

THIS IS A DRAFT DOCUMENT PROVIDED TO MOST APPLICANTS AS AN AID. THE CONTENT IS BEING CONSTRUCTED/REVISED. A COMPLETED VERSION WILL BE POSTED WHEN AVAILABLE.

**DATA HANDBOOK
FOR
THE CAROLINA COST AND QUALITY INITIATIVE¹**

Guidelines and Procedures

CECIL G. SHEPS CENTER FOR HEALTH SERVICES RESEARCH

(Revised Fall 2018)

¹ Handbook format adapted from the 2016-2017 UNC HPM PhD Program Manual

The *Guidelines and Procedures* contains many of the rules, regulations, policies, and procedures for using data affiliated with the Carolina Cost and Quality Initiative (CCQI), as established by the leadership and Director of the Cecil G. Sheps Center for Health Services Research. This handbook also contains various tips and answers to frequently asked questions that can be used to assist with analysis and research involving CCQI data. All researchers requesting the use of CCQI data should become familiar with the material contained in this handbook, and, together with a representative from the Sheps Center, confirm answers to any outstanding questions that remain. The most current *Guidelines and Procedures* manual is posted on the Sheps Center website (<http://www.shepscenter.unc.edu/data/bcbsnc-claims-data-ccqi/>).

Table of Contents

The Carolina Cost and Quality Initiative	
Overview.....	1
Data Access Guidelines.....	1
General Eligibility.....	
Cost of Data.....	
Data Storage.....	
Preliminary Requests.....	1
Requests for Analytic Data.....	1
Documents Required for All Data Requests.....	
Receiving IRB Approval.....	
Researcher Responsibilities.....	1
Contact Information.....	1
North Carolina Medicaid Claims Data	
Database Description.....	
Data Dictionaries.....	
Descriptive Statistics.....	
Additional Documents Required for NC Medicaid Data Requests.....	
Additional Responsibilities When Using NC Medicaid Data.....	
Note on Mental Health Encounters.....	
Additional Frequently Asked Questions About Using NC Medicaid Data.....	
Blue Cross Blue Shield of North Carolina (BCBSNC) Claims Data	
Database Description.....	
Data Dictionaries.....	
Descriptive Statistics.....	
Additional Documents Required for BCBSNC Data Requests.....	
Additional Responsibilities When Using BCBSNC Data.....	
Revenue Code Listing.....	
Additional Frequently Asked Questions About Using BCBSNC Data.....	
Programming & Statistical Analysis Guidelines for Claims Data: Dos and Don'ts, Tips and Tricks	
Comparing Professional and Facility Healthcare Claims	
Overview.....	
Claims.....	
Diagnosis Codes.....	
Procedure Codes and Coding: Evaluation and Management (E&M).....	
Date of Service.....	
Status Indicators.....	

Medical Coding Nomenclatures

Current Procedural Terminology (CPT) Code.....

International Classification of Diseases (ICD) Code.....

 Ninth Revision (ICD-9).....

 Tenth Revision (ICD-10).....

 How to Convert Between ICD-9 Code and ICD-10 Code.....

Healthcare Common Procedure Coding System (HCPCS).....

National Drug Code (NDC).....

Diagnosis-Related Group (DRG) and Major Diagnostic Category (MDC).....

Hospital Revenue Code.....

De-identified and Limited Data Guidelines

De-identified Data.....

Limited Data.....

How to Intake New Data (For Sheps Programmers Only)

Overview.....

North Carolina Medicaid Claims Data.....

Blue Cross Blue Shield of North Carolina (BCBSNC) Claims Data.....

 Extracting Data by Processing/Warehouse Date Vs. Service Date.....

Appendix 1: Summary of Workflow for Project Approval, Processing, and Governance

Figure A1. Workflow Diagram for Project Approval, Processing, and Governance.....

Appendix 2: SAS Library for Analyzing NC Medicaid Data.....

Appendix 3: SAS Library for Analyzing BCBSNC Data.....

The Carolina Cost and Quality Initiative

Overview

The *Carolina Cost and Quality Initiative* is a collaborative partnership between UNC's Gillings School of Global Public Health and the Cecil G. Sheps Center for Health Services Research. Its mission is to promote population-based research on the incidence and prevalence of disease in insured populations, patterns of utilization and treatment, and cost of care in North Carolina in order to improve the delivery and quality of care for its residents.

The Initiative currently maintains claims data from North Carolina Division of Medical Assistance (DMA) Medicaid and Blue Cross Blue Shield of North Carolina (BCBSNC).

Data Access Guidelines

General Eligibility

Only University of North Carolina at Chapel Hill faculty, staff, and graduate students are permitted to access data maintained by the Carolina Cost and Quality Initiative. Project proposals must be submitted to a review committee for approval in order to assure the feasibility of the proposed research.

Cost of Data

For unfunded projects, there are no fees for initial consultations or creation of basic analytic data sets of reasonable size. It is anticipated that these basic services will take up to 10 hours (2 hours of analytic consultation and 8 hours of analytic time). Additional charges per year of data will be made for large or complex requests. These charges cover costs associated with data preparation.

Funded projects will be asked to pay for all analytic costs. We recommend including sufficient funding for analytic time in all grant proposals. Funding to pay for data management and programming time must be arranged and documented before the data request is submitted. Data management or programming work will be estimated based on a needs assessment using an hourly rate of \$70/hour.

Data Storage

All data, including analysis subsets and extracts, must be stored and analyzed on approved secure servers at the Sheps Center. Server access fee of \$3,050 per user per year will apply. This fee supports systems maintenance and security controls.

Preliminary Requests

Requests for preliminary data for grant submissions or planning purposes can be submitted by [email](#). (Click the embedded link or see the Contact Information section, page 3.) Include a paragraph describing your project and what information is needed. Preliminary information is limited to counts and basic descriptive statistics.

Requests for Analytic Data

Phone consults are required prior to submitting a data request. Please see the Contact Information section on page 3.

The CCQI Oversight Committee reviews all requests to use NC Medicaid data and BCBSNC claims data. This committee is made up of UNC faculty as well as representatives of the data owners. The full approval process can take several months, depending on whether revisions are required. The entire review process will be coordinated by the CCQI project manager. A full summary of the workflow for project, approval, processing, and governance is provide in Appendix 1 (p. ...).

