Chronic Condition Data Warehouse

User Manual

Version 1.5
May 2009

Overview

One of the goals of Section 723 of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003 is to make Medicare data readily available to researchers who are studying chronic illness in the Medicare population. To support this effort, the Centers for Medicare & Medicaid Services (CMS) contracted with the Buccaneer Computer Systems and Services, Inc. (BCSSI) to establish the Chronic Condition Data Warehouse (CCW). The CCW contains existing CMS beneficiary data (from multiple data sources) linked by a unique identifier, allowing researchers to analyze information across the continuum of care. The CCW currently contains data from fee-for-service Institutional and Non-institutional claims, enrollment/eligibility, and assessment (all payers) data (*Minimum Data Set, Outcome and Assessment Information Set, Swing bed assessments,* and *Inpatient Rehabilitation Facility Patient Assessment Instrument*) from January 1, 1999 forward for a random 5% Medicare beneficiary population, and from January 1, 2005 forward for 100% Medicare population. (See Chapter 1 for a description of the CCW population.)

CCW data are available upon request for specific chronic condition cohorts. Data may also be requested for other cohort(s) of interest. The specific chronic condition cohorts were defined by CMS, in collaboration with the Research Data Assistance Center (ResDAC) and BCSSI. Much of the supporting documentation is provided with assistance from ResDAC (e.g., chronic condition definitions, standard data dictionary files, etc.).

The intended use of the CCW data is to identify areas for improving the quality of care provided to chronically ill Medicare beneficiaries, reduce program spending, and make current Medicare data more readily available to researchers studying chronic illness in the Medicare population. By predefining the chronic conditions, data extraction from the CCW is very efficient, allowing for data requests to be fulfilled in a timelier and more cost efficient manner.

Medicare beneficiary Health Insurance Claim numbers (HICs) will be removed from the data files delivered to researchers (unless otherwise specified/approved in the Data Use Agreement). A unique CCW beneficiary identifier will be included in each data file delivered as part of the output package (see Chapter 3 for details), thus allowing linkage of an individual's data across data sources/types described above. A separate file will be provided for those requests requiring beneficiary identifiable data. If a researcher needs to obtain the HIC in order to link to outside data sources or extract claims not part of the CCW database, then the researcher will need to submit justification for this information in the study protocol and request identifiable variables.

The unique CCW beneficiary identifier field is specific to the CCW (not applicable to any other identification system or data source). This identifier is encrypted prior to delivering the data files to researchers. In addition, all data files delivered to researchers are encrypted (see Encryption Information in Chapter 4 for details). Each research

request will employ a different encryption key for the beneficiary identifier field and the data files.

This manual provides users with information that may be helpful in understanding and analyzing the CCW data.

Chapter 1 describes the CCW sample population.

Chapter 2 describes the chronic condition definitions.

Chapter 3 outlines the various types of CMS data contained in the CCW and unique beneficiary identifier.

Chapter 4 describes the content, format, and encryption of the CCW output package.

Chapter 5 explains known limitations of the CCW data.

Chapter 6 provides resources for further assistance with CCW data.

Attachment A outlines the chronic condition definitions.

Chapter 1. CCW Population

The CCW is populated with CMS data for a sample of Medicare beneficiaries eligible for coverage during a specified time period. The CCW data are available for services beginning January 1, 1999 through the most current year of Medicare data available, for a 5% random sample of Medicare beneficiaries.

The 5% sample initially loaded to the CCW includes those eligible and enrolled for Medicare on or after January 1, 1999 through the most current period covered by the release, who had a Health Insurance Claim (HIC) number equal to the Claim Account Number (CAN) plus Beneficiary Identity Code (BIC) (HIC=CAN+BIC) where the last two digits of the CAN are in the set {05, 20, 45, 70, 95} at some time on or before December 31, 2004. All claims and eligibility information from 1999 forward for this 5% sample are included in the CCW. This means that if a beneficiary had a change in their HIC and the change caused this beneficiary to fall out of the original CMS 5% sample, the CCW will continue to include this individual and his/her associated claims, assessment, and eligibility information in the database. In other words, this 5% sample includes *ever* enrolled beneficiaries from January 1, 1999 forward.

Beginning in 2005, Medicare beneficiaries who enter the CMS 5% sample for the first time will be added to the CCW 5% sample. Data for services provided starting with the month that the beneficiary joined the CMS 5% sample will be included in the CCW. For these beneficiaries, the CCW will not include data for services provided from January 1, 1999 forward, but from the first month that they were included in the CMS 5% sample. Once the beneficiary is included in the CCW 5% sample, they will not be removed, regardless of future HIC changes.

Since the CCW is an *ever* enrolled 5% data file, in order to extrapolate results to obtain a national estimate, the researcher will need to refer to the Beneficiary Summary File included in the extracted data provided by the CCW. This file contains a flag that indicates whether a beneficiary was included in the CMS 5% sample for the year (i.e., the cross-sectional annual sample) or if the beneficiary was included as a member of the CCW 5% sample (i.e., all inclusive, *ever* enrolled sample). The CMS 5% annual sample can be multiplied by 20 to extrapolate results to obtain a national estimate.

In addition, CCW includes data for 100% Medicare beneficiaries for enrollment and services provided from January 1, 2005 forward. This allows for greater flexibility in defining cohorts or populations of interest. For services occurring in 2005 forward, researchers may request specific cohorts from the 100% Medicare beneficiary population. This may include chronic condition cohorts, finder files of populations previously used by researchers, or unique cohorts as defined by the researcher.

