Vital and Health Statistics
Series 1, Number 58
September 2015

CDC
Monitoring the Nation's Health

SAFER • HEALTHIER • PEOPLE™

Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Suggested citation

Golden C, Driscoll AK, Simon AE, et al. Linkage of NCHS population health surveys to administrative records from Social Security Administration and Centers for Medicare & Medicaid Services. National Center for Health Statistics. Vital Health Stat 1(58). 2015.

Library of Congress Cataloging–in–Publication Data

Names: National Center for Health Statistics (U.S.), issuing body.
Title: Linkage of NCHS population health surveys to administrative records from Social Security Administration and Centers for Medicare & Medicaid Services.
Other titles: Linkage of National Center for Health Statistics population health surveys to administrative records from Social Security Administration and Centers for Medicare and Medicaid Services | Vital and health statistics. Series 1, Programs and collection procedures ; no. 58. | DHHS publication ; no. (PHS) 2015–1334. 0276–4733
Description: Hyattsville, Maryland : U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, September 2015. | Series: Vital and health statistics, series 1, programs and collection procedure ; number 58 | Series: DHHS publication ; no. (PHS) 2015–1334 | Includes bibliographical references.
Identifiers: LCCN 2015031985 | ISBN 0840606737 (alk. paper)
Subjects: | MESH: United States. Social Security Administration. | Health Surveys--methods--United States. | Records as Topic--United States. | Centers for Medicare & Medicaid Services (U.S.)--United States. | United States Government Agencies--United States.
Classification: LCC RA407.3 | NLM W2 A N148va no.58 2015 |
DDC 614.4/273--dc23 LC record available at http://lccn.loc.gov/2015031985

For sale by the U.S. Government Printing Office
Superintendent of Documents
Mail Stop: SSOP
Washington, DC 20402–9328
Printed on acid-free paper.

Vital and Health Statistics

Series 1, Number 58

Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services

Programs and Collection Procedures

U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES
Centers for Disease Control and Prevention
National Center for Health Statistics

Hyattsville, Maryland
September 2015
DHHS Publication No. 2015–1334

Contents

Acknowledgments v

Abstract 1

Introduction 1

Data Sources 2
- NCHS Population-based and Health Care Surveys 2
- Social Security Administration Files 4
- CMS Medicare Data Files 5
- CMS Medicaid Data Files 6
- End-stage Renal Disease Files 8

Linkage Eligibility 8

Linkage Methods 9
- Linkage Process: SSA 9
- Linkage Process: Medicare and ESRD 10
- Linkage Process: Medicaid 11

Access to Linked Data Files 11
- Research Data Center 11
- Public-use Feasibility Files 11

Analytic Guidelines and Considerations 11
- General Analytic Guidelines 12
- Analysis Using Linked NCHS–SSA Data 12
- Analysis Using Linked NCHS–Medicare Data 16
- Analysis Using Linked NCHS–MAX Data 20

Conclusions 24

Published Research Papers Using NCHS Survey-linked SSA, CMS, and USRDS Data 24
- NCHS–SSA Linked Data 24
- NCHS–CMS Medicare Linked Data 24
- NCHS–CMS Medicaid (MAX) Linked Data 25
- NCHS–USRDS Linked ESRD Data 25

References 26

Appendix I. Social Security Administration Programs 35

Appendix II. Centers for Medicare & Medicaid Services Programs 36

Appendix III. Adjusting Sample Weights for Linkage-eligibility Using SUDAAN 38

Appendix IV. First NCHS–SSA–CMS Linkage: NSCL–I 43

Appendix V. List of Resources for NCHS–SSA–CMS–USRDS Linked Data Files 44

Text Figures

1. Algorithm used for linking NCHS survey data to administrative data sources 9
2. Temporal alignment of National Health Interview Survey and Medicaid Analytic eXtract files 13

Detailed Tables

1. NCHS–Social Security Administration linked file sample sizes and percentage linked, by survey and age at interview 27
2. NCHS–Medicare linked file sample sizes and percentage linked, by survey and age at interview: 1991–2007 29
3. NCHS–Medicare linked file sample sizes and percentage linked, by survey and age at interview: 1999–2007. 30
4. NCHS–Medicaid linked file sample sizes and percentage linked, by survey and age at interview: 1999–2009. 31
5. Unweighted percentage of NCHS–Medicare linked sample of ages 65 and over at time of interview enrolled in managed care (for at least 1 month during year) for surveys containing 1991–2007 Medicare enrollment and claims data, by survey 33
6. Unweighted percentage of NCHS–Medicare linked sample of ages 65 and over at time of interview enrolled in managed care (for at least 1 month during year) for surveys containing 1999–2007 Medicare enrollment and claims data, by survey 33
7. Unweighted percentage of NCHS–Medicaid linked sample enrolled in Medicaid managed care programs for at least 1 month or full year, during 1999–2009 Medicaid Analytic eXtract data years, by survey. 34

Appendix Figure

Plot of National Health Interview Survey annual sample weight compared with ineligibility-adjusted weight 41

Appendix Table

SUDAAN output for weight-adjusted marginal model. 40

Acknowledgments

The authors gratefully acknowledge Christine S. Cox, Kimberly A. Lochner, Gloria Wheatcroft, Stephanie Bartee, Lisa B. Mirel, and Hannah R. Day for their contributions to previous versions of sections of this report. The authors also acknowledge James Brittain, Rashmi Tandon, Catherine R. Duran, Eileen Call, Keith Zevallos, and Jesse Bassich for providing programming support related to the production of the NCHS survey-linked data files and tables provided in this report. The authors gratefully acknowledge the editor, Jane Sudol; the typesetter, Zung T. Le; and the NCHS Office of Information Services, Information Design and Publishing Staff, for their dedication and great attention to detail during the production of this report.

Abstract

Background

As the nation's principal health statistics agency, the National Center for Health Statistics (NCHS) provides 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 socioeconomic circumstances. Although the survey data collected provide information on a wide range of health-related topics, they often lack information on longitudinal outcomes.

Methods

This report describes the second linkage conducted between several NCHS surveys and administrative records from the Centers for Medicare & Medicaid Services, the Social Security Administration, and the United States Renal Data Systems. A brief overview of the data sources, the methods used for linkage, descriptions of the resulting linked data files, and analytic guidance is provided.

Conclusions

Through its record linkage program, NCHS has been able to enhance the survey data it collects by augmenting survey information with information from administrative data sources. These linkages 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 low-income families with children, the elderly, and the disabled.

Keywords: record linkage • end-stage renal disease • National Health Interview Survey • National Health and Nutrition Examination Survey

Linkage of NCHS Population Health Surveys to Administrative Records From Social Security Administration and Centers for Medicare & Medicaid Services

Cordell Golden, B.A.; Anne K. Driscoll, Dr.P.H.; Alan E. Simon, M.D.; Dean H. Judson, Ph.D.; Eric A. Miller, Ph.D.; and Jennifer D. Parker, Ph.D.

Introduction

Federally sponsored health surveys are a critical source of information on public health in the United States. The National Center for Health Statistics (NCHS) is the nation's principal health statistics agency and is responsible for collecting accurate, relevant, and timely data. NCHS conducts several population-based national surveys as well as collecting vital statistics data, which are used by a broad range of users (researchers and policy makers, among others) to evaluate and profile the health of the American people. These national health surveys provide rich cross-sectional information on risk factors such as smoking, height and weight, health status, and socioeconomic circumstances, but information on longitudinal outcomes is often missing. Demand is increasing to incorporate information from additional sources in order to enhance the availability and quality of information on exposures and outcomes.

NCHS' record linkage program is designed to maximize the value of NCHS population-based surveys by augmenting survey information with information from administrative data sources, such as vital statistics, Medicare and Medicaid enrollment and claims, and Social Security benefit history. Beyond enhancing survey information, record linkage also reduces the cost burden of obtaining additional information, given the expense of active follow-up of survey respondents.

Under an interagency agreement among NCHS, the Centers for Medicare & Medicaid Services (CMS), the Social Security Administration (SSA), and the Office of the Assistant Secretary for Planning and Evaluation of the Department of Health and Human Services, two linkages have been conducted between several NCHS population-based surveys and CMS and SSA administrative records. The linkages have been undertaken to support various research initiatives of the participating agencies. The linked survey files provide a unique opportunity to examine the factors that influence health status, disability, chronic disease, health care utilization, and expenditures in the elderly, disabled, and poor populations of the United States. The linked survey files allow examination of the administrative data for some NCHS survey respondents during the year the NCHS survey was conducted, for some respondents in years after the survey was conducted, and for some respondents in years prior to the NCHS survey.

The first linkage under this interagency agreement was initiated in July 2001. Respondents from several NCHS surveys were matched to SSA benefit history data covering 1962–2003 and Medicare enrollment and claims data for 1991–2000. A detailed

description of the 2001 linkage of NCHS surveys to SSA and CMS administrative data, referred to as NSCL–I, is provided later in this report. (NSCL is an acronym drawn from the federal agencies conducting the record linkages: NCHS, SSA, and CMS.)

This report is an overview providing guidelines for using the NCHS survey-linked data files resulting from the second linkage, NSCL–II, which was initiated in 2007. The second linkage of NCHS surveys to SSA and CMS administrative data continues linkages to SSA benefit history data through 2007 and Medicare enrollment and claims data for 1999–2007, and it expands the linkage to Medicaid enrollment and claims data for 1999–2009. The report describes the data sources and linkage methods used to create the NSCL–II data files, followed by a description of the resulting data files and guidance on analytic methods. Detailed appendices provide information on SSA (Appendix I) and CMS (Appendix II) programs. Appendix III provides step-by-step instructions for adjusting sample weights for linked data files. Appendix IV describes the first linkage of NCHS, SSA, and CMS data. Finally, Appendix V includes a list of resources that may be helpful when using the NCHS survey-linked data.

Data Sources

NCHS Population-based and Health Care Surveys

Although NCHS conducts several population-based health surveys and establishment surveys of health care facilities, not all NCHS surveys can be linked to administrative data because some NCHS surveys do not collect the information necessary for linkage. Even within NCHS surveys that have been linked to administrative records, not all survey respondents are ''linkage-eligible,'' meaning that they did not supply sufficient information to enable the linkage. Thus, linked administrative data are available for those NCHS respondents who were enrolled in the Medicare, Medicaid, or SSA programs, who provided personal identification data to NCHS, and for whom NCHS was able to match identifying information with CMS or SSA administrative records. Not all NCHS surveys or survey years have been linked to all CMS, SSA, and United States Renal Data Systems (USRDS) administrative records. For NSCL–II, linked SSA, CMS, and USRDS administrative data are available for the following NCHS surveys:

- National Health Interview Survey (NHIS), 1994–2005
- National Health and Nutrition Examination Survey (NHANES)
  - NHANES 1999–2000, 2001–2002, and 2003–2004
  - NHANES III (1988–1994)
  - NHANES I Epidemiologic Follow-up Study (NHEFS)
    - 1971–1975 baseline
    - 1982–1984 follow-up
    - 1986 follow-up
    - 1987 follow-up
    - 1992 follow-up
- National Nursing Home Survey (NNHS)
  - 1985 (SSA only)
  - 1995 (SSA only)
  - 1997 (SSA only)
  - 2004 (SSA, CMS, and USRDS)
- Second Longitudinal Study of Aging (LSOA II)
  - 1994 baseline
  - 1997–1998 follow-up
  - 1999–2000 follow-up

Each of the NCHS surveys linked to SSA, CMS, and USRDS administrative data are briefly described below. More information about each can be obtained from the NCHS website: http://www.cdc.gov/nchs/index.htm.

National Health Interview Survey

NHIS is a cross-sectional household interview survey designed to monitor the health of the U.S. population through the collection and analysis of data on a broad range of health topics. It has been conducted continuously since 1957, and the survey content is periodically updated. The sampling plan follows a multistage area probability design that permits the representative sampling of households and noninstitutional group quarters (e.g., college dormitories). The sampling plan is redesigned after every decennial census.

Prior to 1997, NHIS included a set of basic health questions and demographic items, known as the Core questionnaire. The Core questionnaire was used for everyone in the household, with supplemental questionnaires used for more specific health topics. In 1997, NHIS underwent a sample redesign and questionnaire revision. While the demographic and basic health information were still collected on everyone in the household, one adult and one child within the household were randomly selected to complete a more detailed survey on specific health topics. Since 1997, the Core questions have remained largely unchanged from year to year. This allows for trends analysis and for data from more than 1 year to be pooled to increase sample size for analytic purposes. The Core contains four major components: Household, Family, Sample Adult, and Sample Child. The Household component collects limited demographic information on all of the individuals living in a particular house. The Family component verifies and collects additional demographic information on each member from each family in the house, and collects data on topics including health status and limitations, injuries, health care access and utilization, health insurance, and income and assets.

From each family, one sample adult and one sample child (if any children under age 18 are present) are randomly selected. Information is collected about each selected person using the Sample Adult Core and the Sample Child Core questionnaires. The Supplements are used to respond to new public health data needs as they arise. Some are fielded only once; others are repeated as needed. For example, in 2004, supplements were fielded on childhood mental health and adult immunizations. A complete list of supplements can be

found at: http://www.cdc.gov/nchs/nhis/supplements_cosponsors.htm.

Data are collected through a personal household interview conducted by interviewers employed and trained by the U.S. Census Bureau according to procedures specified by NCHS. A computer-assisted personal interviewing, or CAPI, mode using a laptop computer has been used since 1997. More information on NHIS is available from: http://www.cdc.gov/nchs/nhis.htm.

National Health and Nutrition Examination Survey

NHANES combines interviews, laboratory tests, and physical examinations to assess the health and nutritional status of adults and children in the United States. The survey currently examines a nationally representative sample of about 5,000 persons in 15 counties across the country each year.

Health interviews are conducted in respondents' homes. Health measurements are performed in specially designed and equipped mobile examination centers, which travel to locations throughout the country. The study team consists of a physician, medical and health technicians, and dietary and health interviewers. All participants visit the physician and have dietary interviews and body measurements taken. All but the very young have a blood sample taken and a dental screening.

The NHANES interview includes demographic, socioeconomic, dietary, and health-related questions. The examination component consists of medical, dental, and physiological measurements, as well as laboratory tests. Survey findings are used to determine the prevalence of major diseases and risk factors for diseases. NHANES findings are also the basis for national standards for such measurements as height, weight, and blood pressure. Data are also collected on chronic and previously undiagnosed conditions.

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

Third National Health and Nutrition Examination Survey

Prior to becoming a continuous survey in 1999, NHANES was conducted periodically, with the last periodic survey, NHANES III, being conducted from 1988 to 1994. NHANES III is a national probability sample of 33,994 persons aged 2 months and over. It was designed to provide national estimates of the health and nutritional status of the civilian noninstitutionalized U.S. population aged 2 months and over. NHANES III included a standardized physical examination, laboratory tests, and questionnaires that covered various health-related topics.

More information on NHANES III is available from: http://www.cdc.gov/nchs/nhanes/nh3data.htm.

NHANES I Epidemiologic Follow-up Study

NHEFS is a national longitudinal study that was jointly initiated by NCHS and the National Institutes of Health's National Institute on Aging in collaboration with other agencies of the Public Health Service. NHEFS was designed to investigate the relationships among clinical, nutritional, and behavioral factors assessed in the first NHANES (NHANES I) and subsequent morbidity, mortality, and hospital utilization, as well as changes in risk factors, functional limitation, and institutionalization.

The NHEFS cohort includes all persons aged 25–74 who completed a medical examination as part of NHANES I during 1971–1975 (n = 14,407). It comprises a series of follow-up waves, four of which have been conducted to date: 1982–1984, 1986, 1987, and 1992. The first follow-up wave included an in-person visit with brief examination. The subsequent waves consisted of telephone interviews.

Tracing and data collection rates in NHEFS have been very high. Ninety-six percent of the study population has been successfully traced at some point through the 1992 follow-up. Tracing rates for each completed wave ranged from 90% to 94%, and interview rates ranged from 91% to 96% of those traced. More information on NHEFS can be obtained from: http://www.cdc.gov/nchs/nhanes/nhefs/nhefs.htm.

National Nursing Home Survey

NNHS is part of a continuing series of nationally representative sample surveys of U.S. nursing homes and their services, staff, and residents. NNHS was first conducted during 1973–1974 and repeated in 1977, 1985, 1995, 1997, 1999, and, most recently, 2004.

NNHS provides information on nursing homes from two perspectives—the provider of services and the recipient of care. Data about the facilities include characteristics such as size, ownership, Medicare–Medicaid certification, services provided and specialty programs offered, and charges. For recipients, data were obtained on demographic characteristics, health status and medications taken, services received, and sources of payment.

Survey data 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. More information on NNHS is available from: http://www.cdc.gov/nchs/nnhs.htm.

Second Longitudinal Study of Aging

LSOA II is a nationally representative prospective study comprising 9,447 civilian non-institutionalized persons aged 70 and over at the time of their 1994 NHIS interview, which served as the baseline for the study. It followed this cohort of older persons through two follow-up interviews, conducted in 1997–1998 and 1999–2000. Thus, it follows a cohort of the elderly who began the survey while functioning effectively within the community.

LSOA II provides information on the sequence and consequences of health events, including medical care and services employed for assisted

community living, changes in social activities, living arrangements, familial support, the use of community services, the deployment of assisted living strategies, use of assistive devices, and the accessibility of technological and environmental adaptations. It also provides information on the causes and correlates of changes in functioning, including social and demographic characteristics, pre-existing and emerging physical illnesses, cognitive and emotional status, and social and environmental support, as well as new information on individual health risks and behaviors in the elderly including alcohol and cigarette use, use of hormone replacement therapy, receipt of important health screenings such as mammography and prostate examinations, body mass and weight loss, physical activity, and diet and nutrition. More information on LSOA II is available from: http://www.cdc.gov/nchs/lsoa/lsoa2.htm.

Social Security Administration Files

The NCHS surveys were linked to five SSA administrative data files: Master Beneficiary Record (MBR) file, Supplemental Security Record (SSR) file, Payment History Update System (PHUS) file, 831 Disability Master File (831 DMF), and a special extract of summarized quarters of coverage (QOC) from the Master Earnings File. Benefit history data from 1962 through December 2007 were provided to NCHS by SSA. Brief descriptions of each of the SSA administrative data files are provided below; the data dictionary for the linked NCHS–SSA data files is available from: http://www.cdc.gov/nchs/data/datalinkage/nchs_ssa_data_codebook_2009.pdf.

NCHS surveys that have been linked to SSA benefits history data are:

- NHIS 1994–2005
- NHANES 1999–2004
- NHEFS
- NHANES III
- LSOA II
- NNHS 1985, 1995, 1997, and 2004

Master Beneficiary Record File

The MBR file contains the data that generate Social Security benefit checks under the SSA Old Age, Survivors, and Disability Insurance (OASDI) program. The OASDI program provides income support to persons who have retired or are disabled, or to a spouse, widow(er), or other relative of the primary beneficiary. An MBR record is created when a person applies for benefits, including those whose claims were denied. The record includes information regarding the OASDI benefit amount, payment status, dual entitlement (i.e., whether the person is entitled to benefits based on more than one person's work history), and, if applicable, information about disability entitlement, estimates and reports of earnings, and student entitlement.

Supplemental Security Record File

Title XVI of the Social Security Act, or Supplemental Security Income (SSI), is a needs-based program that provides cash assistance to the elderly, the blind, and the disabled. The SSR file maintains information on all persons who have ever applied for Title XVI SSI and on persons who, prior to implementation of the federal SSI program in 1974, received state benefits and were converted to the SSI program. Each record contains monthly eligibility and payment history from the beginning of the SSI program, providing a retrospective look at the recipient's benefits amounts.

Payment History Update System File

With the passage of the 1983 Amendments to the Social Security Act, a portion of Social Security benefits have been subject to federal income taxes. Starting in 1984, the actual amount of the check or direct deposit that the respondent received can be determined. The aggregate amount of benefit payments, repayments, and reductions for each person in each calendar year was collected in the PHUS file.

831 Disability Master File

The Disability Determination Services (DDS) renders the initial medical determination for persons applying for disability benefits under Title II Social Security and Title XVI SSI. An 831 DMF record is established as soon as DDS completes its initial disability decision. The 831 DMF is primarily used for research on initial disability or continuing disability diagnoses and is available dating to 1988.

The 831 DMF contains data from subsequent decisions, such as those by the SSA Office of Hearings and Appeals. However, only minimal data are stored regarding subsequent appeals. Subsequent decisions, or corrections to earlier decisions, result in additional records with the same identifiers but with new values and dates. When more than one record exists for a given application, the field AL identifies the level of decision within a progression of appeals (e.g., initial, reconsideration, and others). The Date of SSA Decision, or DODEC, variable holds the date of each decision and can be used to identify the most current outcome.

Quarters of Coverage File

Due to Internal Revenue Service (IRS) regulations, NCHS was able to extract only a small set of summarized annual variables regarding quarters of coverage from the Master Earnings File (MEF). MEF is SSA's primary repository of earnings data for the U.S. population. The summarized QOC variables describe a person's ''insured status'' based on their earnings history and can be found in this QOC file developed specifically for the NCHS linkage. Insured status is the minimum number of credits or quarters of coverage a worker must earn to become eligible for his or her own Social Security benefit under the Title II program, or to have benefits paid to his or her family. The QOC file contains data regarding these credits dating to 1953.

CMS Medicare Data Files

The Medicare data files contain information on the enrollment status and health care utilization and expenditures of Medicare beneficiaries. Nearly all Medicare beneficiaries receive Part A hospital insurance benefits, which help cover inpatient hospital care, skilled nursing facility stays (not custodial or long-term care), home health care, 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. Brief descriptions of each of the Medicare files that were linked to NCHS survey data are provided below.

The Medicare Standard Analytical Files, or SAFs, are research-oriented files containing final action claims for Medicare beneficiaries. Each file contains information collected by Medicare to pay for health care services provided to a Medicare beneficiary. These files are generated from raw claims through final action algorithms that match the original claim with adjusted claims to resolve any adjustments. Finalized claim files are available from CMS approximately 9 months into the following calendar year (for example, 2007 data would become available in September 2008).

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

- NHIS 1994–1998
- NHEFS
- NHANES III
- LSOA II

NCHS surveys that have been linked to Medicare enrollment and claims data covering 1999–2007 are:

- NHIS 1999–2005
- NHANES 1999–2004
- 2004 NNHS

Medicare Denominator File

The Denominator File provides data on all Medicare beneficiaries enrolled or entitled to Medicare benefits in a given year. Monthly information on the enrollment status and type of enrollment of linked Medicare beneficiaries is provided, 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. The Denominator File contains one record per person per year of Medicare enrollment. The data dictionary for the Denominator File is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/Denominator%20(edited).pdf.

Carrier File

The Carrier File (formerly the Physician/Supplier Part B file) contains final action claims data submitted by noninstitutional providers. The data are largely made up of physician claim records, although the file also includes claims from other providers such as physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, ambulance providers, and stand-alone ambulatory surgical centers. One record is in the file for each claim, although some episodes of care include more than one claim. The data dictionary for the Carrier File is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/carrier.pdf.

Durable Medical Equipment File

The Durable Medical Equipment (DME) File contains final action claims data submitted by DME regional carriers. Information in the DME file includes up to 10 ICD–9–CM diagnosis codes, service type codes, dates of service, and reimbursement amount. One record is in the file for each claim, although some episodes of care include more than one claim. The data dictionary for the DME file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/Durable%20Medical%20Equipment.pdf.