The CCQI Oversight Committee uses the following criteria to evaluate data requests:

1. Is the research question one that is in keeping with the overall focus of the Initiative and important to be answered, for North Carolinians or the US health system as a whole, either from a policy or practice perspective?
2. Can the research question be adequately answered using the requested database? Are there an adequate number of cases to be studied for the proposed analysis? Are the right data elements available and specified? For the key data elements needed, is the quality of the data adequate?
3. Is the proposed methodology sound and viable with respect to the research questions, population and data requested?
4. Does the research team possess the requisite skills/experience and proposed resources to conduct the research with high standards and in an adequately specified time frame?

Documents Required for All Data Requests

The following two documents must be completed prior to all data requests:

- [CCQI Data Request Form](#) (To access, click the embedded link)
- IRB approval (in PDF, on UNC IRBIS letterhead)

Completed data request forms may be emailed to [Abigail Haydon](#) (Contact Information, p. 3).

Researcher Responsibilities

Data use agreements (DUAs) must be kept up to date. All staff and researchers with access to the data must sign a DUA.

The following funding acknowledgment must be included in all journal articles or conference abstracts using CCQI databases:

The database infrastructure used for this project was funded by the Cecil G. Sheps Center for Health Services Research; the Department of Health Policy and Management, UNC Gillings School of Global Public Health; the CER Strategic Initiative of UNC's Clinical and Translational Science Award (UL1TR001111); and the UNC School of Medicine.

In addition, as a condition of CCQI data use, investigators must adhere to the following requirements:

- a) Submit periodic project updates (e.g., quarterly updates).
- b) Leave all data on the designated server.
- c) Submit manuscripts to the Sheps Center for review at least 30 days prior to submission of a peer-reviewed publication (14 days for an abstract or presentation). These are shared with the data owner who the right to comment and suppress any inappropriate release of identifying information that violates the DUA.
- d) Prepare a final "report" that provides value to the data owner. This may take the form of a technical memo, a short policy brief, or another mutually agreeable form.

Contact Information

For more information on the Carolina Cost and Quality Initiative, please contact:

[Sandra B. Greene, DrPH](#)

Co-Director, Program on Health Economics and Finance & Senior Research Fellow

919 966-0993

sandrab_greene@unc.edu

[Abigail Haydon, PhD](#)

Research Program Manager, NC TraCS Institute

(919) 966-6879

ahaydon@email.unc.edu



North Carolina Medicaid Claims Data

Database Description

North Carolina Medicaid claims data are available as a limited dataset. Patient identifiers are encrypted and no identification of individual patients is possible. Data are updated approximately every 3 months.

The NC Medicaid database contains claims data from the North Carolina Department of Health and Human Services Division of Medical Assistance.

The database includes the following:

- Claims information (including doctor visits, inpatient and outpatient care, prescription medications, dental treatments, charge amounts, dates of service, and diagnoses and procedures).
- Provider information (including specialty and location).
- Member information (including date of birth, gender, county/ZIP code of residence, race and ethnicity).

Dates available: January 2011 – December 2017

Data Dictionaries

Data dictionaries for NC Medicaid data can be found at:

<http://www.shepscenter.unc.edu/data/bcbsnc-claims-data-ccqi/data-dictionaries/>

Excel versions of data dictionaries and additional look-up tables are available upon request

Descriptive Statistics

Descriptive statistics for NC Medicaid data can be found at:

<http://www.shepscenter.unc.edu/data/bcbsnc-claims-data-ccqi/descriptive-statistics/>

Additional Documents Required for NC Medicaid Data Requests

The following two documents (in addition to the documents listed on page 2) must be completed prior to all data requests:

- [Request for Access to Health Information for Research](#)
- [NC Medicaid Data Use Acknowledgment](#)

Completed data request forms may be emailed to [Abigail Haydon](#) (Contact Information, p. 3).

Additional Responsibilities When Using NC Medicaid Data

In addition to the researcher responsibilities listed on pages 2 and 3, all journal articles and conference abstracts must be submitted to the Sheps Center to send on to the NC Division of Medical Assistance at least 30 days prior to publishing the research findings.

Note on Mental Health Encounters

Since 2013, there have been mental health encounters at NC Local Management Entities / Managed Care Organizations that do not appear in the Sheps NC Medicaid database. Often, these excluded encounters were originally processed by the DMA as denied encounters (due to some technicalities), when in fact the encounters were legitimate paid claims. The Sheps NC Medicaid database does not include denied encounters, so paid claims that were originally processed by the DMA as denied encounters are not recorded by Sheps. The Sheps Center is currently in the process of asking the DMA to provide these denied encounters so that they can be added to the NC Medicaid database.

Additional Frequently Asked Questions About Using NC Medicaid Data

Q: What are the differences between a member file, a claim file, and a member reference file?

A: To put it simply, a *member file* keeps track of enrollment and eligibility start / end dates for Medicaid enrollees. Each member is assigned a unique member identification number, known as an MID. A given Medicaid enrollee will have the same MID as long as he or she is continuously enrolled / covered. However, if an individual drops in and out of coverage, they may be assigned a new MID for each continuous coverage period. It is also important to note that MID is often labeled as "AIT_MBR_XREF_ID_ENCRYPT" in member files.

A *claim file* keeps track of the claims associated with Medicaid enrollees over time. Claim files use the variable "ALT_MBR_ID_ENCRYPT" to match claims to a given Medicaid enrollee. Note that the ALT_MBR_ID_ENCRYPT variable is similarly named to the AIT_MBR_XREF_ID_ENCRYPT (MID) variable from the member files. However, these variables are NOT the same. A Medicaid enrollee will have the same MID as long as Medicaid continuously covers the individual. In contrast, a single Medicaid enrollee can have multiple values for the ALT_MBR_ID_ENCRYPT variable over time. The MID from the member file and the ALT_MBR_ID_ENCRYPT value from the claim file will sometimes be the same, but such is not always the case.

