The Tuva Project

2024-12-22

Table of contents

Welcome

Welcome to the Tuva Project Knowledge Base! This is an open-source book on healthcare data and analytics.

Code Sets

Introduction

Healthcare data is made up of many different types of code sets. Each code set is a classification system comprised of anywhere from dozens to hundreds of thousands of distinct codes. Different code sets identify and classify different types of things. For example, ICD-10 is a code set used to classify diseases and procedures while RxNorm is used to identify and classify drugs.

Clinicians do not typically interact with code sets. Rather, they use the EHR to order diagnostics and therapies and enter notes. These orders and notes are then transformed into code sets behind the scenes, often for billing purposes.

Below we describe the various code sets and nuances that are important to understand when using them in analytics.

NDC

The most important piece of information included on pharmacy claims is the information about the actual medication being prescribed. The National Drug Code (NDC) on a pharmacy claim describes the actual drug being prescribed. NDC is a complex data element so we will spend some time describing it here.

The NDC code set was first introduced in 1972 by the U.S. Food and Drug Administration (FDA). The original NDC consisted of 10 digits broken up into 3 segments:

- 1st Segment: Labeler
- 2nd Segment: Product
- 3rd Segment: Package

The Labeler segment is the only segment assigned by the FDA and it identifies the drug manufacturer i.e. the organization that produced the drug. The product segment identifies specific information about the drug. And the package segment identifies specific information about the package e.g. number of pills.

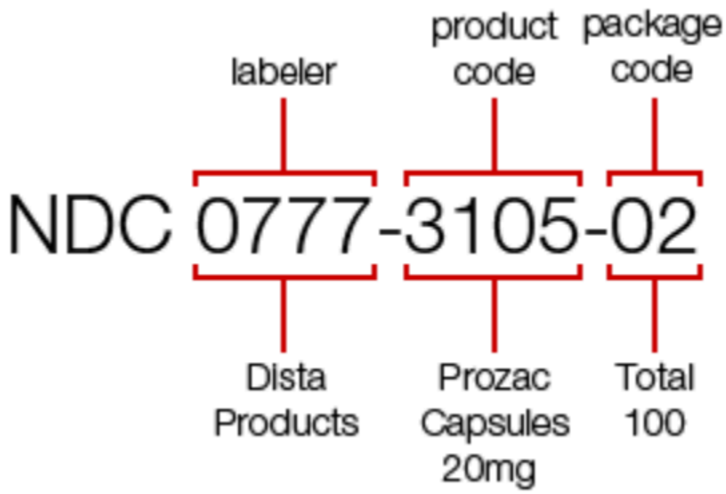


Figure 1: NDC

Today, NDC codes are written as a 10-digit number on drug packaging. You typically find this number near the bar code on the packaging. An additional digit is added, bringing the total to 11 digits, when billing an NDC on a healthcare claim. The 11-digit number follows a 5-4-2 format i.e. 5 digits in the first segment, 4 digits in the second segment, and 2 digits in the third segment. The rules for which segment the additional digit is added to are as follows:

- 4-4-2 becomes 5-4-2
- 5-3-2 becomes 5-4-2
- 5-4-1 becomes 5-4-2

Essentially you add a leading zero to whichever segment needs it.

Not Just NDCs - Multiple Code Sets

The NDC is a complicated data element to work with, in part because the field contains entries from other code sets. For example, the NDC field in a claims dataset will also contain:

- NDCs that have not been fully approved by the FDA
- Drug Supply or Medical Device Codes e.g. UPC (Universal Product Code) or HRI (National Health Related Item Codes)

The latter is included because retail pharmacies often sell and bill health insurers for drug supplies and other medical equipment e.g. syringes for insulin.

One Drug, Many NDCs

Another thing that makes NDC a complex data element to work with is that there are often many NDCs for the same drug or active ingredient. So answering a question like "which patients have received Drug X?" often requires looking up dozens of NDCs.

Part I

Claims Data

1 Intro to Claims

Healthcare claims data is the oldest and most widely analyzed type of healthcare data. In this section we provide an overview of the healthcare entities and overall process involved in the creation of claims data.

