Suggested citation: National Center for Health Statistics, Office of Analysis and Epidemiology. Linkages between Survey Data from the National Center for Health Statistics and Medicare Program Data from the Centers for Medicare and Medicaid Services. December 2012. Hyattsville, Maryland. (Available at the following address: http://www.cdc.gov/nchs/data/datalinkage/cms_medicare_methods_report_final.pdf)

Last Updated: December 3, 2012

Linkages between Survey Data from the National Center for Health Statistics and Medicare Program Data from the Centers for Medicare and Medicaid Services

Introduction

Under an interagency agreement including the National Center for Health Statistics (NCHS), the Centers for Medicare and Medicaid Services (CMS), the Social Security Administration (SSA), and the Office of the Assistant Secretary for Planning and Evaluation, DHHS (ASPE), several NCHS population-based surveys are linked to Medicare administrative records. The linkages are undertaken to support various research initiatives of the participating agencies. The NCHS-Medicare linked files combine health and socio-demographic information from the surveys with 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 provides researchers with an overview of the updated NCHS-Medicare linked data. This updated version of the NCHS-Medicare linkage includes linkage to Medicare records for additional survey years for some surveys that were not previously linked to Medicare files, linkage to Medicare records for the 2004 National Nursing Home Survey and linkage to additional years of Medicare records for surveys previously linked to Medicare records. The NCHS survey respondents whose records were linked in the prior linkage and were determined to be a successful match in the current linkage had CMS data extracted from each of the Medicare files for 1999-2007, and had their CMS records for the years 1991-1998 maintained from the previous linkage. The NCHS respondents whose records were linked for the first time to the Medicare claims data, had CMS data extracted from each of the Medicare files for 1999-2007.

This report describes the NCHS surveys and the Medicare administrative data files followed by a discussion of the linkage processes, linkage rates, and the linked data files. The linked NCHS-Medicare data files are large and complex. Researchers are advised to read the documentation and supporting tabular data in order to understand the complexity

of the data files. For confidentiality reasons, the linked data files are restricted and are only accessed through the Research Data Center (RDC). In order to access the data, researchers need to submit a proposal to the RDC that fully describes the data needed for an analysis.   Please see the RDC website, http://www.cdc.gov/rdc/ (accessed December 3, 2012).

Data Sources

National Center for Health Statistics

The following NCHS surveys were linked to Medicare data: the National Health Interview Survey (NHIS), the Second Longitudinal Study of Aging (LSOA II), the NHANES I Epidemiologic Follow-Up Study (NHEFS), the Second National Health and Nutrition Examination Survey (NHANES II), the Third National Health and Nutrition Examination Survey (NHANES III), the continuous National Health and Nutrition Examination Survey (NHANES) and the National Nursing Home Survey (NNHS).

The 1994-1998 **NHIS** data were linked to Medicare enrollment and claims data that covered services for the years 1991 to 2007. The 1999-2005 NHIS were linked to Medicare enrollment and claims data that covered services for the years 1999-2007. The **NHIS** has been conducted since 1957. In the current design, each year data are collected from approximately 35,000 households, including about 88,000 persons. The NHIS collects data on basic social and demographic items, health conditions and health behaviors, as well as health insurance, access to health care and utilization. In addition, the 1994 and 1995 NHIS included a supplement on disability. For detailed information on the NHIS's contents and methods and the NHIS Disability Survey, refer to the NHIS website, http://www.cdc.gov/nchs/nhis.htm (accessed December 3, 2012).

The **LSOA II** data were linked to Medicare enrollment and claims data that covered services for the years 1991-2007. The **LSOA II** is 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, http://www.cdc.gov/nchs/lsoa.htm (accessed December 3, 2012).

The **NHEFS** data were linked to Medicare enrollment and claims data that covered services for the years 1991-2007. **NHEFS** is a national longitudinal study that includes participants who were 25-74 years of age when first examined in NHANES I (1971-75), which served as the baseline for the longitudinal follow-up study. 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 on the NHEFS contents and methods, refer to the NHEFS website, http://www.cdc.gov/nchs/nhanes/nhefs/nhefs.htm (accessed December 3, 2012).

