Suggested citation: National Center for Health Statistics, Office of Analysis and Epidemiology. Analytic Issues in Using the Medicare Enrollment and Claims Data Linked to NCHS Surveys. December 2012. Hyattsville, Maryland. (Available at the following address: http://www.cdc.gov/nchs/data/datalinkage/cms_medicare_analytic_issues_final.pdf)

Last Updated: December 3, 2012

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 and claims data that are linked to NCHS survey data. Users also should refer to the Matching Methodology Report (accessed December 3, 2012) and the Description of the Linked CMS Medicare and USRDS ESRD Data (accessed December 3, 2012) 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. This document 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 http://www.resdac.org/ (accessed December 3, 2012) 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 analytic data file and conducting analyses.

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

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 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 plans are also

referred to as Medicare Advantage (MA) and include Health Maintenance Organizations (HMOs), Preferred Provider Organizations (PPOs), Private Fee-for-Service (PFFS) Plans, Special Needs Plans, and Medicare Medical Savings Account Plans.

2. Exclusion of claims paid by a source other than Medicare (e.g., Medicare Part C plans)

CMS generally does not receive claims data for Medicare beneficiaries who enroll in Medicare Part C plans (including private fee-for-service plans paid on a capitation basis). 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 a Medicare Part C plan. 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.

As noted above the Medicare Denominator file can be used to identify Medicare Part C enrollees. A summary of the percent of NCHS survey respondents who were enrolled in a Medicare Part C plan by year and survey can be found at http://www.cdc.gov/NCHS/data/datalinkage/nchs-cms_medicare_linked_data_managed_care_enrollment_tables.pdf (accessed December 3, 2012)

Researchers should consider the percent 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 the Medicare Utilization Files and how to address them in analyses:

- http://www.resdac.org/resconnect/articles/114 (accessed December 3, 2012)
- 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 (1991-2007)

Although Medicare provides coverage for a wide range of services, there are health care services not covered by Medicare. Examples of services not covered include routine physical exams, long-term care, and some cancer screening procedures.

These gaps in coverage mean that there are no claims records for these services or for certain time periods. You may find more information on what is not currently covered by Medicare in the *Medicare and You Handbook* at www.medicare.gov (accessed December 3, 2012).

In addition, Medicare data contains little information on prescription drugs for years prior to 2006. However, beginning in 2006 prescription drug coverage for Medicare beneficiaries became available through the Medicare Part D program. Prescription drug information paid by Medicare for 2006-2007 is available on 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.

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 and Discrepancies in Coverage Periods

Medicare enrollment and claims data linked to NCHS data are available for the years 1991-2007. Several of the surveys linked to the 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.

Researchers should be aware that there may be differences in Medicare data by age depending on the survey year (s). For example, an NHIS respondent who is 80 years old in 1999 at their interview and matched at some point to the 1999-2007 Medicare Denominator file would only have Medicare data for the time period 1999 to 2007. A similar NHIS respondent, who was 80 years old at their interview in 1998 and linked to the 1999-2007 Medicare Denominator file and linked in the previous 2001 Medicare linkage would hypothetically have Medicare data for 1991-2007 (see Appendix A in methodology report). This issue is particularly important when combining data across survey and Medicare coverage years and researchers need to determine how to address these discrepancies in coverage periods in their analyses.

6. Records with only 1991-1998 Medicare claims data

Several NCHS surveys (1994-1998 NHIS, NHEFS, NHANES III and LSOA II) were included a previous 2001 Medicare linkage. There are some instances, where these survey respondents, with a verified SSN from SSA, had been successfully linked to CMS records in a previous linkage, but did not match the 1999-2007 Medicare Denominator file. The majority of these respondents died prior to 1999 and did not match to the 1999-2007 Denominator file because the CMS denominator file includes individuals who are entitled to benefits and enrolled in Medicare for that calendar year. Once a person dies they are no longer entitled to Medicare benefits. The other cases remained as non-matches and were dropped since their vital status could not be verified.

Please refer to the NCHS-CMS Medicare Linkage Methodology Report (accessed December 3, 2012) for more detailed information about linkage eligibility for both the previous 2001 and most recent Medicare Linkages.

7. On Denominator file with no claims data

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

8. 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) and End Stage Renal Disease (ESRD).