Chapter 2. Chronic Condition Definitions

While all administrative claims, assessment, and enrollment/eligibility data for each beneficiary in the enhanced 5% sample are available (note: Medicare administrative claims data are not available if the beneficiary is enrolled in managed care), researchers may request data for a specific, predefined cohort based on a set of twenty-one common chronic condition categories. (Note that Medicare claims information contain primarily fee-for-service population.) The preset definitions will assist in simplifying and expediting the fulfillment of research requests, and allow for more cost efficient methods of delivering readily available data files to researchers for chronic illness research.

Research Populations

Researchers may request data files for cohorts based on *standard*, *modified standard*, or *custom* definitions:

1. *Standard* definitions include specific criteria for reference time periods, diagnosis and procedure codes, number/type of qualifying claims (e.g., must have 2 Carrier claims during reference time period), coverage (see Chronic Condition Flag descriptions below), geography, and exclusions. Researchers may request CCW data for any of the 21 predefined chronic conditions as defined by CMS in collaboration with ResDAC and BCSSI. These common chronic disease classifications include the following (see Attachment A for complete definitions):

    - 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
    - Female Breast Cancer
    - Colorectal Cancer
    - Prostate Cancer
    - Lung Cancer
    - Endometrial Cancer

    The reference time period is the look-back period during which the other criteria must be met. It is possible for a beneficiary to meet the criteria for a given year and not the next year. For example, Request A is submitted for 2001 claims for a cohort of beneficiaries with chronic kidney disease. The cohort is identified by applying the chronic kidney disease criteria to the universe of applicable claims for service provided on or before December 31, 2001 back through January 1, 2000 (a two-year reference period). A beneficiary meets the cohort inclusion criteria with one qualifying claim occurring in 2000 and has no subsequent claim meeting the specified criteria. Request B is submitted for 2002 claims for a chronic kidney disease cohort, which has a look-back period of December 31, 2002 back through January 1, 2001. Since the beneficiary's only qualifying claim

occurred in 2000, the beneficiary does not meet the inclusion criteria. The beneficiary is included in the cohort for Request A but not Request B.

2. *Modified Standard* definitions use the diagnosis and procedure codes defined in the standard request but allow researchers to modify other parameters such as beneficiary demographics, geography, and Medicare coverage status.

3. *Custom* definitions allow researchers to request a cohort based on unique criteria provided by the researcher (i.e., all claims for a particular procedure, diagnosis, or specified population). This approach would also be used if the researcher used a different definition for one of the chronic conditions already defined by the CCW classifications. This type of request may also include data requested based on diagnosis and procedure categorizations schemes or assigned comorbidities, such as those developed by the Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software and Comorbidity Software.

Chronic Condition Flags

The chronic condition flags consider clinical criteria (from administrative claims), coverage criteria (from enrollment data), and specified time periods. The clinical criteria consider variations of the following (see Attachment A for definitions of each chronic condition):

- ICD-9, CPT4, or HCPCS codes
- Claim type(s) and count(s)
- Date(s) of service (e.g., claim thru dates at least one day apart)

The coverage criteria consider variations of Medicare Part A, B, and no HMO coverage. The specified time periods, or *reference time periods*, consider the length of time during which the clinical and coverage criteria are considered.

Three types of chronic condition flags are included in the Chronic Condition Summary File. This includes:

1. **Yearly flags**
   Algorithm criteria applied, using December 31 as the end of the reference year (e.g., 2007 yearly flag for algorithm with 1-year reference period includes services between 01/01/07-12/31/07). The following are valid values for the yearly flag:
   0 = Neither claims nor coverage met
   1 = Claims met, coverage not met
   2 = Claims not met, coverage met
   3 = Claims and coverage met

2. **Mid-year flags**
   Algorithm criteria applied, using July 1 as the end of the reference year (e.g., 2007 mid-year flag for algorithm with 1-year reference period includes services between 07/01/06-06/30/07). Researchers can use the Beneficiary Summary File

to determine whether alive and enrolled on July 1, for the purposes of producing statistics. The following are valid values for the mid-year flag:
0 = Neither claims nor coverage met
1 = Claims met, coverage not met
2 = Claims not met, coverage met
3 = Claims and coverage met

3. ***Ever* date** (first occurrence of condition - YYYYMMDD)
   Date the beneficiary first met the *clinical* criteria of the algorithm (no coverage criteria applied), with the earliest possible date of 01/01/99.

Consideration of claims criteria for the algorithms includes a 7-day grace period for claim thru dates occurring within:

- 7 days prior to first date of coverage
- 7 days after date of death

NOTE: Unless otherwise specified by the researcher, standard data requests will include (by default) all beneficiaries with yearly (or mid-year, if requested) flags = 1 or 3.

Control Populations

Requests for control populations should be made at the time of the initial data request. The inclusion/exclusion criteria for the control population should be specified by the researcher completing the data request form. The *standard, modified standard*, or *custom* definitions explained above for research populations can also be applied to control populations. Alternatively, the researcher can request a control population lacking in any chronic conditions, if desired. Researchers can request the entire 5% file or customize the control population as needed. Specifications should include type(s) of data files, applicable diagnosis or procedure codes or DRGs, time periods, and any related demographic selection criteria.