To illustrate the difference between an MID and the ALT_MBR_ID_ENCRYPT variable, consider the following table:

MID	ALT_MBR_ID_ENCRYPT	HDR_SVC_BGN_DT	HDR_SVC_END_DT
OEPb4MIAUv	OEPboTPoUv	18DEC2012	18DEC2012
OEPb4MIAUv	OEPboTPoUv	18DEC2012	18DEC2012
OEPb4MIAUv	OEPboTPoUv	18DEC2012	18DEC2012
OEPb4MIAUv	OEPb4MIAUv	15AUG2013	15AUG2013
OEPb4MIAUv	OEPb4MIAUv	15AUG2013	15AUG2013
OEPb4MIAUv	OEPb4MIAUv	15AUG2013	15AUG2013

OEPb4MIAUv	OEPb4MIAUv	17APR2014	17APR2014
OEPb4MIAUv	OEPb4MIAUv	17APR2014	17APR2014
OEPb4MIAUv	OEPb4MIAUv	17APR2014	17APR2014
OEPb4MIAUv	OEPb4MIAUv	16FEB2015	16FEB2015
OEPb4MIAUv	OEPb4MIAUv	16FEB2015	16FEB2015
OEPb4MIAUv	OEPttotyhA	28DEC2015	28DEC2015
OEPb4MIAUv	OEPttotyhA	28DEC2015	28DEC2015
OEPb4MIAUv	OEPttotyhA	28DEC2015	28DEC2015
OEPb4MIAUv	OEPttotyhA	30JUN2016	30JUN2016
OEPb4MIAUv	OEPttotyhA	30JUN2016	30JUN2016
OEPb4MIAUv	OEPttotyhA	30JUN2016	30JUN2016
OEPb4MIAUv	OEPttotyhA	30JUN2016	30JUN2016
OEPb4MIAUv	OEPttotyhA	30JUN2016	30JUN2016
OEPb4MIAUv	OEPttotyhA	30JUN2016	30JUN2016
OEPb4MIAUv	OEPttotyhA	07SEP2016	07SEP2016
OEPb4MIAUv	OEPttotyhA	07SEP2016	07SEP2016
OEPb4MIAUv	OEPttotyhA	10OCT2016	10OCT2016
OEPb4MIAUv	OEPttotyhA	10OCT2016	10OCT2016
OEPb4MIAUv	OEPttotyhA	14NOV2016	14NOV2016
OEPb4MIAUv	OEPttotyhA	14NOV2016	14NOV2016

The above table contains the claims information for a single Medicaid enrollee with MID "OEPb4MIAUv." This individual is continuously enrolled from March 1st, 2011 until August 31st, 2018 and has various claims submitted on his/her behalf during this time period. Note that the MID stays constant across the entire time period (the MID column is usually not included in a claim file, but is included in the above table for illustrative purposes). However, the ALT_MBR_ID_ENCRYPT variable changes over time. ALT_MBR_ID_ENCRYPT "OEPboTPoUv" was applied to the claims on December 18th, 2012, ALT_MBR_ID_ENCRYPT "OEPb4MIAUv" was applied to the claims from August 2013 until February 2015, and "OEPttotyhA" was applied to the claims from December 2015 until November 2016.

Finally, a *member reference* file, also known as a *mem_xref* file or a "lookup table," can be used in conjunction with the aforementioned files to track a continuously enrolled individual patient across time. The member reference file contains a "REF_MBR_ID" variable that will match the MID variable from the member file (remember, the *member reference* file and the *member* file are two separate files). Somewhat confusingly, the member reference file also contains its own "MID" variable that records the different ALT_MBR_ID_ENCRYPT values associated with an individual patient. Thus, a member reference file would be formatted similarly to the following table:

MID	REF_MBR_ID	ALT_MBR_RCRD_STATUS	RCRD_STS_CD
OEPB4MIAUv	OEPB4MIAUv	A	A
OEPboTPoUv	OEPB4MIAUv	M	A

OEPTtotyhA	OEPB4MIAUv	M	A
OEPsLolXYq	OEPsLolXYq	M	A
OEIW031X2H	OEPsLolXYq	A	A
OEl0yhTNwA	OEPsLolXYq	M	A

Note that the first three observation rows in the above table all correspond to the same individual patient, and the three values in the "MID" column for this patient correspond to the three ALT_MBR_ID_ENCRYPT values in the example claims file table. The last three observation rows in the example member reference file table correspond to a different individual patient with a REF_MBR_ID of "OEPsLolXYq."

Q: What are the differences between an MID, ALT_MBR_ID_ENCRYPT, and Ref_MBR_ID?

A: See above question.

Q: How to I track an individual patient across time?

A: To track an individual patient's continuous enrollment across time, simply use the information contained in the member file.

To track an individual patient's claims history, use the member reference file in conjunction with the claim file. First, match each observation in the claim file to the corresponding observation in the member reference file using the ALT_MBR_ID_ENCRYPT (claim file) and MID (member reference file) variables. In a correct match, the ALT_MBR_ID_ENCRYPT variable and MID variable will have the same sequence of letters and numbers. Next, group all matched observations by the REF_MBR_ID variable in the member reference file. All observations belonging to a single REF_MBR_ID will represent the claims history for an individual patient.

To link the claims history with the data from the member file, match the REF_MBR_ID variable from the member reference file to the MID variable from the member file (Note: the member reference file and the member file are two separate files).

Q: What algorithm is used to link member data across files (e.g., claim and member files)?

A: See above question.

Q: How do I check for multiple claim records on one event?

A:

Blue Cross Blue Shield of North Carolina Claims Data

Database Description

Blue Cross Blue Shield of North Carolina (BCBSNC) claims data are available as a limited dataset. Patient identifiers are encrypted and no identification of individual patients is possible. Data are updated approximately every 6 months.

The BCBSNC database contains claims data on health care services reimbursed by BCBSNC. The database includes claim records from three types of insured groups: 1. Insured and multiple employer welfare arrangement (small groups), 2. Administrative services only (ASO) groups (large groups), and 3. Individual market and Affordable Care Act Exchange Plans (insurance plans purchased by individuals).

The database includes the following:

- Claims information (including doctor visits, inpatient and outpatient care, prescription medications, dental treatments, charge amounts, dates of service, and diagnoses and procedures).
- Provider information (including specialty and location).
- Member information (including date of birth, gender, county/ZIP code of residence).

Dates available: January 2006 – December 2017.

Data Dictionaries

Data dictionaries for BCBSNC data can be found at:

<http://www.shepscenter.unc.edu/data/bcbsnc-claims-data-ccqi/data-dictionaries/>

Excel versions of data dictionaries and additional look-up tables are available upon request

Descriptive Statistics

Descriptive statistics for BCBSNC data can be found at:

<http://www.shepscenter.unc.edu/data/bcbsnc-claims-data-ccqi/descriptive-statistics/>

Additional Documents Required for BCBSNC Medicaid Data Requests

The following document (in addition to the documents listed on page 2) must be completed prior to all data requests:

- [CCQI Data Use Agreement](#)

Completed data request forms may be emailed to [Abigail Haydon](#) (Contact Information, p. 3).

Additional Responsibilities When Using BCBSNC Data

In addition to the researcher responsibilities listed on pages 2 and 3, all journal articles and conference abstracts must be submitted to the Sheps Center to send on to BCBSNC and/or the State Health Plan at the time of submission to journal or conference, or earlier.