A beneficiary's *current reason* for Medicare entitlement is found in the variable CURR_REASON_FOR_ENTITLEMENT. Possible values include: Old Age and Survivors Insurance (OASI), Disability Insurance Benefits (DIB) and End Stage Renal Disease (ESRD). This variable is populated from the Medicare Enrollment Data Base (EDB). The EDB is a master enrollment file of all people ever entitled to

Medicare. Many of the variables on the Denominator file are extracted from the EDB. The EDB is not available to researchers.

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.

9. 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
http://www.cms.gov/ProspMedicareFeesSvcPmtGen/ (accessed December 3, 2012)

10. Child survey participants

Survey participants under 18 years of age at the time of the survey, are considered linkage-eligible, the criteria by which survey participants can be potentially linked to CMS data, if consent is provided by their parent or guardian. Linkages to CMS administrative data are conducted linking survey data to multiple years of administrative data. Consequently, linkage-eligible child survey participants can be under 18 years of age for some years of linked administrative data and 18 years of age or older for later years. For example, a 15-year old 2003-2004 NHANES participant can be linked to CMS data for 2006 and earlier years as a child, but would be an adult in 2007 (approximately) and later years.

In accordance with NCHS Ethics Review Board (ERB) guidelines, for survey participants younger than 18 years of age at the time of the survey, NCHS will only provide linked CMS data generated for program participation, claims and other events that occurred prior to the participant's 18th birthday. The linkage of NHANES to the CMS Medicaid data potentially has a large number of child survey participants linked to one or more years of Medicaid data collected after age 18. This should be taken into consideration by analysts when estimating their potential sample size for RDC proposals. Analysts requiring more information about potential sample sizes for RDC proposals or more information on this NCHS ERB guidance should contact the NCHS Data Linkage team (datalinkage@cdc.gov).

11. Weighting of Linked Data files

The survey weights provided in NCHS population health survey data files adjust for oversampling of specific subgroups and differential non-response, and are post-stratified to annual population totals for specific population domains to provide nationally representative estimates. The properties of these weights for CMS Medicare linked data files with incomplete linkage due to ineligibility for linkage and non-matches are unknown. In addition, methods for using the survey weights for some longitudinal analyses require further research. Because this is an important and complex methodological topic, ongoing work 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 survey weights. The survey weights available on NCHS population health survey data files can be adjusted for incomplete linkage and non-matches, hereafter referred to as non-response, using standard weighting domains to reproduce population counts within these domains -- gender, age, and race/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) allows auxiliary information to be used to adjust the statistical weights for non-response. This approach is promising and this software is provisionally recommended for adjusting survey weights for the linked files. As 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. Additional information will be posted on the NCHS Data Linkage webpage as it becomes available, including sample SUDAAN code and a detailed literature review of weighting methods.

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, (http://www.cdc.gov/nchs/data_access/data_linkage/mortality.htm, accessed December 3, 2012) which ascertains deaths from probabilistic matches to the National Death Index, death certificates, or longitudinal survey re-contacts for the 1986-2004 National Health Interview Survey (NHIS), NHANES I Epidemiologic Follow-up Survey (NHEFS), Second National Health and Nutrition Examination Survey (NHANES II), Third National Health and Nutrition Examination Survey (NHANES III),1999-2004 National Health and Nutrition Examination Survey, The Second Longitudinal Study of Aging (LSOA II), and 1985, 1995, 1997, and 2004 National Nursing Home Survey (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.

Documentation for the Denominator File (accessed December 3, 2012) is available in PDF format.

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

The MedPAR File contains inpatient hospitalization and skilled nursing facility (SNF) final action claim records. All Medicare Part A short and long stay hospitalization claims and SNF claims for each calendar year are included on the MedPAR File. Each MedPAR claim record includes up to 10 ICD-9-CM diagnoses and 6 ICD-9-CM procedures associated with each hospital or SNF stay. The MedPAR File will include 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 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 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 File (accessed December 3, 2012) is available in PDF format.

3. 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-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 (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 (accessed December 3, 2012) is available in PDF format.

4. 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-CM 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 (accessed December 3, 2012) is available in PDF format.

5. 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-CM 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 (accessed December 3, 2012) is available in PDF format.