The **NHANES III**[1] data were linked to Medicare enrollment and claims data that covered services for the years 1991-2007. **NHANES III** is a national probability sample of persons aged 2 months and older and was conducted from 1988 to 1994. It 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. NHANES III included a standardized physical examination, laboratory tests, and questionnaires that covered various health-related topics. For detailed information on the NHANES III contents and methods, refer to the NHANES III website, http://www.cdc.gov/nchs/nhanes/nh3data.htm (accessed December 3, 2012).

The 1999-2004 **NHANES** data were linked to Medicare enrollment and claims data that covered services for the years 1999-2007. In 1999, the National Health and Nutrition Examination Survey (NHANES) became a continuous annual survey. For a variety of reasons, including disclosure issues, the NHANES survey data are released on public-use

---

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

data files in two-year increments (**NHANES 1999-2000, NHANES 2001-2002, NHANES 2003-2004**)*.* The survey examines a nationally representative sample of about 5,000 persons each year. These persons are located in counties across the country, 15 of which are visited each year. The survey includes a standardized physical examination, laboratory tests, and questionnaires that cover various health-related topics. For detailed information on the NHANES content and methods, refer to the NHANES website, http://www.cdc.gov/nchs/nhanes/nhanes_questionnaires.htm (accessed December 3, 2012).

The 2004 **NNHS** data were linked to Medicare enrollment and claims data that covered services for the years 1999-2007. The National Nursing Home Survey 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. For more information on the NNHS content and methods, refer to the NNHS website, http://www.cdc.gov/nchs/nnhs.htm (accessed December 3, 2012).

Medicare Data

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 care[2].

---

[2] http://www.cms.hhs.gov/MedicareGenInfo/ (accessed December 3, 2012)

Medicare claims information was extracted for the years 1991-2007. **Denominator** files are annual files 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 Denominator file is used to determine beneficiary demographic characteristics, entitlement, and beneficiary participation in a Medicare Part C plan also referred to as Medicare Advantage (MA).

There are three Medicare utilization files, two files that are grouped as **MedPAR** files and the **Part D Prescription Drug Event (PDE)** file that contains the utilization records for beneficiaries enrolled in the new Part D Prescription Drug program. 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.

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. The **Part D Denominator** files contains all of the demographic and enrollment variables available on the standard Medicare Denominator file, plus additional variables such as 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 Part D Denominator file contains one record per person.

The Outpatient standard analytic files contain Part B claims data submitted by institutional outpatient providers, such as hospital outpatient departments, rural health clinics, comprehensive outpatient rehabilitation facilities, community mental health centers, and ambulatory surgical centers for each calendar year. **Home Health Agency** files 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** files 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.

Carrier standard analytic files (formerly known as the Physician/Supplier Part B file) contain claims data submitted by non-institutional providers, such as physicians, 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. Separate **Durable Medical Equipment** (DME) files containing claims processed by authorized DME Regional Carriers (DMERC's) are 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 and DME files are for separate services and are not duplicates. For more information on DME claims in the Carrier and DME files, please refer to the documentation ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/Durable Medical Equipment.pdf (accessed December 3, 2012).

The **Chronic Condition (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 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; and Endometrial Cancer. The CC Summary File is constructed, based on the specified reference period for each condition. The CC Summary File contains one record per person. Please refer to Chronic Conditions definitions document for more details regarding reference periods and clinical specifications for each condition: http://www.ccwdata.org/cs/groups/public/documents/document/ccw_conditioncategories.pdf (accessed December 3, 2012).

The CC Summary file is only available for 2005-2007. Yearly, mid-year and date of occurrence flags for the 21 conditions are only available for participants who linked during these three years. However, for participants with a condition identified in these three years, the date of first occurrence is available from 1999 forward.

Researchers should refer to the data documentation and usage for more information on each file ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/ (accessed December 3, 2012). In addition, researchers are encouraged to refer to the Research Data Assistance Center (ResDAC) http://www.resdac.org/ (accessed December 3, 2012); a CMS contractor providing free assistance to researchers interested in using Medicare data for their research.

NCHS has created a **Summary Medicare Enrollment and Claims (SMEC)** file 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) http://www.cms.gov/MCBS/ (accessed December 3, 2012) 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 of the NCHS surveys linked to Medicare enrollment and claims data. For more information on the SMEC file, please refer to the documentation:
ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/summary_medicare_enrollment_and_claims_files.pdf (accessed December 3, 2012)

NCHS has also 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) http://www.usrds.org/ (accessed December 3, 2012). 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. For more information about the data available on the linked ESRD files, please refer to the documentation: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/nchs-usrds_linked_esrd_data_files_documentation.pdf (accessed December 3, 2012)

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:
http://www.cdc.gov/nchs/data_access/data_linkage/cms_medicaid.htm (accessed December 3, 2012).