Revenue Code Listing

A list of commonly used revenue codes can be found at:

<http://valuehealthcareservices.com/education/understanding-hospital-revenue-codes/>

Revenue codes must be filed with appropriate CPT and HCPCS codes for Outpatient and Ambulatory Surgery UB-04 or 837-Institutional Healthcare Claims.

Additional Frequently Asked Questions About Using BCBSNC Data

Q: What is the difference between a "policy member ID" and a "unique member ID"?

A: The policy member ID (PMI) is defined as the member assigned to a particular employer group for a coverage term. If a particular member were to change employers, that member (if they still had BCBSNC employer coverage with the new employer) would get a new PMI assignment. Similarly, if the member were to change from their insurance policy to become a dependent on another person's policy, their PMI assignment would change as well. To summarize, a PMI accurately identifies the combination of a member and their enrollment.

The unique member ID (UMI), also called the PersonID, is occasionally subject to change as well. The UMI is an attempt to identify a person regardless of whether they change coverage or employer groups. The difficulty with this assignment is that it is not necessarily static across time. At a given moment in time, this UMI is relatively accurate (profiling by BCBSNC shows > 95% accuracy), but as new information arrives from source data, the UMI may change.

As an example, a UMI could change when earlier information leads BCBSNC to believe that two people are the same person, and then later information causes BCBSNC to decide that they are not the same person. In the instance of twins, where the birth date, gender, and address are the same and the SSN is blank, BCBSNC has sometimes assigned the same UMI to both people. An update to the member demographics at a later time may show two different SSNs for these people, which causes BCBSNC to assign a new UMI to one of the twins.

In general, UMIs are useful when examining high-level trends, but may not allow the accurate tracking of a specific person across time.

Q: How to I track an individual patient across time using policy member IDs and unique member IDs?

A:

Q: What algorithm is used to link data?

A:

Q: One row for a particular enrolled member shows a coverage termination date of '12/31/9999,' and another row for the same member shows the same enrollment begin date but a termination date that is not '12/31/9999.' How do I interpret this change in coverage termination date?

A: As an example of coverage month changes, consider a member whose record for April 2015 coverage showed an enrollment period of '1/1/2015' through '12/31/9999.' This means that, as of April 2015, this member had no stated termination date. Now, assume the records for May, June, July, etc. showed an enrollment period of '1/1/2015' through '12/31/2015.' This means that, sometime in May, this member's enrollment was updated to show that it will terminate at the end of the year. The point is that the enrollment period info is accurate AS OF the stated coverage month.

Q: How are data recorded for a member with enrollment in multiple lines of business?

A: Consider a member with enrollment in both PPO1 (Blue Options medical coverage) and GIS1 (Group Insurance Services life insurance). For each coverage month, the member has 2 rows, one for each line of business. The enrollment period for each type of coverage may be different.

Q: How does a new user access the ... folder?

A: In order to access the ... folder, a new user must have the following:

- 1) A connection to the UNC Virtual Private Network (VPN)
- 2) Access to base SAS or SAS Studio

To connect to the Sheps server from SAS, use the following syntax:

```
%let server=n2.schsr.unc.edu 1234;  
options comamid=tcp remote=server;  
signon username=_prompt_;
```

Q: How do I check for duplicate claims?

A:

Programming & Statistical Analysis Guidelines for Claims Data: Dos and Don'ts, Tips and Tricks

How to Handle Claims Records with Service Dates in the Future

By law, BCBSNC must process all claims submitted by providers. If there is a date error in a submitted claim, the error could be retained in our data. As an example, a claim submission may have a data error with a service date in the future. Those claims are in our data but they should be "Denied." We have found a small number of claims with service dates beyond 2018. Always check the Denied Indicator field (DEN_IND) to determine if a claim was denied. Be similarly vigilant when checking claims with service dates much earlier than 2006 (these may also be data errors).

How to Determine the Correct Time Period for Hospitalizations

Extended hospital stays can result in multiple claim filings by providers, particularly if the hospital stay goes over multiple billing cycles for a provider and the provider submits interim bills. This could result in an inaccurate value for the Date_Service_End variable because the end date for a submission may be related to the billing cycle date instead of the actual discharge date for the patient. In such situations, multiple claims are filed to cover the entire hospital time period. These multiple claims should have the same value for the Date_Service_Start variable but will have different Date_Service_End values. Remember to follow the chain of claims to get the correct Date_Service_End value.

How to Accurately Count Number of Procedures for a Patient or Population

How to Calculate Hospital Discharge Rates

The numerator of the rate is the number of discharges in a specific time period, typically 12 months. The numerator is derived from inpatient institutional claims. The key is to make sure one does not over count. If there are multiple inpatient institutional claims for an individual, look at admission dates to make sure you are counting unique discharges. The denominator of the rate is the average number of person years that were covered members during the 12-month period. Once you divide the numerator (number of discharges) by the denominator (population), multiply by 1000. This will give the discharge rate per 1000 persons.

How to Determine the Total Cost of an Episode of Care

(Note: We do not have cost data on claims, so the best we can do is determine charges) If the episode is an inpatient admission, the first step is to identify the institutional claims related to the admission. There may be multiple institutional claims for the same episode, related to interim bills or corrections/additions to previous bills. Once you identify the correct set of institutional claims, determine the admission and discharge dates. Then look at the professional claims and identify all claims for the patient where the date of service was between the admission and discharge dates. Once all the claims, both institutional and professional, are identified, sum up charges on all of these claims. This should represent the charges related to the episode.

How to Access Stata and R on the Sheps Server

Before accessing Stata or R on the Sheps server, first connect to the Sheps VPN.

Documentation for connecting to the Sheps VPN is provided here:

<https://intranet.shepscenter.unc.edu/services/computer-support/connecting-to-the-sheps-vpn/>

Accessing Stata (Mac)

On your Mac, you'll need XQuartz if you don't already have it (see here for download details: <https://www.xquartz.org>).

Once XQuartz is installed and opened, a Terminal window should appear. If a window does not appear automatically, go to the Applications menu on your Mac and choose "Terminal."

In the Terminal window, type "xhost +n2.schsr.unc.edu" and hit return. This tells your Mac to allow connections from the server.

Then, type "ssh -Y onyen@n2.schsr.unc.edu" into the Terminal window, replacing "onyen" with your UNC onyen.

After you authenticate, you'll have a command line on the server. To run Stata with the X-Windows system, type "xstata" and hit return.

Stata should start up in a windowing environment. You are now working remotely on the server.