6. 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 (accessed December 3, 2012, with additional codes specific to CMS called Level II HCPCS. 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. 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 DME File. There is no overlap between the DME claims on the Carrier and the DME Files. See the section on Durable Medical Equipment (DME) 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-CM 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 (accessed December 3, 2012).

7. Durable Medical Equipment (DME) File

The DME 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 DME includes diagnosis (10 ICD-9-CM diagnosis codes), service type codes, dates of service, and reimbursement amount. There can be multiple DME claim records per person on the DME 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 DME 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 (accessed December 3, 2012) is available in PDF format.

8. Part D Denominator and Prescription Drug Event (PDE) Files[1]

The Medicare Part D Denominator File contains demographic information and enrollment/eligibility status for each Medicare-eligible beneficiary during the calendar year. This information is present regardless of the type of Medicare Part D plan the beneficiary might select (i.e., enrollment data are present for managed care participants as well as those enrolling in stand-alone prescription drug plans). Information is also available for Medicare beneficiaries who did not obtain Part D

[1] The majority of this text in this section was excerpted from the Chronic Condition Data Warehouse Part D Data User Manual, Version 3.1, December 2010. (Buccaneer, A Vangent Company, under contract with the Centers for Medicare & Medicaid Services) http://www.ccwdata.org/cs/groups/public/documents/document/ccw_partddata_userguide.pdf (accessed December 3, 2012).

coverage. In addition to the variables available on the standard Medicare Denominator File, the Part D Denominator File contains a derived race/ethnicity code, an indicator for Other Credible Drug Coverage, and monthly indicators for Medicare Advantage Prescription Drug Plans (MA-PD), prescription drug plan (PDP) enrollment, Low Income Subsidy (LIS) enrollment, Retiree Drug Subsidy (RDS), and State Reported Dual Eligibility Status. In a small number of cases there may be differences in the enrollment information on the standard Medicare Denominator File and the Part D Denominator File. This is due to updates made by CMS after the standard Medicare Denominator File was delivered.

The Part D Prescription Drug Event (PDE) File contains a summary of prescription drug costs and payment data used by CMS to administer benefits for Medicare Part D enrollees. The PDE File does not contain individual drug claims, but are summary extracts submitted to CMS by Medicare Part D prescription drug plan providers.

All Medicare Part D prescription drug benefits are provided through private plans (a.k.a. plan sponsors). 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 they provide to Medicare beneficiaries under Part C.

Claims for prescription drugs are submitted by pharmacies to the Part D health plans for beneficiaries enrolled in Medicare Part D. The PDE data are created from point-of-service transactional data at the time a prescription is filled. Data for prescriptions which are ordered but not filled do not exist in this database (i.e., data are not prescribing data, but rather reflect filled prescriptions). The PDE data are considered "final action", as 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. It is important to note that PDE data are not submitted by plans which receive RDS, or for other types of plans which are considered to be Part D creditable coverage (e.g., Veterans Administration (VA) , TRICARE).

The 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 a 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 which is rejected due to a prior authorization requirement and later, when the prior authorization criteria are met, resubmits the claim which is 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 in the CCW represent "final action", all PDE adjustments received by CMS through the PDE submission deadline for payment

reconciliation are accounted for in the data, including PDE adjustments, resubmissions, and deletions.

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.

There are several situations where Benefit Phase and Utilization Management (UM) values cannot be determined:

- The PDE is for a non-covered 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 will be "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 but did submit formulary information (e.g., employer direct) will have "NA" assigned for Benefit Phase. UM variables will be 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 HMPS plan information, the value of "XX" is applied for all five of the event characteristics variables.

Documentation for the Medicare Part D Denominator (accessed December 3, 2012) and Part D PDE (accessed December 3, 2012) Files are available in PDF format.

9. Chronic Condition (CC) Summary File[2]

The 2005-2007 Chronic Condition (CC) Summary File includes claims, enrollment, and assessment data for Medicare beneficiaries extracted based on the following set of 21 chronic health conditions:

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

The CC Summary File provides a summary of clinical information, including date of first occurrence, yearly, and mid-year flags for each of the 21 chronic conditions.

