR||HHX||FMX||FPX));`

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

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

2005-2013 NHIS

<u>Variable</u>	Public-use		<u>Description</u>
	<u>Location</u>	<u>Length</u>	
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household number
FMX	16-17	2	Family number
FPX	18-19	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

`length publicid $14;`

`PUBLICID = trim(left(SRVY_YR||HHX||FMX||FPX));`

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

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

Linked NCHS-CMS Medicare Data (1999-2013)
Linkage Methodology and Analytic Considerations

II. National Health and Nutrition Examination Survey I Epidemiologic Follow-up Study (NHEFS)

<u>Item</u>	<u>Length</u>	<u>Description</u>
SEQN	5	Participant identification number

All of the NHEFS public-use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHEFS Files to the NHEFS- CMS Medicare linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

III. 1999-2012 National Health and Nutrition Examination Survey (NHANES)

<u>Item</u>	<u>Length</u>	<u>Description</u>
SEQN	5	Participant identification number

All of the NHANES public-use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHANES Files to the NHANES-CMS Medicare linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

IV. Third National Health and Nutrition Examination Survey (NHANES III)

<u>Item</u>	<u>Length</u>	<u>Description</u>
SEQN	5	Participant identification number

All of the NHANES III public-use data files are linked with the common survey participant identification number (SEQN). Merging information from multiple NHANES III Files to the NHANES III-CMS Medicare linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

Linked NCHS-CMS Medicare Data (1999-2013)
Linkage Methodology and Analytic Considerations

V. The Second Longitudinal Study of Aging (LSOA II)

On the LSOA II survey, researchers need to construct the LSOA II public id from the following variables.

LSOA II

<u>Variable</u>	Public-use <u>Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
QUARTER	5	1	Calendar quarter of interview
PSUNUMR	6-8	3	Random recode of PSU #
WEEKCEN	9-10	2	Week of interview within quarter
SEGNUM	11-12	2	Segment number
HHNUM	13-14	2	Household number within quarter
PNUM	15-16	2	Person number within household

Note: Concatenate all variables to get the unique person identifier.

SAS example:

```
length publicid $14;  
PUBLICID =  
trim(left(YEAR||QUARTER||PSUNUMR||WEEKCEN||SEGNUM||HHNUM||PNUM));
```

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

```
egen PUBLICID = concat(YEAR QUARTER PSUNUMR WEEKCEN SEGNUM  
HHNUM PNUM)
```

VI. 2004 National Nursing Home Survey (NNHS)

<u>Item</u>	<u>Length</u>	<u>Description</u>
RESNUM	6	Resident Record (Case) Number

All of the 2004 NNHS public-use data files are linked with the common resident record (case) number (RESNUM). Merging information from the 2004 NNHS Files to the 2004 NNHS-CMS Medicare linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.

Linked NCHS-CMS Medicare Data (1999-2013)
Linkage Methodology and Analytic Considerations

VII. 2007 National Home and Hospice Care Survey (NHHCS)

<u>Item</u>	<u>Length</u>	<u>Description</u>
PATNUM	6	Patient/Discharge Record (Case) Number

All of the 2007 NHHCS public-use data files are linked with the common patient/discharge record (case) number (PATNUM). Merging information from the 2007 NHHCS Files to the 2007 NHHCS-CMS Medicare linked files using this variable ensures that the appropriate information for each survey participant is linked correctly.