Data Linkage with 1999-2007 Medicare records[3]

The linkage of NCHS survey respondents to their Medicare claims records was performed by an interagency agreement, including NCHS, CMS, SSA, and ASPE, and is not the responsibility of researchers using the data. The linkage had approval from NCHS's Research Ethics Review Board (ERB)[4] and was conducted for all NCHS survey respondents who were eligible for linkage. Respondents eligible for linkage included those who had sufficient personal identifier information (PII) and/or had not refused to provide their Social Security Number (SSN) or Health Insurance Claim (HIC) number and had a SSN verified by the SSA Enumeration Verification System. In July 2009 NCHS survey respondents who met these linkage eligibility criteria were sent to CMS to match to the CMS Denominator file for the years 1999-2007. The file used for matching did not contain the NCHS survey public-use ID, nor did it contain any information that could identify the original survey source. Note that here and throughout this document, the term *eligible* is used to indicate linkage eligibility, not Medicare benefits entitlement.

To link NCHS survey respondents with their Medicare data, the following individual identifiers were used:

- SSN
- Date of birth (month, day, year)
- Sex

To be considered a successful match the NCHS survey respondent's record needed to match exactly to the CMS Denominator file on all three items listed above.

For the 1999-2005 NHIS, 1999-2004 NHANES and the 2004 NNHS, the records that were determined to be a successful match to the 1999-2007 Denominator file had CMS data extracted from each of the Medicare files from 1999-2007, where available.

---

[3] Depending on the NCHS survey, some respondents may have Medicare claims data dating back to 1991.
[4] 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.

For the 1994-1998 NHIS, NHEFS, NHANES III and LSOA II, the records that were determined to be a successful match to the 1999-2007 Denominator file had CMS data extracted from each of the Medicare files for 1999-2007, where available, and had CMS records for the years 1991-1998 maintained from a previous 2001 linkage (see Appendix A for a detailed description of the previous 2001 linkage).

There are some instances, where NCHS respondents, with a verified SSN from SSA, had been successfully linked to CMS records in a previous linkage, but did not match the 1999-2007 Medicare Denominator file. The majority of these respondents died prior to 1999 and did not match to the 1999-2007 Denominator file because the CMS denominator file includes individuals who are entitled to benefits and enrolled in Medicare for that calendar year. Once a person dies they are no longer entitled to Medicare benefits. The other cases remained as non-matches and were dropped since their vital status could not be verified. For more information on the previous linkage please see Appendix A.

NCHS survey participants were considered ineligible for matching to the Medicare Denominator file if they refused to provide their SSN or HIC number at the time of the interview and/or did not have a verified SSN from SSA. Non-matches include those eligible for linkage that were not enrolled in Medicare as well as those not successfully matched to the Denominator file.

Survey participants who were under 18 years of age at the time of the survey are considered linkage-eligible, if the listed linkage-eligibility criteria is met and consent is provided by their parent or guardian.   However, in accordance with NCHS ERB guidelines, NCHS will only provide linked CMS data generated for program participation, claims and other events that occurred prior to the participant's 18th birthday. Please see the data linkage website for more information on this NCHS ERB guidance:
http://www.cdc.gov/nchs/data/datalinkage/nchs_survey_participants_under_age_18.pdf
(accessed December 3, 2012)

Linkage Rates

Linkage rates include successful matches to the Medicare Denominator file, which indicate if a NCHS survey participant was enrolled in Medicare during the calendar years of the matching interval (e.g. for the 2004 NNHS the matching interval was 1999-2007) and the small number of participants who matched in the previous 2001 Medicare linkage but died prior to 1999. The Linkage Rate table is divided into Table 1 and Table 2. Table 1 includes surveys that have Medicare data from 1991-2007 and Table 2 includes surveys with Medicare data from 1999-2007. Table 1 and 2 show for each survey, the total survey sample size, the sample size eligible for 1999-2007 Medicare linkage, the number of eligible survey respondents linked to any Medicare Denominator file, including those who died and had Medicare records maintained from the previous 2001 Medicare linkage, and two linkage rates.

