Analytic Issues in Using the Medicare Enrollment and Claims Data Linked to NCHS Surveys

I. General Notices to Users

This document provides additional information about Medicare enrollment & claims data that are linked to NCHS survey data. Users also should refer to the Matching Methodology report and the Description of NCHS-CMS Linkage document. This document is not an exhaustive or systematic review of the analytic issues researchers may encounter while using the NCHS-Medicare Linked Data Files and it will be updated as additional analytic issues are identified and brought to the attention of the NCHS Research Data Center or Data Linkage Unit. Users are encouraged to visit the ResDAC website www.resdac.umn.edu for more information on Medicare data.

The advantages of Medicare data are that they are population-based, not subject to recall bias, and can be linked to NCHS population health surveys to expand their analytic potential. However because Medicare data were collected for the purpose of making healthcare payments, and not for research, there are limitations to the data that researchers should consider when constructing their samples and conducting analyses.

II. Analytic issues

1. Importance of the Denominator File

All applications to the NCHS Research Data Center (RDC) should include a request for the Denominator File for the years that the researcher is examining claims data. The Denominator record contains basic demographic and enrollment information about each beneficiary entitled to Medicare during each calendar year and is needed to help construct your analytic sample, particularly to identify Medicare beneficiaries enrolled in a managed care plan (see below).

2. Exclusion of claims paid by a source other than Medicare (e.g., managed care organizations or HMOs)

CMS generally does not receive claims data for Medicare beneficiaries who enroll in managed care plans (including private fee-for-service plans paid on a capitation basis). Please note that exceptions to this do exist. For example, all Hospice claims are processed as Medicare claims regardless of whether the beneficiary is in a Fee for Service (FFS) or managed care plan. During the time covered by the linked database, enrollment in managed care increased from approximately 6% of beneficiaries in 1991 to 17% in 1999. A number of managed care plans withdrew from Medicare beginning in 1999, resulting in a decrease in enrollment in 2000 to 16% of beneficiaries.

In general, studies based on analysis of claims data should exclude managed care 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 managed care enrollees is to include them for the time period prior to entering a managed care plan and then censor them at the time they enter a managed care plan.

The following documents and citations provide detailed information about Medicare managed care enrollees and the Medicare utilization Files and how to address them in analyses:

- www.resdac.umn.edu/Tools/TBs/TN-009b.pdf
- www.resdac.umn.edu/Tools/TBs/TN-009.asp.
- Virnig BA et al. Survival analysis using Medicare data: example and methods. *Health Services Research* 2000 Dec;35(5 Pt 3):86-101.

3. Services not covered (1991-2000)

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

In addition, Medicare data contains little information on prescription drugs for years prior to 2006. Prescription drug information paid for by Medicare and available for the data years 1991-2000 includes:

- Medication given in an inpatient/hospice/SNF setting - but note that the 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 breaks 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.

4. Cost Sharing

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

5. Gaps in data

Medicare enrollment and claims data linked to NCHS data are available for the years 1991-2000. Several of the surveys linked to the Medicare data, such as NHEFS (1971-1992), NHANES II (1976-1980), and NHANES III (1988-1994) have gaps of several years between the end of the study period and the beginning of the Medicare data.

6. Records linked to Medicare enrollment database but no Denominator or claims records

As described in the Matching Methodology report, the process of linking each NCHS survey participant with Medicare data began by matching individual survey respondents with Medicare's Enrollment Database (EDB). The EDB is a master enrollment File of all people ever entitled to Medicare. For the NCHS records determined to be matched to the Medicare EDB, CMS extracted data from each of the Medicare claims Files for those records for the years 1991-2000. However, not all survey participants matched to the EDB will have claims information for the years 1991-2000. There are several reasons a survey participant would have linked to the EDB but not have a Denominator File record or Medicare claims information for the years 1991-2000:

- Entitled to Medicare but lost entitlement for the years 1991-2000.
- Entitled to Medicare but died prior to 1991.
- Entitled to Medicare but not utilizing Medicare services during 1991-2000.
- Entitled to Medicare but Medicare services being used are not reimbursable claims.
- A small fraction of linked participants may have claims data but no associated Denominator record, this may be due to a change in the Medicare HIC number from the date of processing the Medicare claims data compared to the date when the Denominator data was extracted.
- CMS record keeping inconsistencies, i.e. the CMS data are collected for administrative, not research purposes.