Home Health Agency File

The Home Health Agency (HHA) File contains final action claims for home health services. Information 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, and reimbursement amount. An HHA claim may cover services provided over a period of time, rather than a single day. One record is in the file for each claim, although some episodes of care include more than one claim. The data dictionary for the HHA file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/Home%20Health%20Agency.pdf.

Hospice File

The Hospice File contains final action claims data submitted by hospice providers. The data in this file include the type of hospice care received (e.g., routine home care or inpatient respite care). The Hospice File contains data fields for up to 10 ICD–9–CM diagnosis codes and up to 6 ICD–9–CM procedure codes, dates of service, reimbursement amount, and some demographic information (such as date of birth, race, and sex). One record is in the file for each claim although some episodes of care include more than one claim. The data dictionary for the Hospice File is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/hospice.pdf.

Outpatient File

The Outpatient File 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. One record is in the file for each claim, although some episodes of care include more than one claim. The data dictionary for the Outpatient File is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/outpatient.pdf.

Medicare Provider Analysis and Review 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 in the MedPAR file. Each MedPAR claim record includes up to 10 ICD–9–CM diagnosis codes and up to 6 ICD–9–CM procedure codes 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. Multiple claims records are possible per person in the MedPAR file. The data dictionary for the MedPAR file can be found at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/MedPAR%20(edited).pdf.

Medicare Part D Denominator File and Medicare Part D Prescription Drug Event File

The Medicare Part D Denominator File contains demographic and enrollment information for each beneficiary enrolled in Medicare during the calendar year. In addition to the variables available on the standard Denominator File, the Part D Denominator File contains a derived race and 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 contains one record per person per year of enrollment.

The Medicare 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, including payments to the plan providers. It does not contain individual drug claims, but rather summary extracts submitted to CMS by Medicare Part D prescription drug plan providers. The file contains one record per event. Multiple records are possible per person in the Part D PDE File.

The Medicare Part D Denominator and PDE files are available for Medicare claim years 2006 and 2007. The data dictionary for the Part D Denominator File can be found at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/nchs-cms_medicare_part_d_denominator_file.pdf.

The data dictionary for the Part D PDE file can be found at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/nchs-cms_medicare_part_d_prescription_drug_event_file.pdf.

Medicare Summary Files

Chronic Condition Summary File

The Chronic Condition (CC) Summary File includes claims, enrollment, and assessment data 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:

- Acute myocardial infarction
- Alzheimer's disease
- Alzheimer's disease, related disorders, or senile dementia
- Atrial fibrillation
- Cataract
- Chronic kidney disease
- Chronic obstructive pulmonary disease (COPD)
- Depression
- Diabetes
- Glaucoma
- Heart failure
- Hip/pelvic fracture
- Ischemic heart disease
- Osteoporosis
- Rheumatoid arthritis/Osteoarthritis (RA/OA)
- Stroke/Transient ischemic attack (TIA)
- Breast cancer
- Colorectal cancer
- Prostate cancer
- Lung cancer
- Endometrial cancer

The file provides a summary of clinical information, including date of first diagnosis or occurrence, and yearly and midyear flags for each of the 21 conditions. The files are available for Medicare claim years 2005–2007. However, the date of first diagnosis or occurrence can be prior to 2005. The data dictionary for the CC Summary file can be found at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/nchs-cms_medicare_chronic_condition_summary_file.pdf.

NCHS Summary Medicare Enrollment and Claims File

NCHS created a Summary Medicare Enrollment and Claims (SMEC) file to assist researchers 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 months of Medicare entitlement per year, 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 cost and use files. In addition, summary variables related to the beneficiary's total number of emergency room visits and Part D prescription drug costs have been created. Each SMEC file has a fixed-length record format, with one record for each survey respondent successfully linked to Medicare administrative record files. The data dictionary for the SMEC file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/summary_medicare_enrollment_and_claims_files.pdf.

CMS Medicaid Data Files

Medicaid Analytic eXtract Files

Medicaid data have been collected by each state since 1999 and provided to CMS through the Medicaid Statistical Information System (MSIS). These data contain enrollee eligibility information and Medicaid claims paid in each quarter of the federal fiscal year. Data are extracted from MSIS to create the Medicaid Analytic eXtract (MAX) Files which were used to link to NCHS surveys. Descriptions of the MSIS and MAX files follow.

MSIS collects, manages, analyzes, and disseminates information on

enrollees, including utilization and payment for services covered by state Medicaid or the Children's Health Insurance Program (CHIP). CHIP provides health coverage to low-income uninsured children and pregnant women in families with incomes too high to qualify for most state Medicaid programs. CHIP is administered by the states but is jointly financed by the federal and state governments. Each state has the option of expanding Medicaid eligibility to children who previously had been ineligible due to their income, or creating a program distinct from its existing Medicaid program. For this report, expansions of existing Medicaid programs are referred to as M–CHIP, and programs distinct from existing Medicaid programs are referred to as S–CHIP. Additional information on the distinctions between M–CHIP and S–CHIP can be found in Appendix II.

In general, each state MSIS file contains one record for each person covered by Medicaid/CHIP for at least 1 day during the reporting quarter. The files contain specified data elements for: 1) persons covered by Medicaid/CHIP (enrollment files), which consist of demographic and monthly enrollment data; and 2) adjudicated claims (paid claims files), which consist of information from medical service-related claims and capitation payments. States submit four types of claims files representing inpatient (IP), institutional long-term care (LT), prescription drugs (RX), and other noninstitutional services (OT). Claims records contain information on the types of services provided, service providers, and dates, costs, and types of reimbursement. The files serve as the historical source of detailed Medicaid eligibility and paid claims data maintained by CMS.

MAX data are research extracts of MSIS. The MAX data system was developed to provide calendar-year utilization and expenditure information and serve as a research tool for examining Medicaid/CHIP enrollment, service utilization, and expenditures by subgroup and over time. MAX contains person-level information on demographics, monthly enrollment status, eligibility group, and use and costs of services during the year. It also includes claims-level records for more detailed analysis of patterns of service utilization, diagnoses, and cost of care among Medicaid/CHIP enrollees.

The MAX data systems have been modified from MSIS in a number of ways to enhance their usefulness for research purposes and to enable individual enrollee-level analysis. MSIS claims files contain separate records for initial claims, voided claims, and positive or negative adjustments; such records are combined to reflect final service event records in MAX. Changes in program eligibility reported retroactively are incorporated in MAX monthly enrollment measures. MAX data are linked to the Medicare Enrollment Database to help identify people dually enrolled in Medicare and Medicaid.

MAX includes summary information and claims data for all Medicaid enrollees in each state and the District of Columbia but not those in Puerto Rico or other U.S. territories. The MAX data system consists of a person summary (PS) file and four claims files: IP, LT, RX, and OT. Each of these files is described in greater detail below.

MAX data covering 1999–2009 are available for the following NCHS surveys:

- NHIS 1994–2005
- NHANES 1999–2004
- NNHS 2004
- NHEFS
- NHANES III
- LSOA II

Person Summary File

The PS file for each year of MAX data contains one record for each person included in the MAX file, including persons enrolled in Medicaid, M–CHIP, and some (but not all) people enrolled in S–CHIP. In some cases, as described further below, a beneficiary may have more than one record on the PS file during the same year. This might happen, for example, if a person was enrolled in Medicaid in more than one state during the same year. The PS file contains basis of eligibility, monthly enrollment data, type of coverage, demographic information, and summary information regarding expenditures and service use. The data dictionary for the PS file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_PS_Dictionary.pdf.

Inpatient Hospital File

The IP hospital file contains complete stay records for Medicaid enrollees who used inpatient hospital services. Data include admission and discharge dates, diagnosis-related groups (DRG), Medicaid payment for fee-for-service records, third-party payments, Medicaid-paid Medicare copayment and deductible amounts, up to nine ICD–9–CM diagnosis codes, and principal and additional procedure codes. The data dictionary for the IP file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_IP_Dictionary.pdf.

Long-term Care File

The LT file includes institutional long-term care (LTC) records for services provided by four types of long-term care facilities: mental hospitals for the aged, inpatient psychiatric facilities for persons under age 21, intermediate care facilities for the mentally disabled, and nursing facilities (NF). Information in the LT file includes start and end dates of services, patient status at discharge, Medicaid payment amounts for fee-for-service records, third-party payments, Medicaid-paid Medicare copayment and deductible amounts, and up to five ICD–9–CM diagnosis codes. These records do not include procedure codes. Other community-based LTC services (e.g., many home-based and personal care services) are included in the OT file. The data dictionary for the LT file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_LT_Dictionary.pdf.

Prescription Drug File

The RX file contains prescribed drugs, over-the-counter drugs, and other items dispensed by a freestanding pharmacy (nonhospital-based). Information in the RX file includes

prescription fill date, new or refill indicator, National Drug Code, and quantity and day supply. Also included are payment amounts, third-party payments, and Medicaid-paid Medicare copayment and deductible amounts. The data dictionary for the RX file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_RX_Dictionary.pdf.

Other Services File

The OT file contains two major types of records: 1) records for all noninstitutional services delivered that are not reported in other files, and 2) payment records for premiums paid to the following types of Medicaid managed care plans: HMOs, health insurance organizations, prepaid health plans (PHPs), and primary care case management plans (PCCMs). The service types in the OT file include physician and professional services, outpatient and clinic services, DME, hospice, home health care, and laboratory and X-ray. Information in the OT files includes dates and types of service, Medicaid payment for fee-for-service enrollees, third-party payments, Medicaid-paid Medicare copayment and deductible amounts, a procedure code, and up to two ICD–9–CM diagnosis codes. The data dictionary for the OT file is available from: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/medicaid/MAX_2009_OT_Dictionary.pdf.

End-stage Renal Disease Files

NCHS surveys were linked to a separate set of data files containing information on patients diagnosed with end-stage renal disease (ESRD) obtained from USRDS. USRDS is a national data system funded by the National Institute of Diabetes and Digestive and Kidney Diseases, or 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.

ESRD patient data covering information for 1974–2008 are available for the following NCHS surveys:

- NHIS 1994–2005
- NHANES 1999–2004
- NNHS 2004
- NHEFS
- NHANES III
- LSOA II

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 (form 2746) File contains general demographic information, primary and contributing causes of ESRD diagnosis, date of first dialysis, and information on kidney transplants. It also contains death information, if a Death Notification form was completed. The file has one record for each successfully linked NCHS survey respondent entered into USRDS.

ESRD Medical Evidence Report File

The ESRD Medical Evidence Report (form 2728) File contains information obtained by USRDS from the completed ESRD Medical Evidence Report, Medicare Entitlement, or Patient Registration forms. The file contains general demographic information, health insurance status, primary and contributing causes of ESRD diagnosis, values and dates of ESRD-related test results, dialysis information (e.g., date of first dialysis treatment), and kidney transplant status. The file may contain multiple records for each successfully linked NCHS survey respondent entered into USRDS.

ESRD Treatment History File

The ESRD Treatment History file contains information about ESRD-related events, such as dialysis or transplants. It includes the treatment start and end dates, and modality of ESRD treatment. The file may contain multiple records for each successfully linked NCHS survey respondent entered into 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 may contain multiple records for each successfully linked NCHS survey respondent entered into USRDS. The data dictionary for the NCHS–USRDS Linked ESRD Data Files can be found at: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/nchs-usrds_linked_esrd_data_files_documentation.pdf.

Linkage Eligibility

Linkage of NCHS survey respondents to administrative records was performed for respondents who were linkage-eligible. Linkage eligibility refers to the potential ability to link data from an NCHS survey respondent to administrative data. It is distinct from program eligibility, which defines whether a person meets federal and state-specific eligibility criteria for a specific government-administered or-funded program.

Linkage eligibility for NCHS survey respondents is generally determined by whether respondents agreed to provide personal identification data to NCHS and for whom NCHS was able to validate and match with administrative records. Personal identification information includes Social Security number (SSN) and date of birth. Respondents eligible for linkage included those who had not refused to provide their SSN or health insurance

claim (HIC) number (in the case of CMS and USRDS linkages); had sufficient personal identifier information (PII); and had an SSN verified by the SSA Enumeration Verification System (EVS).

NCHS' process for the collection of personal identification information and the determination of linkage eligibility has changed over time and differs across surveys. Note that the proportion of NHIS respondents who provide their SSN varies by survey year and by respondent characteristics. The linkage-eligible records in any given NCHS survey comprise a subset of the survey respondents. Of those linkage-eligible, the actual number of records linked depends on both program enrollment and the linkage process.

These linkages had approval from NCHS' Research Ethics Review Board (ERB) and were conducted for all NCHS survey respondents who were eligible for linkage. The NCHS ERB, also known as an Institutional Review Board or IRB, is an administrative body of scientists and nonscientists that is established to protect the rights and welfare of human research subjects.

NCHS survey participants who were under age 18 at the time of the survey are considered linkage-eligible, if the listed linkage eligibility criteria are met and consent is provided by their parent or guardian. However, in accordance with NCHS ERB guidelines, NCHS can provide linked data generated for program participation, claims, and other events only if they occurred prior to the participant's 18th birthday. 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.

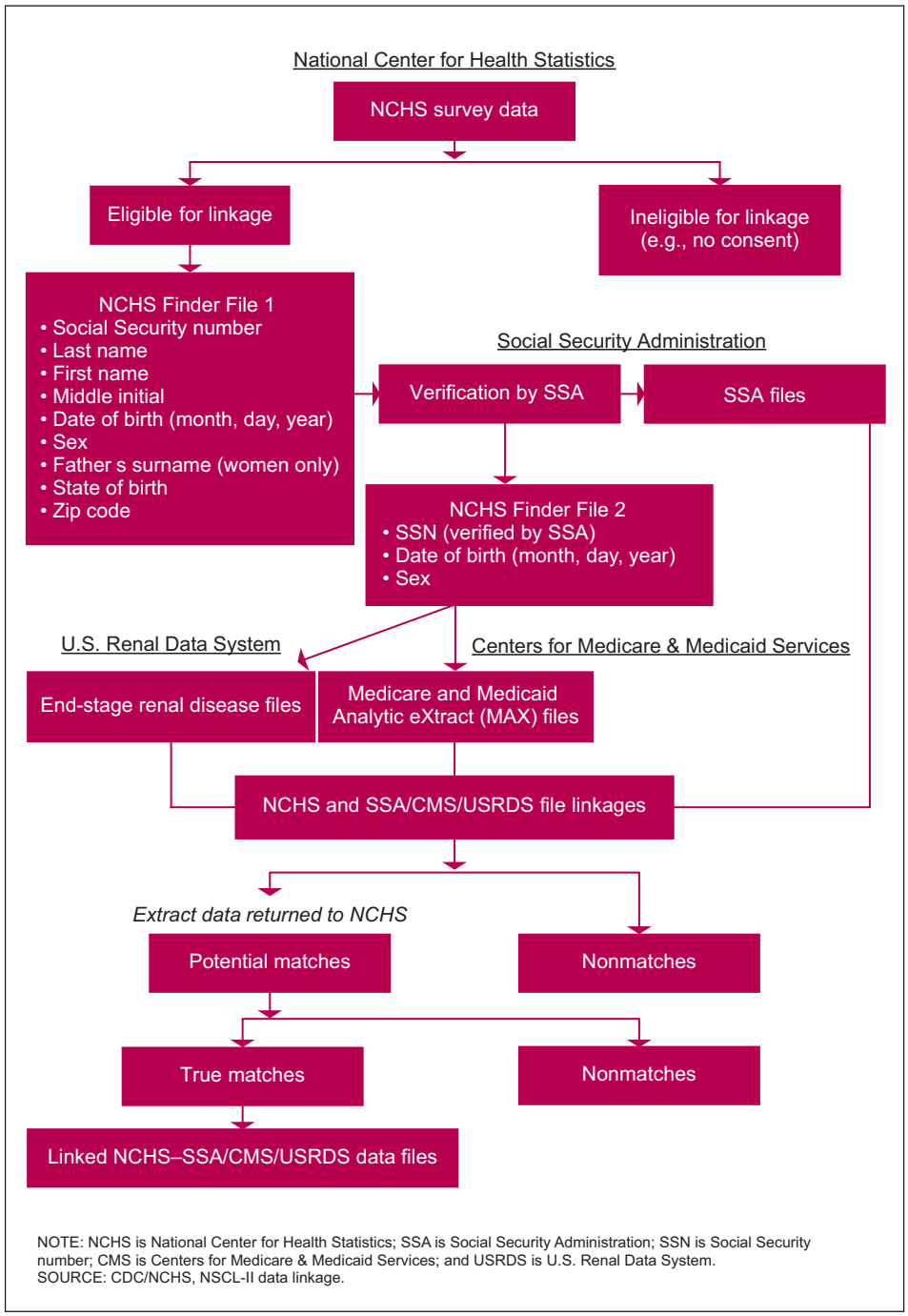
Linkage Methods

Figure 1 presents the overall process for linking NCHS survey data to administrative data from SSA, CMS, and USRDS. The linkage process for the first linkage, NSCL–I, was similar to the process for the second, although differences occurred (Appendix IV).

Linkage Process: SSA

This section provides an overview of the process that linked Social Security benefit history records with NCHS survey data. A more detailed description is available from: http://www.cdc.gov/nchs/data/datalinkage/ssa_methods_report_2009.pdf.

The process of linking NCHS survey data with Social Security data began by matching individual survey respondents with Social Security's Numident file. The Numident file is a record of applications for Social Security cards maintained by SSA. Persons are assigned unique, lifelong SSNs based on these applications. A full record of any changes to the information (such as change of name) is maintained within the Numident file. The data elements on a Numident record include name, date and place of birth, parents'



NOTE: NCHS is National Center for Health Statistics; SSA is Social Security Administration; SSN is Social Security number; CMS is Centers for Medicare & Medicaid Services; and USRDS is U.S. Renal Data System.
SOURCE: CDC/NCHS, NSCL-II data linkage.

Figure 1. Algorithm used for linking NCHS survey data to administrative data sources

names, and date of death (1). To carry out this linkage, NCHS provided a submission file to SSA with as many of the individual identifiers listed in Figure 1 as were available on the survey record for all linkage-eligible survey respondents. This file is referred to as NCHS finder file #1. NCHS survey participants were considered ineligible for matching to the Numident file if they refused to provide their SSN at the time of the interview. Additional ineligibility criteria included refused, missing, or incomplete information on last name and date of birth. The file provided to SSA did not contain the NCHS survey public-use ID, nor did it contain any information that could identify the original survey source.

The first step of the match process consisted of verifying SSNs received from NCHS. SSA verified whether the SSN received from NCHS was correct using EVS. In cases where the SSN of a survey participant was missing or could not be verified, SSA utilized an enhanced EVS matching algorithm to try to determine the correct SSN. The enhanced EVS matching algorithm was developed by SSA's Office of the Actuary to utilize additional identifying data elements collected during the survey interview and contained in other administrative records held by SSA. It features a scoring system with a threshold score used to determine which potential matches are acceptable and, thus, provided the opportunity to increase the number of successful matches. For NCHS records determined to be matched to the Numident file, SSA extracted data, where available, from the benefit history files.

Table 1 presents the total number of survey respondents for each survey and year, the number who were eligible for linkage, and the number who were linked to the Numident file, as well as the number of respondents available in the OASDI, SSR, and 831 Disability Master files. Since not all survey participants matched to the Numident have Social Security benefit history data, the number of records with benefit history data is less than the number of records matched to the Numident. NCHS–SSA linkage rates are based upon successful matches to the Numident file, not to the individual SSA administrative benefit history files. Due to the significant variation in the proportion of eligible survey respondents across surveys, two linkage rates are provided: a total survey sample linkage rate (the percentage of all survey respondents who were actually linked) and an eligible sample linkage rate (the percentage of linkage-eligible respondents who were actually linked). Additionally, linkage rates for each survey were examined overall and by two age groups—18–64 and 65 and over. Age was defined as the survey participant's age at the time of the interview.

Linkage Process: Medicare and ESRD

This section provides an overview of the process that links Medicare program and ESRD data records with NCHS survey data. A more detailed description is available from: http://www.cdc.gov/NCHS/data/datalinkage/cms_medicare_methods_report_final.pdf.

Following the SSN verification process at SSA, a second finder file (NCHS finder file #2) containing the verified or assigned SSN, date of birth (month, day, year), and sex was created by SSA and sent directly to CMS for linkage to Medicare and Medicaid program data (Figure 1). After processing by SSA, most NCHS records in the resulting NCHS finder file #2 contained an SSN verified or assigned by SSA, although a small subset of records had a missing SSN (0.1%). The survey respondents in NCHS finder file #2 were matched to the Medicare Denominator File for 1999–2007. To be considered a successful match, the NCHS survey respondent's record needed to match SSN exactly with only one beneficiary in the Medicare Denominator File. If the survey respondent matched SSN with multiple beneficiaries, sex and date of birth were used to identify a single beneficiary among all other candidates.

For the all of the surveys included in the linkage, CMS extracted Medicare enrollment and claims data for 1999–2007 for the records that were determined to be a successful match to the Medicare Denominator File for the corresponding year. For the 1994–1998 NHIS, NHEFS, NHANES III, and LSOA II, Medicare enrollment and claims data for 1991–1998 were also maintained from the previous NCHS–CMS linkage (NSCL–I).

In some instances, 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 because the Denominator File includes only persons who are entitled to benefits and enrolled in Medicare for that calendar year. Other cases were considered to be nonmatches because their vital status could not be verified.

NCHS–Medicare 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 2004 NNHS, the matching interval was 1999–2007), or if a participant matched in the previous 2001 Medicare linkage (NSCL–I) but died prior to 1999 (a small number of participants). The NCHS–Medicare linkage rate table is divided into two parts. The first part includes surveys that have Medicare data from 1991–2007; the second includes surveys with Medicare data from 1999–2007. For each survey, Tables 2 and 3 show 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 (NSCL–I); and two linkage rates, both the percentage of all survey respondents who were linked and the percentage of linkage-eligible survey respondents who were linked.

For linkage purposes in these tables, age was defined as the survey participant's age at interview. Medicare has age-based entitlement at age 65.

NCHS survey participants who died in the period between the survey and the

linkage interval were eligible for linkage if they had sufficient PII and had not refused SSN or HIC. Linkage rates for those under age 65 were about 5%, because Medicare entitlement for that age group is limited to persons meeting the SSA criteria for disability benefit entitlement or persons diagnosed with ESRD.

As part of the Medicare linkage, ESRD data for the linked Medicare beneficiaries are available to researchers. Although nearly all NCHS respondents linked to USRDS records are also linked to Medicare records, a small number of the linked USRDS records are not linked to Medicare records (fewer than 5%).

Linkage Process: Medicaid

NCHS finder file #2 was also used by CMS for linkage to the MAX PS file. For NCHS records from the verified file to link to the MAX PS file, an exact match on SSN, month and year of birth, and sex was required. From the linked NCHS–PS file, CMS extracted the associated claims from each of the four MAX claims files. CMS linked NCHS data separately for each state and then combined the subsets into one national file.

MAX records with missing SSNs are common because many states are unable to collect SSNs from some enrollees. Most of those without an SSN are children, persons who qualify only for family planning benefits, or aliens who qualify only for emergency coverage. CMS removed records without SSNs from the MAX PS file before linking it to the NCHS finder file.

The NCHS–MAX linkage process was conducted in four rounds. The first round included MAX data from 1999–2004, round 2 included MAX data from 2005–2007, round 3 included 2008 MAX data, and round 4 included 2009 data. Unfortunately, the 2009 data contained data for only 43 states and the District of Columbia. Seven states were not included because their MSIS files, the source of the MAX files, were unavailable at the time of the linkage or contained significant data problems. The excluded states are Hawaii, Idaho, Missouri, New Hampshire, Oklahoma, Utah, and Wisconsin.