The two linkage rates provided in Tables 1 and 2 are: A) a total survey sample linkage rate (column 3 divided by column 1); and B) an eligible sample linkage rate (column 3 divided by column 2). The eligible sample for linkage is based upon (1) having sufficient PII and/or not refusing to provide SSN or HIC numbers and (2) having a verified SSN from SSA. 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 death, if the participant had died, or an assumed age at the end of the matching interval (December 31, 2007). Medicare has age-based entitlement at 65. The age variable used in Tables 1 and 2 captures the survey participants who became entitled to Medicare at age 65 after the time of their initial survey interview. For example, a respondent in the 2000 NHIS who was 59 years of age at the interview and not entitled to Medicare based on his or her age would be 66 or 67 years as of December 31, 2007, the time of the linkage.

The proportion of NCHS survey participants who were ineligible for linkage varied by survey, ranging from 7% of participants being ineligible for linkage from the NHANES III to 55% of participants being ineligible from the 2005 NHIS. 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 had sufficient PII and had not refused SSN. Linkage rates for those less than 65 years were about 5% since Medicare entitlement for those less than age 65 is limited to persons meeting the Social Security Administration criteria for disability benefit entitlement or persons diagnosed with end-stage renal disease.

Data Limitations[5]

Although Medicare provides coverage for a wide range of services, there are health care services not covered by Medicare as well as a number of cost sharing requirements for Medicare beneficiaries. Examples of services not covered include routine physical exams, long-term care, and some cancer screening procedures. These gaps in coverage and required cost-sharing mean that there are no claims records for these services or for certain time periods. You may find more information on what is not covered by Medicare at www.cms.gov (accessed December 3, 2012) or www.aarp.org/health/medicare (accessed December 3, 2012).

CMS generally does not receive claims data for Medicare beneficiaries who enroll in Medicare Part C plans (including private fee-for-service plans paid on a capitation basis). Medicare Part C plans are also referred to as Medicare Advantage (MA) 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 database, Medicare Part C enrollment increased from approximately 6% of beneficiaries in 1991 to 20% in 2007. A summary of the percent of NCHS survey respondents who were enrolled in a Medicare Part C plan by year and survey can be found at http://www.cdc.gov/NCHS/data/datalinkage/nchs-cms_medicare_linked_data_managed_care_enrollment_tables.pdf (accessed December 3, 2012).

---

5 Please see the analytic issues document for a more comprehensive look at these issues.

Medicare Part C enrollment can be identified on the Denominator file from the HMO indicator variables (HMO_INDICATOR01- HMO_INDICATOR12) one for each month. During periods of Medicare Part C 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 Part C enrollment. Therefore, in general, studies based on analysis of claims data should exclude Medicare Part C enrollees from their beneficiary samples. For more information on how to create an analytic sample that excludes Medicare beneficiaries enrolled in a Medicare Part C plan, refer to a document written by ResDAC http://www.resdac.org/resconnect/articles/114 (accessed December 3, 2012) or contact ResDAC, which provides free consultation for researchers using Medicare files, www.resdac.org (accessed December 3, 2012).

Data Confidentiality

The NCHS must provide safeguards for the confidentiality of its survey respondents. To ensure confidentiality, all personal identifiers have been removed from the NCHS-Medicare linked data files. However, there remains the small possibility of re-identification and for this reason; the linked NCHS-Medicare data are not available as public-use files. NCHS has provided a Feasibility Study Data File that includes a limited set of variables for researchers to use in determining the feasibility and sample sizes of their proposed research projects. The files can be accessed at http://www.cdc.gov/NCHS/data_access/data_linkage/cms/cms_medicare_feasibility.htm (accessed December 3, 2012). Researchers who want to obtain the NCHS-Medicare linked data must submit a research proposal to the Research Data Center http://www.cdc.gov/rdc/ (accessed December 3, 2012).