Chapter 3. CMS Data Available through the CCW

The CCW contains various types of CMS data from multiple care settings. This includes Medicare Institutional and Non-institutional claims, assessment (*Minimum Data Set* and *Outcome and Assessment Information Set*), and Medicare enrollment and eligibility data. These data are available from various provider settings, including the following:

Files	Years	Linking Key or Stand Alone File
1999-2004 Claims and Assessment Files for 5% FFS sample population 2005-forward Claims and Assessment Files for 100% Medicare FFS population		
Institutional Claims Files		
Inpatient Base Claim Files	1999 – current	CCW BENE_ID
Inpatient Revenue Center Files	1999 – current	CCW CLM_ID*
Outpatient Base Claim Files	1999 – current	CCW BENE_ID
Outpatient Revenue Center Files	1999 – current	CCW CLM_ID*
Skilled Nursing Facility Base Claim Files	1999 – current	CCW BENE_ID
Skilled Nursing Facility Revenue Center Files	1999 – current	CCW CLM_ID*
Home Health Agency Base Claim Files	1999 – current	CCW BENE_ID
Home Health Agency Revenue Center Files	1999 – current	CCW CLM_ID*
Hospice Base Claim Files	1999 – current	CCW BENE_ID
Hospice Revenue Center Files	1999 – current	CCW CLM_ID*
Institutional Condition Code File	1999 – current	CCW BENE_ID
Institutional Occurrence Code File	1999 – current	CCW BENE_ID
Institutional Span Code File	1999 – current	CCW BENE_ID
Institutional Value Code File	1999 – current	CCW BENE_ID
Non-Institutional Claims Files		
Carrier Claim Files (Physician/Supplier)	1999 – current	CCW BENE_ID
Carrier Line Files (Physician/Supplier)	1999 – current	CCW BENE_ID
Durable Medical Equipment (DMERC) Claim Files	1999 – current	CCW BENE_ID
Durable Medical Equipment (DMERC) Line Files	1999 – current	CCW BENE_ID
Assessment		
Minimum Data Set (MDS)	1999 – current	CCW BENE_ID
Outcome and Assessment Information Set (OASIS)	2000 – current	CCW BENE_ID
Swing Bed	2002 – current	CCW BENE_ID
Inpatient Rehabilitation Facility Patient Assessment Instrument (IRF-PAI)	2002 – current	CCW BENE_ID
Other CCW Files		
Beneficiary Summary File	1999 – current	CCW BENE_ID
Chronic Condition Summary File	1999 – current	CCW BENE_ID
CMS 100% Denominator File	1999 – current	CCW BENE_ID
CMS 5% Denominator File	1999 – current	CCW BENE_ID
Research Identifiable File (CCW BENE_ID to HIC crosswalk)	1999 – current	CCW BENE_ID
MCBS to BENE_ID crosswalk	1999 – current	CCW BENE_ID
Medicare Current Beneficiary Survey (MCBS) Access to Care	1999 – current	Stand Alone File
MCBS Cost and Use	1999 – current	Stand Alone File

* The CCW_CLM_ID is the unique key to link revenue center information to a specific claim, which is then linked to the CCW_BENE_ID.

The data files delivered from the CCW contain a subset of the original source files (see data dictionaries included in documentation information). Variables used infrequently or not applicable to a particular setting have been removed. A description of the variables contained within each file can be found within the “File Record Layout” folder contained within the CCW documentation. Within this folder are CCW Variables files that contain descriptions of the variables.

CCW Unique Identifier (CCW BENE_ID)

For each Medicare beneficiary enrolled and eligible for Medicare during the given time period, a unique CCW identifier provides a common link across all applicable types of data available. Based on the approved research request, the CCW data delivered may or may not include patient identifying information. Regardless of whether patient identifying information is included, the unique patient identifier provides researchers with the ability to analyze information across the continuum of care for a particular beneficiary or chronic condition cohort.

The unique CCW BENE_ID is created from the unloaded EDB file, using the EDB Link Number, HIC number, and other beneficiary identifiers (i.e., gender, SSN, date of birth) from the RIC 'A' for each unloaded EDB beneficiary. Analysis is performed to ensure that the beneficiary is not represented multiple times in the CCW HIC history table. If the beneficiary already exists in the CCW HIC history table, the table is updated with the new HIC information corresponding to the existing CCW BENE_ID. If they do not already exist in the CCW HIC history table, the record will be added and the beneficiary will be assigned a new unique CCW BENE_ID. All cross reference records for this beneficiary will be assigned this unique ID.

Beneficiary Summary File

The Beneficiary Summary File is created annually and contains demographic entitlement and enrollment data for beneficiaries who 1) were part of the CCW 5% sample, 2) were documented as being alive for some part of the reference year of the Beneficiary Summary File, and 3) were enrolled in the Medicare program and entitled to its benefits during the file’s reference year. *Reference year* refers specifically to the calendar year accounted for in the Beneficiary Summary File. For example, the 2004 Beneficiary Summary File covers the year 2004; therefore the year 2004 is the reference year for the 2004 Beneficiary Summary file.

Most fields contained in the Beneficiary Summary File are derived through the twelve-month reference year; however, other fields reflect information from outside of the reference year. The twelve-month or fifteen-month data derivation period determinations are specified in the Beneficiary Summary File data dictionary. The data sources for the Beneficiary Summary File are the March Unload of the Enrollment Database and the March Third Party Master File. The Beneficiary Summary File data sources are designed to have a three-month period of time after the end of the reference year in order to accommodate late-arriving data to the data sources.

Chronic Condition Summary File

The Chronic Condition Summary File is a summary of clinical information for all beneficiaries included in the requested cohort. This file includes the *yearly*, *mid-year*, and *ever* flags described in Chapter 2 for each of the 21 chronic conditions, as well as the encrypted beneficiary ID and reference time period.

Researchers may find it useful to request and analyze information from the Chronic Condition Summary File (and/or the Beneficiary Summary File) *prior to* requesting/purchasing administrative claims or assessment information.