Note: This approach is great for interactive use, but if you run very long jobs (i.e., jobs that take hours or days), you may want to work from the command line directly in a terminal. This is because an idle network connection will eventually be cut by the campus firewall, killing your job. To get jobs running on the server from the command line without the need to stay connected, use commands such as "nohup," "batch," "at," or "tmux" and type "stata" on the command line rather than "xstata."

Accessing Stata (Windows)

Accessing R (Mac & Windows):

For R, the easiest way to work is using RStudio. Once you've connected to the Sheps VPN, all you need to do is open a browser to:

<http://n2.schsr.unc.edu:8787>

You may see a security warning about lack of SSL, but this is okay since you are already working over an encrypted VPN tunnel.

DRAFT

Medical Coding Nomenclatures

Current Procedural Terminology (CPT) Code

From the website of the American Academy of Professional Coders

(<https://www.aapc.com/resources/medical-coding/cpt.aspx>):

“CPT® codes are the United States’ standard for how medical professionals document and report medical, surgical, radiology, laboratory, anesthesiology, and evaluation and management (E/M) services. All healthcare providers, payers, and facilities use CPT® codes.

The five-character CPT® codes are used by insurers to help determine the amount of reimbursement that a practitioner will receive for services provided.

Current Procedural Terminology (CPT®) codes were first published in 1966 and are developed, maintained, and copyrighted by the American Medical Association (AMA). Thousands of CPT® codes are in use, and they are updated annually. They fall into three categories:

- Category I – These five-digit codes have descriptors which correspond to a procedure or service. Codes range from 00100 - 99499.
- Category II – These alphanumeric tracking codes are used for execution measurement. Using them is often optional.
- Category III – These are provisional codes for new and developing technology, procedures, and services. The codes were created for data collection and assessment of new services and procedures.

Two-character modifiers are appended to CPT® codes to report special circumstances and to clarify or modify the description of the procedure. These modifiers are developed by the AMA and the Centers for Medicare & Medicaid Services (CMS).

The rules for assigning appropriate code(s) are complex, and although CPT® codes are standardized, the amount various practitioners are reimbursed for their services or procedures are not necessarily the same. In Medicare, CPT® is Level I of the Healthcare Common Procedure Coding System ([HCPCS](#)) code system and accompanied by the HCPCS Level II supply and service codes.

[Medical terminology](#), anatomy, insurance payer rules, procedural bundling rules, and proper modifier usage are all essential elements to know when reporting CPT® codes. To ensure coding accuracy, individuals who are responsible for determining and reporting CPT® codes on medical claims should receive proper training and credentials.”

For a list of CPT codes, see here: <https://coder.aapc.com/cpt-codes/>

International Classification of Diseases (ICD) Code Ninth Revision (ICD-9)

From the website of the American Academy of Professional Coders (<https://www.aapc.com/resources/medical-coding/icd9.aspx>):

“The International Classification of Diseases (ICD) is the oldest method of tracking diseases and mortality in the world. It was first developed in Europe, and several versions have evolved over the years. The first edition, known as the International List of Causes of Death, was adopted by the [International Statistical Institute](#) in 1893. The current version used in the United States was established by the [World Health Organization \(WHO\)](#) and has seen regular modifications. ICD-9-CM (Clinically Modified) was adopted in United States in 1979. The code set is updated at least once a year, based on the input of providers, payers, and other key stakeholders. A new generation and much larger code set, [ICD-10](#), [replaced] ICD-9 codes on Oct 1, 2015.

ICD-9-CM was mandated in 2003 by the [Health Insurance Portability and Accountability Act of 1996 \(HIPAA\)](#). A comprehensive understanding of the ICD-9-CM code set is essential for all medical coders and billers who work for health plans, healthcare clearinghouses, and healthcare providers transmitting any electronic health information.

The ICD-9-CM code set consists of:

- Volume 1: The numeric listing of diseases, classified by etiology and anatomical system, along with as a classification of other reasons for encounters and causes of injury. This is called the tabular section of ICD-9-CM. Volume 1 is used by all health care providers and facilities.
- Volume 2: The alphabetic index used to locate the codes in Volume 1. Volume 2 is used by all healthcare providers and facilities.
- Volume 3: A procedural classification with a tabular section and an index. This set of [procedure codes](#) is used only by hospitals to report services.

All ICD-9-CM changes and modifications are the responsibility of the Centers for Disease Control (CDC), the National Center for Health Statistics (NCHS), and the Centers for Medicare & Medicaid Services (CMS).”

ICD-9-CM Diagnoses Codes adhere to the following guidelines:

- 3-5 digits
- First digit is alpha (E or V) or numeric
- Digits 2-5 are numeric
- Decimal is after the third digit

ICD-9-CM Procedure Codes adhere to the following guidelines:

- 3-4 digits
- All digits are numeric
- Decimal is after the second digit

For a list of ICD-9 codes, see here: <https://coder.aapc.com/icd9-codes>

Tenth Revision (ICD-10)

From the website of the American Academy of Professional Coders

(<https://www.aapc.com/resources/medical-coding/icd-10.aspx>):

"ICD-10 codes refer to the International Classification of Diseases, Tenth Revision (ICD-10) standard coding system used by physicians and other healthcare providers to classify and report medical diagnoses. The ICD-10-CM code set was developed by the World Health Organization (WHO) in 1996 and adopted by the United States in its modified form on Oct. 1, 2015. Like its predecessor ICD-9 with over 17,000 codes, ICD-10 uses alphanumeric codes to identify diseases, related health problems, abnormal findings, signs and symptoms of, external causes of injury or diseases, and social circumstances; only ICD-10 codes far exceed ICD-9 codes in the number of concepts and codes at over 141,000 codes. When most people talk about ICD-10 codes, they are referring to ICD-10-CM (Clinical Modification), not ICD-10-PCS (Procedure Coding System). Here are the differences in the two ICD-10 code sets:

- ICD-10-CM diagnosing coding was developed by the Centers for Disease Control and Prevention (CDC) for use in all healthcare treatment settings in the United States.
- ICD-10-PCS is a United States code set developed by 3M Health Information Systems and the Centers for Medicare & Medicaid Services (CMS) exclusively for the United States to report inpatient procedures in the hospital setting only.

Both sets use seven character, alphanumeric "smart code" systems that indicate the type of illness or surgery, the location, the severity, and any other information needed to tell the whole story of the patient's diagnosis.

Clinical staff, coders, health information managers, and other healthcare professionals use ICD-10-CM to assist in storage and retrieval of diagnostic information. ICD-10 codes provide better data than ICD-9 codes for measuring and tracking healthcare use and the quality of patient care and in the compilation of national mortality and morbidity statistics.