Table 4 provides the full sample numbers for the NCHS surveys, the numbers that are linkage-eligible in each year, and the number actually linked because they are in the MAX file.

Access to Linked Data Files

Research Data Center

Due to confidentiality requirements, the NCHS survey-linked data files are restricted for research use, and the data are accessible only through the Research Data Center (RDC). RDC is a center housed on-site at CDC facilities in Hyattsville, Md., and Atlanta, Ga. In addition, NCHS data can be accessed from RDCs housed in U.S. Census Bureau offices in several locations across the country. Researchers need to be on-site at one of the RDCs; no remote access to the linked NCHS–CMS and NCHS–SSA data files is allowed.

All researchers must submit a research proposal to gain access to these restricted data files. The proposal provides a framework which allows RDC staff to identify potential disclosure risk. More information regarding RDC and instructions for the data can be found at: http://www.cdc.gov/rdc/.

Public-use Feasibility Files

To facilitate the use of the linked data files, NCHS has created public-use feasibility files that can be downloaded directly from the NCHS website. These files provide information about survey respondents' eligibility for the SSA and CMS linkages, final match status, and the linked data files (and years, in the case of CMS) available for each successfully linked survey respondent. These files can be used to determine the maximum available sample size to assess the feasibility of conducting analyses using the SSA and CMS linked data. These files make it possible to identify demographic and health status variables collected in the survey that may be of interest in subdividing the survey sample, calculating estimates of numbers of people in categories of interest, and determining cell sizes by related categories. This process helps prospective applicants for an NCHS RDC proposal establish the likelihood of sufficient numbers of persons and a sufficient number of degrees of freedom in the category of interest among those who have been successfully linked to SSA or CMS data for an NCHS RDC proposal.

Note that the feasibility files do not provide any information pertaining to benefits or payments; instead, they contain record status variables and record counts to assist researchers who are considering whether to initiate a RDC proposal. Due to risk of disclosure, the CMS Medicare Feasibility Files do not contain any variables that indicate linkage for ESRD data.

Researchers should be aware that all NCHS surveys linked to the SSA and CMS administrative data have complex survey designs. Therefore, considerations of statistical power should account for the survey design, in addition to the unweighted number of observations available for a particular project.

Documentation and instructions are available for downloading the NCHS–SSA (http://www.cdc.gov/nchs/data_access/data_linkage/ssa/ssa_feasibility.htm), NCHS–Medicare (http://www.cdc.gov/NCHS/data_access/data_linkage/cms/cms_medicare_feasibility.htm), and NCHS–Medicaid (http://www.cdc.gov/NCHS/data_access/data_linkage/cms/cms_medicaid_feasibility.htm) Feasibility files.

Analytic Guidelines and Considerations

This section describes general considerations and guidelines for analysis using the linked data files identified by NCHS programmers and analysts. Potential new analytic issues discovered in the course of analysis can be reported to the NCHS Special Projects Branch at datalinkage@cdc.gov.

General Analytic Guidelines

Variables to request in RDC applications

To create analytic files for use in 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–SSA file variables. The restricted variables from NCHS surveys and the exact variables from the CMS or SSA files that the researcher will use also need to be specifically requested as part of a researcher's application to RDC. RDC staff 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—Users who required information on geography should request data from the NCHS survey. However, geographic information is also available on the administrative data for linked participants and may differ from that in the survey.
- Linked mortality data for NCHS surveys—All NCHS surveys that have been linked to the CMS and SSA data have also been linked to mortality files that provide date and cause of death for each survey respondent who has died. These may be of use to some researchers but must be specifically requested as part of the researcher's proposal to RDC.
- NHANES month and year of examination and interview—The NHANES surveys combine at least 2 years of data. The exact year (and month) of a respondent'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 CMS or SSA data and the NHANES interview or examination. The variables that indicate the month and year of NHANES interview or examination must be requested specifically.

Researchers are advised to request the following variables, available from the public-use NCHS survey files, for inclusion in analytic files:

- NCHS sample weights and design variables—Sample weights are needed to create nationally representative estimates. Similarly, design variables are required to account for the survey design in analyses. 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, these weights should be adjusted for linkage eligibility.
- 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 CMS and SSA files. Therefore, where possible, the NCHS data should be used for demographic variables.

Sample weights

The sample weights provided in NCHS population health survey data files adjust for oversampling of specific subgroups and differential nonresponse, and are poststratified 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 nonmatches, 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 at NCHS and elsewhere is examining the use of survey weights for linked data in multiple ways.

Until specific recommendations are available, preliminary guidance 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 and nonmatches (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.

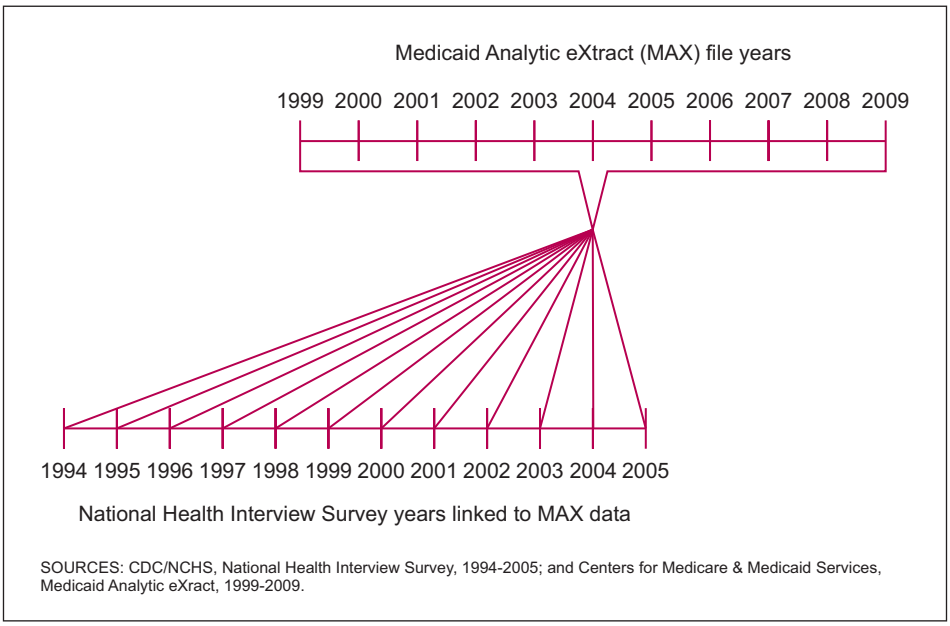
A model-based calibration approach developed within the SUDAAN software package (Procedure WTADJUST or WTADJX [SUDAAN 11]) allows auxiliary information to be used to adjust the statistical weights for nonresponse. This approach is promising, and this software is provisionally recommended for adjusting survey 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. More detailed information on adjusting sample weights for linkage eligibility using SUDAAN can be found in Appendix III.

Temporal alignment of survey and administrative data

Each NCHS survey has been linked to multiple years of administrative data. Depending on the survey year, administrative 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, discontinuous program coverage, and residential mobility of the survey participant. More information on each of these issues is provided in the program-specific analytic considerations sections that follow. The temporal alignment between NHIS and MAX data is illustrated in Figure 2.

Analysis Using Linked NCHS–SSA Data

Social Security data in these records, as with many administrative data, are extracted from files designed



SOURCES: CDC/NCHS, National Health Interview Survey, 1994-2005; and Centers for Medicare & Medicaid Services, Medicaid Analytic eXract, 1999-2009.

Figure 2. Temporal alignment of National Health Interview Survey and Medicaid Analytic eXtract files

for program administration and not for research. The following summary is a quick overview and explanation of some of the more complicated data usage issues that may arise when using SSA data.

The Rand Corporation has published an SSA Program Data User's Manual (2). The majority of information contained in this section is excerpted from the Rand SSA User's Manual.

MBR File

All applications to the NCHS RDC for linked NCHS–SSA data should include a request for the MBR file. An MBR record is created when a person applies for benefits, whether or not the application was approved. Because the MBR is quite large, researchers are advised to request only the MBR variables required for their project.

NCHS–MBR data files contain one record for each survey respondent who was linked to an MBR record from the SSA master file. Each MBR record can have up to three segments of OASDI benefit history data. The first record segment describes the benefit, if any, for which the respondent applied through December 2007 and payment history data for that entitlement. Persons who receive benefits through a dual entitlement situation will have benefit data on a second or third record segment. In a dual entitlement situation, a person is entitled to benefits as a retired or disabled worker and is entitled as well to benefits as the survivor of a spouse or parent who was entitled to benefits (2). Therefore, the second or third record segments describes the benefit, if any, for which the respondent applied through December 2007 and payment history data for that entitlement as an auxiliary beneficiary, based on another person's earnings. A person may also have benefit data on the second or third record segments if denied benefits on one account but entitled on another, because denial information is also documented. These second and third benefit record segments have the same file layout as the first record segment, with the variables identified by the extension _rp2 for the second record segment and _rp3 for the third. In addition, the variable OTBEN_CODE will have a value of "2" if a person has a second record segment, or "3" as indication that a person has a third benefit record segment.

The NCHS–MBR linked files extracted from SSA retrieve the most recent data, as well as past values for certain data elements. Therefore, some historic information exists on the MBR file, which was expanded after June 2004; these files contain several occurrences of historical values for certain data groups. Three main multiple-occurrence entry fields exist in the MBR: primary insurance history (PIH), disability insurance benefits (DIB), and historical payment data (HIST) entries. As the values of the fields change, new entries are added to the MBR account record. The PIH entry is part of the variable account data and records data about the Primary Insurance Amount (PIA) for this account. The PIA is the benefit a person would receive if he or she elects to begin receiving retirement benefits at his or her normal retirement age. At this age, the benefit is neither reduced for early retirement nor increased for delayed retirement. Information related to the disability benefits for an individual is recorded in the DIB entries. The HIST entry records actual payments made on this account over time. Multiple occurrences of entries can be used to track events of importance to the beneficiaries (2).

PIH entries record the earnings of the wage earner with the PIA and the family maximum dollar amount, or FMAX. Changes in PIA are the result of new earnings data or corrections to old earnings data for active accounts. The PIH entries occur in an MBR record up to 50 times. If more than 50 entries are needed, the oldest are dropped. This information shows changes in the basis of dollar benefit amounts over time. Changes in PIA are the result of cost-of-living increases, new earnings data, or corrections to old earnings data for active accounts. This information shows changes in the basis of dollar benefit amounts over time (2).

Information relating to disabilities is recorded in the DIB entries. Specific data elements include: Date of entitlement to disability (DOED), disability adjudication date (DSD), date of disability offset (DDO), disability award code (DAC), and drug addiction/alcoholism code, or DAA. Most of these fields contain date information, but the variable DAC identifies the type of disability. Using the multiple occurrences of DIB, it is possible to analyze changes in types of disabilities over time and the duration of entitlements (2).

HIST entries can have up to 552 occurrences (one for each month from January 1962 through December 2007) for each beneficiary, and each records the benefit amounts actually paid or withheld for specific months. If the benefit amount changes, the reason for the change is given. Changes can occur when earnings change or with a suspension or termination. Data elements related to historical payment data are further described below and include monthly benefit amount (O_MBA), ledger account file (O_LAF), monthly benefit paid (O_MBP), and monthly benefit credited (O_MBC) (2).

Types of claimants

In identifying a particular type of beneficiary or claimant (e.g., retirement, disability, or others), users may need to examine several fields (2). For example:

- Beneficiary Identification Code (BIC)
- Historical Beneficiary Identification Code (BCLM–BIC)
- Historical Beneficial Identification Code—Denials (DENY–BIC)
- Ledger Account File (LAF)
- Type of Claim—retirement/disability, primary/auxiliary (TOC)

Dates related to filing and eligibility

Date fields are critical to determining eligibility of SSA benefits (2). The following list includes a few variables of particular interest:

- Date of birth of primary (DOBP)
- Date of death of primary (DODP)
- Date of filing—through June 2004 (DOF)
- Date on which the beneficiary files—after June 2004 (BCLM–DOF)
- Date of filing on a denied claim—after June 2004 (DENY–DOF)
- Date of disability onset, or DDO
- Date of current entitlement (DOEC)
- Date of entitlement to disability, or DOED
- Date of initial entitlement (DOEI)
- Date of suspension/termination (DOST)
- Date of effective offset—through June 2004 (OED)
- Date of offset termination—through June 2004 (OTD)

DOBP is used to determine early and late retirement and the amount of outside income allowed without the loss of benefits. DODP determines the time at which survivor benefits are allowed to spouses and children. BDOB and BDOD are beneficiary date of birth and death (2).

Disability dates are complex and should be used carefully. To determine the date that a person is eligible to begin receiving benefits, or DOED, one must determine when the person became disabled, or DDO, as well when the application was filed, or DOF (2).

Identified for each date is the beneficiary identification code (BIC) and the type of benefit (TOB), which show why the benefit was received. For example, DOEC_BIC=A (Primary Claimant/ number holder) and DOEC_TOB=1 would indicate a retired-worker beneficiary as of the current entitlement. DOEI_BIC=A and DOEI_TOB=2 would indicate a disabled-worker beneficiary as of the initial entitlement (2).

The Dual Entitlement Data (DED) group contains fields that are used only in dual entitlement (2). The most pertinent fields from the DED group are:

- Dual Entitlement Status Code (DESC)
- Other Beneficiary Identification Code (OTBIC)
- Other date of entitlement (OTDOE)
- Type of dual entitlement (TOD)

Supplemental Security Record File

SSR files' history data can be used to determine duration of receipt of SSI benefits. However, take care when using eligibility history data because it can be overlaid by new data, making it difficult to reconstruct specific events.

Types of claimants

To identify a particular type of beneficiary or claimant, users may need to examine several fields (2). Below are some fields to be particularly aware of:

- Type of Action—Program category of unit (TOA)
- Master File Type—Program category of person (MFT)
- Payment status—Eligibility status (PSTAT)

SSA program eligibility

SSA program eligibility is determined based on citizenship, income, financial resources, and age or disability. Data elements related to income and financial resources are listed in the Payment History Update System File section below; for other criteria, the following fields describe or affect a person's program eligibility for SSI:

- Birth date (BIRTH_JD)
- Denial Code (DENCDE)
- Payment Status Code (PSTAT)
- State Program Category Eligibility Code (STCONCATM)

The variable PSTAT shows eligibility status for SSI benefits and provides information about why a person leaves current pay status. The variable STCONCATM can be used to identify state program categories for which the SSI recipient is eligible (2).

Benefit determination

SSR contains a number of fields related to the amount of benefits received on a given account (2). Some fields of interest are:

- Claim filed date (CLM_FIL_JD)
- Disability Payment Code (DISPAYCDE)
- Disability onset date (DIS_ONST_JD)
- Residence state and county (PDSCC)
- Payment Status Code, or PSTAT

Payment History Update System File

The linked NCHS–PHUS files were extracted from the PHUS master file by SSA. These linked files contain two variables that provide the information needed to correctly compute benefits paid: DIRECT_PAY and MEDICARE. Each DIRECT_PAY variable is the actual amount that a beneficiary received in a check or direct deposit in a specific month. Payment history is stored as a PHUS event with up to

264 occurrences, since 1984. The PHUS file is strictly a payment history system and not a transaction history file.

Data usage considerations

For months prior to January 1996, SSA had a policy under which MBR was not updated for single beneficiaries going into and out of payment suspense/termination status when the total payment to a family was unchanged. This process was called Facility of Payment. For these persons, PHUS cannot accurately determine the benefits paid to an individual, and no PHUS record is provided. As a result, a small number of survey respondents will have an MBR record with no PHUS record. In these cases, the only available payment data are the amounts shown in the Payment History section of the MBR: monthly benefit amount (O_MBA), monthly benefit credited (O_MBC), and monthly benefit paid (O_MBP).

831 Disability File

Diagnosis coding

The 831 DMF primary impairment code (PDX) and the secondary impairment code (SDX) identify the medical conditions considered in the medical decision about disability or blindness. Prior to 1985, the diagnosis codes used were from the *International Classification of Diseases, Ninth Revision* (ICD–9). Starting in 1985, impairment codes are assigned SSA-developed Impairment Codes, also known as Fussell codes. On occasion, the PDX and SDX will still contain ICD–9 codes set by the examining doctor (2).

The 831 DMF also contains a recoded primary impairment code (RPDX) and a recoded secondary impairment code (RSDX) that contain only SSA impairment codes for the medical conditions. These are derived fields that look at the PDX and SDX and, if necessary, convert ICD–9 codes to their SSA equivalent. The National Disability Determination Services System (NDDSS) is the system used by SSA primarily for processing Title II disability (SI), Title XVI SSI disability, and black lung (BL) claims. NDDSS propagates the impairment codes to the 831 DMF data elements PDX and SDX; the MBR data elements Diagnosis Code (DIG) and Secondary Diagnosis Code (SDIG); and the SSR data elements Disability Diagnosis Code (DIBDIG) and Secondary Disability Diagnosis Code (DIBDIG2). Because program files (in aforementioned variables listing MBR and SSR files) receive the original diagnosis codes, not the recoded values, care must be taken when using disability fields in the program files (2).

Linking 831 data to MBR and SSR

The 831 DMF can provide more detailed information on the medical conditions and severity of illness associated with a beneficiary in the Title II MBR or Title XVI SSR data for SSI benefits (2).

To link 831 DMF data to MBR records, the 831 records that apply to MBR cases must be selected by finding records with a RID value of ‘‘2’’ on the 831 DMF. The date values in the MBR may not always exactly match the date variables in the 831 DMF for the same disability application. A tolerance of up to plus-or-minus 3 months must be used (2).

To match 831 DMF data with SSR data, SSR records must be matched to the 831 DMF using the data element RID (value of ‘‘R’’) found on the 831 DMF. The date values in the SSR may not always exactly match the date variables in the 831 DMF for the same disability application. SSR records may require a plus-or-minus 1 month tolerance to reconcile the difference in dates between the linked SSR and 831 DMF (2).

QOC File

Insured status

The Master Earnings File or MEF is SSA’s primary repository of earnings data for the U.S. population (2). The summarized QOC variables are extracted from the MEF and describe individuals’ ‘‘insured status’’ based on their earnings history. Insured status is the minimum number of credits or quarters of coverage a worker must earn to become eligible for his or her own Social Security benefit under Title II. The QOC File contains data regarding these credits dating to 1953.

To determine whether a worker has insured status, Social Security considers the amount of the worker’s earnings (employment or self-employment) covered under Social Security and assigns ‘‘credits’’ for those earnings. These credits are called quarters of coverage (QCs). In 2007, one QC is credited for each $1,000 in annual covered earnings, up to a maximum of four QCs for the year. Earnings of $4,000 or more in 2007 will give the worker the maximum four QCs for the year, regardless of when the money is actually paid during the year (3). The amount of earnings required for a QC is adjusted automatically each year in proportion to increases in the national average wage level.

Fully insured

SSA program eligibility for most types of benefits requires that the worker be fully insured. To be fully insured, a worker must have a number of QCs at least equal to the number of calendar years elapsing between the year in which the worker is age 21 (or 1950, if later) and the year in which he or she reaches age 62, becomes disabled, or dies—whichever occurs first. To compute ‘‘elapsed’’ years, Social Security does not count the year in which the worker attains age 21 (or 1950, if later) or the year in which the worker attains age 62, becomes disabled, or dies. If the resulting number of elapsed years is less than 6, the number is raised to 6. All workers need at least 6 QCs to be insured. Workers who reach age 62 in 1991 or later need 40 QCs to be fully insured. Special rules may apply if the worker had a prior period of disability. For workers who become disabled or die before age 62, the number of QCs needed for fully insured status depends on their age at the time of disability or death.

Currently insured

Generally, if a worker dies before meeting fully insured status, benefits can still be paid to certain survivors if the worker was ‘‘currently insured’’ at the

time of death. Survivors' benefits are potentially payable to a worker's children and to a widow(er) who takes care of the deceased's child who is under age 16 or disabled and receiving Social Security benefits. To be currently insured, the worker must have earned 6 QCs in the 13 quarters ending with the quarter of death.

Disability insured

To qualify for disability benefits, a nonblind worker must have recent work activity in addition to being fully insured. Under the requirement involving recent work, a nonblind worker who is age 31 or over must have earned at least 20 QCs during the 40 calendar-quarter period ending with the quarter in which the disability began. In general, workers disabled at ages 24–30 must have earned QCs in one-half of the calendar quarters, beginning with the quarter following the quarter in which age 21 is attained and ending with the calendar quarter in which the disability began. In this case, the quarters counted will date to before the quarter in which the worker turned age 21. Workers under age 24 need 6 QCs in the 12-quarter period ending with the quarter in which the disability began. Workers who qualify for benefits based on blindness need only be fully insured. Special rules may apply if the worker had a prior period of disability. More information on the basic provisions concerning insured status can be found at: http://www.ssa.gov/regulations/index.htm.

Analysis Using Linked NCHS–Medicare Data

All applications to NCHS RDC for linked NCHS–Medicare data should include a request for the Denominator File for the years that the researcher is examining data. The Denominator File contains basic demographic and enrollment information about each beneficiary entitled to Medicare during each calendar year and is needed to construct an analytic data file, particularly to identify Medicare beneficiaries enrolled in a Medicare Part C plan.

Medicare Part C or Medicare Advantage plans

Medicare Part C plans are also referred to as Medicare Advantage or MA and include HMOs, preferred provider organizations (PPOs), private fee-for-service (PFFS) plans, special needs plans, and Medicare Medical Savings Account Plans. CMS generally does not receive claims data for Medicare beneficiaries who enroll in Medicare Part C plans (including PFFS plans paid on a capitation basis). There are exceptions; for example, all hospice claims are processed as Medicare claims regardless of whether the beneficiary is in a fee-for-service (FFS) or a Medicare Part C plan (4,5). During the time covered by the linked database, enrollment in Medicare Part C plans increased from approximately 6% of beneficiaries in 1991 to 20% in 2007.

In general, studies based on analysis of claims data should exclude Medicare Part C enrollees from their beneficiary samples. For health outcome or epidemiologic studies (as opposed to utilization or cost studies), an alternative approach for dealing with Medicare Part C enrollees is to include them for the time period prior to entering a Medicare Part C plan and then censor them at the time they enter a Medicare Part C plan.

A summary of the percentage of NCHS survey respondents who were enrolled in a Medicare Part C plan by year and survey can be found in Tables 5 and 6. Researchers should consider the percentage of respondents enrolled in a Medicare Part C program when determining the feasibility and sample sizes of their proposed research projects.

The following documents and citations provide detailed information about Medicare Part C enrollees and Medicare Utilization Files, and how to address them in analyses:

- Medicare Managed Care Enrollees and Medicare Utilization Files, available from: http://www.resdac.org/resconnect/articles/114.
- Virnig BA, Ash A, Kind S, Mesler DE. Survival analysis using Medicare data: Example and methods. Health Serv Res 35(5 Pt 3):86–101. 2000.

Services not covered (1991–2007)

Although Medicare provides coverage for a wide range of services, there are health care services not covered by Medicare. Examples include routine physical examinations, long-term care, and some cancer screening procedures. Therefore, no claims records exist for these services or for certain time periods.

In addition, Medicare data contain little information on prescription drugs for years prior to 2006. In 2006, prescription drug coverage for Medicare beneficiaries became available through the Medicare Part D program. Prescription drug information paid by Medicare for 2006–2007 is available in the Part D Prescription Drug Event (PDE) File. Prescription drug information for data years 1991–2005 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 Medigap insurance.