7. Medicare entitlement variables

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

A beneficiary's *original reason* for Medicare entitlement is found in the variable ORIG_REASON_FOR_ENTITLEMENT. This variable is coded by CMS using information provided by the Social Security Administration and/or Railroad Retirement Board. Knowing a beneficiary's original reason for entitlement can be useful for identifying which aged beneficiaries were formerly Medicare disabled,

since their cost and utilization profiles tend differ from other aged beneficiaries, especially at ages 65-74. ORIG_REASON_FOR_ENTITLEMENT values include: Old Age and Survivors Insurance (OASI), Disability Insurance Benefits (DIB) & End Stage Renal Disease (ESRD).

A beneficiary's *current reason* for Medicare entitlement is found in the variable CURR_REASON_FOR_ENTITLEMENT. This variable is populated from the Medicare Enrollment Data Base (EDB). Possible values include: Old Age and Survivors Insurance (OASI), Disability Insurance Benefits (DIB) & End Stage Renal Disease (ESRD).

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

8. Medicare's Prospective Payment System (PPS)

Medicare's PPS refers to a method of reimbursement where the Medicare payment is made based upon a predetermined, fixed amount. Medicare uses a separate PPS for several services, where the particular payment amount is derived based upon the classification system for that particular service. Please note that for outpatient and Home Health Agency reimbursable claims, the PPS was implemented in July 2000, meaning that claims submitted for reimbursement before this date will be different than after July 2000.

For more information on the PPS, please visit
www.cms.hhs.gov/ProspMedicareFeesSvcPmtGen/

III. Additional analytic issues specific to each of the Medicare administrative Files

1. Denominator File

The Denominator File provides data on all Medicare beneficiaries entitled to Medicare benefits in a given year. Monthly information on the enrollment status of linked Medicare beneficiaries including managed care enrollment information is provided by the HMO Indicator code. However, the Denominator File does not include specific plan information for those beneficiaries enrolled in managed care. The Denominator File is fixed length and contains one record per person.

Date of death information obtained by CMS is available on the Denominator File. CMS updates the Denominator File with death information collected through the first three 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. For example, a CMS recorded death occurring on 02/01/1995 will have a date

recorded in variable DOD 'Date of Death' on the 1995 Denominator File and may also have a death day recorded in variable DOD 'Date of Death' on the 1994 Denominator File.

Death information is occasionally mis-reported to CMS but included on the yearly Denominator File. This erroneous information is not corrected by CMS; however, these cases can be identified as 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 insure that deaths are not over-counted. In addition, the *actual* date of death information is occasionally mis-reported to CMS. Cases can be identified by examining the variable, "Valid Date of Death Switch", where a value of "V" indicates that CMS has validated the actual date the beneficiary died, whereas a "blank" indicates that it was not validated. In the event, that 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, which ascertains deaths from probabilistic matches to the National Death Index, death certificates, or longitudinal survey re-contacts for the 1994-1998 NHIS, NHANES I/NHEFS, NHANES II, NHANES III, and LSOA II. 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.

Documentation for the Denominator File is available in PDF format. The variable names used in this data file come from the suggested SAS alias variable name provided by CMS in the Denominator File documentation.

2. Medicare Provider Analysis and Review (MedPAR) Hospital Stay File

The MedPAR Hospital Stay File contains inpatient hospitalization records. All Medicare Part A short and long stay hospitalization claims for each calendar year are included on the MedPAR Hospital Stay File. Each MedPAR Hospital Stay record includes up to 10 ICD-9 diagnoses and 6 ICD-9 procedures associated with each hospital stay. The MedPAR Hospital Stay File will include all hospitalizations that had a discharge 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 stay or short stay hospitalization use variable 'MEDPAR_SS_LS_SNF_IND_CD' - Short Stay/Long Stay/SNF Indicator' which 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 upon the length of a beneficiary's stay and the amount of inpatient services used throughout the stay.

The following fields on the MedPAR Hospital Stay 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, the 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.

Documentation for the MedPAR Hospital Stay File is available in PDF format. The variable names used in this data file come from the suggested standard alias variable names provided by CMS in the MedPAR File documentation.