Because ICD-10 codes are so specific and clinical, individuals assigning them must be trained as medical coding professionals, with a strong understanding and experience in anatomy and physiology."

ICD-10-CM Diagnoses Codes adhere to the following guidelines:

- 3-7 digits
- Digit 1 is alpha
- Digit 2 is numeric
- Digits 3-7 are alpha or numeric (alpha digits are not case sensitive)
- Decimal is after the third digit

ICD-10-PCS Procedure Codes adhere to the following guidelines:

- 7 digits

- Each digit is either alpha or numeric (alpha digits are not case sensitive and letters O and I are not used to avoid confusion with numbers 0 and 1)
- No decimal

To look-up ICD-10 codes, see here: <https://icdlookup.com/icd-10/codes>

How to Convert Between ICD-9 Code and ICD-10 Code

To convert between ICD-9 and ICD-10 code, use the following online tool:

<https://www.aapc.com/icd-10/codes/>

Healthcare Common Procedure Coding System (HCPCS)

From the website of the American Academy of Professional Coders

<https://www.aapc.com/resources/medical-coding/hcpcs.aspx>:

“HCPCS is an acronym for Healthcare Common Procedure Coding System (HCPCS). Standardized code sets are necessary for Medicare and other health insurance providers to provide healthcare claims that are managed consistently and in an orderly manner. HCPCS Level II coding system is one of several code sets used by healthcare professionals, including medical coders and billers. The Level I HCPCS code set includes CPT® (Current Procedural Terminology) codes. CPT is developed and owned by the American Medical Association (AMA).

Sometimes described as the "hall closet of coding," HCPCS Level II meets several needs. The code set is made up of five-character, alpha-numeric codes mainly representing medical supplies, durable medical goods, non-physician services, and services not represented in the Level I code set (CPT®). HCPCS Level II includes services such as ambulance, durable medical equipment, prosthetics, orthotics and supplies (DMEPOS), when used outside a physician's office. This coding system is also used as an official code set for outpatient hospital care, chemotherapy drugs, Medicaid, and other services. The Blue Cross Blue Shield Association and the American Dental Association (ADA) post their procedure codes as part of HCPCS Level II. The Centers for Medicare & Medicaid Services (CMS) regularly uses HCPCS Level II to post codes for the tracking of demonstration projects and new technologies.”

The development and use of HCPCS Level II codes began in the 1980s. In 2003, the Secretary of Health and Human Services (HHS) delegated authority (under the Health Insurance Portability & Accountability Act of 1996 [HIPAA] legislation) to CMS to maintain and distribute HCPCS Level II codes. The code set is revised quarterly based on public input, which includes feedback from providers, manufacturers, vendors, specialty societies, the ADA, Blue Cross, and others.”

For a list of HCPCS codes, see here: <https://coder.aapc.com/hcpcs-codes>

National Drug Code (NDC)

For an overview of NDCs, see the following document provided by the Idaho Medicaid Management Information System:

<https://www.idmedicaid.com/Reference/NDC%20Format%20for%20Billing%20PAD.pdf>

To look up NDCs, see here: <https://ndclist.com/search>

Diagnosis-Related Group (DRG) and Major Diagnostic Category (MDC)

For an overview of DRGs and MDCs, see the following document provided by the Centers for Medicare and Medicaid Services:

[https://www.cms.gov/ICD10Manual/version34-fullcode-cms/fullcode_cms/Design_and_development_of_the_Diagnosis_Related_Group_\(DRGs\)_PBL-038.pdf](https://www.cms.gov/ICD10Manual/version34-fullcode-cms/fullcode_cms/Design_and_development_of_the_Diagnosis_Related_Group_(DRGs)_PBL-038.pdf)

Hospital Revenue Code

The following information comes directly from ValueHealthcareServices.com

(<http://valuehealthcareservices.com/education/understanding-hospital-revenue-codes/>):

“Medical coding and billing is a complex procedure. That’s where Revenue Codes come into play to help make the whole process smoother and universal.

In short, Revenue Codes are descriptions and dollar amounts charged for hospital services provided to a patient. The revenue code tells an insurance company whether the procedure was performed in the emergency room, operating room or another department.

Hospitals run on three sets of universal codes:

1. ICD-9 & ICD-10 – for diagnoses
2. CPT – for procedures
3. HCPCS – for medical goods and services.

These codes are located in various places on the UB-04 forms (bill form used by Hospitals and other providers), the purpose of Revenue Codes, however, is to group similar types of charges onto one line in the form.

If a revenue code is attached to a supply code, it identifies the equipment and whether the equipment was used in the hospital or taken home by a patient. This is important because many of the procedures done in the hospital may be done in different areas. For example, stitches may be given to a patient in the emergency room, or in a completely different area of the hospital like the maternity ward. Depending on where the procedure was done the price for the procedure can be drastically different. We therefore see the tremendous role that Revenue Codes have in the revenue cycle of a hospital.

A valid procedure code must be accompanied by a revenue code for it to be accepted by the insurance provider. There are 81 fields on the UB-04 and the Revenue Codes are located by field 42-49 (FL42-49). The Revenue Codes were developed for the Medicare system but were

soon adopted as standard for hospitals. Every item in a hospital's chargemaster (catalog of all services performed by that hospital) must have one revenue code attached to it.

Originally Revenue Codes were 3 digits, however, the need for additional codes required the addition of a new digit. Nowadays, the code is always 4 digits, the first digit is typically a zero and many insurance agencies will assume a zero for the first digit rather than require it to be listed. When the last digit of a Revenue Code is a zero this indicates that the service was unspecified and a more specific assignment could not be made. When the last digit is a nine this indicates that "other" services within the specified category were performed which do not have a specific code assigned to them. A claim form sent to an insurance provider without a Revenue Code will be rejected, and only Revenue Codes that are covered by the insurance provider will be paid."

The list of Revenue Codes is quite extensive. A list of commonly used Revenue Codes can be found at: <http://valuehealthcareservices.com/education/understanding-hospital-revenue-codes/>.

Comparing Professional and Facility Healthcare Claims

Overview

The choice of whether to bill on a professional claim (CMS 1500) or facility claim (UB-04) is based on insurance coverage. If the insurance allows the billing of a facility fee, a facility claim may be filed. For example, if a physician provides a service in an emergency department, both a facility fee and a professional fee will be filed. But if the physician provides the same service in a clinic, only a professional fee is allowed.