Gaps and discrepancies in coverage periods

Medicare enrollment and claims data linked to NCHS data are available for 1991–2007. Several of the surveys linked to Medicare data, such as NHEFS (1971–1992) and NHANES III (1988–1994), have gaps of several years between the end of the study period and the beginning of the Medicare data. Note that Medicare data may differ by age depending on the survey year(s). For example, an NHIS respondent who is 80 years old when interviewed in 1999 and matched at some point to the 1999–2007 Medicare Denominator File would have Medicare data for only 1999–2007, when that NHIS respondent was aged 80–88. A similar NHIS respondent, who was 80 years old when interviewed in 1994 and linked to the 1999–2007 Medicare Denominator File as well as linked in the previous 2001 Medicare linkage (NSCL–I) would hypothetically have Medicare data for 1991–2007, when that NHIS respondent was aged 77–93. This issue is particularly important when combining data across survey and Medicare coverage years, and researchers must determine how to address these discrepancies in coverage periods in their analyses.

Records with only 1991–1998 Medicare claims data

Several NCHS surveys (1994–1998 NHIS, NHEFS, NHANES III, and LSOA II) were included in the 2001 NCHS–SSA–CMS linkage (NSCL–I). In some instances, survey respondents with a verified SSN from SSA were successfully linked to CMS records in the 2001 linkage, but did not match the 1999–2007 Medicare Denominator File. The majority of these respondents died prior to 1999 and, thus, did not match to the 1999–2007 Denominator File. The other cases remained as nonmatches and were dropped because their vital status could not be verified.

Refer to the NCHS–CMS Medicare Linkage Methodology Report for more detailed information about linkage eligibility for both the first and second Medicare linkages, available from: http://www.cdc.gov/nchs/data/datalinkage/cms_medicare_methods_report_final.pdf.

Instances may occur in which an NCHS survey respondent is on the Denominator File but no claims data exist. It is possible to be enrolled in Medicare but not utilize Medicare services during the coverage period. In addition, some recordkeeping inconsistencies may occur because CMS data are collected for administrative, not research, purposes.

Medicare entitlement variables

The Denominator File includes three variables indicating Medicare entitlement: original reason for entitlement (ORIG_REASON_FOR_ENTITLEMENT), current reason for entitlement (CURR_REASON_FOR_ENTITLEMENT), and Medicare status code (MEDICARE_STATUS_CODE), which specifies the **most recent status** of the beneficiary's entitlement to Medicare benefits. The values for these three codes follow.

ORIG_REASON_FOR_ENTITLEMENT and CURR_REASON_FOR_ENTITLEMENT values include:

- OASI (Old Age and Survivors Insurance)
- DIB (Disability Insurance Benefits)
- ESRD, or end-stage renal disease

MEDICARE_STATUS_CODE values include:

- Aged without ESRD
- Aged with ESRD
- Disabled without ESRD
- Disabled with ESRD

Medicare's Prospective Payment System

Medicare's Prospective Payment System (PPS) is a method of reimbursement in which the Medicare payment is made based on a predetermined, fixed amount. Medicare uses a separate PPS for several services, in which the particular payment amount is based on the classification system for that particular service. The PPS for outpatient and HHA reimbursable claims was implemented in July 2000, so claims submitted for reimbursement before this date will differ from those after July 2000. More information on PPS is available from: http://cms.hhs.gov/Medicare/Medicare-Fee-for-Service-Payment/ProspMedicareFeeSvcPmtGen/index.html.

Employer-sponsored insurance

While almost all adults are enrolled in Medicare as they turn age 65, some working adults may still have employer-sponsored insurance, which is primary to Medicare. Although Medicare data do not contain an enrollment variable to identify Medicare beneficiaries who have primary employer-sponsored insurance, the Medicare data do show that beneficiaries with only Part A enrollment (and not parts B, C, or D) have lower health care utilization rates. This may be due to claims being submitted to only primary insurers, not Medicare. Therefore, for particular analyses, researchers may want to consider excluding beneficiaries with only Medicare Part A.

Additional analytic issues specific to each of the Medicare Administrative Files

Denominator File

Date of death information obtained by CMS is available in the Denominator File. CMS updates the Denominator File with death information collected through the first 3 months of the following calendar year. Deaths to Medicare-eligible beneficiaries occurring in the first quarter of the year will be recorded on that year's Denominator File but may also be recorded on the previous year's Denominator File.

Death information is occasionally misreported to CMS and included on the yearly Denominator File. This erroneous information is not corrected by CMS; however, these cases can be identified because they continue to be eligible for Medicare benefits in later years, or they

have new death information recorded in a later Denominator File. Analysts should use extra caution in analyzing Medicare death information to ensure that deaths are not overcounted. In addition, the **actual** date of death information is occasionally misreported to CMS. Cases can be identified by examining the variable, ''Valid Date of Death Switch,'' in which a value of ''V'' indicates that CMS has validated the actual date the beneficiary died, whereas a blank indicates that it was not validated. If the date of death is not validated, CMS assigns the date of death as the last day of the month.

Mortality information also is available from the NCHS Linked Mortality Files (http://www.cdc.gov/nchs/data_access/data_linkage/mortality.htm), which ascertain deaths from probabilistic matches to the National Death Index, death certificates, or longitudinal survey recontacts for the 1985–2009 NHIS, NHEFS, NHANES II, NHANES III, NHANES 1999–2010, LSOA II, Supplement on Aging (SOA), 2007 National Home and Hospice Care Survey (NHHCS), and 1985, 1995, 1997, and 2004 NNHS. No attempt has been made to reconcile inconsistent death information from CMS and these other sources. RDC research proposals that intend to analyze mortality outcomes should utilize death information from both the Medicare data and the NCHS Linked Mortality Files.

Carrier File

The claims on the Carrier File are processed by private carriers working under contract to CMS. Each carrier claim includes a Health Care Procedure Classification Code (HCPCS) to describe the nature of the billed service. The HCPCS are composed primarily of Level I HCPCS or CPT–4 codes developed by the American Medical Association, with additional codes specific to CMS called Level II HCPCS. Level II HCPCS are used to identify products, supplies, and services that are not included in CPT codes. These may include ambulance services, durable medical equipment or DME, prosthetics, and orthotics. Each HCPCS code on the carrier claim must be accompanied by an ICD–9–CM diagnosis code, providing a reason for the service. In addition, each record includes the date of service and reimbursement amount.

The Carrier File includes records for noninstitutional claims; however, this does not mean that they are outpatient claims. Providers, such as physicians, can bill for services provided in the office, hospital, or other sites. PLCSRVC ''Line Place of Service Code'' indicates where the service was provided, but it is not required for payment purposes, is not a validated code, and may contain inaccuracies.

The Carrier File contains DME claims processed by carriers who also process physician claims. The DME line items on the Carrier File can be identified by Claim Type Code (CLM_TYPE) equal to 72. DME claims processed through DME regional carriers are found on the DME Files, not on the Carrier File. DME claims on the Carrier File are for separate services. For additional information on DME regional carrier claims, see the DME File description below.

The Carrier File has two pairs of date fields. FROM_DT ''Claim From Date'' and THRU_DT ''Claim Through Date'' generally cover a period of service (but not always a single date of service), while EXPNSDT1 ''Line First Expense Date'' and EXPNSDT2 ''Line Last Expense Date'' represent the specific day of service.

For every billed procedure (using an HCPCS code), a corresponding ICD–9–CM diagnosis code (LINEDGNS) should appear providing the reason for the billed service. In the case of laboratory tests, the diagnosis will often be XX000, because the outside laboratory has no information from the physician about the reason for the test. In addition, the Carrier File contains space for up to four diagnoses, DGNS_CD1 through DGNS_CD4. These are not necessarily linked with any of the billed procedures and may reflect coexisting health conditions.

Some services may not appear in the Carrier claims, although they may have been received by the beneficiary. For example, CMS pays physicians a fixed amount for surgeries; this practice is called bundling. As part of bundling, CMS expects that certain care will be included in the payment amount, such as the first one or two office visits following surgery, or a biopsy just before surgery. Bundled services will not appear in the physician data. Interpretation of the rules on bundling varies by carrier (physician).

DME File

Durable medical equipment or DME can be billed through either a) the carriers who also process physician claims, or b) DME Regional Carriers (DMERCs), who process only DME claims. Each year, CMS distributes a jurisdiction list, available from the CMS website, which specifies whether a carrier or a DMERC can process a claim for a particular service. Often, both carriers and DMERCs are allowed to process and pay a DME claims service, depending on whether the DME was provided as ''incident to the physician's service.''

As noted previously, DME claims processed by suppliers who also process physician claims are included only on the Carrier File. These claims can be identified by Claim Type Code (CLM_TYPE) equal to 72 on the Carrier File. DME claims processed by regional carriers are included only on the DME File. Researchers should examine both files for 1993–2000, because approximately 90% of DME claims data are found on the DMERC File. However, for 1991 and 1992, nearly 100% of DME claims data are found on the Carrier File.

Hospice File

All beneficiaries in the Hospice File have a primary diagnosis, but most have no secondary diagnosis (90%). Although data fields exist for procedure codes, such information generally is not found in the Hospice File. Physician claims are for services provided by physicians employed or receiving payment from the hospice facility. All hospice claims are processed as Medicare claims regardless of whether the beneficiary is in a fee-for-service (FFS) or managed care plan.

Outpatient File

Same-day surgeries performed in a hospital are included in the Outpatient File. However, claims for surgeries

performed in freestanding surgical centers appear in the Carrier File, not in the Outpatient File. Information in the Outpatient File includes diagnosis and procedure codes, dates of service, reimbursement amounts, revenue center codes, and some demographic information (such as date of birth, race, and sex). The Outpatient File contains data fields for 10 ICD–9–CM diagnosis and 6 procedure codes, but the reporting of these codes is sporadic. Services provided can be obtained from the Health Care Procedure Classification Codes or HCPCS (HCPSCD01 through HCPSCD45), which can occur 10 times for a total of 450 occurrences. Additional information can be found in the revenue center codes (REV_CNTR). Definitions for revenue center codes can be found in the file documentation for the Outpatient File.

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 MEDPAR_SS_LS_SNF_IND_CD. It is coded S for short stay or L for long stay.

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 admission (MEDPAR_SRC_IP_ADMSN_CD)
- Group health organization payment code (MEDPAR_GHO_PD_CD)

In addition, MedPAR files include a mortality variable. However, if the outcome of interest is mortality, users should use the mortality indicator, DATE OF DEATH, on the Denominator File, or mortality status from the NCHS Linked Mortality Files.

Medicare Part D Denominator File and Medicare Part D Prescription Drug Event File

General Dynamics Information Technology, or GDIT (formerly Buccaneer, Inc.) has published a manual on understanding and analyzing Medicare Part D data. The majority of the text in this section was excerpted from the Chronic Condition Data Warehouse Part D Data User Manual, Version 3.1 (6).

Information is available on the Medicare Part D Denominator File for both Medicare beneficiaries who did not obtain Part D coverage and those who did obtain Part D coverage. In a small number of cases, the enrollment information in the standard Medicare Denominator File and the Part D Denominator File may differ. This is due to updates made by CMS after the standard Medicare Denominator File was delivered.

Generally, Part D coverage is provided under prescription drug plans (PDPs), which offer only prescription drug coverage, or through Medicare Advantage Prescription Drug Plans (MA–PD) plans, which offer prescription drug coverage that is integrated with the health care coverage provided to Medicare beneficiaries under Part C. The PDE file includes data from beneficiaries in both types of plans. The PDE file contains summary extracts submitted to CMS by Medicare Part D prescription drug plan providers. All Medicare Part D prescription drug benefits are provided through private plans (plan sponsors) (6).

Claims for prescription drugs are submitted by pharmacies to the Part D health plans for beneficiaries enrolled in Medicare Part D. PDE data are created from point-of-service transactional data at the time a prescription is filled. Data for prescriptions that are ordered but not filled do not exist in this database (i.e., data are not ''final action''), because they represent the final status of a drug claim at the time of CMS' payment reconciliation process (i.e., the records account for post-transaction adjustments). Not all Medicare-enrolled beneficiaries elect to purchase Part D coverage. Note that PDE data are not submitted by plans that receive retiree drug subsidies (RDS), or for other types of plans that are considered to be Part D creditable coverage (e.g., Veterans Administration [VA] or TRICARE) (6).

PDE differs from a pharmacy claim in several ways. Each PDE record is a summary record containing the final status of a drug claim sent by a pharmacy to Part D sponsors, accounting for any subsequent adjustments. Pharmacy claims rejected by the sponsor are not included in PDE data. For example, if a pharmacy submits an original claim to a plan sponsor that is rejected due to a prior authorization requirement, and later, when the prior authorization criteria are met, resubmits the claim which is then accepted by the sponsor, the sponsor would then submit only one PDE record to CMS reflecting the final status of the accepted claim. Similarly, if a pharmacy submits a claim to a plan sponsor and then soon after reverses (cancels) the claim, the sponsor would not submit a PDE record to CMS. Additionally, since the PDE data represent ''final action,'' all PDE adjustments received by CMS through the PDE submission deadline for payment reconciliation is accounted for in the data, including PDE adjustments, resubmissions, and deletions (6).

Not all drugs used by Part D enrolled beneficiaries are included in the PDE files. Data generally do not include Part D-excluded prescription drugs (unless the plan covers excluded drugs as a supplemental benefit). Prescriptions obtained through a third party (e.g., VA) or those for which a claim is not submitted (e.g., if a beneficiary pays cash out of pocket) are not available. In addition, over-the-counter (OTC) drugs are excluded from Part D and typically are not included in the PDE files, unless

they are part of an approved step therapy protocol (6).

Several situations occur in which Benefit Phase (i.e., pre-initial coverage limit, coverage gap, and catastrophic [post-out-of-pocket threshold]) and Utilization Management (UM) (i.e., tier, step therapy, quantity limits, and prior authorization) values cannot be determined:

- The PDE is for a noncovered drug. In this situation the Benefit Phase value will be blank. If the drug is found in the plan's formulary, then the UM variables will be assigned based on the formulary values. Otherwise, the UM variables will be assigned "NA" if the drug is not found in the formulary.
- Due to special waivers, some organization types are not required to submit details of their drug benefit package (e.g., employer direct and Programs of All-inclusive Care for the Elderly [PACE] plans). The Benefit Phase value for PDEs associated with these plans is "NA."
- Some plans may not utilize or be required to submit a formulary to CMS as part of the plan/formulary approval process (e.g., PACE plans). PDEs occurring for beneficiaries enrolled in these types of plans will have the least restrictive values for each of the four utilization management variables.
- Plans waived from submitting plan benefit information that did submit formulary information (e.g., employer direct) have "NA" assigned for the Benefit Phase. UM variables are assigned according to the plans' submitted formulary.
- If the drug on the PDE is not found on the plan's formulary, then all UM variables are assigned values of "NA."
- If the plan information on the PDE cannot be linked to the Health Plan Management System (HMPS) plan information, the value of "XX" is applied for all five of the event characteristics variables (6).

Chronic Condition Summary File

GDIT has also published a technical guidance manual for researchers using data from CMS' Chronic Condition Data Warehouse (CCW) for calculating population statistics. This manual is a useful resource for information on the CC Summary file, as well other CMS data containing information on Medicare beneficiaries. The majority of the text in this section was excerpted from the CMS Chronic Condition Data Warehouse, "Technical Guidance for Researchers Calculating Population Statistics" (7).

The CC Summary file is constructed each year, based on the specified reference period for each condition. The three CC variables for each of the 21 CCs have values that signify whether the pattern of utilization (i.e., FFS claims) indicated the presence of the condition for the beneficiary during the surveillance period ending with the last month of the reference period (December for the yearly indicators, June for the midyear indicators). Note that claims prior to the reference year may have been examined to make this determination if the CC definition was a 2- or 3-year condition (e.g., diabetes, congestive heart failure, or Alzheimer's disease) (7). Refer to the CC definitions document for more details regarding reference periods and clinical specifications for individual CC definitions: http://www.ccwdata.org/cs/groups/public/documents/document/ccw_conditioncategories.pdf.

Analysis Using Linked NCHS–MAX Data

All applications to NCHS RDC for linked Medicaid (MAX) data should include a request for the Person Summary or PS file for the years that the researcher is examining data. For each NCHS survey and survey year, a file has been created that includes MAX PS file observations for every MAX data year (1999–2009) that link to observations in that NCHS survey file. For example, for NHIS 2000, a PS file exists that has PS file observations from 1999–2009 that link to NHIS 2000. Similar IP, LT, RX, and OT files are available for each NCHS survey and survey year.

PS File

Each Medicaid enrollee is classified by two eligibility groups, a maintenance of assistance status (MAS) group and a basis of eligibility (BOE) group. MAS describes the financial criteria that allow an enrollee to be eligible for Medicaid: receiving cash assistance, being a parent with income below the 1996 Aid to Families with Dependent Children (AFDC) income thresholds, being part of a demonstration project, or other. BOE describes the group to which the enrollee belongs that is categorically eligible for Medicaid (children, elderly, disabled, and other). These measures are combined into the variable EL_MAX_ELGBLTY_CD_LTST, which concatenates MAS and BOE. MAS is in the first position; BOE is in the second position.

In general, across the years of linked MAX files, variables have changed little, and variable names have been largely standardized by CMS and NCHS. However, occasional additions or changes in variables have been made, so careful examination of the data dictionaries prior to analyses is suggested.

Multiple PS records

Many NCHS survey respondents will be linked to multiple MAX file PS records. Most often, this is because a respondent is linked to several years of MAX data. However, less frequently, a survey respondent may be linked to multiple PS records within the same year. For example, for NHIS 1999–2005, 7.9% of NHIS survey respondents who were linked to the MAX files had linkages to more than one PS record in at least 1 year. For NHANES 1999–2004, 9.5% of observations that linked to MAX files had linkages to more than one PS record in at least 1 year. There are multiple explanations for this situation:

- Medicaid enrollees moving between states in a given year
- Eligibility changes resulting in survey respondents disenrolling and re-enrolling in Medicaid within the same year

- Administrative changes or errors with Medicaid reporting
  - Some administrative changes and errors can be state- and year-specific. Certain record anomalies in each state have been identified and are provided by CMS (https://www.cms.gov/MedicaidDataSourcesGenInfo/downloads/anomalies1.pdf).

Most NCHS survey respondents with multiple PS records per year had records in multiple states. Among observations in NHIS 1999–2005 linked to multiple PS records in the same year, 86.6% came from different states. For NHANES 1999–2004, 81.7% came from different states.

Another source of multiple PS records in the same year is false matches due to misreporting of PII or issues with linkage methodology. The validity of multiple records in the same year can be difficult to ascertain. While some records show eligibility in different states in nonoverlapping months, others show eligibility in different states in the same months of the same year. A researcher may choose to exclude these records, depending on the research question being explored.

The existence of multiple PS records within 1 year with overlapping months of Medicaid enrollment data between the PS records can complicate analyses. In considering how to assess Medicaid enrollment in the presence of multiple PS records within a year, researchers may consider the use of variables that indicate enrollment by month in each record. By determining whether a person was enrolled in each month across the multiple records within a year, the number of total months of enrollment across records can be obtained. The variables MAX_ELG_CD_MO_1 through MAX_ELG_CD_MO_12 indicate whether an enrollee was eligible for Medicaid in a given month and, if so, under what criteria.

To help identify enrollees with multiple PS records within the same year, NCHS has added a set of flag variables to the PS file to identify these observations. FLG_MULT_RECS identifies whether enrollees have multiple records in any year. Additional variables identify enrollees who had any multiple PS records in a year and whether they occurred in the same or different state overall and in a given year. FLG_PRSN_MULT_RECS identifies enrollees with multiple records in any year; FLG_YEAR_MULT_RECS identifies enrollees with multiple records in a given year. The response categories for both variables are:

- 0 = No multiple records
- 1 = Multiple records in the same state
- 2 = Multiple records in different states
- 3 = Multiple records in the same and different states

If multiple PS files within a year link to an NCHS survey observation, and both a PS and an IP file, for example, were linked to the NCHS survey, which PS observation is related to a specific IP observation would be unclear. Thus, to merge data from both the PS file and another file to the NCHS survey data, the following steps must be taken to identify which PS observations are related to a specific claims file observation:

1. Merge NCHS survey observations for an individual year to the PS file from that year using the identification variable on the NCHS survey and the PS file.
2. Merge the desired claims file (IP, LT, RX, or OT) to the already merged NCHS–PS file using all three of the following variables:
   - FILE_YEAR4 on the claims file (IP, LT, RX, or OT) and the already merged PS file
   - MSIS_SEQN on the claims file and the already merged PS file
   - Identification variable on the claims file and the already merged PS–NCHS survey file.

Identifying Medicaid enrollees

The best method for identification of NCHS survey respondents who were Medicaid enrollees depends on the exact research question. In general, however, MSNG_ELG_DATA on the PS file provides information on enrollment status. The values for MSNG_ELG_DATA include:

- • = Enrolled in Medicaid during the year
- 2 = Enrolled in S–CHIP
- 1 = Enrolled in neither Medicaid nor S–CHIP

MSNG_ELG_DATA exists on each of the MAX files (PS, RX, IP, LT, OT) and at times, different values are assigned in the different files for the same person (in the same year). However, the value on the PS file is the most valid of these and should be used for all data for that observation.

Non-Medicaid/Non-S–CHIP observations

A small number of observations in the MAX files are coded as having been enrolled in neither Medicaid nor S–CHIP. It is not possible to determine why an individual is in the MAX files if he or she was not enrolled in either Medicaid or S–CHIP for any months of the year. Therefore, it is advisable to eliminate these observations from data analyses. These observations represent a small portion of the total data. They cluster during the 1999–2002 period; however, they remain less than 1.5% of records that link to the NHIS files in every year. Additionally, these observations cluster by state; 7 states had rates greater than 2%.

For MAX PS files between 1999–2002, some observations for which the enrollee was coded as having neither Medicaid nor S–CHIP during the year have contradictory data when examined at the monthly enrollment level. Specifically, some of these observations are coded as having had S–CHIP during some individual months, despite the overall yearly variable (MSNG_ELG_DATA) indicating they had neither S–CHIP nor Medicaid during the year. These data have not been changed, as it is not clear which data are correct. Treatment of these records is up to the researcher.

Days in the month

In a small number of observations, the Monthly Days of Eligibility exceed the total number of days in the month. For example, these variables may suggest that the enrollee was eligible for 31 days in September. These values are caused by administrative errors by the state and should be corrected if an analysis includes these values. An individual can enroll and disenroll in Medicaid at any time during a month. As a result, enrollees may have fewer days of Medicaid eligibility in a given month than exist in the month. The number days of Medicaid eligibility per month are captured in EL_DAYS_EL_CNT_1 through EL_DAY_EL_CNT_12. The values range from 0 to 31.

Dual eligible

Interest exists among researchers and policy makers in the group of people who receive benefits from both Medicaid and Medicare, often referred to as ''dual eligible.'' In the NCHS–MAX linked files, this group can be identified several ways. On the PS file, EL_MDCR_DUAL_ANN identifies dual eligibles. Values 50–59 signify that the enrollee was found in the Medicare database. Values 01–09 signify that the enrollee was not found in the Medicare database but was believed to be Medicare-eligible by the state. Values 50–59 should be used to identify dual eligibles because the Medicare enrollment database is the preferred indicator of dual enrollment.

