

NCHS-CMS MEDICARE AND NCHS-USRDS LINKED DATA

NCHS has linked NCHS population based survey data with Medicare enrollment and claims data collected from the Centers for Medicare and Medicaid Services (CMS) and the United States Renal Data System (USRDS). These linked survey files provide the data needed to formulate and answer critical public health research questions regarding Medicare service use and health care costs and End Stage Renal Disease (ESRD).

Medicare enrollment and claims and ESRD data are available for NCHS respondents who agreed to provide personal identification data to NCHS and for whom NCHS was able to validate and match with Medicare and USRDS administrative records. Although nearly all NCHS respondents linked to USRDS records are also linked to Medicare records, a small number (less than 5%) of the linked USRDS records are not linked to Medicare records. As part of the Medicare linkage, ESRD data for the linked Medicare beneficiaries are available to researchers.

Important notes regarding analytic issues with the NCHS-CMS Medicare and USRDS linked data files can be found in [Medicare Enrollment and Claims Data Analytic Issues](#).

Medicare enrollment and claims data covering service from 1991-2007 are available for the following NCHS surveys:

- **1994-1998 National Health Interview Survey (NHIS)**
- **NHANES I Epidemiologic Follow-up Study (NHEFS)**
- **Third National Health and Nutrition Examination Survey (NHANES III)¹**
- **The Second Longitudinal Study of Aging (LSOA II)**

Medicare enrollment and claims data covering service from 1999-2007 are available for the following NCHS surveys:

- **1999-2005 National Health Interview Survey (NHIS)**
- **1999-2004 National Health and Nutrition Examination Survey (NHANES)**
- **2004 National Nursing Home Survey (NNHS)**

ESRD patient data covering service from 1974-2008 are available for the following NCHS surveys:

- **1994-2005 National Health Interview Survey (NHIS)**
- **1999-2004 National Health and Nutrition Examination Survey (NHANES)**
- **2004 National Nursing Home Survey (NNHS)**
- **NHANES I Epidemiologic Follow-up Study (NHEFS)**
- **Third National Health and Nutrition Examination Survey (NHANES III)¹**
- **The Second Longitudinal Study of Aging (LSOA II)**

¹ In addition to NHANES III, a subset of NHANES II sample was linked to data from CMS and USRDS. Inquiries about this linkage should be directed to the NCHS Data Linkage Team, datalinkage@cdc.gov.

The following Medicare enrollment and claims files are available for analytic use through the [NCHS Research Data Center \(RDC\)](#):

Medicare Enrollment File

Denominator File

The Denominator file provides data on all Medicare beneficiaries enrolled and/or entitled to Medicare benefits in a given year. Monthly information on the enrollment status of linked Medicare beneficiaries including third party payer information and Medicare Part C also referred to as Medicare Advantage (MA), Managed Care Organization (MCO), Group Health Organization (GHO) or Health Maintenance Organization (HMO) plan enrollment information is provided. The Denominator file is fixed length and contains one record per person.

Documentation for the [Denominator File](#) is available in PDF format.

Standard Analytical Files (SAF)

Carrier SAF

The Carrier SAF (formerly the Physician/Supplier Part B file) contains final action claims data submitted by non-institutional providers. The data are largely made up of physician claim records, although the file also includes claims from other non-institutional providers such as physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, ambulance providers, and stand-alone ambulatory surgical centers. Due to the large number of carrier claim variables, CMS provides the Carrier data in variable length files. There can be multiple carrier claims per person on a file. The Carrier files are provided in the CMS Standard Analytical File format.

Documentation for the [Carrier SAF](#) is available in PDF format

Outpatient SAF

The Outpatient SAF contains Medicare Part B final action claims from institutional outpatient providers for each calendar year. Hospital outpatient departments, rural health clinics, renal dialysis facilities, outpatient rehabilitation facilities, comprehensive outpatient rehabilitation facilities, and community mental health centers are examples of institutional outpatient providers. Same day surgeries performed in a hospital are also included in the Outpatient file. There can be multiple outpatient claims records per person on the Outpatient files. The Outpatient files are provided in the CMS Standard Analytical File format.

Documentation for the [Outpatient SAF](#) is available in PDF format.