3. Medicare Provider Analysis and Review (MedPAR) Skilled Nursing Facility (SNF) File

The MedPAR Skilled Nursing Facility File contains skilled nursing facility stays. Skilled Nursing Facility (SNF) stays for each calendar year are provided on the MedPAR SNF File. Each MedPAR SNF record includes up to 10 ICD-9 diagnoses and 6 ICD-9 procedures provided association with a SNF stay. Inclusion in the MedPAR SNF is based on year of admission into the facility.

Each MedPAR SNF record represents a stay in a skilled nursing facility. Each MedPAR SNF may represent one claim or multiple claims, depending upon the length of the beneficiary's stay and the amount of services used throughout the stay. In many cases, SNF records have no discharge date as persons remain in institutions beyond the period of Medicare coverage for that year.

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

- source of admission (SRC_IP_ADMSN_CD)
- group health organization payment code (GHO_PD_CD)

In addition, the 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.

Documentation for the MedPAR Skilled Nursing Facility File is available in PDF format. The variable names used in this data file come from the suggested standard

alias variable names minus the beginning text string “MEDPAR_” in the MedPAR File documentation provided by CMS.

4. 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. Same day surgeries performed in a hospital will be in the Outpatient File. However claims for surgeries performed in freestanding surgical centers appear in the Carrier File, not in the Outpatient File.

Some of the information provided 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 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 (HCPCS) (HCPSCD01-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. There can be multiple outpatient claims records per person on the Outpatient Files. The Outpatient Files are provided in the CMS Standard Analytic File (SAF) format.

Documentation for the Outpatient SAF is available in PDF format. The variable names used in this data file come from the suggested SAS alias variable name provided by CMS in the Outpatient SAF documentation.

5. Home Health Agency (HHA) File

The Home Health Agency File contains final action claims for home health services. Some of the information contained in this file includes the number of visits, type of visit (skilled-nursing care, home health aides, physical therapy, speech therapy, occupational therapy, and medical social services), diagnosis (10 ICD-9 diagnosis codes), dates of visits, reimbursement amount.

There can be multiple HHA claim records per person in the HHA Files. The HHA Files are provided in the CMS Standard Analytic File (SAF) format.

Documentation for the Home Health Agency SAF is available in PDF format. The variable names used in this data file come from the suggested SAS alias variable names provided by CMS in the Home Health Agency SAF documentation.

6. Hospice File

The Hospice File contains final action claims data submitted by Hospice providers. The data contained in this file include the type of hospice care received (e.g., routine home care, inpatient respite care). The Hospice File contains data fields for 10 ICD-9 diagnosis and 6 procedure codes, dates of service, reimbursement amount, and some demographic information (such as date of birth, race, and sex).

- All beneficiaries have a primary diagnosis, but most (90%) have no secondary diagnosis.
- Although there are data fields for procedure codes, in general, such information is not found on 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. There can be multiple Hospice claims records per person on the Hospice File. The Hospice Files are provided in the CMS Standard Analytic File (SAF) format. Documentation for the Hospice SAF is available in PDF format. The variable names used in this data file come from the suggested SAS alias variable names provided by CMS in the Hospice SAF documentation.

7. Carrier File

The Carrier File (formerly the Physician/Supplier Part B File) contains final action claims data submitted by non-institutional providers. The data are largely made up of physician claim records, although the file also includes claims from other non-institutional providers such as physician assistants, clinical social workers, nurse practitioners, independent clinical laboratories, ambulance providers, and stand-alone ambulatory surgical centers.

The claims 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. Each HCPCS code on the carrier claim must be accompanied by an ICD-9 diagnosis code, providing a reason for the service. In addition, each record includes the date of service and reimbursement amount. Due to the large number of carrier claim variables, CMS provides the Carrier data in variable length Files. There can be multiple carrier claims per person on a File. The Carrier Files are provided in the CMS Standard Analytic File (SAF) format.

The Carrier File includes records for non-institutional 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. The variable PLCSRVC ‘Line

Place of Service Code' indicates where the service was provided, but it is not required for payment purposes and is not a validated code and may contain inaccuracies.

The Carrier File contains Durable Medical Equipment (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 DMERC Files not on the Carrier File. The DME claims on the Carrier File are for separate services than those on the DMERC File. There is no overlap between the DME claims on the Carrier and the DMERC Files. See the section on Durable Medical Equipment (DMERC) below for additional information on DME regional carrier claims.