To assess whether sample size will be adequate for a particular analysis, as discussed above, using the feasibility files is recommended. While there is no flag for dual eligibles on the feasibility files, researchers can use the feasibility files for the Medicare linked data files and the feasibility files for the MAX linked data files to create their own dual eligible flag. This method approximates the number identified as dual eligible on the MAX file.

For analyses of dual eligibles, some researchers will choose to use NCHS surveys linked to both Medicare data and the MAX files. Researchers should note that the methodology used for linking the NCHS surveys to MAX data differs from the methodology used to link Medicare data to the NCHS surveys. The Medicare linkage required that the records match exactly on SSN, sex, and day, month, and year of birth, whereas the MAX linkage required matching exactly on SSN, sex, and month and year of birth, but not necessarily on day of birth. A flag in the MAX files (EXACT_DOB_FLAG) indicates records that did match exactly on month, day and year of birth. Using this flag, researchers may choose to select records from the MAX file that used the same linkage methodology as the Medicare linkage and exclude the other records.

CHIP enrollment

Two issues related to S–CHIP should be considered when using MAX data. First, states have the option of not reporting information on S–CHIP enrollees to MSIS. Therefore, whereas the data on persons with Medicaid or M–CHIP can be considered universal, the MAX files do not include all S–CHIP enrollees. Variables provide monthly information on CHIP eligibility, as well as whether a person was enrolled in M–CHIP or S–CHIP.

For S–CHIP enrollees in the files, some data elements contain no information. Therefore, variables that are counts of months may not be accurate for persons enrolled in S–CHIP for one or more months, because those months are not counted in the total counts. Although S–CHIP enrollees may be a group of particular interest for some researchers, note that they account for a small percentage of NCHS survey respondents linked to the MAX files. For example, among those linked to the NHIS in 2005, less than 1.5% of enrollees in the MAX files are in S–CHIP in any given month. The variables EL_CHIP_FLAG_1 through EL_CHIP_FLAG_12 document CHIP eligibility monthly, and whether an enrollee was in M–CHIP or S–CHIP. The values include:

- 0 = Not eligible for Medicaid or CHIP during this month
- 1 = Enrolled in Medicaid during this month
- 2 = M–CHIP during this month
- 3 = S–CHIP during this month

For enrollees with a value of ''3'' (S–CHIP) for EL_CHIP_FLAG_1– through EL_CHIP_FLAG_12, no information is recorded for other monthly variables for that month.

Diagnosis and procedure coding

Diagnoses are uniformly provided in the MAX files for inpatient (IP), outpatient (OT), and long-term care (LT) using *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD–9–CM) codes. However, the coding systems used in the IP and OT files for procedures vary (no procedure codes are provided in the LT files). In the IP files, PRCDR_CD_SYS_1 through PRCDR_CD_SYS_6 describe the type of codes used for procedure codes in PRCDR_CD_1 through PRCDR_CD_6, respectively. In the OT file, the variable PRCDR_CD_SYS describes the type of codes used in the procedure code variable PRCDR_CD. For each procedure code variable, a separate variable is provided that identifies the coding system used for that code. For the IP file, the majority of procedure codes are ICD–9–CM. However, for the OT file, the majority of codes are either in CPT–4 or HCPCS codes.

State differences in Medicaid

Although Medicaid is administered under general federal guidelines, Medicaid varies substantially at the state level. Medicaid program eligibility, services offered, provider reimbursement, and other factors vary greatly from state to state. Consideration of these differences by state may be necessary for many analyses. State identifiers for each NCHS survey need to be specifically requested in these circumstances.

State-specific data issues

Because the data for the MAX files are obtained from each state, data quality differs between states. Prior to conducting analyses, researchers should consult the CMS website on the MAX files at: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/MAXGeneralInformation.html. This website provides data dictionaries, data anomalies for the whole MAX file and by state within the MAX files, and summary information by state from the MAX files. For any given analysis, there may be states or variables that present problematic data, and careful examination of the resources on the CMS website may reveal these issues before attempting data analysis.

Managed care, fee for service

Many Medicaid and CHIP enrollees are enrolled in managed care plans, and enrollment in these programs has grown over time. Rates of managed care enrollment also vary markedly across states. The PS file contains variables that identify beneficiaries enrolled in any type of managed care plan and the number of months that they were enrolled in the plan. EL_PHP_TYPE_1_1 through EL_PHP_TYPE_4_12 identify up to four types of plans for each month. Types of managed care plans identified in EL_PHP_TYPE_1_1 through EL_PHP_TYPE_4_12 are:

- Medical or comprehensive managed care plan
- Dental managed care plan
- Behavioral managed care plan
- Prenatal/delivery managed care plan
- Long-term care managed care plan
- All-inclusive care for the elderly or PACE plan
- Primary care case management or PCCM plan
- Other managed care plan

Waivers

Section 1115 of the Social Security Act provides the Secretary of Health and Human Services broad authority to authorize experimental, pilot, or demonstration projects requested by the states that are likely to assist in promoting the objectives of the Medicaid statute. These projects are intended to test and evaluate a policy or approach that has not been widely used. Some states expand eligibility to persons not otherwise eligible under the Medicaid program, provide services that are not typically covered, or use innovative service delivery systems. Examples include expanding care for children in foster care, providing specialty mental health care, and expanding Medicaid eligibility for family planning services to women of childbearing age not otherwise eligible for Medicaid. General information about Medicaid waivers can be found at: https://www.cms.gov/MedicaidStWaivProgDemoPGI/03_Research&DemonstrationProjects-Section1115.asp.

Waivers for specific groups make up one of the Maintenance Assistance Status or MAS categories. The variables MAX_ELG_CD_MO_1 through MAX_ELG_CD_MO_12 can be used to identify MAS and BOE monthly enrollment information, although the specific type of waiver cannot be identified before 2005. Starting in 2005, MAX files include three elements for each month (MAX_WAIVER_TYPE_1_MO_1 through MAX_WAIVER_TYPE_3_MO_12) that give detailed information on the type of waivers under which enrollees are eligible for Medicaid.

Prescription Drug File

The RX file does not include drugs provided during an inpatient hospital stay. Injectable drugs administered by a health professional are included in the Other Services or OT file described below. Beginning in 2006, full-benefit dual eligibles receive the Medicare Part D drug benefit, and their utilization for Part D-covered drugs is in the Medicare Part D Event data. Drugs provided to patients in LTC facilities are included in the Part D file for most states. However, a small number of states bundle these drugs in their LTC facility payment rate; in these cases, no details on drugs are available.

As with the other files, occasional additions or changes in variables have been made to the RX files over time; therefore, careful examination of the data dictionaries prior to analyses is suggested.

Other Services File

Because of how certain types of services are billed, multiple claims may exist for the same person and service dates in the OT file. These are not errors or data anomalies. What appear to be duplicate records are not true duplicates, but instead distinct services or portions of a service provided. For example, a patient who has a visit to an outpatient physician and then is sent to an outside laboratory to have a blood test on the same day would have a separate record for each of these services, but both would have the same date of service.

General limitations of MAX

There are some limitations to the information contained in the MAX files. Because these files contain only Medicaid-paid services, they do not capture service use or expenditures during periods of nonenrollment, services paid by other payers, or services provided at no charge. Because MAX files consist only of enrollee-level information, they do not include prescription drug rebates received by Medicaid; Medicaid payments made to disproportionate share hospitals (DSH)—hospitals that serve a disproportionate share of low-income patients with special needs; payments made through upper payment limit (UPL) programs; and payments to states to cover administrative costs.

In addition, service information in MAX may be missing or incomplete for certain groups of enrollees. This is particularly important for individuals enrolled in both Medicaid and Medicare (dual eligibles), persons enrolled in Medicaid managed care plans (either comprehensive or partial plans), and children with S–CHIP coverage.

Because Medicare is the first payer for services used by dual enrollees covered by both Medicare and Medicaid, MAX captures such service use only if additional Medicaid payments are made on behalf of the enrollee for Medicare cost sharing or for shared services. Medicare premiums paid by Medicaid on behalf of duals are not included in MAX.

For enrollees in Medicaid managed care plans, information in MAX is restricted to premium payments and some service-specific utilization information. While records for services delivered (including diagnoses and procedures) are uniformly provided for recipients with FFS coverage, encounter records for those with comprehensive managed care plans are not provided by all states. In some states, only a portion of managed care recipients have encounter data recorded. When included in the files, managed care encounter data list $0 as the amount paid for the services provided, even when the services are covered by the managed care plan.

A summary of the percentage of NCHS survey respondents who were enrolled in a Medicaid managed care plan by year and survey can be found in Table 7. Researchers should consider the percentage of respondents enrolled in a Medicaid managed care plan when determining the feasibility and sample sizes of their proposed research projects.

Conclusions

The SSA and CMS programs are complex, as are the resulting administrative data. Although this complexity can complicate analyses of linked data files, the wide range of information added to the NCHS population health surveys through linkage to SSA and CMS records provides a unique opportunity to study changes in health status, health care utilization, and expenditures in specialized populations, such as low-income families with children, the elderly, and the disabled.

Published Research Papers Using NCHS Survey-linked SSA, CMS, and USRDS Data

NCHS–SSA Linked Data

Gindi R, Cohen RA. Assessing measurement error in Medicare coverage from the National Health Interview Survey. Medicare Medicaid Res Rev 2(2):E1–E15. 2012.

Livermore GA, Stapleton DC, Claypool H. Health care when workers need it most: Before and after entry into the Social Security Disability Insurance Program. Inquiry 47(2):135–49. 2010.

Livermore G, Stapleton D, Claypool H. Costs and benefits of eliminating the Medicare waiting period for SSDI beneficiaries. Mathematica Policy Research Inc., Disability Policy Research Brief, no 09–02. 2009. Available from: http://www.mathematica-mpr.com/our-publications-and-findings/publications/costs-and-benefits-of-eliminating-the-medicare-waiting-period-for-ssdi-beneficiaries.

Livermore G, Stapleton D, Claypool H. Health insurance and health care access before and after SSDI entry. The Commonwealth Fund. 2009. Available from: http://www.commonwealthfund.org/publications/fund-reports/2009/may/health-insurance-and-health-care-access-before-and-after.

Rasch EK, Huynh M, Ho PS, Heuser A, Houtenville A, Chan L. First in line: Prioritizing receipt of Social Security disability benefits based on likelihood of death during adjudication. Med Care 52(11):944–50. 2014.

Riley GF. Health insurance and access to care among Social Security Disability Insurance beneficiaries during the Medicare waiting period. Inquiry 43(3):222–30. 2006.

NCHS–CMS Medicare Linked Data

Cai L, Lubitz J, Flegal KM, Pamuk ER. The predicted effects of chronic obesity in middle age on Medicare costs and mortality. Med Care 48(6):510–7. 2010.

Day HR, Parker JD. Self-report of diabetes and claims-based identification of diabetes among Medicare beneficiaries. National health statistics reports; no 69. Hyattsville, MD: National Center for Health Statistics. 2013. Available from: http://www.cdc.gov/nchs/data/nhsr/nhsr069.pdf.

Decker SL, Doshi JA, Knaup AE, Polsky D. Health service use among the previously uninsured: Is subsidized health insurance enough? Health Econ 21(10):1155–68. 2012.

Gorina Y, Kramarow EA. Identifying chronic conditions in Medicare claims data: Evaluating the Chronic Condition Data Warehouse algorithm. Health Serv Res 46(5):1610–27. 2011. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1475-6773.2011.01277.x/pdf.

Gorina Y, Pratt L, Kramarow E, Elgaddal N. Hospitalization, readmission, and death experience of noninstitutionalized Medicare fee-for-service beneficiaries aged 65 and over. National health statistics reports; no 84. Hyattsville, MD: National Center for Health Statistics. 2015 [Forthcoming].

Honeycutt AA, Segel JE, Zhuo X, Hoerger TJ, Imai K, Williams D. Medical costs of CKD in the Medicare population. J Am Soc Nephrol 24(9):1478–83. 2013.

Livermore GA, Stapleton DC, Claypool H. Health care when workers need it most: Before and after entry into the Social Security Disability Insurance Program. Inquiry 47(2):135–49. 2010.

Livermore G, Stapleton D, Claypool H. Costs and benefits of eliminating the Medicare waiting period for SSDI beneficiaries. Mathematica Policy Research Inc., Disability Policy

Research Brief, no 09–02. 2009. Available from: http://www.mathematica-mpr.com/our-publications-and-findings/publications/costs-and-benefits-of-eliminating-the-medicare-waiting-period-for-ssdi-beneficiaries.

Livermore G, Stapleton D, Claypool H. Health insurance and health care access before and after SSDI entry. The Commonwealth Fund. 2009. Available from: http://www.commonwealthfund.org/publications/fund-reports/2009/may/health-insurance-and-health-care-access-before-and-after.

Lloyd JT, Blackwell SA, Wei II, Howell BL, Shrank WH. Validity of a claims-based diagnosis of obesity among Medicare beneficiaries. Eval Health Prof 2014.

Looker AC, Dawson-Hughes B, Tosteson AN, Johansson H, Kanis JA, Melton LJ 3rd. Hip fracture risk in older US adults by treatment eligibility status based on new National Osteoporosis Foundation guidance. Osteoporos Int 22(2):541–9. 2011.

Looker AC, Eberhardt MS, Saydah SH. Diabetes and fracture risk in older U.S. adults. Bone 2015.

Looker AC. Femur neck bone mineral density and fracture risk by age, sex, and race or Hispanic origin in older US adults from NHANES III. Arch Osteoporos 8(141):141. 2013.

Looker AC. Hemoglobin and hip fracture risk in older non-Hispanic white adults. Osteoporos Int 25(10): 2389–98. 2014. Available from: http://link.springer.com/article/10.1007/s00198-014-2769-3.

Looker AC, Mussolino ME. Serum 25-hydroxyvitamin D and hip fracture risk in older U.S. white adults. J Bone Miner Res 23(1):143–50. 2008.

Looker AC. Serum 25-hydroxyvitamin D and risk of major osteoporotic fractures in older U.S. adults. J Bone Miner Res 28(5):997–1006. 2013.

Mainous AG 3rd, Diaz VA, Knoll ME, Hulihan MM, Grant AM, Wright RU. Transferrin saturation and hospital length of stay and mortality in Medicare beneficiaries. J Am Geriatr Soc 61(1):132–6. 2013.

Miller EA, Decker SL, Parker JD. Characteristics of those who choose Medicare advantage over fee-for-service upon Medicare enrollment at age 65. J Ambul Care Manage. In press 2015.

Mirel LB, Wheatcroft G, Parker JD, Makuc DM. Health characteristics of Medicare traditional fee-for-service and Medicare Advantage enrollees: 1999–2004 National Health and Nutrition Examination Survey linked to 2007 Medicare data. National health statistics reports; no 53. Hyattsville, MD: National Center for Health Statistics. 2012. Available from: http://www.cdc.gov/nchs/data/nhsr/nhsr053.pdf.

Polsky D. How the newly insured use health services: A lesson for the U.S. from Medicare. LDI Issue Brief 17(4):1–4. 2011.

Riley GF. Health insurance and access to care among Social Security Disability Insurance beneficiaries during the Medicare waiting period. Inquiry 43(3):222–30. 2006.

NCHS–CMS Medicaid (MAX) Linked Data

Capo-Ramos DE, Duran C, Simon AE, Akinbami LJ, Schoendorf KC. Preventive asthma medication discontinuation among children enrolled in fee-for-service Medicaid. J Asthma 51(6):618–26. 2014.

Hedley Dodd A, Gleason PM. Using the MAX–NHANES merged data to evaluate the association of obesity and Medicaid costs. MAX Medicaid Policy Brief #16. Washington, DC: Mathematica Policy Research. 2013. Available from: https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/MAX_IB16_MAX_NHANES.PDF.

Lee JS. Food insecurity and healthcare costs: Research strategies using local, state, and national data sources for older adults. Adv Nutr 4(1):42–50. 2013.

Lloyd PC, Simon AE, Parker JD. Characteristics of children in Medicaid managed care and Medicaid fee-for-service, 2003–2005. National health statistics reports; no 80. Hyattsville, MD: National Center for Health Statistics. 2015. Available from: http://www.cdc.gov/nchs/data/nhsr/nhsr080.pdf.

Mirel LB, Golden C, Wheatcroft G, Simon AE, Schoendorf KC. Linking children from the National Health and Nutrition Examination Survey to Medicaid enrollment and claims data. In: 2012 Proceedings of the Federal Committee on Statistical Methodology Research Conference. Washington, DC: Federal Committee on Statistical Methodology. 2012.

Mirel LB, Simon AE, Golden C., et al. Concordance between survey report of Medicaid enrollment and linked Medicaid administrative records in two national studies. National health statistics reports; no 72. Hyattsville, MD: National Center for Health Statistics. 2014. Available from: http://www.cdc.gov/nchs/data/nhsr/nhsr072.pdf.

Simon AE, Driscoll A, Gorina Y, Parker JD, Schoendorf KC. A longitudinal view of child enrollment in Medicaid. Pediatrics 132(4):656–62. 2013. Available from: http://pediatrics.aappublications.org/content/132/4/656.full.html.

Simon AE, Schoendorf KC. Medicaid enrollment gap length and number of Medicaid enrollment periods among US children. Am J Public Health 104(9):e55–61. 2014. Available from: http://ajph.aphapublications.org/doi/pdf/10.2105/AJPH.2014.301976.

Tao G, Hua J, Chen JL. Understanding sexual activity and Chlamydia testing rate based on linked national survey and Medicaid claims data. PLoS One 10(4). 2015.

NCHS–USRDS Linked ESRD Data

Muzaale AD, Massie AB, Wang MC, Montgomery RA, McBride MA, Wainright JL, Segev DL. Risk of end-stage renal disease following live kidney donation. JAMA 311(6):579–86. 2014.

References

1. McNabb J, Timmons D, Song J, Puckett C. Uses of administrative data at the Social Security Administration. Social Security Bulletin 69(1):74–84. 2009.
2. Panis C, Euller R, Grant C, Bradley M, Peterson CE, Hirscher R, Steinberg P. SSA program data user's manual. Rand Corporation Contract PM–973–SSA. 2000.
3. Social Security Administration. Annual statistical supplement to the Social Security Bulletin, 2007. SSA Publication No. 13–11700. 2008. Available from: http://www.ssa.gov/policy/docs/statcomps/supplement/2007/supplement07.pdf.
4. Asper F, Mann E. Medicare managed care enrollees and the Medicare utilization files. ResDAC KnowledgeBase Article #114 (under contract with Centers for Medicare & Medicaid Services). 2011. Available from: http://www.resdac.org/resconnect/articles/114.
5. Virnig BA, Ash A, Kind S, Mesler DE. Survival analysis using Medicare data: Example and methods. Health Serv Res 35(5 Pt 3):86–101. 2000.
6. Chronic Conditions Data Warehouse. Part D Data User Guide (Version 3.1). (Buccaneer, Inc., under contract with Centers for Medicare & Medicaid Services). 2010.
7. Chronic Condition Data Warehouse. Technical guidance for researchers calculating Medicare population statistics. (Buccaneer, Inc., under contract with Centers for Medicare & Medicaid Services). 2008.
8. Judson DH, Parker JD, Larsen MD. Adjusting sample weights for linkage-eligibility using SUDAAN. Hyattsville, MD: National Center for Health Statistics. 2013. Available from: http://www.cdc.gov/nchs/data/datalinkage/adjusting_sample_weights_for_linkage_eligibility_using_sudaan.pdf.
9. Miller D, Gindi R, Parker JD. Trends in record linkage refusal rates: Characteristics of National Health Interview Survey participants who refuse record linkage. Presented at: Joint Statistical Meeting 2011, Miami Beach, FL, July 30–August 4, 2011.
10. Judson DH, Parker JD. On dealing with ''incompletely linked'' data in linked survey/administrative databases: An empirical comparison of alternative methods. Presented at: Joint Statistical Meeting 2012, San Diego, CA, July 28–August 2, 2012.
11. Lohr SL. Sampling: Design and analysis. Brooks/Cole Publishing: Pacific Grove CA. 1999.
12. RTI International. SUDAAN language manual (Release 10.0). 2008.
13. Witt MB. Overview of software that will produce sample weight adjustments. Proceedings of the Section on Survey Research Methods, Joint Statistical Meeting 2009. Washington, D.C., August 1–6, 2009.
14. Little RJA, Rubin DB. Statistical analysis with missing data. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc. 2002.

Table 1. NCHS–Social Security Administration linked file sample sizes and percentage linked, by survey and age at interview

Survey and age group (years)[1]	Sample size			Percent linked		Respondents with benefit history data		
	Total sample	Eligible for linkage	Linked to Numident File	Total sample	Eligible sample	OASDI[2] from MBR[3]	SSI[4] payments from SSR[5]	831 disability File[6]
National Health Interview Survey								
NHIS 1994	83,719	67,916	63,835	76.2	94.0	32,304	8,078	6,977
18–64	69,148	55,844	52,121	75.4	93.3	20,869	6,473	6,797
65 and over	14,571	12,072	11,714	80.4	97.0	11,435	1,605	180
NHIS 1995	72,756	57,719	53,390	73.4	92.5	26,165	7,228	6,205
18–64	60,801	47,994	43,988	72.3	91.7	16,984	5,806	5,959
65 and over	11,955	9,725	9,402	78.6	96.7	9,181	1,422	246
NHIS 1996	63,402	47,262	30,926	48.8	65.4	15,015	4,196	3,691
18–64	38,140	28,244	25,523	66.9	90.4	9,718	3,402	3,564
65 and over	7,134	5,591	5,403	75.7	96.6	5,297	794	127
NHIS 1997	73,685	50,848	46,207	62.7	90.9	22,059	6,378	5,564
18–64	61,794	42,067	37,749	61.1	89.7	13,797	5,091	5,266
65 and over	11,891	8,781	8,458	71.1	96.3	8,262	1,287	298
NHIS 1998	70,663	44,072	39,451	55.8	89.5	18,378	5,306	4,723
18–64	59,243	36,696	32,447	54.8	88.4	11,507	4,220	4,413
65 and over	11,420	7,376	7,004	61.3	95.0	6,871	1,086	310
NHIS 1999	69,788	41,847	36,669	52.5	87.6	16,478	4,903	4,390
18–64	58,788	34,891	30,236	51.4	86.7	10,182	3,886	4,038
65 and over	11,000	6,956	6,433	58.5	92.5	6,296	1,017	352
NHIS 2000	72,123	41,890	36,141	50.1	86.3	15,560	4,855	4,314
18–64	60,917	35,198	30,000	49.2	85.2	9,554	3,854	3,985
65 and over	11,206	6,692	6,141	54.8	91.8	6,006	1,001	329
NHIS 2001	72,188	39,873	34,912	48.4	87.6	14,791	4,605	4,230
18–64	61,171	33,492	28,993	47.4	86.6	9,015	3,633	3,798
65 and over	11,017	6,381	5,919	53.7	92.8	5,776	972	432
NHIS 2002	67,195	46,278	36,531	54.4	78.9	14,967	4,522	4,162
18–64	56,610	39,025	30,417	53.7	77.9	8,978	3,620	3,739
65 and over	10,585	7,253	6,114	57.8	84.3	5,989	902	423
NHIS 2003	66,611	43,435	33,645	50.5	77.5	13,408	4,139	3,832
18–64	56,338	36,775	28,057	49.8	76.3	7,914	3,299	3,366
65 and over	10,273	6,660	5,588	54.4	83.9	5,494	840	466
NHIS 2004	68,299	41,410	32,611	47.7	78.8	12,712	4,021	3,721
18–64	57,541	34,920	27,058	47.0	77.5	7,265	3,181	3,195
65 and over	10,758	6,490	5,553	51.6	85.6	5,447	840	525
NHIS 2005	71,835	41,683	32,243	44.9	77.4	12,136	3,871	3,605
18–64	60,605	35,364	26,942	44.5	76.2	6,928	3,086	3,114
65 and over	11,230	6,319	5,301	47.2	83.9	5,208	785	491
National Health and Nutrition Examination Survey								
NHEFS[7]	14,407	13,540	12,974	90.1	95.8	11,434	2,382	1,033
18–64	10,551	9,932	9,577	90.8	96.4	8,134	1,625	1,033
65 and over	3,856	3,608	3,397	88.1	94.2	3,300	757	–
NHANES III	19,618	19,183	18,586	94.7	96.9	11,339	3,977	2,481
18–64	14,366	14,047	13,510	94.0	96.2	6,385	2,768	2,443
65 and over	5,252	5,136	5,076	96.6	98.8	4,954	1,209	38
NHANES 1999–2000	5,448	4,610	4,413	81.0	95.7	2,300	878	619
18–64	4,056	3,423	3,252	80.2	95.0	1,173	565	531
65 and over	1,392	1,187	1,161	83.4	97.8	1,127	313	88
NHANES 2001–2002	5,993	5,393	5,081	84.8	94.2	2,457	823	663
18–64	4,530	4,101	3,805	84.0	92.8	1,205	569	560
65 and over	1,463	1,292	1,276	87.2	98.8	1,252	254	103
NHANES 2003–2004	5,620	5,101	4,891	87.0	95.9	2,476	838	697
18–64	4,126	3,747	3,551	86.1	94.8	1,169	576	561
65 and over	1,494	1,354	1,340	89.7	99.0	1,307	262	136

See footnotes at end of table.