CMS 100% or 5% Denominator File

The CMS Denominator File combines Medicare beneficiary entitlement status information from administrative enrollment records with entitlement and Part B state buy-in information and Group Health Plan enrollment information. The Denominator File contains data on all Medicare beneficiaries enrolled and/or entitled in a given year. It does not contain data on all beneficiaries ever entitled to Medicare. The file contains data only for beneficiaries who were entitled during the year of the data. The Denominator File is often used by researchers who want to compare Medicare entitlement with Medicare utilization.

The source of the Denominator File is the Enrollment Data Base (EDB). (The data used to create and update the Denominator File are validated prior to this step.) The CMS denominator file is available in May for the previous calendar year. The cut off date for any updated transactions is generally the last Friday in March.

The Denominator File is sorted by HIC and contains other variables such as state and county codes, birth and death dates, sex, race, age, etc. Please see the CCW Variables – CMS Denominator File in the File Record Layout folder for a complete list of variables included in this file.

A copy of the CMS 100% or 5% Denominator File is available through the CCW. This information is provided with the BENE_ID. The CMS Denominator file containing the BENE_ID is necessary to link this data to any other CCW data.

Medicare Current Beneficiary Survey

The Medicare Current Beneficiary Survey (MCBS) is a continuous, multipurpose survey of a representative national sample of the Medicare population, conducted by the Office of Research, Development, and Information of CMS through a contract with Westat. The central goals of the MCBS are to determine expenditures and sources of payment for all services used by Medicare beneficiaries, including co-payments, deductibles, and noncovered services; to ascertain all types of health insurance coverage and relate coverage to sources of payment; and to trace processes over time, such as changes in health status and spending down to Medicaid eligibility and the impacts of program changes, satisfaction with care, and usual source of care. The MCBS produces two files annually, the Access to Care and Cost and Use files.

The information used to create these files is obtained through interviews with selected Medicare beneficiaries. These data are collected, submitted to logical checks, and edited by Westat prior to the creation of the MCBS file.

The Access to Care file, which contains summaries of use and expenditures for the year from Medicare files along with survey data on insurance coverage, health status and functioning, access to care, information needs, satisfaction with care, and income is the first MCBS file to be released. This occurs approximately one year after the completion of the Medicare beneficiary interviews.

The Cost and Use file is available approximately 2 years after the completion of the Medicare beneficiary interviews. This file contains reconciled information on events, charges, and payments from both survey and claims for the year as well as most of the variables from the Access to Care file. The sample represents the "ever-enrolled" Medicare population: all persons who were in the program at any time during the calendar year.

MCBS data is available through the CCW. The MCBS data are not linked to the CCW BENE_ID, therefore the MCBS data cannot be linked to other CCW data at the beneficiary level. A crosswalk of identifiers used in the MCBS data and the CCW data files is available upon request (for linkage purposes). This information will allow researchers to merge MCBS data with other CCW data at the beneficiary level.

Chapter 4. Content, Format and Encryption of CCW Output Files

This section describes the content and format of the CCW output package (the CCW data that is delivered to researchers). The files that are delivered to the researcher are organized in the following format. Descriptions of each of these items are detailed in the tables that follow.

- 1999 (one folder will be created for each year of data requested)
  - READ_ME_FIRST_R0000_YYYY.txt
  - R0000_YYYY_OS.exe (.exe extension for Windows only)
- File Record Layouts
  - Data Dictionaries (contains text files that are hyperlinks in the following .xls files)
  - CCW Variables - Assessments.xls
  - CCW Variables - Chronic Condition Summary File.xls
  - CCW Variables - Claims Files.xls
  - CCW Variables -Beneficiary Summary File.xls
- CCW User Manual.pdf
- Code Reference Sets.xls
- Decryption Instructions.pdf

File	File Description
READ_ME_FIRST_R0000_YYYY.txt	This is a text file that describes the files contained in the output package. File Name Example: READ_ME_FIRST_R0014_2001.txt
R0000_YYYY_OS.exe	This is the executable that must be run to decrypt and uncompress the data files. The file naming convention is R0000 = CCW request number, YYYY = year of data being decrypted and uncompressed, OS = the target operating system (e.g., win32) File Name Example: R0014_2001_win32.exe
CCW Variables - Assessments.xls	Spreadsheet containing the variables that are provided in the assessment files. File contains SAS names, descriptions and valid codes for each variable.
CCW Variables - Chronic Condition Summary File.xls	Spreadsheet containing the variables that are provided in the Chronic Condition Summary File. Spreadsheet contains SAS names, descriptions and valid codes for each variable.
CCW Variables - Claims Files.xls	Spreadsheet containing the variables that are provided in the claims files. File contains SAS names and descriptions for each variable.

File	File Description
CCW Variables -Beneficiary Summary File.xls	Spreadsheet containing the variables that are provided in the Beneficiary Summary File. File contains SAS names, descriptions, valid codes, and derivation information for each variable.
CCW User Manual.pdf	An electronic copy of this document.
Code Reference Sets.xls	Code lists for ICD-9 diagnosis and procedure codes, HCPCS codes, Revenue Center and other codes contained in the extracted files.
Decryption Instructions.pdf	This document contains instructions for decrypting/uncompressing the data files.

Depending on the type(s) of data requested, the output package delivered to researchers will contain the following files in fixed column flat files (see data dictionaries for variables included). These files are accessed by performing the executable described above.