There are two pairs of date fields on the Carrier File. The variables 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 the variables EXPNSDT1 'Line First Expense Date' and EXPNSDT2 'Line Last Expense Date' represent the specific day of service.

For every billed procedure (using a HCPCS code), there should be a corresponding ICD-9 diagnosis code (LINEDGNS) that provides the reason for the billed service. In the case of lab tests, the diagnosis will often be XXOOO because the outside lab has no information from the physician about the reason for the test. In addition, the Carrier File contains space for up to 4 diagnoses, DGNS_CD1 – DGNS_CD4. These are not necessarily linked with any of the billed procedures and may reflect co-existing 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. How the rules on bundling are interpreted vary by carrier (physician).

Documentation for the Carrier SAF is available in PDF format. The variable names used in this data file come from the suggested SAS alias variable name provided by CMS in the Carrier SAF documentation.

8. Durable Medical Equipment (DMERC) File

The DMERC contains final action claims data submitted by Durable Medical Equipment (DME) regional carriers. Durable Medical Equipment can be billed through either a) the Carriers who also process physician claims or b) the DME Regional Carriers (DMERC's) who process only DME claims. Each year CMS distributes a jurisdiction list, available on 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 or not the DME was provided "incident to the physician's service".

Some of the information contained in the DMERC includes diagnosis (10 ICD-9 diagnosis codes), service type codes, dates of service, and reimbursement amount. There can be multiple DME claim records per person on the DMERC File.

DME claims processed by suppliers who also process physician claims are only included 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 only included on the DMERC File. Researchers should examine both the Carrier File and the DMERC File to obtain information about all DME claims. For years 1993-2000 approximately 90% of DME claims data are found on the DMERC File. However, for years 1991 and 1992 nearly 100% of the DME claims data are found on the Carrier File.

Documentation for Durable Medical Equipment is available in PDF format. The variable names used in this data file come from the suggested SAS alias variable names provided by CMS in the Durable Medical Equipment File documentation.

9. Public-use Linkage Summary File

NCHS has provided a public-use File that includes a limited set of variables for researchers to use in determining the feasibility and sample sizes of their proposed research projects. This File includes:

(1) A public ID variable (PUBLICID) so that users can merge variables from NCHS public use data to the Linkage Summary File.
(2) An indicator (CMS_MATCH) of whether the NCHS participant was eligible for the matching and whether he/she linked to the Medicare Enrollment Database. CMS_MATCH contains values 1, 2, or 3.
    a. 1 indicates that the participant is linked; 2 indicates that the participant is not linked; and 3 indicates that the participant was ineligible for matching.
    b. NCHS survey participants were considered ineligible for matching to the EDB, if they refused to provide their SSN or HIC number at the time of the interview. Additional ineligibility criteria included refused, missing, or incomplete information on last name and date of birth.
    c. Ineligible participants must be excluded from all analyses.[1]
(3) Information indicating if a survey respondent has a linked data record on *any* of the eight Medicare administrative record Files for *any* of the ten years of Medicare benefit coverage.
(4) A continuous variable, LINK_AGE, indicates the assumed age of participants at the time of the Medicare extraction (July 1, 2001), because some survey participants became age eligible for Medicare after the time of their baseline survey interview.

[1] A similar variable for whether the participant was linked to Social Security data is also included on the File.

Documentation for the Linkage Summary File and how to download it can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/feasibility_study_data.

10. Summary Medicare Enrollment and Claims (SMEC) Files

The Summary Medicare Enrollment and Claims Files include enrollment information taken from the Denominator Files plus summary variables of several expenditure and claims variables. The SMEC Files are modeled after the Medicare Current Beneficiary Survey (MCBS) cost and use Files. SMEC Files will be available for each of the NCHS-CMS linked surveys by year of Medicare data (1991-2000). Documentation for the SMEC Files can be found at ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/summary_medicare_enrollment_and_claims_files.pdf.

Acknowledgments

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

Centers for Medicare & Medicaid Services (CMS)
www.cms.hhs.gov

Research Data Assistance Center (ResDAC)
www.resdac.umn.edu

National Cancer Institute SEER-Medicare Linked Database
http://healthservices.cancer.gov/seermedicare/

Date Last Updated: April 17, 2007