Table 1. NCHS–Social Security Administration linked file sample sizes and percentage linked, by survey and age at interview—Con.

Survey and age group (years)[1]	Sample size			Percent linked		Respondents with benefit history data		
	Total sample	Eligible for linkage	Linked to Numident File	Total sample	Eligible sample	OASDI[2] from MBR[3]	SSI[4] payments from SSR[5]	831 disability File[6]
National Nursing Home Survey								
NNHS 1985	11,166	10,542	9,945	89.1	94.3	9,544	3,051	49
18–64	1,203	1,147	1,059	88.0	92.3	905	735	49
65 and over	9,963	9,395	8,886	89.2	94.6	8,639	2,316	–
NNHS 1995[8]	8,045	7,342	6,524	81.1	88.9	6,299	1,933	249
18–64	674	583	522	77.4	89.5	448	391	176
65 and over	7,371	6,759	6,002	81.4	88.8	5,851	1,542	73
NNHS 1997[8]	14,739	13,663	12,554	85.2	91.9	12,113	3,582	772
18–64	1,377	1,282	1,193	86.6	93.1	1,019	879	579
65 and over	13,362	12,381	11,361	85.0	91.8	11,094	2,703	193
NNHS 2004[8]	13,492	13,372	13,165	97.6	98.5	12,737	4,001	1,572
18–64	1,553	1,532	1,504	96.8	98.2	1,299	1,158	921
65 and over	11,939	11,840	11,661	97.7	98.5	11,437	2,843	651
Longitudinal Study of Aging								
LSOA II[9]	9,447	7,635	7,499	79.4	98.2	7,304	1,042	26

– Quantity zero.

[1]Excludes information for participants under age 18 at time of survey. For this age group, in accordance with NCHS Ethics Review Board guidelines, NCHS provides linked administrative data generated only for program participation, claims, and other events occurring prior to the participant's 18th birthday. NCHS no longer releases linked administrative data generated on or after such participants' 18th birthday. Availability of information in Social Security Administration (SSA) Files for these participants depends on the variables requested.

[2]Old Age, Survivors, and Disability Insurance.

[3]Master Beneficiary Record.

[4]Supplemental Security Income.

[5]Supplemental Security Record.

[6]Containing data on decisions rendered by the SSA Disability Determination Services and on subsequent decisions by the SSA Office of Hearings and Appeals for persons applying for disability benefits under Title II (Social Security) and Title XVI (SSI).

[7]NHANES I Epidemiologic Followup Study.

[8]Age at interview missing (1995: $n = 2$; 1997: $n = 57$; 2004: $n = 1$). Total survey population count is less than reported by health survey due to lack of age data for match rate tables.

[9]All participants were aged 70 and over at time of interview.

NOTES: NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; NNHS is National Nursing Home Survey; and LSOA is Longitudinal Study of Aging. NHANES III was conducted in 1988–1994, and NHEFS in 1971–1975, with continued follow-up of the NHEFS population in 1982–1984, 1986, 1987, and 1992. LSOA II was conducted in 1994, with continued follow-up of the LSOA II population in 1997–1998 and 1999–2000.

Table 2. NCHS–Medicare linked file sample sizes and percentage linked, by survey and age at interview: 1991–2007

Survey and age group (years)	Sample size			Percent linked	
	Total sample	Eligible for linkage	Linked to Medicare Denominator File	Total sample	Eligible sample
National Health Interview Survey					
NHIS 1994	116,179	87,079	23,662	20.4	27.2
Under 65	101,608	75,387	12,235	12.0	16.2
65 and over	14,571	11,692	11,427	78.4	97.7
NHIS 1995	102,467	73,809	18,816	18.4	25.5
Under 65	90,512	64,410	9,576	10.6	14.9
65 and over	11,955	9,399	9,240	77.3	98.3
NHIS 1996	63,402	42,407	10,482	16.5	24.7
Under 65	56,268	37,036	5,202	9.2	14.0
65 and over	7,134	5,371	5,280	74.0	98.3
NHIS 1997	103,477	62,689	14,972	14.5	23.9
Under 65	91,586	54,675	7,102	7.8	13.0
65 and over	11,891	8,014	7,870	66.2	98.2
NHIS 1998	98,785	53,535	12,080	12.2	22.6
Under 65	87,365	47,028	5,673	6.5	12.1
65 and over	11,420	6,507	6,407	56.1	98.5
National Health and Nutrition Examination Survey					
NHEFS[1]	14,407	12,879	7,258	50.4	56.4
Under 65	10,551	9,491	6,254	59.3	65.9
65 and over	3,856	3,388	1,004	26.0	29.6
NHANES III	33,994	31,607	8,900	26.2	28.2
Under 65	28,742	26,531	4,026	14.0	15.2
65 and over	5,252	5,076	4,874	92.8	96.0
Longitudinal Study of Aging					
LSOA II[2]	9,447	7,390	7,235	76.6	97.9

[1]NHANES I Epidemiologic Followup Study.

[2]All participants were aged 70 and over at time of interview.

NOTES: NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; and LSOA is Longitudinal Study of Aging. NHANES III was conducted in 1988–1994, and NHEFS in 1971–1975, with continued follow-up of the NHEFS population in 1982–1984, 1986, 1987, and 1992. LSOA II was conducted in 1994, with continued follow-up of the LSOA II population in 1997–1998 and 1999–2000.

Table 3. NCHS–Medicare linked file sample sizes and percentage linked, by survey and age at interview: 1999–2007

Survey and age group (years)	Sample size			Percent linked	
	Total sample	Eligible for linkage	Linked to Medicare Denominator File	Total sample	Eligible sample
National Health Interview Survey					
NHIS 1999[1]	97,059	49,467	10,577	10.9	21.4
Under 65	86,059	43,663	4,879	5.7	11.2
65 and over	11,000	5,804	5,698	51.8	98.2
NHIS 2000	100,618	49,127	9,750	9.7	19.8
Under 65	89,412	43,510	4,237	4.7	9.7
65 and over	11,206	5,617	5,513	49.2	98.1
NHIS 2001	100,760	47,452	9,238	9.2	19.5
Under 65	89,743	42,027	3,910	4.4	9.3
65 and over	11,017	5,425	5,328	48.4	98.2
NHIS 2002	93,386	53,074	9,336	10.0	17.6
Under 65	82,801	47,325	3,683	4.4	7.8
65 and over	10,585	5,749	5,653	53.4	98.3
NHIS 2003	92,148	49,095	8,052	8.7	16.4
Under 65	81,875	43,881	2,915	3.6	6.6
65 and over	10,273	5,214	5,137	50.0	98.5
NHIS 2004	94,460	45,805	7,617	8.1	16.6
Under 65	83,702	40,537	2,443	2.9	6.0
65 and over	10,758	5,268	5,174	48.1	98.2
NHIS 2005	98,649	44,835	6,926	7.0	15.4
Under 65	87,419	39,972	2,142	2.5	5.4
65 and over	11,230	4,863	4,784	42.6	98.4
National Health and Nutrition Examination Survey					
NHANES 1999–2000	9,965	7,852	1,715	17.2	21.8
Under 65	8,573	6,737	615	7.2	9.1
65 and over	1,392	1,115	1,100	79.0	98.7
NHANES 2001–2002	11,039	9,274	1,790	16.2	19.3
Under 65	9,576	8,033	571	6.0	7.1
65 and over	1,463	1,241	1,219	83.3	98.2
NHANES 2003–2004	10,122	8,624	1,737	17.2	20.1
Under 65	8,628	7,328	463	5.4	6.3
65 and over	1,494	1,296	1,274	85.3	98.3
National Nursing Home Survey					
NNHS 2004[1]	13,506	13,180	12,518	92.7	95.0
Under 65	1,567	1,518	1,012	64.6	66.7
65 and over	11,939	11,661	11,505	96.4	98.7

[1]Age at interview missing ($n = 1$). Total survey population count is less than reported by health survey due to lack of age data for match rate tables.

NOTE: NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; and NNHS is National Nursing Home Survey.

Table 4. NCHS–Medicaid linked file sample sizes and percentage linked, by survey and age at interview: 1999–2009

Survey and age group (years)	Sample size				Percent linked			
			Linked to MAX PS[1] File		Total sample		Eligible sample	
	Total sample	Eligible for linkage	In survey year	In any year	In survey year	In any year	In survey year	In any year
National Health Interview Survey								
NHIS 1994	116,179	83,738	. . .	16,187	. . .	13.9	. . .	19.3
Under 18	32,460	19,939	. . .	5,636	. . .	17.4	. . .	28.3
18–64	69,148	52,107	. . .	7,986	. . .	11.5	. . .	15.3
65 and over	14,571	11,692	. . .	2,565	. . .	17.6	. . .	21.9
NHIS 1995	102,467	70,922	. . .	15,524	. . .	15.2	. . .	21.9
Under 18	29,711	17,540	. . .	5,846	. . .	19.7	. . .	33.3
18–64	60,801	43,983	. . .	7,555	. . .	12.4	. . .	17.2
65 and over	11,955	9,399	. . .	2,123	. . .	17.8	. . .	22.6
NHIS 1996	63,402	40,973	. . .	9,452	. . .	14.9	. . .	23.1
Under 18	18,128	10,086	. . .	3,539	. . .	19.5	. . .	35.1
18–64	38,140	25,516	. . .	4,616	. . .	12.1	. . .	18.1
65 and over	7,134	5,371	. . .	1,297	. . .	18.2	. . .	24.1
NHIS 1997	103,477	60,862	. . .	15,058	. . .	14.6	. . .	24.7
Under 18	29,792	15,172	. . .	5,952	. . .	20.0	. . .	39.2
18–64	61,794	37,676	. . .	7,121	. . .	11.5	. . .	18.9
65 and over	11,891	8,014	. . .	1,985	. . .	16.7	. . .	24.8
NHIS 1998	98,785	52,296	. . .	13,435	. . .	13.6	. . .	25.7
Under 18	28,122	13,424	. . .	5,585	. . .	19.9	. . .	41.6
18–64	59,243	32,365	. . .	6,187	. . .	10.4	. . .	19.1
65 and over	11,420	6,507	. . .	1,663	. . .	14.6	. . .	25.6
NHIS 1999	97,059	48,568	7,173	13,157	7.4	13.6	14.8	27.1
Under 18	27,271	12,584	3,762	5,559	13.8	20.4	29.9	44.2
18–64	58,788	30,180	2,721	6,139	4.6	10.4	9.0	20.3
65 and over	11,000	5,804	690	1,459	6.3	13.3	11.9	25.1
NHIS 2000	100,618	48,427	7,699	13,913	7.7	13.8	15.9	28.7
Under 18	28,495	12,884	4,040	6,108	14.2	21.4	31.4	47.4
18–64	60,917	29,926	2,893	6,359	4.7	10.4	9.7	21.2
65 and over	11,206	5,617	766	1,446	6.8	12.9	13.6	25.7
NHIS 2001	100,760	46,939	7,931	14,036	7.9	13.9	16.9	29.9
Under 18	28,572	12,599	4,239	6,358	14.8	22.3	33.6	50.5
18–64	61,171	28,915	2,979	6,326	4.9	10.3	10.3	21.9
65 and over	11,017	5,425	713	1,352	6.5	12.3	13.1	24.9
NHIS 2002	93,386	52,498	9,267	15,367	9.9	16.5	17.7	29.3
Under 18	26,191	16,388	5,256	7,723	20.1	29.5	32.1	47.1
18–64	56,610	30,361	3,250	6,389	5.7	11.3	10.7	21.0
65 and over	10,585	5,749	761	1,255	7.2	11.9	13.2	21.8
NHIS 2003	92,148	48,665	9,115	14,576	9.9	15.8	18.7	30.0
Under 18	25,537	15,451	5,270	7,375	20.6	28.9	34.1	47.7
18–64	56,338	28,000	3,143	6,086	5.6	10.8	11.2	21.7
65 and over	10,273	5,214	702	1,115	6.8	10.9	13.5	21.4
NHIS 2004	94,460	45,556	8,772	14,101	9.3	14.9	19.3	31.0
Under 18	26,161	13,291	4,863	6,811	18.6	26.0	36.6	51.2
18–64	57,541	26,997	3,144	6,109	5.5	10.6	11.6	22.6
65 and over	10,758	5,268	765	1,181	7.1	11.0	14.5	22.4
NHIS 2005	98,649	44,689	8,765	13,792	8.9	14.0	19.6	30.9
Under 18	26,814	12,939	4,810	6,669	17.9	24.9	37.2	51.5
18–64	60,605	26,887	3,214	6,083	5.3	10.0	12.0	22.6
65 and over	11,230	4,863	741	1,040	6.6	9.3	15.2	21.4

See footnotes at end of table.

Table 4. NCHS–Medicaid linked file sample sizes and percentage linked, by survey and age at interview: 1999–2009—Con.

Survey and age group (years)	Sample size				Percent linked			
			Linked to MAX PS[1] File		Total sample		Eligible sample	
	Total sample	Eligible for linkage	In survey year	In any year	In survey year	In any year	In survey year	In any year
National Health and Nutrition Examination Survey								
NHEFS[2]	14,407	12,879	. . .	1,570	. . .	10.9	. . .	12.2
Under 65	10,551	9,491	. . .	1,422	. . .	13.5	. . .	15.0
65 and over	3,856	3,388	. . .	148	. . .	3.8	. . .	4.4
NHANES III	33,994	29,300	. . .	8,614	. . .	25.3	. . .	29.4
Under 18	14,376	10,714	. . .	4,128	. . .	28.7	. . .	38.5
18–64	14,366	13,510	. . .	3,323	. . .	23.1	. . .	24.6
65 and over	5,252	5,076	. . .	1,163	. . .	22.1	. . .	22.9
NHANES 1999–2000	9,965	7,594	. . .	3,431	. . .	34.4	. . .	45.2
Under 18	4,517	3,232	. . .	1,989	. . .	44.0	. . .	61.5
18–64	4,056	3,247	. . .	1022	. . .	25.2	. . .	31.5
65 and over	1,392	1,115	. . .	420	. . .	30.2	. . .	37.7
NHANES 2001–2002	11,039	9,100	. . .	3,908	. . .	35.4	. . .	42.9
Under 18	5,046	4,059	. . .	2,419	. . .	47.9	. . .	59.6
18–64	4,530	3,800	. . .	1,136	. . .	25.1	. . .	29.9
65 and over	1,463	1,241	. . .	353	. . .	24.1	. . .	28.4
NHANES 2003–2004	10,122	8,522	. . .	4,020	. . .	39.7	. . .	47.2
Under 18	4,502	3,677	. . .	2,451	. . .	54.4	. . .	66.7
18–64	4,126	3,549	. . .	1,200	. . .	29.1	. . .	33.8
65 and over	1,494	1,296	. . .	369	. . .	24.7	. . .	28.5
National Nursing Home Survey								
NNHS 2004[3]	13,506	13,179	9,029	9,827	66.9	72.8	68.5	74.6
Under 65	1,567	1,518	1,280	1,343	81.7	85.7	84.3	88.5
65 and over	11,939	11,661	7,749	8,484	64.9	71.1	66.5	72.8
Longitudinal Study of Aging								
LSOA II[4]	9,447	7,390	. . .	1,782	. . .	18.9	. . .	24.1

. . . Category not applicable.

[1]Medicaid Analytic eXtract Person Summary.

[2]NHANES I Epidemiologic Followup Study.

[3]Missing age at interview ($n = 1$). Total survey population count is less than reported by health survey due to lack of age data for match rate tables.

[4]All participants were aged 70 and over at time of interview.

NOTES: NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; NNHS is National Nursing Home Survey; and LSOA is Longitudinal Study of Aging. NHANES III was conducted in 1988–1994, and NHEFS in 1971–1975, with continued follow-up of the NHEFS population in 1982–1984, 1986, 1987, and 1992. LSOA II was conducted in 1994, with continued follow-up of the LSOA II population in 1997–1998 and 1999–2000.

Table 5. Unweighted percentage of NCHS–Medicare linked sample of ages 65 and over at time of interview enrolled in managed care (for at least 1 month during year) for surveys containing 1991–2007 Medicare enrollment and claims data, by survey

Survey	Medicare enrollment and claims year																
	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
NHIS 1994	7.1	7.8	8.5	9.8	11.9	14.6	17.6	19.9	20.6	20.6	19.2	18.0	17.1	17.1	18.3	20.7	22.5
NHIS 1995	6.9	7.6	8.4	9.8	12.1	15.0	17.8	19.9	20.6	20.2	18.7	17.3	16.2	16.5	17.4	20.4	22.5
NHIS 1996	8.0	8.5	9.1	10.3	12.7	15.3	18.2	20.5	21.3	20.8	19.1	17.4	16.0	16.6	17.7	20.7	22.8
NHIS 1997	6.4	7.1	7.5	8.6	11.0	14.0	17.0	19.4	19.8	19.6	17.8	16.3	15.4	15.8	17.0	20.5	23.3
NHIS 1998	7.6	8.4	8.8	10.0	12.4	14.9	18.1	20.8	21.5	21.1	19.3	16.8	15.9	16.4	17.8	21.1	24.5
NHANES III	6.3	7.0	8.2	9.2	11.9	14.3	16.2	19.2	21.2	20.8	19.7	17.8	17.4	17.9	18.5	21.7	24.5
LSOA II	7.3	8.0	8.5	9.8	11.4	13.8	16.4	18.7	19.5	19.5	18.3	17.1	15.9	16.2	17.4	19.8	21.2

NOTES: NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; and LSOA is Longitudinal Study of Aging. NHANES III was conducted in 1988–1994; LSOA II was conducted in 1994, with continued follow-up of the LSOA II population in 1997–1998 and 1999–2000.

Table 6. Unweighted percentage of NCHS–Medicare linked sample of ages 65 and over at time of interview enrolled in managed care (for at least 1 month during year) for surveys containing 1999–2007 Medicare enrollment and claims data, by survey

Survey	Medicare enrollment and claims year								
	1999	2000	2001	2002	2003	2004	2005	2006	2007
NHIS 1999	21.2	21.1	19.3	16.8	15.8	16.1	17.2	20.7	23.2
NHIS 2000	21.9	21.8	19.5	17.7	16.9	17.5	19.1	22.4	25.3
NHIS 2001	21.1	20.6	18.4	16.2	15.6	15.6	17.0	20.5	23.4
NHIS 2002	22.3	21.6	19.5	17.4	16.7	16.5	18.3	21.9	24.5
NHIS 2003	21.8	20.9	18.4	16.2	15.2	15.8	17.5	21.1	23.2
NHIS 2004	21.2	20.4	18.7	16.0	15.3	15.6	17.4	21.8	25.0
NHIS 2005	20.4	19.9	17.9	15.6	14.8	14.8	16.5	20.4	23.8
NHANES 1999–2000	23.3	23.8	21.8	20.5	19.8	19.0	20.1	26.5	30.1
NHANES 2001–2002	23.7	23.0	20.2	19.0	19.1	19.7	20.8	23.7	26.9
NHANES 2003–2004	29.5	28.5	26.0	23.4	22.4	22.4	22.3	27.2	29.2
NNHS 2004	12.0	11.3	10.0	8.6	7.8	7.2	6.7	9.2	9.9

NOTE: NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; and NNHS is National Nursing Home Survey.

Table 7. Unweighted percentage of NCHS–Medicaid linked sample enrolled in Medicaid managed care programs for at least 1 month or full year, during 1999–2009 Medicaid Analytic eXtract data years, by survey

Survey	MAX data (months)																					
	1999		2000		2001		2002		2003		2004		2005		2006		2007		2008		2009	
	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months	1 month or more	12 months
NHIS 1994	37.9	19.0	35.4	18.6	34.4	19.3	35.3	16.5	30.7	16.7	29.8	16.2	30.2	16.2	31.4	16.1	32.6	16.1	32.1	16.8	33.3	18.8
NHIS 1995	38.3	19.6	37.1	19.1	36.7	20.0	36.8	17.8	33.6	17.9	33.0	17.6	33.6	18.4	33.9	17.2	34.6	16.7	33.4	17.5	35.4	19.2
NHIS 1996	39.0	20.1	37.3	19.8	36.9	20.0	37.7	17.6	34.8	19.1	34.6	18.7	34.8	20.0	35.6	18.8	35.9	17.8	36.2	19.0	38.7	21.5
NHIS 1997	41.2	20.0	39.1	20.0	37.8	20.9	38.9	17.8	35.5	18.7	35.0	18.2	35.1	19.0	35.8	18.1	37.2	18.3	36.5	18.9	39.0	20.5
NHIS 1998	42.3	21.4	40.7	21.7	39.1	22.2	40.2	19.3	37.7	20.3	36.8	19.0	36.9	19.3	36.7	18.3	38.2	19.1	38.5	20.4	40.8	22.1
NHIS 1999	41.7	19.6	40.4	19.3	39.4	21.1	41.0	19.1	37.2	20.0	37.2	19.4	37.5	20.2	38.4	18.8	40.6	19.9	39.9	21.3	42.2	23.2
NHIS 2000	42.8	20.9	42.6	20.2	41.8	22.0	43.4	20.1	40.4	21.4	40.2	20.8	39.8	21.4	41.2	20.7	41.3	21.5	42.6	21.3	45.0	23.9
NHIS 2001	43.2	20.3	42.1	20.6	43.4	21.7	44.0	20.7	40.3	21.3	39.3	20.5	38.7	20.2	40.3	19.7	42.6	20.1	43.7	22.1	46.2	24.9
NHIS 2002	45.0	20.5	44.1	20.6	43.7	21.7	44.8	20.5	42.3	21.9	41.3	21.6	40.8	22.3	42.8	21.2	44.4	20.5	44.1	22.8	46.2	24.3
NHIS 2003	42.9	18.9	42.4	19.0	41.8	21.5	44.1	19.8	42.7	21.5	42.2	22.4	42.1	22.6	45.1	22.0	46.1	22.8	45.8	22.2	48.0	25.2
NHIS 2004	42.8	19.4	41.9	20.3	42.5	20.9	43.9	19.3	42.0	21.3	41.5	21.2	42.4	21.8	43.8	20.9	44.3	20.8	44.9	22.1	46.8	25.2
NHIS 2005	45.3	21.6	44.1	20.9	42.7	20.9	45.4	18.8	42.0	21.2	41.5	21.3	42.3	21.8	45.1	21.6	46.0	21.7	45.6	22.9	48.1	26.3
NHEFS[1]	17.5	10.7	16.3	12.4	16.5	11.7	15.6	4.1	7.0	4.1	7.0	4.2	8.1	5.1	8.6	4.7	10.4	4.9	11.1	6.7	15.1	9.2
NHANES III	36.5	19.4	35.3	19.3	33.7	18.4	33.8	16.9	31.6	17.8	30.8	16.7	29.3	17.2	30.5	14.0	29.1	14.6	28.2	14.4	28.1	15.6
NHANES 1999–2000	35.2	16.4	37.3	17.5	37.2	18.0	39.5	19.2	40.1	21.7	41.4	23.4	41.8	25.1	43.5	22.8	42.4	21.2	43.1	21.9	45.5	24.3
NHANES 2001–2002	46.7	24.1	46.2	22.9	49.4	25.5	49.6	22.7	46.6	23.8	44.9	23.5	43.9	23.6	49.3	22.7	54.1	23.5	53.9	30.0	56.0	31.9
NHANES 2003–2004	54.4	30.1	53.6	28.6	52.9	29.6	55.1	27.5	53.0	28.9	50.9	27.6	50.0	29.5	50.0	27.6	54.1	28.0	54.1	30.3	58.1	33.1
NNHS 2004	10.0	6.8	9.1	6.2	8.4	5.9	7.6	3.0	4.4	2.7	4.2	2.7	4.2	2.5	4.2	2.7	5.0	2.6	6.4	3.4	7.8	6.0
LSOA II	9.0	5.4	8.3	4.9	7.8	4.7	6.9	2.2	3.4	2.1	4.4	2.9	5.4	3.4	5.0	3.0	6.1	3.6	7.2	4.9	8.4	5.1

[1]NHANES I Epidemiologic Followup Study.