Claims Files		
File Name	**File Description**	**Unit of Analysis**
inpatient_base_claim_(date/time stamp).dat[1]	Inpatient Base Claim File	Claim
inpatient_revenue_center_(date/time stamp).dat	Inpatient Revenue Center File	Revenue Center Detail
outpatient_base_claim_(date/time stamp).dat	Outpatient Base Claim File	Claim
outpatient_revenue_center_(date/time stamp).dat	Outpatient Revenue Center File	Revenue Center Detail
snf_base_claim_(date/time stamp).dat	SNF Base Claim File	Claim
snf_revenue_center_(date/time stamp).dat	SNF Revenue Center File	Revenue Center Detail
hospice_base_claim_(date/time stamp).dat	Hospice Base Claim File	Claim
hospice_revenue_center_(date/time stamp).dat	Hospice Revenue Center File	Revenue Center Detail
hha_base_claim_(date/time stamp).dat	HHA Base Claim File	Claim
hha_revenue_center_(date/time stamp).dat	HHA Revenue Center File	Revenue Center Detail
dme_line_(date/time stamp).dat	DMERC Line File	Line Item
dme_claim_(date/time stamp).dat	DMERC Claim Summary File	Claim

Claims Files (cont.)		
File Name	**File Description**	**Unit of Analysis**
bcarrier_line_(date/time stamp).dat	Carrier (Physician/Supplier) Line File	Line Item
bcarrier_claims_(date/time stamp).dat	Carrier Claim Summary File	Claim
institutional_condition_codes_(date/time stamp).dat	Institutional Condition Code File[2]	Code Detail
institutional_occurrence_codes_(date/time stamp).dat	Institutional Occurrence Code File[2]	Code Detail
institutional_span_codes_(date/time stamp).dat	Institutional Span Code File[2]	Code Detail
institutional_value_codes_(date/time stamp).dat	Institutional Value Code File[2]	Code Detail
Assessment Files		
mds_assessment_summary_(date/time stamp).dat	MDS	Assessment
hha_assessment_summary_(date/time stamp).dat	OASIS	Assessment
irf_assessment_summary_(date/time stamp).dat	Inpatient Rehabilitation Facility (IRF-PAI)	Assessment
sb_assessment_summary_(date/time stamp).dat	Swing Bed	Assessment

Additional Files	
File Name	**File Description**
chronic_condition_summary_(date/time stamp).dat	Chronic Condition Summary File: Summary of a beneficiary's chronic conditions by year. This file will be specific to beneficiaries included in the requested cohort and will include the time period requested by the researcher.
beneficiary_summary_(date/time stamp).dat	CCW Beneficiary Summary File[3]: This file contains variables that are very similar to the Denominator file. The source for the Beneficiary Summary File is the Unloaded EDB. This file will be specific to beneficiaries included in requested cohort (and control group, if applicable). This is an annual file.
Denominator_(date/time stamp).dat	CMS 100% Denominator File with CCW BENE_ID

[1] The following is an example of an actual file name with the date/time stamp information: inpatient_base_claim_20060501151013.dat

[2] The institutional code files include codes for all institutional claim types.

[3] The data contained in the CCW Beneficiary Summary File is very similar to that provided in the CMS 100% Denominator File.

In addition to the data files described above, the following files are also generated by each executable in the output package.

File Name	File Description
(file name).fts	For each extracted data file there will be a corresponding transfer summary file. The names of these files will correspond with the data file name [e.g., inpatient_base_claim_(date/time stamp).fts]. This file transfer summary files contain: • File name • File transfer format • File source • # Columns • File transfer mode • # Rows • Row length • File size
(file name)_v6.sas	Program to read data into a SAS version 6.x environment. For example, the file inpatient_base_claim_(date/time stamp)_v6.sas reads the inpatient base claims data into a SAS version 6.x environment.
(file name)_v8.sas	Program to read data into a SAS version 8.x environment.

Encryption Information

The encryption technique for files extracted from the CCW uses PGP Command Line 9.0 with the Self-Decrypting Archive (SDA) method. This method builds a compressed, encrypted, password protected file using a FIPS 140-1/140-2 approved AES256 cipher algorithm. The SDA is built on the CCW production server, downloaded to a desktop PC, and burned to a CD, DVD, or USB hard drive depending on the size of the files.

After the data media is shipped to the researcher, the password to decrypt the archive is sent to the researcher by electronic mail. The password is specific to a Data Use Agreement (DUA). Therefore each researcher request will have a unique encryption. The password and the data media will never be packaged together. To decrypt the data files, the researcher will need to access the e-mail containing the decryption password. Detailed instructions for using this password are included with the data.

The CCW beneficiary identifier field (BENE_ID) is specific to the CCW (not applicable to any other identification system or data source). All requested data is linked using this field. It is encrypted using an BCSSI-developed cipher prior to delivery of data files to researchers. The CLM_ID (Claim ID) and ASMT_ID (Assessment ID) will also be encrypted using the same cipher since these identifiers are also unique to a beneficiary. (These fields will not be decrypted upon receipt by the researcher. Rather it is intended that the encrypted BENE_ID will be used by the researcher to link the data and the encrypted CLM_ID and ASMT_ID will be used to identify records from the same claim/assessment). The cipher used is unique for each DUA and is determined at the time the data is requested. This key is then kept on file for future use if requested by a researcher and approved by CMS. A researcher may stipulate in a new DUA that the

data obtained must be linked to that obtained from a previous DUA. CMS will then evaluate and approve or disapprove the request. If approved, the data obtained from the CCW will be encrypted using the same cipher as the previous DUA allowing data from both requests to be linked.

Chapter 5. Limitations of the CCW Data

There are certain expected anomalies in working with large, national, administrative datasets. Minimal data cleansing has occurred during processing of CCW data. However, some of the known limitations of CMS or CCW data are described below.