Additional information from Health_INS_101.LiveJournal.com:

"A *professional claim* is a claim representing services provided to you by a physician, surgeon, therapist, or other skilled health care professional who is licensed for independent practice (that is to say, they can, if they choose, go start a practice on their own - they don't have to operate under the aegis of a larger organization). This will include pretty much any medical doctor you see in a doctor's office, as well as most doctors you see in a hospital (but not resident physicians, who are still not fully licensed). It also includes "freestanding" health care providers, such as physical therapists with offices (that is, not part of a larger hospital), and independent clinical laboratories (for instance, [Labcorp](#)). Most durable medical equipment is billed on this kind of form. A freestanding urgent care center claim will be on this form.

A *facility claim* is a claim billed by a hospital, ambulatory surgery center, dialysis center, home health care provider, or any other type of acute-care or ancillary health care provider. Hospital-based urgent care center claims will be on this form, as will the claims of physical therapists that you see inside a hospital. Facility claims are billed on a UB-92 form."²

The following sub-sections are taken directly from a 06/18/14 blog post by Edie Hamilton, CPC.³

Claims

- *Professional*: Each provider submits an individual claim.
- *Facility*: The facility submits a claim with all services for a single patient for each date of service at the same facility. A single facility claim may be associated with multiple professional claims.

Diagnosis Codes

- *Professional*: The claim includes diagnosis codes relevant to the encounter by a single provider, and pointers are used to link specific codes to each procedure.

² <https://health-ins-101.livejournal.com/1444.html>

³ <https://blog.veriskhealth.com/comparing-professional-and-facility-healthcare-claims>

- *Facility*: The claim includes diagnosis codes relevant to all services for a single patient for each date of service at the same facility. This may include services performed in multiple areas (e.g., lab, radiology, clinic). There are no pointers to assign diagnosis codes to individual lines.

Procedure Codes and Coding: Evaluation and Management (E&M)

- *Professional*: The 1995 or 1997 Documentation Guidelines for Evaluation and Management Services are used to select the appropriate type of visit and level of service [**Professional claims use CPT codes**].
- *Facility*: For the Centers for Medicare & Medicaid Services (CMS), there is a single E&M code, G0463, for related facility claims, except for the emergency department. For other payers, hospitals develop and follow their own internal guidelines that should, per CMS, “reasonably relate the intensity of hospital resources to the different levels of effort represented” [Facility claims used ICD-9 and ICD-10 codes. While CPT codes are often on ambulatory surgery facility claims, they are only sometimes used for hospital outpatient facility claims.].

Date of Service

- *Professional*: Lines may have multiple units and a “from and thru” date span.
- *Facility*: The header includes “from and thru” dates, but each line is specific to a single date of service.

Status Indicators

- *Professional*: The Medicare Physician Fee Schedule Relative Value File defines professional status indicators.
- *Facility*: Addendum B from CMS defines facility status indicators.

De-identified and Limited Data Guidelines

De-identified Data

A de-identified data set **may not include** any of the following direct identifiers of the individual or of the individual's relatives, employers, or household members.

- Names
- Geographic subdivisions smaller than a state
- Zip codes
- All elements of dates except year directly related to an individual, including birth or death or dates of health care services or health care claims
- Telephone numbers
- Fax numbers
- Electronic mail addresses
- Social security numbers
- Medical record numbers
- Health plan beneficiary identifiers
- Account numbers
- Certificate/license numbers
- Vehicle identifiers and serial numbers, including license plate numbers
- Device identifiers and serial numbers
- Web universal resource locators (URL)
- Internet protocol (IP) address numbers
- Biometric identifiers, including finger and voice prints
- Full face photographic images
- Any other number, characteristic, or code that could be used by the researcher to identify the individual

Note: Although a de-identified data set cannot contain a birth date, it may contain the individual's age expressed in years, months, days, or hours, as appropriate, except for individuals who are aged 90 years or more. For persons aged 90 years and above, the age in a de-identified data set can only be stated as being within the category of age 90 or above. [L] [SEP]

Limited Data

The following information comes directly from <https://hipaa.wisc.edu/hipaa-faqs.htm#WhatIsLDS>:

In contrast to a de-identified data set, a limited data set can contain dates related to the individual (birth date, death date, etc.) and dates of services as well as geographic information at the level of town or city, State, and zip code. A limited data set is protected health information (PHI) that excludes the following direct identifiers of the individual or of relatives, employers, or household members of the individual:

- Names;

- Postal address information, other than town or city, State, and zip code;
- Telephone numbers;
- Fax numbers;
- Electronic mail addresses;
- Social security numbers;
- Medical record numbers;
- Health plan beneficiary numbers;
- Account numbers;
- Certificate/license numbers;
- Vehicle identifiers and serial numbers, including license plate numbers;
- Device identifiers and serial numbers;
- Web Universal Resource Locators (URLs);
- Internet Protocol (IP) address numbers;
- Biometric identifiers, including finger and voice prints; and
- Full face photographic images and any comparable images.

DRAFT

How to Intake New Data (For Sheps Programmers Only)

Overview

North Carolina Medicaid Claims Data

The following instructions explain how Medicaid data is processed at Sheps. CCQI loads NC Medicaid data from the NC Division of Medical Assistance (DMA) monthly to the folder "Cecil:\DMA_Transfer\". The Sheps intake process is quarterly.

Final data sets should be compressed. Other datasets that were created during the data intake process should be cleaned in order to save space on the server.

Member files are accumulated by CCQI. New member files should replace previous files once they become available

Claim Data Process

STEP ONE: Transfer and Verify Data Load

- 1) Create a folder named "RAW_YYYYMM" under "/nearline/data/projects/NCDMA_Medicaid/Data/", where "YYYYMM" is replaced with the year and month corresponding to the data. For example, we would create a folder named "RAW_201801" for NC Medicaid data from January 2018.
- 2) Move all data files from "Cecil:\DMA_Transfer\" to "/nearline/data/projects/NCDMA_Medicaid/Data/RAW_YYYYMM/"
- 3) Verify that the number of rows in each data file matches its corresponding audit file.

Example SAS code for STEP ONE is saved under "/nearline/data/projects/NCDMA_Medicaid/SAS/SAS_201801/DMA_data_load_vrf.sas".

STEP TWO: Merge Data by Data Type and Name

- 1) Create a temporary folder to merge all files by data type and name. Name the folder "DMA_BZYR_YYYYMM" and save it under "/nearline/data/projects/NCDMA_Medicaid/Data/". Note that "YYYYMM" should be replaced with the year and month corresponding to the given data.
- 2) Merge all files by data type and save the result in the folder "DMA_BZYR_YYYYMM".

- 3) In the merged file, create a new variable named "last_xtrct_dt" and set it equal to "YYYYMM."