The 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 which 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 (e.g., December 2005 for the yearly indicators in the 2005 CC Summary File; June 2005 for the mid-year indicators in the 2005 CCW Summary File). It is important to note that claims prior to the reference year (e.g., 2005) may have been examined to make this determination, if the CC definition was a 2- or 3- year condition (e.g., diabetes, CHF, Alzheimer's). Refer to

[2] The majority of this text in this section was excerpted from the CMS Chronic Condition Data Warehouse: Technical Guidance for Researchers Calculating Population Statistics (O'Donnell, Brian; Schneider, Kathy; Dean, Debbie. September 2008. Buccaneer, A Vangent Company, under contract with the Centers for Medicare & Medicaid Services)
http://www.ccwdata.org/cs/groups/public/documents/document/ccw_techguideresearchers.pdf (accessed December 3, 2012)

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 (accessed December 3, 2012)

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

Documentation for the CC Summary File (accessed December 3, 2012) is available in PDF format.

10. Summary Medicare Enrollment and Claims (SMEC) File

NCHS has created a Summary Medicare Enrollment and Claims file to assist researchers who are interested in analyzing Medicare cost and claims data from multiple Medicare service files. The SMEC file contains data on the beneficiary's reason for Medicare entitlement, total number of months of Medicare entitlement, Medicare Part C plan enrollment, and summarized Medicare service charges, total expenditures, and reimbursement amounts. These summarized (or summary) variables are modeled after the Medicare Current Beneficiary Survey (MCBS) (accessed December 3, 2012) cost and use files. In addition, summary variables related to the beneficiary's total number of Emergency Room (ER) visits and Part D prescription drug costs have been created.

A SMEC file is available for each of the NCHS surveys linked to Medicare enrollment and claims data. Documentation for the SMEC File (accessed December 3, 2012) is available in PDF format.

11. Feasibility Study Data File

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

(1) A public ID variable (PUBLICID) so that users can merge variables from NCHS public use survey data to the Linkage Summary File.
(2) An indicator (CMS_MEDICARE_MATCH) of whether the NCHS participant was eligible for the matching and whether he/she linked to the Medicare Denominator File. CMS_MEDICARE_MATCH contains values 1, 2, 3, or 9.
    a. 1 indicates that the participant is linked; 2 indicates that the participant is not linked; 3 indicates a child survey participant with partial administrative data available; and 9 indicates that the participant was ineligible for linkage.
    b. NCHS survey participants were considered ineligible for matching to the Denominator, if they refused to provide their SSN or Health Insurance Claim (HIC) number at the time of the interview or their SSN was not verified by the SSA Enumeration Verification System. 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 using the linked CMS data.

(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 years of Medicare benefit coverage.

Documentation for the Feasibility Study Data File and how to download it can be found at http://www.cdc.gov/NCHS/data_access/data_linkage/cms/cms_medicare_feasibility.htm (accessed December 3, 2012).

12. Additional linked data files

NCHS has also linked to a separate set of data files containing information on patients diagnosed with End Stage Renal Disease (ESRD) obtained from the United States Renal Data System (USRDS) http://www.usrds.org/ (accessed December 3, 2012). The USRDS is a national data system funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) designed to collect, analyze, and distribute information about ESRD in the United States. The linked ESRD data files can be used by researchers interested in conducting analysis specifically related to patients with ESRD. For more information about the data available on the linked ESRD files, please refer to the documentation: ftp://ftp.cdc.gov/pub/Health_Statistics/NCHS/datalinkage/nchs-usrds_linked_esrd_data_files_documentation.pdf (accessed December 3, 2012)

NCHS has also linked to CMS Medicaid enrollment and claims data. Linkage of the NCHS survey participants with the CMS Medicaid data provides the opportunity to study changes in health status, health care utilization and expenditures in low income families with children, the elderly and disabled U.S. populations. For more information about the linked CMS Medicaid data, please see the data linkage website: http://www.cdc.gov/nchs/data_access/data_linkage/cms_medicaid.htm (accessed December 3, 2012).

Acknowledgments

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

Centers for Medicare & Medicaid Services (CMS)
http://www.cms.gov/ (accessed December 3, 2012)

Research Data Assistance Center (ResDAC)
http://www.resdac.org/ (accessed December 3, 2012)

National Cancer Institute SEER-Medicare Linked Database
http://healthservices.cancer.gov/seermedicare/ (accessed December 3, 2012)

Buccaneer, A Vangent Company
http://www.ccwdata.org (accessed December 3, 2012)