Home Health Agency (HHA) SAF

The Home Health Agency SAF contains final action claims for home health services. Some of the information contained in this file includes the number of visits, type of visit (skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), diagnosis (10 ICD-9-CM diagnosis codes), dates of visits, reimbursement amount. An HHA claim may cover services provided over a period of time, not a single day. There can be multiple HHA claims records per person on the HHA files. The HHA files are provided in the CMS Standard Analytical File format.

Documentation for the [Home Health Agency SAF](#) is available in PDF format.

Hospice SAF

The Hospice SAF contains final action claims data submitted by Hospice providers. The data contained in this file include the type of hospice care received (e.g., routine home care, inpatient respite care), The Hospice file contains data fields for 10 ICD-9-CM diagnosis and 6 procedure codes, dates of service, reimbursement amount, and some demographic information (such as date of birth, race, and sex). There can be multiple Hospice claims records per person on the Hospice file. The Hospice files are provided in the CMS Standard Analytic File format.

Documentation for the [Hospice SAF](#) is available in PDF format.

Durable Medical Equipment (DME) SAF

The DME SAF contains final action claims data submitted by Durable Medical Equipment (DME) regional carriers. Some of the information contained in the DME file includes diagnosis (10 ICD-9-CM diagnosis codes), service type codes, dates of service, and reimbursement amount. There can be multiple DME claim records per person on the DME files. The DME files are provided in the CMS Standard Analytical File format.

Documentation for [Durable Medical Equipment](#) is available in PDF format.

Utilization Files

Medicare Provider Analysis and Review (MedPAR) File

The Medicare Provider Analysis and Review (MedPAR) file contains inpatient hospitalization and skilled nursing facility (SNF) final action claim records. 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 claim record includes up to 10 ICD-9-CM diagnoses and 6 ICD-9-CM procedures 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.

Documentation for the [MedPAR File](#) is available in PDF format.

Medicare Part D Data Files

The following Medicare Part D Data Files are available for all successfully matched NCHS survey participants for service years 2006 and 2007:

- Medicare Part D Denominator File
- Medicare Part D Prescription Drug Event (PDE) File

Medicare Part D Denominator File

The Medicare Part D Denominator File contains demographic and enrollment for each beneficiary enrolled in Medicare during the calendar year. In addition to the variables available on the standard Medicare Denominator file, the Part D Denominator File contains a derived race/ethnicity code, an indicator for Other Credible Drug Coverage, and monthly indicators for Medicare Advantage Prescription Drug Plans (MA-PD) and prescription drug plan (PDP) enrollment, Low Income Subsidy (LIS) enrollment, Retiree Drug Subsidy, and State Reported Dual Eligibility Status. The Medicare Part D Denominator file is fixed length and contains one record per person.

Medicare Part D Prescription Drug Event (PDE) File

The Part D Prescription Drug Event (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 contains one record per event. There can be multiple records per person.

Documentation for the [Medicare Part D Denominator](#) and [Part D PDE](#) Files are available in PDF format.

Summary Files

Chronic Condition (CC) Summary File

The Chronic Condition (CC) Summary File includes claims, enrollment, and assessment data for Medicare beneficiaries extracted based on the following set of 21 chronic health conditions:

- Acute Myocardial Infarction
- Alzheimer's Disease
- Alzheimer's Disease, Related Disorders, or Senile Dementia
- Atrial Fibrillation
- Cataract
- Chronic Kidney Disease
- Chronic Obstructive Pulmonary Disease
- Depression
- Diabetes
- Glaucoma
- Heart Failure
- Hip/Pelvic Fracture
- Ischemic Heart Disease
- Osteoporosis
- Rheumatoid arthritis / Osteoarthritis (RA/OA)
- Stroke / Transient Ischemic Attack
- Breast Cancer
- Colorectal Cancer
- Prostate Cancer
- Lung Cancer
- Endometrial Cancer

The CC Summary File provides a summary of clinical information, including date of first occurrence, yearly, and mid-year flags for each of the 21 chronic conditions. The file provides a summary of clinical information for each successfully matched NCHS survey participant. CC Summary files are available for each of the NCHS-CMS Medicare linked surveys listed above for 2005 through 2007.

Documentation for the [CC Summary File](#) is available in PDF format.