Example SAS code for STEP TWO is saved under
"/nearline/data/projects/NCDMA_Medicaid/SAS/SAS_201801/DMA_clm_svc_201801.sas".

STEP THREE: Extract Claims by Service Start Date

- 1) Create a folder named "DMA_SVC_YYYYMM", where "YYYYMM" is replaced by the year and month corresponding to the given data. Save the folder under "/nearline/data/projects/NCDMA_Medicaid/Data/".
- 2) Extract claims by the variable "HDR_SVC_BGN_DT" and save the output in "DMA_SVC_YYYYMM".

Example SAS code for STEP THREE is saved under
"/nearline/data/projects/NCDMA_Medicaid/SAS/SAS_201801/DMA_clm_svc_201801.sas".

STEP FOUR: Concatenate New Data to Existing Datasets

- 1) Concatenate the claims data extracted in STEP THREE to the existing datasets saved under "/nearline/data/master/NCDMA_SVC_Master/". Save the output in a new folder called "DMA_SVC_temp" and place the new folder under "/nearline/data/projects/NCDMA_Medicaid/Data/".
- 2) Sheps IT will move the existing data (created during the previous data intake process) to the directory archives under "/nearline/data/master/archives/NCDMA_SVC_Master_yyyymm/".
- 3) Sheps IT will move all files from "/nearline/data/projects/NCDMA_Medicaid/Data/DMA_SVC_temp" to "/nearline/data/projects/NCDMA_SVC_Master/".
- 4) Verify the log to check if the intake data has been correctly processed.
- 5) Send email to the group stating that, "The new Medicaid data load for the time period up until [YYYY MM] is ready."

Member Data Process

Member files occasionally contain observations with a null MID. Use the following example code to delete records with null MID then save the updated files to "/nearline/data/projects/NCDMA_Medicaid/Data/DMA_SVC_temp".

SAS Code:

"/nearline/data/projects/NCDMA_Medicaid/SAS/SAS_201801/DMA_clm_svc_201801.sas".

Sheps IT will move all files from the ".../DMA_SVC_temp" folder to "/nearline/data/projects/NCDMA_SVC_Master/".

Provider and mem_xref Data Process

Provider and mem_xref files need to be copied to "/nearline/data/projects/NCDMA_Medicaid/Data/DMA_SVC_temp".

Sheps IT will move all files from the ".../DMA_SVC_temp" folder to "/nearline/data/projects/NCDMA_SVC_Master/".

How to Extract Member Records and Claims Records

The NC Medicaid data from the NC DMA has several fields in different tables that are treated as a member ID.

Examples in "dataset->field" format:

Member -> MID (labeled "ALT_MBR_XREF_ID_ENCRYPT")

Mem_XREF -> REF_MBR_ID

Mem_XREF -> MID

Claims Tables -> ALT_MBR_ID_ENCRYPT

It is possible that an "individual" has more than one Member ID (see page ...). The purpose of the Mem_XREF dataset is to provide the chain of Member IDs for individuals. The key field in the Mem_XREF dataset is REF_MBR_ID. If an individual has multiple Member IDs, there will be one record per REF_MBR_ID / MID combination for that individual in the Mem_XREF dataset.

File "mem_xref" is the middleman between claims records and member records. An individual has only one "REF_MBR_ID" in the mem_xref file but can have more than one MID. The MIDs are used to find corresponding claims records in the claims files.

Extracting claims when you have MID from the Member file

- 1) Follow Member -> MID to Mem_XREF -> REF_MBR_ID
- 2) Gather all Mem_XREF -> MIDs associated with that REF_MBR_ID
- 3) Gather all claims by matching all Mem_XREF -> MIDs to [Claims] -> ALT_MBR_ID_ENCRYPT

Extracting member info when you have ALT_MBR_ID_ENCRYPT from the claims file(s)

- 1) Do the reverse. Follow [Claims] -> ALT_MBR_ID_ENCRYPT to Mem_XREF record
- 2) That can result in multiple ALT_MBR_ID_ENCRYPTs pointing to one REF_MBR_ID

3) Follow REF_MBR_ID to Member -> MID for corresponding Member record(s)

**Blue Cross Blue Shield of North Carolina (BCBSNC) Claims Data
Extracting Data by Processing/Warehouse Date Vs. Service Date**

From Roger Akers, May 6th, 2017, 10:13 AM:

"Gene,

I like the idea of pulling by processing/warehouse date vs service date to get better data. When a claim gets processed and submitted to warehouse, is it considered "final"? In this scenario, what is the likelihood of a record from a future extraction being a replacement for a record in our existing data?

Example: we have claims through Sept 2015 by processing/warehouse date. When we get the next data feed in July, what is the likelihood of some of those claims records replacing records in Sept or earlier because they were processed again? (is that even possible at that point to be processed again?)"

Reply from Gene Williams of BCBSNC, May 6th, 2017, 3:17 PM:

"A claim is only loaded to our warehouse when it is "finalized," meaning that adjudication has completed and we have made an official entry to our accounting records for the claim.

Subsequent events may require that we either back out the claim, or that we replace it with a revised version. That happens to about 7% of claims overall. For both a backout and a revision, we will first load a negative claim into our warehouse, which is a mirror image of the first claim in that all of the monetary and utilization amounts have been multiplied by negative 1. If the original claim had \$70 in payment and 2 service units, the reversal claim will have -\$70 in payment and -2 service units. The idea is to show a net amount of 0 for everything.

If there is a revised version of the claim, we load it into our warehouse as a 3rd claim (1st = original, 2nd = reversal).

Across all types of claims for all types of business, we find that about 7% of claims will be backed out. There is no time limit on how long after original adjudication that a claim may be reversed and/or revised. However, about 70% of those 7% get reprocessed within 180 days of the original claim's arrival in the warehouse. And, only about 20% of those 7% ever get a revised version."

Appendix 1: Summary of Workflow for Project Approval, Processing, and Governance

Figure A1 describes the current workflow for the approval, processing, and governance of projects involving CCQI data. As a reminder, a project must receive IRB approval before an associated data request is submitted to the Sheps Center.

Figure A1. Workflow Diagram for Project Approval, Processing, and Governance

As described in Figure A1, the Sheps Center must give explicit approval for any project involving CCQI data to begin. The only exceptions are

- a) specific *ad hoc* requests by the Sheps Center and
- b) “case-finding” sample size counts (e.g., “How many members have condition X”) in order to ascertain whether a project is feasible.

Appendix 2: SAS Library for NC Medicaid Data

Appendix 3: SAS Library for BCBSNC Data