Table 1. Sample Size and Unweighted Linkage Rates for Linkages of NCHS Surveys Containing 1991-2007 Medicare Enrollment and Claims Data by Survey and Age Group[1]

	Total Person Sample (Column 1)	Sample Eligible for the 1999-2007 Medicare Linkage[2] (Column 2)	Respondents with Information on any Medicare Denominator File[3] (Column 3)	Linkage Rate for Total Sample (Column 3/Column 1)	Linkage Rate for Eligible Sample (Column 3/Column 2)
NHIS 1994	116,179	87,079	23,819	20.5%	27.4%
<65	89,116	65,794	2,938	3.3%	4.5%
>=65	27,063	21,285	20,881	77.2%	98.1%
NHIS 1995	102,467	73,809	18,930	18.5%	25.6%
<65	80,861	57,113	2,535	3.1%	4.4%
>=65	21,606	16,696	16,395	75.9%	98.2%
NHIS 1996	63,402	42,407	10,537	16.6%	24.8%
<65	50,926	33,261	1,547	3.0%	4.7%
>=65	12,476	9,146	8,990	72.1%	98.3%
NHIS 1997	103,477	62,689	15,028	14.5%	24.0%
<65	83,694	49,588	2,171	2.6%	4.4%
>=65	19,783	13,101	12,857	65.0%	98.1%
NHIS 1998	98,785	53,535	12,128	12.3%	22.7%
<65	80,604	43,017	1,797	2.2%	4.2%
>=65	18,181	10,518	10,331	56.8%	98.2%
NHEFS[5]	14,407	12,879	7,258	50.4%	56.4%
<65	3,233	2,860	315	9.7%	11.0%
>=65	11,174	10,019	6,943	62.1%	69.3%
NHANES III	33,994	31,612	8,995	26.5%	28.5%
<65	25,408	23,309	971	3.8%	4.2%
>=65	8,586	8,303	8,024	93.5%	96.6%
LSOA II[4]	9,447	7,390	7,235	76.6%	97.9%

[1]Age is based on either the survey respondent's age at death or the assumed age at the end of the linkage interval (December 31, 2007)

[2]Eligibility for linkage is based upon having sufficient personal identifier information and/or not refusing to provide SSN or HIC numbers and having a verified SSN from SSA.

[3]This group includes respondents who matched in any or multiple years of the 1999-2007 Medicare Denominator file or had died prior to 1999 and had matched in the 2001 linkage

[4]All persons in LSOA II are older than 65 years

[5]NHEFS= NHANES I Epidemiologic Follow Up Study

Table 2. Sample Size and Unweighted Linkage Rates for Linkages of NCHS Surveys Containing only 1999-2007 Medicare Enrollment and Claims Data by Survey and Age Group[1]

	Total Person Sample (Column 1)	Sample Eligible for the 1999-2007 Medicare Linkage[2] (Column 2)	Number of Respondents with Information on any Medicare Denominator File[3] (Column 3)	Linkage Rate for Total Sample (Column 3/Column 1)	Linkage Rate for Eligible Sample (Column 3/Column 2)
NHIS 1999	97,059	49,467	10,613	10.9%	21.5%
<65	80,076	40,286	1,608	2.0%	4.0%
>=65	16,983	9,181	9,005	53.0%	98.1%
NHIS 2000	100,618	49,127	9,772	9.7%	19.9%
<65	83,985	40,718	1,552	1.8%	3.8%
>=65	16,633	8,409	8,220	49.4%	97.8%
NHIS 2001	100,760	47,452	9,255	9.2%	19.5%
<65	85,110	39,629	1,590	1.9%	4.0%
>=65	15,650	7,823	7,665	49.0%	98.0%
NHIS 2002	93,386	53,074	9,353	10.0%	17.6%
<65	79,109	45,122	1,551	2.0%	3.4%
>=65	14,277	7,952	7,802	54.6%	98.1%
NHIS 2003	92,148	49,095	8,061	8.7%	16.4%
<65	79,019	42,323	1,414	1.8%	3.3%
>=65	13,129	6,772	6,647	50.6%	98.2%
NHIS 2004	94,460	45,805	7,618	8.1%	16.6%
<65	81,401	39,348	1,281	1.6%	3.3%
>=65	13,059	6,457	6,337	48.5%	98.1%
NHIS 2005	98,649	44,835	6,926	7.0%	15.4%
<65	85,656	39,075	1,282	1.5%	3.3%
>=65	12,993	5,760	5,644	43.4%	98.0%
NHANES 1999-2000	9,965	7,852	1,722	17.3%	21.9%
<65	8,037	6,282	178	2.2%	2.8%
>=65	1,928	1,570	1,544	80.1%	98.3%
NHANES 2001-2002	11,039	9,274	1,799	16.3%	19.4%
<65	9,130	7,634	189	2.1%	2.5%
>=65	1,909	1,640	1,610	84.3%	98.2%
NHANES 2003-2004	10,122	8,624	1,738	17.2%	20.2%
<65	8,328	7,055	198	2.4%	2.8%
>=65	1,794	1,569	1,540	85.8%	98.2%
NNHS[6] 2004[7]	13,482	13,180	12,518	92.8%	95.0%
<65	1,310	1,271	786	60.0%	61.8%
>=65	12,172	11,909	11,732	96.4%	98.5%