Summary Medicare Enrollment and Claims (SMEC) File

A Summary Medicare Enrollment and Claims file has been created by NCHS to assist researchers who are interested in analyzing Medicare cost and claims data from multiple Medicare service files. The SMEC file contains data on the beneficiary's reason for Medicare entitlement, total number of months of Medicare entitlement, Medicare Part C plan enrollment, and summarized Medicare service charges, total expenditures, and reimbursement amounts. These summarized (or summary) variables are modeled after the [Medicare Current Beneficiary Survey \(MCBS\)](#) cost and use files. In addition, summary variables related to the beneficiary's total number of Emergency Room (ER) visits and Part D prescription drug costs have been created.

A SMEC file is available for each year (1991-2007) of available linked Medicare data for the following NCHS surveys:

- **1994-1998 National Health Interview Survey (NHIS)**
- **NHANES I Epidemiologic Follow-up Study (NHEFS)**
- **Third National Health and Nutrition Examination Survey (NHANES III)**
- **The Second Longitudinal Study of Aging (LSOA II)**

A SMEC file is available for each year (1999-2007) of available linked Medicare data for the following NCHS surveys:

- **1999-2005 National Health Interview Survey (NHIS)**
- **1999-2004 National Health and Nutrition Examination Survey (NHANES)**
- **2004 National Nursing Home Survey (NNHS)**

Each SMEC file has a fixed length record format with one record for each survey respondent successfully linked to Medicare administrative record files.

Documentation for the [SMEC File](#) is available in PDF format.

End Stage Renal Disease Data Files

NCHS has linked to a separate set of data files containing information on patients diagnosed with End Stage Renal Disease (ESRD) obtained from the United States Renal Data System (USRDS). 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.

The ESRD data are divided into four files:

- **Combined ESRD Patient Profile and Death Notification (form 2746) File**
- **ESRD Medical Evidence Report (form 2728) File**
- **ESRD Treatment History File**
- **ESRD Payment History File**

Combined ESRD Patient Profile and Death Notification File

The Combined ESRD Patient Profile and Death Notification file contains general demographic information, primary and contributing causes of ESRD diagnosis, date of first dialysis, and information on kidney transplants. The Combined ESRD Patient Profile and Death Notification file also contains death information, if a Death Notification form (form 2746) was completed.

Each Combined ESRD Patient Profile and Death Notification record has a fixed length format with one record for each successfully linked NCHS survey respondent entered into the USRDS.

ESRD Medical Evidence Report File

The ESRD Medical Evidence Report file contains information obtained by the USRDS from completed ESRD Medical Evidence Report Medicare Entitlement and/or Registration forms (form 2728). The ESRD Medical Evidence Report file contains general demographic information, health insurance status, primary and contributing causes of ESRD diagnosis, values from ESRD-related test results, dialysis information, and kidney transplant status.

Each ESRD Medical Evidence Report record has a fixed length format and may contain up to 3 records for each successfully linked NCHS survey respondent entered into the USRDS.

ESRD Treatment History File

The ESRD Treatment History file contains information about ESRD related events, such as dialysis or transplants. The ESRD Treatment History files include the treatment start and end dates, and modality of ESRD treatment.

The ESRD Treatment History file has a fixed length record format and may contain up to 117 records for each successfully linked NCHS survey respondent entered into the USRDS.

ESRD Payment History File

The ESRD Payment History file contains information on the primary insurance payer for each ESRD treatment period. The file also includes information on Medicare and Medicaid eligibility status.

The ESRD Payment History file has a fixed length record format and may contain up to 14 records for each successfully linked NCHS survey respondent entered into the USRDS.