Since claims for most services provided to Medicare beneficiaries in managed care do not reach the claim data files, the CCW Medicare claims should be viewed as providing utilization information primarily for the fee-for-service population.

Completeness of CCW Data

Data files are received from CMS and loaded to the CCW on a monthly basis. The 1999-2004 yearly CMS 5% SAF files were loaded to the CCW with a 6-month maturity (i.e., any changes to claims data made within 6 months of the end of the year are reflected in the CCW). Also, CMS claims requests were submitted to pull all 1999-2004 claims for any beneficiary in the 5% sample whose HIC changed prior to the end of 2004. Therefore, if a beneficiary's HIC changed during the 1999-2004 period resulting in inclusion in the CCW 5% sample, all 1999-2004 claims for that beneficiary were loaded to the CCW regardless of the HIC at the time of the claim. These transient claims were added to the SAF file claims in the CCW and de-duplicated. No additional final action processing was performed on 1999-2004 claims.

For 2005 forward, claims processed more than 365 days after the date of service are flagged in the data warehouse as *late arriving* records. These late arriving claims will not be delivered to researchers unless specifically requested. Updated assessment records are loaded to the CCW for 365 days (prior records are not maintained – only the updated assessments are maintained in the CCW). CMS estimates that over 99% of claims and assessments are processed within nine months of service (this processing time may vary by care setting).

Assessment data was obtained by creating a finder file containing all of the beneficiaries loaded into the CCW after the initial claims load. This finder file was submitted to the Quality Improvement and Evaluation System (QIES) to obtain the assessment records for those beneficiaries from 1999 forward. Final action processing was performed on these records to ensure that the most recent version of the assessment record was loaded to the CCW.

For 2005 forward, the CCW contains current claims and assessment data (contact CCW Research Coordinator for current status or level of completeness). Claims are not considered final or complete until one year after the claim thru date. Similarly, assessments may be updated until one year after the assessment date. Since data files are requested based on calendar years, researchers should consider the claims maturity or "completeness" of claims processing when requesting CCW data.

Invalid Values

Some of the CCW data files may contain invalid values, or values not conforming to the valid values provided in the CCW supporting documentation. The CCW data files

contain data as received and processed from the original CMS processing source. Invalid values are processed, stored, and delivered as they are received. No modifications or conversions are made to "correct" for invalid variable values.

One exception is the removal of spaces or decimals to the left of diagnosis or procedure codes. Any periods or blank spaces occurring to the left of the first valid numeric value within a diagnosis or procedure code field have been removed. Diagnosis and procedure codes are stored without periods. No edits have been applied to remove any blank spaces occurring within a diagnosis or procedure code (between valid numeric values).

Chapter 6. Further Assistance with CCW data

Research Data Assistance Center

The Research Data Assistance Center (ResDAC) offers free assistance to researchers using Medicare data for research. The ResDAC web site provides links to descriptions of the CMS data available, request procedures, supporting documentation, such as record layouts and SAS input statements, workshops on how to use Medicare data, and other helpful resources. Visit the ResDAC web site at http://www.resdac.umn.edu for additional information.

ResDAC is a CMS contractor and requests for assistance in the application, obtaining, or using the CCW data should first be submitted to ResDAC. Researchers can reach ResDAC by phone at 1-888-973-7322, e-mail at resdac@umn.edu, or online at https://resdac.oit.umn.edu/.

In the event that a ResDAC technical advisor is not able to answer the question, the technical advisor will direct the researcher to the appropriate person at CMS or BCSSI. If additional CMS data (data not available from the CCW) is required to meet research objectives, or the researcher has any questions about other data sources, the researcher should first visit the ResDAC website.

AHRQ

Additional information regarding the AHRQ Clinical Classifications Software and Comorbidity Software can be obtained by accessing http://www.ahrq.gov/data/hcup.

Attachment A: Chronic Condition Definitions

Algorithms	Reference Time Period (# of years)	Valid ICD-9/CPT4/HCPCS Codes[1]	Number/Type of Claims to Qualify[2]	Exclusions	Comments
Acute Myocardial Infarction	1 year	DX 410.01, 410.11, 410.21, 410.31, 410.41, 410.51, 410.61, 410.71, 410.81, 410.91 (ONLY first or second DX on the claim)	At least 1 inpatient claim with DX codes during the 1-yr period		
Alzheimer's Disease	3 years	DX 331.0 (any DX on the claim)	At least 1 inpatient, SNF, HHA, HOP or Carrier* claim with DX codes during the 3-yr period		
Alzheimer's Disease and Related Disorders or Senile Dementia	3 years	DX 331.0, 331.1, 331.11, 331.19, 331.2, 331.7, 290.0, 290.1, 290.10, 290.11, 290.12, 290.13, 290.20, 290.21, 290.3, 290.40, 290.41, 290.42, 290.43, 294.0, 294.1, 294.10, 294.11, 294.8, 797 (any DX on the claim)	At least 1 inpatient, SNF, HHA HOP or Carrier* claim with DX codes during the 3-yr period		
Atrial Fibrillation	1 year	DX 427.31 (ONLY first or second DX on the claim)	At least 1 inpatient claim or 2 HOP or Carrier* claims with DX code during the 1-yr period		Any combination of 2 HOP/Carrier claims at least one day apart.
Cataract	1 year	DX 366.01, 366.02, 366.03, 366.04, 366.09, 366.10, 366.12, 366.13, 366.14, 366.15, 366.16, 366.17, 366.18, 366.19, 366.20, 366.21, 366.22, 366.23, 366.30, 366.32, 366.33, 366.34, 366.41, 366.42, 366.43, 366.44, 366.45, 366.46, 366.50, 366.51, 366.52, 366.53, 366.8, 366.9, 379.26, 379.31, 379.39, 743.31, 743.32, 743.33, 996.53, V431, (ONLY principal DX on the claim)	At least 1 HOP or Carrier* claim with DX codes during the I-yr period		
Chronic Kidney Disease	2 years	DX 016.00, 016.01, 016.02, 016.03, 016.04, 016.05, 016.06, 095.4, 189.0, 189.9, 223.0, 236.91, 249.40, 249.41, 250.40, 250.41, 250.42, 250.43, 271.4, 274.1, 274.10, 283.11, 403.01, 403.11, 403.91, 404.02, 404.03, 404.12, 404.13, 404.92, 404.93, 440.1, 442.1, 572.4, 580.0, 580.4, 580.81, 580.89, 580.9, 581.0, 581.1, 581.2, 581.3, 581.81, 581.89, 581.9, 582.0, 582.1, 582.2, 582.4, 582.81, 582.89, 582.9, 583.0, 583.1, 583.2, 583.4, 583.6, 583.7, 583.81, 583.89, 583.9, 584.5, 584.6, 584.7, 584.8, 584.9, 585, 585.1, 585.2, 585.3, 585.4, 585.5, 585.6, 585.9, 586, 587, 588.0, 588.1, 588.81, 588.89, 588.9, 591, 753.12, 753.13, 753.14, 753.15, 753.16, 753.17, 753.19, 753.20, 753.21, 753.22, 753.23, 753.29, 794.4 (any DX on the claim)	At least 1 inpatient, SNF or HHA claim or 2 HOP or Carrier* claims with DX codes during the 2-yr period		Any combination of 2 HOP/Carrier claims at least one day apart.