[1]Age is based on either the survey respondent's age at death or the assumed age at the end of the linkage interval (December 31, 2007)

[2]Eligibility for linkage is based upon having sufficient personal identifier information and/or not refusing to provide SSN or HIC numbers and having a verified SSN from SSA.

[3]This group includes respondents who matched in any or multiple years of the 1999-2007 Medicare Denominator file

[4]All persons in LSOA II are older than 65 years

[5]NHEFS= NHANES I Epidemiologic Follow Up Study

[6] NNHS=National Nursing Home Survey

[7] Total survey population count is less than reported by health survey. This is due to lack of age data for purposes of these match rate tables.

Missing age at time of linkage for NNHS 2004: n=25

Appendix A

There are some instances, where NCHS respondents, with a verified SSN from SSA, had been successfully linked to Medicare records in a previous 2001 Medicare linkage, but did not match the 1999-2007 Medicare Denominator file. The majority of these respondents died prior to 1999 and did not match to the 1999-2007 Denominator file because the Medicare Denominator file includes individuals who are entitled to benefits and enrolled in Medicare for that calendar year. Once a person dies they are no longer entitled to Medicare benefits. The other cases remained as non-matches and were dropped since their vital status could not be verified. The previous 2001 Medicare linkage methodology is described below.

Description of Previous 2001 Linkage between Survey Data from the National Center for Health Statistics and Medicare Program Data from the Centers for Medicare and Medicaid Services

The linkage of NCHS survey respondents to their Medicare claims records was performed by NCHS and CMS and is not the responsibility of researchers using the data. The linkage was conducted in July 2001 and had approval from NCHS's Research Ethics Review Board. The process of linking each NCHS survey with Medicare data began by matching individual survey respondents with Medicare's Enrollment Database. Medicare's Enrollment Database (EDB) is a master enrollment file of all people ever entitled to Medicare. The EDB's records are comprehensive and updated daily.
To link NCHS survey respondents with their Medicare data, NCHS provided CMS as many as of the following individual identifiers that were available on the survey record for all eligible survey respondents:

- SSN
- Health Insurance Claim (HIC) number
- Last name
- First name

- Middle initial
- Date of birth (month, day, year)
- Sex
- Father's surname (women only)
- State of birth
- Zip code

NCHS survey participants were considered ineligible for matching to the EDB, if they refused to provide their SSN or HIC number at the time of the interview. Additional ineligibility criteria included refused, missing, or incomplete information on last name and date of birth.

CMS identified potential matches between NCHS survey participants and records in the EDB. CMS based potential matches on whether NCHS records matched EDB records on (1) HIC number, (2) SSN, or (3) name and date of birth. For these potential matches, NCHS employed a deterministic matching algorithm to determine which matches were correct. All potential matches were assigned a score based upon whether the identifying information provided matched between the NCHS and CMS records. For example, if NCHS provided a participant's SSN and it matched the SSN in a potential EDB match record, then a predetermined point value was added. If a SSN was submitted and the potential EDB match record for that participant did not match the SSN provided, no points were assigned for that identifying data element. This process was done for each of the identifying data elements submitted. Based upon this process, a total match score was established for all potential EDB matches.

Next, the scored matches were classified according to which identifying data elements matched, reflecting the fact that concordance between some identifying data elements, e.g. SSN, is more important than others. Within each class category, matches above a threshold score were considered acceptable, while matches below a threshold were considered non-matches. Matches falling between these thresholds were manually reviewed to determine the match status. For the NCHS records determined to be matched

to the Medicare EDB, CMS extracted data, where available, from each of the Medicare claims files for those records for the years 1991-2000. Since not all survey participants matched to the EDB have claims information for the years 1991-2000, the number of records with available Medicare claims data is less than the number matched to the EDB.