NOTES: The number of months of enrollment in a Medicaid comprehensive managed care plan (Medicaid Analytic eXtract [MAX] element EL_PPH_PLN_MO_CNT_CMCP) was used to compute data shown. Data represent persons having only one MAX record per year. NHIS is National Health Interview Survey; NHANES is National Health and Nutrition Examination Survey; NNHS is National Nursing Home Survey; and LSOA is Longitudinal Study of Aging. NHANES III was conducted in 1988–1994, and NHEFS in 1971–1975, with continued follow-up of the NHEFS population in 1982–1984, 1986, 1987, and 1992. LSOA II was conducted in 1994, with continued follow-up of the LSOA II population in 1997–1998 and 1999–2000.

Appendix I. Social Security Administration Programs

Old Age, Survivors, and Disability Insurance

The Old Age, Survivors, and Disability Insurance (OASDI) program—Social Security—is the largest income-maintenance program in the United States. The program provides monthly benefits designed to replace, in part, the loss of income resulting from retirement, disability, or death. OASDI has two programs: the Old Age and Survivors Insurance (OASI) program and the Social Security Disability Insurance (DI) program. Coverage is nearly universal, with the two programs covering about 91% of U.S. jobs. Workers finance most of the program through a payroll tax levied under the Federal Insurance Contribution Act or FICA and the Self-Employment Contribution Act or SECA.

Nearly all work performed by citizens and noncitizens is covered if performed within the United States. In addition, the program covers work performed outside the United States by American citizens or residents employed by an American employer.

To qualify for Social Security, a person must be insured for benefits by acquiring a certain number of credits from earnings in covered employment. The number of credits needed depends on the worker's age and type of benefit. For DI entitlement, disability is defined as ''the inability to engage in any substantial gainful activity (SGA) by reason of any medically determinable physical or mental impairment that can be expected to result in death or that has lasted or can be expected to last for a continuous period of not less than 12 months.'' A person's age, education, and work experience are considered along with medical evidence in making a determination of disability. A less strict rule is provided for blind workers aged 55 and over. Such workers are considered disabled if, because of their blindness, they are unable to engage in SGA that requires skills and abilities comparable with those required in their past occupations.

The impairment must be of a degree of severity that renders the person unable to engage in any kind of substantial gainful work in the national economy, regardless of whether such work exists in the immediate area in which the person lives, whether a specific job vacancy exists for that person, or whether that person would be hired upon application for the work. Unlike the OASI program, which is entirely federally administered, the law mandates federal–state cooperation in carrying out the DI program. Each state's Disability Determination Services (DDS) evaluates the medical evidence and makes an initial determination of disability based on federal criteria, after the Social Security Administration (SSA) determines that the applicant is insured for benefits. The federal government reimburses states for DDS costs.

Monthly retirement benefits are payable at age 62 but are permanently reduced if claimed before the normal retirement age (NRA). Before 2000, NRA was age 65; it will gradually increase to age 67. Benefits may also be payable to the spouse and children of retired-worker beneficiaries. A spouse receives benefits at age 62, or at any age if he or she is caring for a child under age 16 or a disabled child. A divorced spouse aged 62 or over married to the worker for at least 10 years is also entitled to benefits. Benefits are payable to unmarried children under age 18, or aged 18 or 19 if they attend school full time.

Monthly benefits are payable to disabled workers after a 5-month waiting period, because DI is not intended to cover short-term disabilities. Benefits terminate if the beneficiary medically improves or returns to work (at an SGA level) despite the impairment. Upon reaching NRA, beneficiaries are transferred to the retirement program. Benefits for family members of a disabled worker are payable under the same conditions as those for retired workers. Monthly benefits are payable to survivors of a deceased worker.

Supplemental Security Income

The Supplemental Security Income (SSI) program consists of two parts: the federal program, and state supplementation of the federal payment. Eligibility requirements and benefit payments for the federal program are uniform across states. Almost 90% of SSI payments are federal. State supplementation can be federally or state administered. States with state-administered programs establish their own eligibility criteria and issue their own checks, but these payments are not captured on the Supplemental Security Record file.

Basic requirements for SSI eligibility involve citizenship, age, disability, income, and financial resources. SSI eligibility is generally restricted to U.S. citizens; however, eligibility is still possible for noncitizen members of certain groups of refugees, military personnel and their families, and some permanent residents. Persons must be at least 65 years old. Qualifying standards for SSI payments based on disability are almost the same as those used for the DI program. Persons do not have to be entirely without income to be eligible for SSI benefits. The maximum SSI payment is reduced dollar-for-dollar to reflect other income and in-kind support and maintenance. In most cases, ownership of financial assets is limited to $2,000 for an individual or $3,000 for a couple.

More information about SSA programs can be found at: http://www.ssa.gov.

Appendix II. Centers for Medicare & Medicaid Services Programs

Medicare

Medicare was created in 1965 under Title XCVIII of the Social Security Act to provide health insurance for people aged 65 and over. In 1972, Medicare eligibility was expanded to younger people with permanent disabilities who receive Social Security Disability Insurance and people of all ages with end-stage renal disease (ESRD), permanent kidney failure requiring dialysis, or a kidney transplant.

Medicare consists of three parts: hospital, medical, and prescription drug insurance. Medicare Part A (hospital insurance) helps cover inpatient care in hospitals, including critical access hospitals, and skilled nursing facilities (not including custodial or long-term care). It also helps cover hospice care and some home health care. Beneficiaries must meet certain conditions to get these benefits. Most people do not pay a premium for Part A because they or a spouse paid for it through their payroll taxes while working. However, they generally pay deductibles and coinsurance and may pay copayments.

Medicare Part B (medical insurance) helps cover doctors' services and outpatient care. It also covers some medical services that Part A does not cover, such as certain physical and occupational therapy and home health care when considered medically necessary. Most people pay a monthly premium for Part B as well as deductibles, coinsurance, and copayments.

Medicare Part C (Medicare Advantage), created in 1997, is a type of Medicare health plan offered by a private company that contracts with Medicare to provide Part A and Part B benefits. Medicare Advantage plans include health maintenance organizations, preferred provider organizations, private fee-for-service plans, special needs plans, and Medicare medical savings account plans. The services provided through Medicare Advantage are covered through the plan and are not paid for under original Medicare. Most Medicare Advantage plans offer prescription drug coverage.

Medicare Part D (prescription drug insurance) became available January 1, 2006, to all Medicare enrollees. Private companies provide various drug insurance plans, and beneficiaries pay a monthly premium for the plan they choose. Beneficiaries who do not enroll in a drug plan when they are first eligible may pay a penalty if they choose to join later.

Medicare funds the vast majority of medical residency training in the United States. Medicare financing covers residents' salaries and benefits through Direct Medical Education payments. Medicare also funds Indirect Medical Education, a subsidy paid to teaching hospitals in exchange for training resident physicians.

Some Medicare beneficiaries, known as ''dual eligibles,'' are also eligible for Medicaid benefits or for a Medicare Savings Program. These beneficiaries have limited income and resources, meeting financial criteria for Medicaid coverage. More than one-third of dual eligibles are people under age 65 with disabilities. Dual eligibles qualify for Qualified Medicare Beneficiary benefits in which Medicare Part A and B premiums are covered by Medicaid, as are deductibles, coinsurance, and copayments.

Medicaid

Medicaid is a means-tested entitlement program that provides health care coverage to some vulnerable U.S. populations, including low-income children and the aged or disabled poor. The program was enacted in 1965 by Title XIX of the Social Security Act and is the third-largest source of health care spending in the United States after Medicare and employer-provided health insurance. Medicaid accounts for almost one-sixth of national spending on personal health care. The number of Medicaid enrollees is greater than the number enrolled in Medicare.

Medicaid is the main payer of nursing home care and long-term care services overall; it is also the largest source of public funding for mental health care. Health centers and safety-net hospitals that serve low-income and uninsured people rely heavily on Medicaid. Seniors and people with disabilities make up approximately 25% of all Medicaid enrollees but account for two-thirds of Medicaid benefit expenditures.

Medicaid is administered by states under guidelines established by the federal government and is financed jointly by federal and state funds. The Federal Medical Assistance Percentage (FMAP), also called the federal match rate, represents the percentage of Medicaid financed by the federal government in each state. FMAP differs by state and takes into account the average per capita income in a state relative to the national average. The 2014 FMAP for individual states can be found at: http://aspe.hhs.gov/health/reports/2014/FMAP2014/fmap2014.cfm. FMAPs for earlier years can be found at: http://aspe.hhs.gov/health/fmap.cfm.

State Medicaid programs must cover mandatory services specified in federal law to receive federal matching funds. Beneficiaries are entitled to the following mandatory services:

- Physicians' services
- Hospital services (inpatient and outpatient)
- Laboratory and X-ray services
- Early and periodic screening, diagnostic, and treatment services for persons under age 21
- Federally qualified health center and rural health clinic services
- Family planning services and supplies
- Pediatric and family nurse practitioner services
- Nurse midwife services
- Nursing facility services for persons aged 21 and over
- Home health care for persons eligible for nursing facility services
- Transportation services
- Medicaid long-term care services

States are permitted to cover services that federal law designates as optional, including dental services, prescription drugs, case management, and hospice services. State variation in Medicaid coverage, with regard to both program eligibility and covered services, results in state differences in enrollment rates and expenditures. Other factors, including the age distribution, poverty rate, and Medicaid provider reimbursement rates, also contribute to variation among states in enrollment, service use, and costs. As a result, Medicaid operates as more than 50 distinct programs—one in each state, the District of Columbia, and each of the territories.

To qualify for Medicaid, a person must belong to one of the mandatory eligibility groups based on age and poverty level as well as, in some groups, eligibility for other programs. Federal law requires states to cover certain mandatory eligibility groups, but the states have broad flexibility to determine their own methods for assessing income and may impose asset tests. Prior to 2014, states were required to cover the following mandatory eligibility groups:

- Pregnant women with family income below 133% of the federal poverty level (FPL)
- Infants (up to age 1 year) born to Medicaid-eligible pregnant women
- Children under age 6 years with family income below 133% FPL
- Children aged 6–18 with family income below 100% FPL
- Limited-income families with dependent children who meet states' July 1996 welfare eligibility levels (often below 50% FPL)
- Supplemental Security Income (SSI) recipients

FPL cutoffs for children and parents vary across states and over time. Most elderly persons and persons with disabilities who receive SSI, and most low-income elderly or disabled Medicare beneficiaries, are eligible for Medicaid. Medicare beneficiaries must meet state income and assets criteria. Under the terms of the Patient Protection and Affordable Care Act of 2010, Medicaid was to be expanded in 2014 to include nearly everyone under age 65 with income up to 138% of FPL (income cutoff of 133% plus a 5% disregard). However, in 2012, the U.S. Supreme Court ruled that state participation in expanded coverage of Medicaid was optional. As of this report, 27 states and the District of Columbia had expanded Medicaid.

Some Medicaid enrollees may be enrolled in both Medicaid and Medicare and are referred to as dual eligible beneficiaries, or ''dual eligibles.'' More information about dual eligibles can be found through Centers for Medicare & Medicaid Services (CMS) at: https://www.cms.gov/Medicare-Medicaid-Coordination/Medicare-MedicaidCoordination.html.

Health care through Medicaid is delivered through fee-for-service and managed care programs. Medicaid managed care programs are insurance plans in which a health care organization provides a defined bundle of health services for a fixed monthly fee paid by the state's Medicaid program. States use an array of different types of managed care arrangements in Medicaid. Medicaid managed care plans include comprehensive plans that cover most (but not necessarily all) enrollee health services. Other plans provide more limited services; coverage varies greatly by plan. Primary care case management plans are the least comprehensive managed care type. Since the 1990s, state Medicaid programs have increasingly relied on managed care to organize and deliver services. Between 1999 and 2009, managed care enrollment rose from 56% to 75% of the Medicaid population.

Children's Health Insurance Program

Implemented in 1997 via Title XXI of the Social Security Act, Children's Health Insurance Program (CHIP) is jointly financed by the federal and state governments and is administered by the states. It targets low-income uninsured children and pregnant women in families with incomes too high to qualify for most state Medicaid programs. Within broad federal guidelines, each state determines its program's design, eligibility groups, benefits, payment levels, and administrative and operating procedures. The federal government provides funding to states for CHIP but caps the amount of funds provided on a matching basis.

Each state has the option of expanding Medicaid eligibility to children who previously had been ineligible due to their income (M–CHIP), or creating a program distinct from its existing Medicaid program (S–CHIP). The federal government matches state spending for CHIP at a higher rate than that for Medicaid. As of 2009, more than 30 states had an S–CHIP program, either by itself or in combination with an M–CHIP program in the same state. S–CHIP programs are allowed to tailor their programs to a greater extent than M–CHIP programs. Data availability is different for M–CHIP and S–CHIP recipients. CHIP was reauthorized in 2009 and has been extended since that time, but authorization for CHIP is set to expire in 2015.

More information about CMS programs can be found at: http://www.cms.gov.

Appendix III. Adjusting Sample Weights for Linkage-eligibility Using SUDAAN

This appendix provides an example of adjusting sample weights to correct for linkage ineligibility. For National Center for Health Statistics (NCHS) data files linked with Centers for Medicare & Medicaid Services (CMS) and Social Security Administration data files, linkage ineligibility refers to a survey respondent who, via action or inaction, excludes himself or herself from the population of persons whose data can be linked to administrative records; linkage eligibility is described in the main report. More detailed information about adjustment methods can be found in an earlier version of this appendix (8), which describes alternative models for adjustment.

For its population health surveys, NCHS creates sample weights based on probabilities of selection into the survey, with adjustments for nonresponse and census poststratification. Post-stratification involves adjusting the sample weights within specific domains to match corresponding population totals. Users of NCHS surveys are encouraged to use these sample weights in their calculations. However, only linkage-eligible survey respondents can be used when analyzing NCHS linked data, not the full sample. Survey respondents who provide sufficient personally identifiable information for linkage are not a random sample of respondents (9). As with other forms of nonresponse, biased analyses of linked data may result when respondents who are linkage eligible differ systematically from respondents who are not linkage eligible for some characteristics (10).

Of critical importance is whether linkage ineligibility varies by sample subsets and whether responses to outcomes of interest vary across these groups (11). Nonresponse weighting adjustments can be effective for handling linkage ineligibility if the nonresponse weighting adjustment is related to the variables that influence the probability of linkage ineligibility and to the variables being studied.

For developing and testing reweighting models, PROC WTADJUST in SUDAAN software (12) is useful. This SUDAAN procedure is designed to work with complex surveys and is flexible enough to implement different approaches to reweighting. As shown below, multiple approaches can also be easily applied and compared. NCHS uses a SAS-callable version of SUDAAN, but a stand-alone version is also available.

This appendix has four sections: The first section begins with a brief overview of SUDAAN's PROC WTADJUST; next, the data files used in the example are described; then, an example using PROC WTADJUST is presented; and in the final section, some suggestions are provided for examining the adjusted sample weights, to enhance confidence that the weight adjustment model chosen will lead to valid inferences.

User familiarity with standard SAS program language conventions (e.g., libnames and filenames) is assumed for the example that follows. Users are advised to read the SUDAAN documentation (12) for a detailed understanding of the WTADJUST procedure and its implementation. Although an example of its use is shown below, the best approach for a particular project may require different applications of the procedure. Furthermore, other software can be used to reweight survey data (13), and other methods can be used to address the issue of linkage eligibility (14).

PROC WTADJUST

SUDAAN is a statistical package designed to correctly handle data analysis for data from complex sample designs. Within SUDAAN, PROC WTADJUST is a module designed specifically for nonresponse and poststratification adjustments. Weight adjustments are created using a model-based calibration approach. There are two ways to use PROC WTADJUST: 1) to correct for nonresponse using a model-based approach (nonresponse option), and 2) to directly poststratify to external control totals (poststratify option). Control totals are population estimates calculated for specific cross-categories of subgroups, often defined by race and ethnicity, age, and sex categories. If the variables that were used to define control totals for the original sample weights are used in PROC WTADJUST, then applying the nonresponse option and fitting a model using all cross-categories of the variables defining the control totals will produce the same adjusted sample weights as using the poststratify option with those control totals.

Data Files Used in Example

Feasibility files

Publicly available feasibility data files for the linkages of NCHS survey data to CMS Medicare claims can be downloaded directly from the NCHS website. The NCHS–CMS Medicare feasibility files provide a limited set of variables that can be used to determine the maximum available sample size for each linked file and assess the potential impact of linkage-ineligible records. These files are especially useful to researchers considering whether to initiate a Research Data Center proposal to analyze the restricted-use linked NCHS–CMS Medicare files. Each feasibility file is NCHS survey- and survey year-specific.

The following information is included in each feasibility study file: NCHS public-use data file identifier (for the National Health Interview Survey [NHIS], this is called PUBLICID), survey respondent eligibility and final match status (CMS_MEDICARE_MATCH), and variables specifying which CMS Medicare data files contain information on the successfully linked survey respondent. The feasibility study

files do not contain any specific information about CMS Medicare benefits. Data users need to use information from the public-use survey or external information to approximate the number of respondents with a specific condition.

The NCHS–CMS Medicare feasibility file can be found at: http://www.cdc.gov/nchs/data_access/data_linkage/cms/cms_medicare_feasibility.htm. For this example, the feasibility file for the 2005 NHIS is placed into SAS library NHIS05 and referred to as NHIS05.FEAS. The linkage eligibility flag, CMS_MEDICARE_MATCH, takes on the following values: one for linkage eligible and linked; two for linkage eligible and not linked; three for linkage eligible but reached age 18 during the administrative follow-up period; and nine for linkage ineligible. The first three values are considered linkage eligible as a group.

2005 NHIS public-use person file

The public-use 2005 NHIS person file can be found at: http://www.cdc.gov/nchs/nhis/nhis_2005_data_release.htm.

From the 2005 NHIS person file, the following public-use variables are used: PUBLICID, AGE_P (respondent's age at survey interview), HISCODI2 (race and Hispanic origin, which is renamed "RACEETH" to be more descriptive), SEX, EDUC1 (education), PHSTAT (which is recoded into a logical variable FAIRPOORHEALTH for assessment), REGION (West, Northeast, South, Midwest), WTFA (the original design-based sample weight), STRATUM, and PSU (primary sampling unit).

In this example, the variable PUBLICID, which is on both the feasibility file and the 2005 NHIS person file, is used. Because public-use NHIS files have had varying variable names in different years, PUBLICID may need to be created in a year-specific way. More information is available from: http://www.cdc.gov/nchs/data/datalinkage/important_information_on_merging_nchs_restricted_and_public_use_survey_data.pdf.