Algorithms	Reference Time Period (# of years)	Valid ICD-9/CPT4/HCPCS Codes[1]	Number/Type of Claims to Qualify[2]	Exclusions	Comments
Chronic Obstructive Pulmonary Disease	1 year	DX 491.0, 491.1, 491.20, 491.21, 491.22, 491.8, 491.9, 492.0, 492.8, 494.0, 494.1, 496 (any DX on the claim)	At least 1 inpatient, SNF, HHA or 2 HOP or Carrier* claims with DX codes during the 1-yr period		Any combination of 2 HOP/Carrier claims at least one day apart.
Depression	1 year	DX 296.20, 296.21, 296.22, 296.23, 296.24, 296.25, 296.26, 296.30, 296.31, 296.32, 296.33, 296.34, 296.35, 296.36, 296.50, 296.51, 296.52, 296.53, 296.54, 296.55, 296.56, 296.60, 296.61, 296.62, 296.63, 296.64, 296.65, 296.66, 296.89, 298.0, 300.4, 309.1, 311 (any DX on the claim)	At least 1 inpatient, SNF, HHA, HOP or Carrier* claim with DX codes during the 1-yr period		
Diabetes	2 years	DX 249.00, 249.01, 249.10, 249.11, 249.20, 249.21, 249.30, 249.31, 249.40, 249.41, 249.50, 249.51, 249.60, 249.61, 249.70, 249.71, 249.80, 249.81, 249.90, 249.91, 250.00, 250.01, 250.02, 250.03, 250.10, 250.11, 250.12, 250.13, 250.20, 250.21, 250.22, 250.23, 250.30, 250.31, 250.32, 250.33, 250.40, 250.41, 250.42, 250.43, 250.50, 250.51, 250.52, 250.53, 250.60, 250.61, 250.62, 250.63, 250.70, 250.71, 250.72, 250.73, 250.80, 250.81, 250.82, 250.83, 250.90, 250.91, 250.92, 250.93, 357.2, 362.01, 362.02, 366.41 (any DX on the claim)	At least 1 inpatient, SNF or HHA claim or 2 HOP or Carrier* claims with DX codes during the 2-yr period		Any combination of 2 HOP/Carrier claims at least one day apart.
Glaucoma	1 year	DX 364.55, 365.10, 365.11, 365.12, 365.15, 365.89, 365.9, 362.85, 365.00, 365.01, 365.02, 365.03, 365.04, 377.14, 365.20, 365.21, 365.22, 365.23, 365.24, 365.41, 365.61, 366.31, V801 (ONLY principal DX on the claim)	At least 1 Carrier* claim with DX codes during the 1-yr period		
Heart Failure	2 years	DX 398.91, 402.01, 402.11, 402.91, 404.01, 404.11, 404.91, 404.03, 404.13, 404.93, 428.0, 428.1, 428.20, 428.21, 428.22, 428.23, 428.30, 428.31, 428.32, 428.33, 428.40, 428.41, 428.42, 428.43, 428.9 (any DX on the claim)	At least 1 inpatient, HOP or Carrier* claim with DX codes during the 2-yr period		
Hip/Pelvic Fracture	1 year	DX 733.98, 808.0, 808.1, 808.2, 808.3, 808.41, 808.42, 808.43, 808.49, 808.51, 808.52, 808.53, 808.59, 808.8, 808.9, 820.00, 820.01, 820.02, 820.03, 820.09, 820.10, 820.11, 820.12, 820.13, 820.19, 820.20, 820.21, 820.22, 820.30, 820.31, 820.32, 820.8, 820.9 (any DX on the claim)	At least 1 inpatient claim with DX code during the 1-yr period		