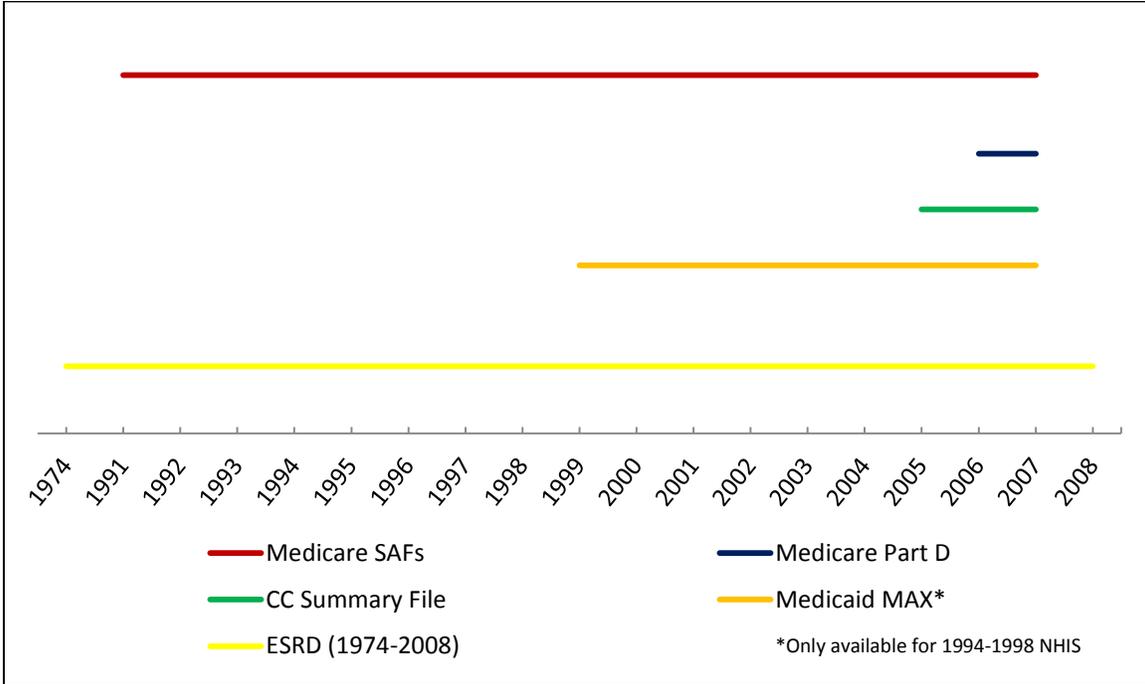
Documentation for the [NCHS-USRDS linked ESRD data files](#) is available in PDF format.

NCHS-CMS Medicare Feasibility Study Data Files

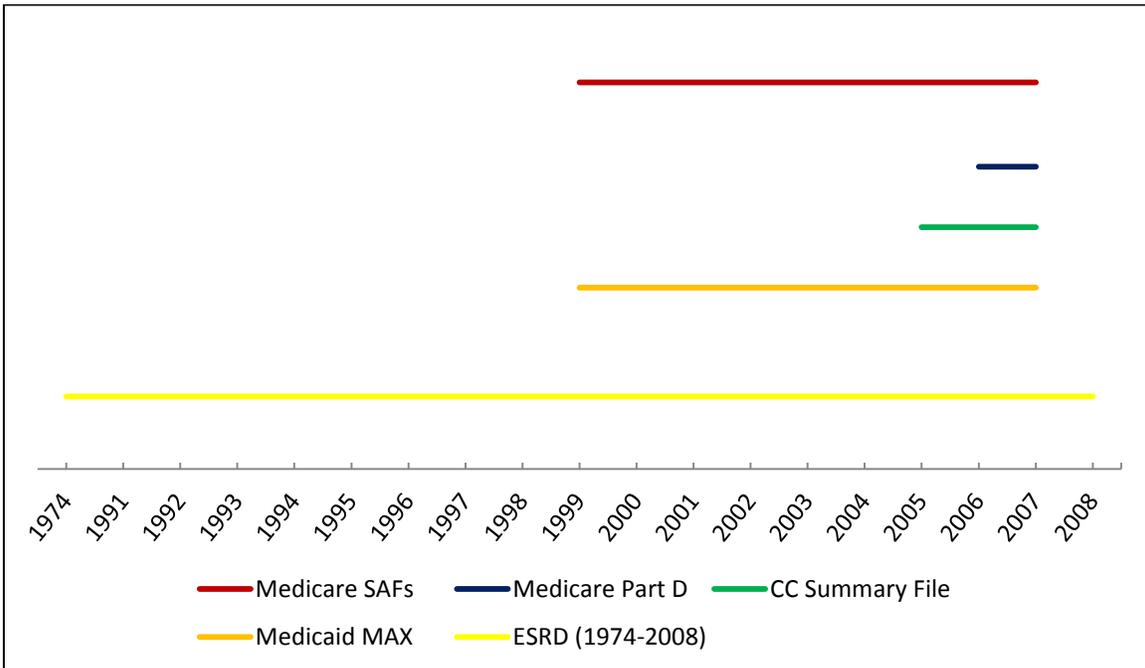
To maximize the use of the linked CMS Medicare data, NCHS has also created public use [NCHS-CMS Medicare Feasibility Study data files](#) that can be downloaded directly from the website. The feasibility files provide a limited set of variables that can be used to determine the maximum available sample size so that the feasibility of conducting analyses utilizing the linked CMS Medicare data can be assessed. Each feasibility file is NCHS survey specific and includes the following: NCHS public id, survey respondent eligibility status for CMS Medicare linkage, final match status, and flags (or indicators) specifying on which CMS Medicare data files the successfully linked survey respondent has a record. It is important to note that the CMS Medicare Feasibility Study files do not provide any information pertaining to benefits or payments; they simply contain record status variables and record counts to assist researchers who are considering whether or not to initiate a RDC proposal. Due to risk of disclosure, the CMS Medicare Feasibility Study files do not contain any variables for the linked ESRD data.