To create the data file, sort and merge the files and code the variables. For this example, the input data files and created data files were kept in SAS library NHIS05. The public-use 2005 NHIS person file is called NHIS05.PERSONSX.

```
/*DEFINE FORMATS FOR LATER USE/*
  PROC FORMAT ;
    VALUE EDUCF
    1 = "< High school"
    2 = "High school or GED or some college"
    3 = "College degree" ;
    VALUE AGECATF
    1 = "UNDER 18 YEARS"
    2 = "18–44 YEARS"
    3 = "45–64 YEARS"
    4 = "65 YEARS OR OLDER" ;
    VALUE REGIONF
    1 = "NORTHEAST"
    2 = "MIDWEST"
    3 = "SOUTH"
    4 = "WEST" ;
    VALUE RACEETHF
    1 = "HISPANIC"
    2 = "NON-HISPANIC WHITE"
    3 = "NON-HISPANIC BLACK"
    4 = "ALL OTHER RACES AND ETHNICITIES" ;
    VALUE FAIRPOORF
    1 = "FAIR OR POOR HEALTH"
    0 = "GOOD, VERY GOOD, EXCELLENT HEALTH"
    . = "MISSING" ;
/*STEP 1: MERGE THE FEASIBILITY AND PUBLIC USE DATA FILES.*/
PROC SORT DATA=NHIS05.FEAS ; BY PUBLICID ;
PROC SORT DATA=NHIS05.PERSONSX ; BY PUBLICID ;
DATA NHIS05.MERGED NOTMERGED1 NOTMERGED2 ;
  MERGE
    NHIS05.FEAS (IN=A KEEP=PUBLICID CMS_MEDICARE_MATCH)
    NHIS05.PERSONSX (IN=B KEEP=PUBLICID AGE_P HISCODI2 PHSTAT SEX REGION EDUC1 WTFA STRATUM PSU) ;
  BY PUBLICID ;
  IF A=1 AND B=1 THEN OUTPUT NHIS05.MERGED ;
  IF A=0 AND B=1 THEN OUTPUT NOTMERGED1 ;
  IF A=1 AND B=0 THEN OUTPUT NOTMERGED2 ;
/*CHECK THE LOG FILE. NO RECORDS SHOULD BE OUTPUT TO NOTMERGED1 OR NOTMERGED2. */
/*STEP 2: DEFINE THE LINKAGE-ELIGIBILITY (MISSING DATA) INDICATOR, LINKABLE. RECODE THE VARIABLES. CREATE VARIABLE FORMATS. */
DATA NHIS05.MERGED ; SET NHIS05.MERGED ;
  LINKABLE=CMS_MEDICARE_MATCH IN (1,2,3) ; /*DEFINES LINKAGE-ELIGIBILTY /*
  FAIRPOORHEALTH=(PHSTAT IN (4,5)) ;
  IF PHSTAT=7 OR PHSTAT=8 OR PHSTAT=9 THEN FAIRPOORHEALTH=. ;
  /*DEFINE THE FAIR OR POOR HEALTH VARIABLE */ ;
  AGE_CAT=. ;
  IF 0 LE AGE_P LE 17 THEN AGE_CAT=1 ;
  IF 18 LE AGE_P LE 44 THEN AGE_CAT=2 ;
  IF 45 LE AGE_P LE 64 THEN AGE_CAT=3 ;
  IF AGE_P > 64 THEN AGE_CAT=4 ;
  EDUC=. ;
  IF EDUC1 IN (0,1,2,3,4,5,6,7,8,9,10,11) THEN EDUC=1 ; < HIGH SCHOOL ;
  ELSE IF EDUC1 IN (12,13,14,15,16,17)
  THEN EDUC=2 ; HIGH SCHOOL OR GED OR SOME COLLEGE ;
  ELSE IF EDUC1 IN (18,19,20,21) THEN EDUC=3 ; COLLEGE DEGREE ;
  /*CONVERT PUBLICID TO NUMERIC FOR USE IN SUDAAN /*
  ID=PUBLICID*1 ;
  RACEETH=HISCODI2; * CREATES A MORE DESCRIPTIVE VARIABLE NAME ;
FORMAT AGE_CAT AGECATF. RACEETH RACEETHF. SEX SEXF. REGION REGIONF. EDUC EDUCF. FAIRPOORHEALTH FAIRPOORF. ;
```

Example

Marginal model

The marginal, or main effects, model is fitted using AGE_CAT, RACEETH, SEX, REGION, and EDUC. By definition of a marginal model, in this model no interaction terms are specified. The factors REGION and EDUC (education, in three categories)

are often related to the propensity to agree to linkage as well as the underlying survey design, where strata are formed based on geographic and socioeconomic characteristics of locations within the United States. After fitting the model, using the output statement, the adjusted weights produced by the model are saved in a new temporary SAS dataset called MATCH1. Next, MATCH1 is merged into the original files, NHIS05.MERGED, so the adjusted weights can be evaluated and used in an analysis.

```
/*MARGINAL MODEL: THIS MODEL
INCLUDES MAIN EFFECTS ONLY/*

PROC WTADJUST DATA=NHIS05.
MERGED DESIGN=WR

ADJUST=NONRESPONSE NOTSORTED ;
   NEST STRATUM PSU ;
   WEIGHT WTFA ;
   CLASS AGE_CAT HISCODI2 SEX
   REGION EDUC / INCLUDE=MISSING;
   REFLEVEL AGE_CAT=2 RACEETH=2
   SEX=1 REGION=1 EDUC=2 ;
   MODEL LINKABLE=AGE_CAT
   RACEETH SEX REGION EDUC ;
   IDVAR LINKABLE AGE_CAT
   RACEETH SEX REGION EDUC ID ;
   PRINT BETA SEBETA P_BETA
   MARGADJ / BETAFMT=F10.4
   SEBETAFMT=F10.4 ;
   OUTPUT /PREDICTED=ALL
   FILENAME=MATCH1 FILETYPE=SAS
   REPLACE ;

RUN ;

PROC SORT DATA=MATCH1 ; BY ID ;

RUN ;

PROC SORT DATA=MERGED ; BY ID ;

RUN ;

/*MERGE THE WEIGHT FILE TO THE
MERGED FILE. RENAME ADJFACTOR
AND WTFINAL, WHICH ARE SUDAAN
INTERNAL VARIABLES, TO
ADJFACT_MARGINAL AND
WTFIN_MARGINAL, SO THAT WE CAN
COMPARE THE ADJUSTMENTS./*

DATA NHIS05.MERGED_WEIGHTED ;
   MERGE NHIS05.MERGED MATCH1
   (KEEP=ID ADJFACTOR WTFINAL) ;
   BY ID ;
   ADJFACT_MARGINAL=ADJFACTOR ;
   WTFIN_MARGINAL=WTFINAL ;

RUN;
```

Assessment of Adjusted Weights

Examine the adjusted weights before using them in analysis. Some approaches are illustrated here, although others exist. The basic SUDAAN model results from the marginal model are shown in the Table.

The default under SAS/SUDAAN is to define the last category as the baseline category and set the coefficient on baseline category to zero. The other coefficients are interpreted as deviations from the baseline category. The baseline can be changed from its default using the REFLEVEL statement, as illustrated above and subsequently (12). The /INCLUDE=MISSING option in the model code instructs SUDAAN to treat missing values as a separate class; thus, EDUC=MISSING gets its own parameter estimate. Although this approach is not generally recommended for statistical analysis, applying other methods (e.g., multiple imputation) or identifying the best way to include nonresponse in the models was beyond the scope of this project.

WTADJUST provides a marginal weight adjustment column, which indicates the average adjustment to the initial weights for records in each category. Examine these values to determine if any one category has undue influence. Because about 50% of respondents were ineligible for linkage in 2005, these weight adjustments generally are expected to be around two. Large differences in the marginal weight

Table. SUDAAN output for weight-adjusted marginal model

Independent variables and effects	β coefficient	SE β	Marginal weight adjustment
Intercept	–0.1693	0.0365	2.1289
AGE_CAT (years)			
Under 18	–0.6819	0.0381	1.9661
18–44 (ref)	. . .	. . .	2.1607
45–64	0.0563	0.0219	2.1891
65 and over	0.1389	0.0350	2.2897
RACEETH			
Hispanic	0.5940	0.0391	2.7480
Non-Hispanic white (ref)	. . .	. . .	1.9927
Non-Hispanic black	0.2767	0.0386	2.1864
Non-Hispanic other	0.4917	0.0617	2.6872
Sex			
Male (ref)	. . .	. . .	2.0730
Female	0.1169	0.0152	2.1854
Region			
Northeast (ref)	. . .	. . .	2.1764
Midwest	–0.2345	0.0429	1.9180
South	–0.2183	0.0423	2.0258
West	0.2873	0.0479	2.6268
EDUC			
Missing response	1.0530	0.0595	2.5872
Less than high school diploma	0.2888	0.0319	2.0303
High school or GED[1], or some college (ref)	. . .	. . .	2.0715
College degree	0.1539	0.0279	2.2230

. . . Category not applicable.

[1]General Educational Development high school equivalency diploma.

NOTES: β coefficient is the estimated effect on adjustment; SE β is the standard error of the estimated effect; marginal weight adjustment is the effect of the category on the original annual weights; and (ref) is the reference group used when comparing each cofactor.

SOURCES: CDC/NCHS, public-use National Health Interview Survey Person File merged with NCHS-Medicare linked Feasibility Data, 2005.

adjustment across groups are to be avoided, because they can increase variance estimates.

After inspecting the results provided by SUDAAN, additional examinations can be performed. For example, the adjustment cell sizes should be examined; a cell size less than 30 is not generally recommended (11). In addition, correlations and scatter plots of adjusted and unadjusted weights should be examined, to qualitatively identify outliers or other anomalies in the adjustment process.

When assessing the results from the WTADJUST procedure, records not eligible for record linkage will have a positive WTFA and a zero for WTFIN_MARGINAL as well as ADJFACT_MARGINAL. These zeroes for the adjusted weights will distort summary statistics, correlations, and graphics, so the zeroes should be removed before performing these steps. The following code implements some of these checks:

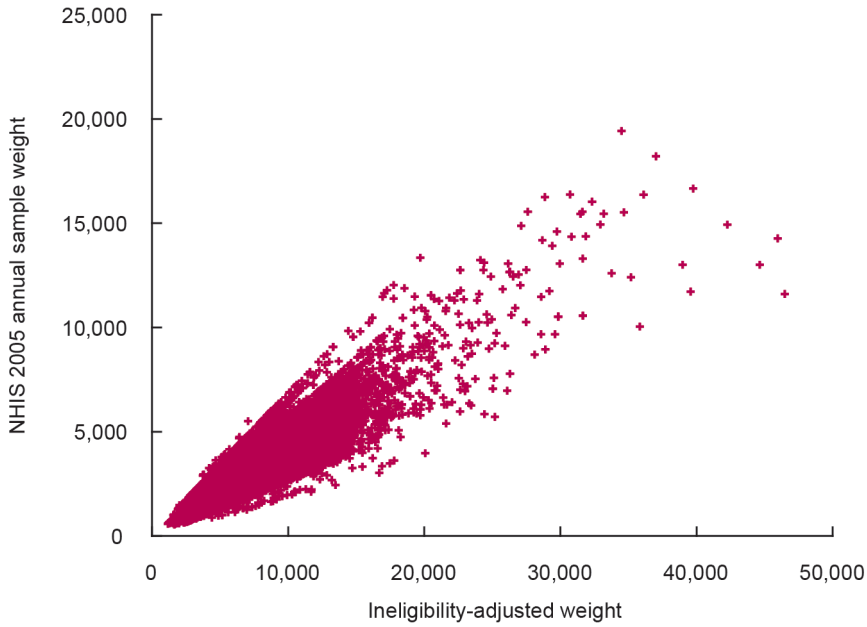
```
DATA NHIS05.MERGED_WEIGHTED ;
   SET NHIS05.MERGED_WEIGHTED ;
   IF WTFIN_MARGINAL=0 THEN
   WTFIN_MARGINAL=. ;
   IF ADJFACT_MARGINAL=0 THEN
   ADJFACT_MARGINAL=. ;

RUN ;
TITLE "ANY ZERO WEIGHT HAS BEEN
CONVERTED TO MISSING" ;

PROC MEANS DATA=NHIS05.MERGED_
WEIGHTED N SUM MIN MAX VAR
SKEW KURT;
   VAR WTFA WTFIN_MARGINAL
   ADJFACT_MARGINAL;

RUN;
```

Based on the results of the PROC MEANS statement above, the original annual weight, WTFA, sums to 291,143,602 persons, and adjusted weights do the same. WTFA has a maximum value of 19,434, and the adjusted weights have a higher maximum (46,481 to 51,772), or about twice as much, reflecting the 50% linkage ineligibility. Variance terms for adjusted weights are larger than the original weight, also reflecting this fact. The kurtosis measures the peakedness and heavy tailedness of a distribution; these results show that the adjusted weights are more peaked and heavier-tailed than the original weight. All of the adjusted weights are relatively close to the same value, but in the analytic phase, some use of various influence statistics would be advisable. Additional code implements other checks:



NOTE: Linkage-ineligible records receive a final weight of zero and are not plotted. NHIS is National Health Interview Survey.
SOURCE: CDC/NCHS, public-use 2005 NHIS Person File merged with linked NHIS-Medicare feasibility data.

Figure. Plot of National Health Interview Survey annual sample weight compared with ineligibility-adjusted weight

```
PROC CORR DATA=NHIS05.MERGED_
WEIGHTED NOSIMPLE ;
   VAR ADJFACT_MARGINAL;
   WHERE WTFIN_MARGINAL> > 0;

RUN ;

PROC CORR
DATA=NHIS05.MERGED_WEIGHTED ;
   VAR ADJFACT_MARGINAL ;
   WHERE WTFIN_MARGINAL> > 0;

RUN ;
```

Ideally, the Pearson correlation coefficient between the original and adjusted weights would be high; using the SAS code above, in this case it takes on the value 0.88. However, the Pearson correlation measures **linear** association and can be misleading if nonlinearities are present, which is why examining plots is desirable as well. A scatterplot of the original weight against the adjusted weight should be examined, as in:

```
PROC PLOT DATA=NHIS05.MERGED_
WEIGHTED ;
   PLOT WTFA*WTFIN_MARGINAL
   WHERE WTFIN_MARGINAL > 0;

RUN ;
```

As shown in the Figure, many values clump in the range of zero to about 15,000, with individual values reaching about 50,000 (thus, these cases may have greater influence in subsequent analytic inference). For these plots, records with high weights are generally singletons, and many more records with lower values are in the lower left corner.

To check for cell sizes, construct unweighted tables that reflect the properties of the chosen reweighting model. For the marginal model, only the margins need to be checked. The following code implements these checks:

```
TITLE "FREQUENCY TABULATION TO
CHECK CELL SIZES CELL SIZES FOR
MODELS" ;

PROC FREQ DATA=NHIS05.MERGED_
WEIGHTED ;
```

```
   WHERE WTFIN_MARGINAL> > 0;
   TABLES AGE_CAT RACEETH SEX
   REGION EDUC /NOPERCENT LIST ;
RUN ;
```

Using the SAS code above, no margin (or category) was close to the ''rule of thumb'' cutoff of 30 cases (not shown).

Finally, the variable FAIRPOORHEALTH is tabulated, which takes on the value 1 if the respondent reported ''poor'' or ''fair'' health; 0 if the respondent reported ''good,'' ''very good,'' or ''excellent'' health; and missing if no response was provided. These results indicate effects of linkage ineligibility. Using the original NHIS 2005 sample and the original sample weight, 9.30% of respondents had fair or poor health. The weighted percentages were elevated for the linkage-eligible subset at 10.43%, using the weights from the MARGINAL model. For comparison, unweighted percentages for the original NHIS 2005 sample were 9.79%, and unweighted percentages for the linkage-eligible subset were 10.49%, respectively.

Appendix IV. First NCHS–SSA–CMS Linkage: NSCL–I

The first linkage of National Center for Health Statistics (NCHS) survey respondents to their Social Security benefit history records and Medicare enrollment and claims records was initiated with NSCL–I in 2001 upon approval of the NCHS Research Ethics Review Board (ERB). NCHS ERB, also known as an institutional review board or IRB, is an administrative body of scientists and nonscientists established to protect the rights and welfare of human research subjects.

The process of linking NCHS survey data with Social Security data began at the Social Security Administration (SSA), where records for individual NCHS survey respondents were matched with SSA's Numident file. The Numident file is a numerically ordered master file for each Social Security number (SSN) ever issued and contains records for approximately 400 million SSNs, including personal identifying information.

To link NCHS survey respondents with their Social Security benefit histories, NCHS provided SSA with as many of the following individual identifiers that were available on the survey record for all eligible survey respondents:

- SSN
- 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 Numident file if they refused to provide their SSN at the time of the interview. Additional ineligibility criteria included refused, missing, or incomplete information on last name or date of birth.

The match process consisted of two steps. First, SSA verified whether the SSN received from NCHS was correct using the Enumeration Verification System (EVS). In cases where the SSN of a survey participant was missing or could not be verified, SSA used an enhanced EVS matching algorithm to try to determine the correct SSN. The enhanced EVS matching algorithm was developed by the SSA Office of the Actuary to apply additional identifying data elements collected during the survey interview and contained in other administrative records held by SSA. It features a scoring system with a threshold score used to determine which potential matches are acceptable, with the potential to increase the number of successful matches. For NCHS records determined to be a match with the Numident file, SSA extracted data, where available, from the benefit history files. Because not all survey participants matched to the Numident have Social Security benefit history data, the records with available benefit history data was less than the number matched to the Numident.

The next part of the process—linking NCHS survey data with Medicare enrollment and claims data—was performed by the Centers for Medicare & Medicaid Services (CMS). NCHS survey respondents were matched with Medicare's Enrollment Database (EDB). EDB is a master enrollment file of all people ever entitled to Medicare. EDB records are comprehensive and updated daily.

To link NCHS survey respondents with their Medicare data, NCHS provided CMS as many of the following individual identifiers as 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 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 EDB records based on on 1) HIC number, 2) SSN, or 3) name and date of birth. For these potential matches, NCHS employed a deterministic matching algorithm to verify which matches were correct. All potential matches were assigned a score based on whether the identifying information provided matched 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 an 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. Using 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 nonmatches. Matches falling between these thresholds were manually reviewed to determine the match status. For those 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 1991–2000. Because not all survey participants matched to the EDB had claims information for 1991–2000, the number of records with available Medicare claims data was less than the number matched to EDB.

Appendix V. List of Resources for NCHS–SSA–CMS–USRDS Linked Data Files

National Center for Health Statistics

National Center for Health Statistics (NCHS) website:
http://www.cdc.gov/nchs/index.htm

NCHS Data Linkage Activities:
http://www.cdc.gov/nchs/data_access/data_linkage_activities.htm

NHANES–CMS Linked Data Tutorial:
http://www.cdc.gov/nchs/tutorials/NHANES-CMS/index.htm

NCHS Research Data Center (RDC):
http://www.cdc.gov/rdc/

Social Security Administration

Social Security Administration (SSA) website:
http://www.ssa.gov/

Social Security Program Rules:
www.ssa.gov/regulations/index.htm

Online Social Security Handbook:
http://ssa.gov/OP_Home/handbook/handbook.html

Centers for Medicare & Medicaid Services

Centers for Medicare & Medicaid Services (CMS) website:
http://www.cms.gov/

Research Data Assistance Center (ResDAC) at the University of Minnesota:
http://www.resdac.org/

Medicare

Medicare and You Handbook:
https://www.medicare.gov/medicaare-and-you/medicare-and-you.html

Medicare Program—General Information:
http://www.cms.gov/Medicare/Medicare-General-Information/MedicareGenInfo/index.html

Medicare Prospective Payment Systems—General Information:
http://www.cms.gov/ProspMedicareFeeSvcPmtGen/

American Medical Association:
http://www.ama-assn.org/ama

Medicaid

Medicaid–Children's Health Insurance Program:
http://www.medicaid.gov/medicaid-chip-program-information/by-population/children/children.html

The Henry J. Kaiser Family Foundation:

State Health Facts
http://www.statehealthfacts.org/

Medicaid Benefits Data Collection
http://medicaidbenefits.kff.org/

National Pharmaceutical Council—Medicaid Pharmaceutical Plan Resources:
http://www.npcnow.org/Public/Issues/i_rel_research/Medicaid_Pharmaceutical_Plan_Resources.aspx

NCHS–Medicaid Analytic eXtract (MAX) linkage reports:

MAX and NCHS Survey Linkage Design Report (2010)
https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/MAX_NCHSLinkageReport.pdf

MAX and NCHS Survey Linkage, 1999–2009 (2012)
https://www.cms.gov/Research-Statistics-Data-and-Systems/Computer-Data-and-Systems/MedicaidDataSourcesGenInfo/Downloads/MAX_NCHS_Linkage_Report1999_2009.pdf

United States Renal Data System

United States Renal Data System (USRDS) website:
http://www.usrds.org/

Vital and Health Statistics Series Descriptions

ACTIVE SERIES

Series 1. **Programs and Collection Procedures**—This type of report describes the data collection programs of the National Center for Health Statistics. Series 1 includes descriptions of the methods used to collect and process the data, definitions, and other material necessary for understanding the data.

Series 2. **Data Evaluation and Methods Research**—This type of report concerns statistical methods and includes analytical techniques, objective evaluations of reliability of collected data, and contributions to statistical theory. Also included are experimental tests of new survey methods, comparisons of U.S. methodologies with those of other countries, and as of 2009, studies of cognition and survey measurement, and final reports of major committees concerning vital and health statistics measurement and methods.

Series 3. **Analytical and Epidemiological Studies**—This type of report presents analytical or interpretive studies based on vital and health statistics. As of 2009, Series 3 also includes studies based on surveys that are not part of continuing data systems of the National Center for Health Statistics and international vital and health statistics reports.

Series 10. **Data From the National Health Interview Survey**—This type of report contains statistics on illness; unintentional injuries; disability; use of hospital, medical, and other health services; and a wide range of special current health topics covering many aspects of health behaviors, health status, and health care utilization. Series 10 is based on data collected in this continuing national household interview survey.

Series 11. **Data From the National Health Examination Survey, the National Health and Nutrition Examination Surveys, and the Hispanic Health and Nutrition Examination Survey**—In this type of report, data from direct examination, testing, and measurement on representative samples of the civilian noninstitutionalized population provide the basis for (1) medically defined total prevalence of specific diseases or conditions in the United States and the distributions of the population with respect to physical, physiological, and psychological characteristics, and (2) analyses of trends and relationships among various measurements and between survey periods.

Series 13. **Data From the National Health Care Survey**—This type of report contains statistics on health resources and the public's use of health care resources including ambulatory, hospital, and long-term care services based on data collected directly from health care providers and provider records.

Series 20. **Data on Mortality**—This type of report contains statistics on mortality that are not included in regular, annual, or monthly reports. Special analyses by cause of death, age, other demographic variables, and geographic and trend analyses are included.

Series 21. **Data on Natality, Marriage, and Divorce**—This type of report contains statistics on natality, marriage, and divorce that are not included in regular, annual, or monthly reports. Special analyses by health and demographic variables and geographic and trend analyses are included.

Series 23. **Data From the National Survey of Family Growth**—These reports contain statistics on factors that affect birth rates, including contraception and infertility; factors affecting the formation and dissolution of families, including cohabitation, marriage, divorce, and remarriage; and behavior related to the risk of HIV and other sexually transmitted diseases. These statistics are based on national surveys of women and men of childbearing age.

DISCONTINUED SERIES

Series 4. **Documents and Committee Reports**—These are final reports of major committees concerned with vital and health statistics and documents. The last Series 4 report was published in 2002. As of 2009, this type of report is included in Series 2 or another appropriate series, depending on the report topic.

Series 5. **International Vital and Health Statistics Reports**—This type of report compares U.S. vital and health statistics with those of other countries or presents other international data of relevance to the health statistics system of the United States. The last Series 5 report was published in 2003. As of 2009, this type of report is included in Series 3 or another series, depending on the report topic.

Series 6. **Cognition and Survey Measurement**—This type of report uses methods of cognitive science to design, evaluate, and test survey instruments. The last Series 6 report was published in 1999. As of 2009, this type of report is included in Series 2.

Series 12. **Data From the Institutionalized Population Surveys**—The last Series 12 report was published in 1974. Reports from these surveys are included in Series 13.

Series 14. **Data on Health Resources: Manpower and Facilities**—The last Series 14 report was published in 1989. Reports on health resources are included in Series 13.

Series 15. **Data From Special Surveys**—This type of report contains statistics on health and health-related topics collected in special surveys that are not part of the continuing data systems of the National Center for Health Statistics. The last Series 15 report was published in 2002. As of 2009, reports based on these surveys are included in Series 3.

Series 16. **Compilations of Advance Data From Vital and Health Statistics**—The last Series 16 report was published in 1996. All reports are available online, and so compilations of Advance Data reports are no longer needed.

Series 22. **Data From the National Mortality and Natality Surveys**—The last Series 22 report was published in 1973. Reports from these sample surveys, based on vital records, are published in Series 20 or 21.

Series 24. **Compilations of Data on Natality, Mortality, Marriage, and Divorce**—The last Series 24 report was published in 1996. All reports are available online, and so compilations of reports are no longer needed.

For answers to questions about this report or for a list of reports published in these series, contact:

Information Dissemination Staff
National Center for Health Statistics
Centers for Disease Control and Prevention
3311 Toledo Road, Room 5419
Hyattsville, MD 20782

Tel: 1–800–CDC–INFO (1–800–232–4636)
TTY: 1–888–232–6348
Internet: http://www.cdc.gov/nchs
Online request form: http://www.cdc.gov/info
For e-mail updates on NCHS publication releases, subscribe online at: http://www.cdc.gov/nchs/govdelivery.htm.

U.S. DEPARTMENT OF HEALTH & HUMAN SERVICES

Centers for Disease Control and Prevention
National Center for Health Statistics
3311 Toledo Road, Room 5419
Hyattsville, MD 20782–2064

For more NCHS
Series Reports, visit:
http://www.cdc.gov/nchs/products/series.htm

DHHS Publication No. 2015–1334, Series 1, No. 58
CS257580