Algorithms	Reference Time Period (# of years)	Valid ICD-9/CPT4/HCPCS Codes[1]	Number/Type of Claims to Qualify[2]	Exclusions	Comments
Ischemic Heart Disease	2 years	DX 410.00, 410.01, 410.02, 410.10, 410.11, 410.12, 410.20, 410.21, 410.22, 410.30, 410.31, 410.32, 410.40, 410.41, 410.42, 410.50, 410.51, 410.52, 410.60, 410.61, 410.62, 410.70, 410.71, 410.72, 410.80, 410.81, 410.82, 410.90, 410.91, 410.92, 411.0, 411.1, 411.81, 411.89, 412, 413.0, 413.1, 413.9, 414.00, 414.01, 414.02, 414.03, 414.04, 414.05, 414.06, 414.07, 414.10, 414.11, 414.12, 414.19, 414.2, 414.3, 414.8, 414.9 Proc 00.66, 36.01, 36.02, 36.03, 36.04, 36.05, 36.06, 36.07, 36.09, 36.10, 36.11, 36.12, 36.13, 36.14, 36.15, 36.16, 36.17, 36.19, 36.2, 36.31, 36.32 HCPCS 33510, 33511, 33512, 33513, 33514, 33515, 33516, 33517, 33518, 33519, 33521, 33522, 33523, 33533, 33534, 33535, 33536, 33542, 33545, 33548, 92975, 92977, 92980, 92982, 92995, 33140, 33141 (any DX, PROC or HCPCS on the claim)	At least 1 inpatient, SNF, HHA, HOP or Carrier* claim with DX, Procedure or HCPC codes during the 2-yr period		
Osteoporosis	1 year	DX 733.00, 733.01, 733.02, 733.03, 733.09 (any DX on the claim)	At least 1 inpatient, HOP or Carrier* claim with DX code during the 1-yr period		
RA/OA (Rheumatoid Arthritis/ Osteoarthritis)	2 years	DX 714.0, 714.1, 714.2, 714.30, 714.31, 714.32, 714.33, 715.00, 715.04, 715.09, 715.10, 715.11, 715.12, 715.13, 715.14, 715.15, 715.16, 715.17, 715.18, 715.20, 715.21, 715.22, 715.23, 715.24, 715.25, 715.26, 715.27, 715.28, 715.30, 715.31, 715.32, 715.33, 715.34, 715.35, 715.36, 715.37, 715.38, 715.80, 715.89, 715.90, 715.98 (any DX on the claim)	At least 2 inpatient, SNF, HHA, HOP or Carrier* claim with DX codes during the 2-yr period		Any combination of claims at least one day apart.
Stroke / Transient Ischemic Attack	1 year	DX 430, 431, 434.00, 434.01, 434.10, 434.11, 434.90, 434.91, 435.0, 435.1, 435.3, 435.8, 435.9, 436, 997.02 (any DX on the claim)	At least 1 inpatient claim or 2 HOP or Carrier* claims with DX codes during the 1-yr period	If any of the qualifying claims have: 800 <= DX Code <= 804.9, 850 <= DX Code <= 854.1 in any DX position OR DX V57xx as the principal DX code, then EXCLUDE.	Any combination of 2 HOP/Carrier claims at least one day apart.

Algorithms	Reference Time Period (# of years)	Valid ICD-9/CPT4/HCPCS Codes[1]	Number/Type of Claims to Qualify[2]	Exclusions	Comments
Female Breast Cancer**	1 year	DX 174.0, 174.1, 174.2, 174.3, 174.4, 174.5, 174.6, 174.8, 174.9, 233.0 (any DX on the claim)	At least 1 inpatient or 2 HOP or Carrier claims with DX codes during the 1-year time period		Any combination of 2 HOP/Carrier claims at least one day apart.
Colorectal Cancer	1 year	DX 154.0, 154.1, 153.0, 153.1, 153.2, 153.3, 153.4, 153.5, 153.6, 153.7, 153.8, 153.9, 230.3, 230.4 (any DX on the claim)	At least 1 inpatient or 2 HOP or Carrier claims with DX codes during the 1-year time period		Any combination of 2 HOP/Carrier claims at least one day apart.
Prostate Cancer	1 year	DX 185, 233.4 (any DX on the claim)	At least 1 inpatient or 2 HOP or Carrier claims with DX codes during the 1-year time period		Any combination of 2 HOP/Carrier claims at least one day apart.
Lung Cancer	1 year	DX 162.0, 162.2, 162.3, 162.4, 162.5, 162.8, 162.9, 231.2 (any DX on the claim)	At least 1 inpatient or 2 HOP or Carrier claims with DX codes during the 1-year time period		Any combination of 2 HOP/Carrier claims at least one day apart.
Endometrial Cancer	1 year	DX 182.0, 182.1, 182.8, 233.2 (any DX on the claim)	At least 1 inpatient or 2 HOP or Carrier claims with DX codes during the 1-year time period		Any combination of 2 HOP/Carrier claims at least one day apart.

[1] Effective dates of these codes vary. Researchers may be interested in confirming the code(s) of interest in accompanying claims or assessment data files.

[2] Carrier claims refers to RIC "O" claims (not DMERC RIC "M" claims), and excludes any claims for which line item Berenson-Eggers Type of Service [BETOS] variable equals D1A, D1B, D1C, D1D, D1E, D1F, D1G, or O1A. The categories with D1 in the first two positions are DME categories. The O1A category includes ambulance services. The intent of the algorithm is to exclude claims where the services do not require a licensed health care professional. SNF refers to skilled nursing facility; HHA refers to home health agency; HOP refers to hospital outpatient.

**The diagnosis codes included in this definition are for female breast cancer only (male breast cancer codes are not included). Researchers may be interested in confirming gender with the accompanying beneficiary data file due to the potential miscoding of diagnosis codes.