Table 1: Coverage period for NCHS-CMS and NCHS-USRDS Linked Data

Surveys: 1994-1998 NHIS, NHEFS, NHANES III, LSOA II



Surveys: 1999-2005 NHIS, 1999-2004 NHANES, 2004 NNHS



Important Information on Merging Restricted and Public Use NCHS Survey Data

The data provided on the restricted NCHS-linked files can be merged with the NCHS public use survey data files using unique survey person identification numbers. However, the identifying variables are different across surveys and years.

Note: The linked data files are only available 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 the correct survey person identification number.

The following table presents the variables needed to merge each public use survey data with restricted NCHS-linked data in the RDC.

NCHS Survey	Survey Person Identification Variable
National Health and Nutrition Examination Survey (NHANES)	SEQN
Third National Health and Nutrition Examination Survey (NHANES III)	SEQN
NHANES Epidemiologic Follow-up Study (NHEFS)	SEQN
National Health Interview Survey (NHIS)	PUBLICID*
Second Longitudinal Study of Aging (LSOA II)	PUBLICID*
National Nursing Home Survey (NNHS)	RESNUM

*PUBLICID must be created for NHIS and LSOA (see below). It is important to note the construction of PUBLICID varies by year in NHIS.

I. National Health Interview Survey (NHIS)

On the NHIS surveys, researchers need to construct the person identification number (PUBLICID) from the following variables. The number and public-use location varies by NHIS survey year.

NHIS 1994

<u>Variable</u>	<u>Public-use Location</u>	<u>Length</u>	<u>Description</u>
YEAR	3-4	2	Year of interview
WTFQ	5	1	Calendar quarter of interview
PSU	6-8	3	Random recode of PSU #
WEEKPROC	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 || WTFQ || PSU || WEEKPROC || SEGNUM || HHNUM || PNUM));
```

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

```
egen PUBLICID = concat(YEAR WTFQ PSU WEEKPROC SEGNUM HHNUM PNUM)
```

NHIS 1995, 1996

<u>Variable</u>	<u>Public-use 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 a string variable)

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

NHIS 1997-2003

<u>Variable</u>	<u>Location</u>	<u>Length</u>	<u>Description</u>
SRVY_YR	3-6	4	Year of interview
HHX	7-12	6	Household serial 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 || PX));
```

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

```
egen PUBLICID = concat(SRVY_YR HHX 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).

NHIS 2004, 2005

<u>Variable</u>	<u>Location</u>	<u>Length</u>	<u>Description</u>
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

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 a string variable)

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

II. NHANES I Epidemiologic Follow-up Study NHEFS

<u>Variable</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. National Health and Nutrition Examination Survey (NHANES)

<u>Variable</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>Variable</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.

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

On the LSOA II survey, researchers need to construct the LSOA II the person identification number (PUBLICID) from the following variables.

LSOA II

<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
PSU	6-8	3	Random recode of PSU #
WEEKPROC	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 || PSU || WEEKPROC || SEGNUM || HHNUM || PNUM));
```

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

```
egen PUBLICID = concat(YEAR QUARTER PSU WEEKPROC SEGNUM HHNUM PNUM)
```

VI. 2004 National Nursing Home Survey (NNHS)

<u>Variable</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.