1.1 Healthcare Entities

Healthcare claims are created by healthcare providers for the purpose of billing health insurance companies for the services and supplies they have rendered to patients. The following types of entities play an important role in the claims data generation process:

- **Providers:** Includes organizations and people that render healthcare services or supplies to patients, including:
  - Individual Providers: Physicians, physician assistants, licensed therapists and social workers, etc.
  - Healthcare Organizations: Health systems, hospitals, skilled nursing facilities, home health organizations, hospice organizations, ambulatory surgery centers, etc.
  - Pharmacies: Retail pharmacies e.g. Walgreens, CVS, etc.
  - Lab Testing Companies: e.g. Labcorp, Quest Diagnostics, etc.
  - Durable Medical Equipment Companies: Companies that sell durable medical equipment to provider organizations or directly to patients.
- **Clearing Houses:** These organizations sit between providers and health insurers. They collect claims from providers in a standard format, perform basic checks and verifications of each claim, and then route the claims that pass these basic checks to the appropriate health insurer for further processing and adjudication.
- **Health Insurers:** These are the organizations patients have their medical and or pharmacy coverage through. These organziations have contracts with providers to pay them for services and supplies at specific prices.
- **Revenue Cycle Management (RCM) Companies:** Help providers manage their entire billing process, including the coding of claims and the collection of payments from providers and patients.

1.2 The Claims Creation Process

1. **Healthcare Service:** A provider renders healthcare service or supply to a patient. For example, a patient has a visit with their primary care physician.
2. **Claim Created:** The provider (or revenue cycle company working on their behalf) creates a claim and submits it to the appropriate clearing house. This is almost always done electronically using either an 837I or 837P EDI transaction.
3. **Clearing House:** The clearing house performs basic checks of each claims. For example, they make sure all the required fields are populated. The clearing house then transmits the claim to the appropriate health insurer.
4. **Health Insurer:** The health insurer receives the claim from the clearing house and adjudicates the claim. Claims adjudication is a process where the insurer determines whether or not to pay the claim and the amount to be paid, if warranted. For example, the insurer will check:

    1. Whether the patient had insurance coverage during the date of service on the claim
    2. Whether the patient's insurance covers the particular services or supplies they received
    3. Whether the patient meets certain prior authorization requirements (e.g. this is common for surgical procedures)

5. **Payment Decision:** Once the claim is fully adjudicated (i.e. a decision about whether to pay the claim has been reached), the health insurer will issue a remittance to the provider. This remittance is another electronic transaction called an 835 or Electronic Remittance Advice (ERA) transaction. This transaction includes information about whether the claim was paid or denied and is sent back to the provider.
6. **Corrections:** If the claim was denied the provider will work to correct the error on the claim which led to the denial, if possible, and re-submit the claim to the health insurer.
7. **Adjustments and Reverals:** Occassionally a claim is submitted and paid in error - these claims are eventually adjusted and reversed. See the section on adjustments and reversals for more information about this process.
8. **Data Warehousing:** Ultimately the health insurer will aggregate all claims (final claims, adjustments, denials, and reversals) in a database. This claims dataset typically includes eligibility information, medical claims, and pharmacy claims. When we discuss analyzing claims, this is the dataset we are referring to. Health insurers often make this claims data available to provider partners (e.g. accountable care organizations) and pharmaceutical companies frequently purchase de-identified copies of this data for drug safety, efficacy, and commercialization research.

1.3 Claims Forms

A healthcare claim is created when a healthcare provider populates a claim form for services or supplies they've rendered to a patient. These days almost all claims are created and submitted electronically. Often this is automated or partially automated by the electronic medical record (EMR) system.

There are 3 main types of claim forms. The type of form used depends on the type of healthcare entity submitting the claim.

CMS-1500 - Also Known As: Professional claim - Electronic Version: 837P - Maintained By: National Uniform Claim Committee (NUCC) - Example Form: https://nucc.org/images/stories/PDF/1500_claim_form_2012_02.pdf - Used By: Physicians, lab test companies, durable medical equipment companies, etc.

CMS-1450 - Also Known As: UB-04, institutional, or facility claim - Electronic Version: 837I - Maintained By: National Uniform Billing Committee (NUBC) - Example Form: https://www.amerihealth.com/pdfs/providers/npi/ub04_form.pdf - Used By: Facilities (e.g. hospitals, SNFs, ambulatory surgery centers), home health agencies, hospice organizations, etc.

NCPDP Universal Claim Form - Also Known As: N/A - Electronic Version: - Maintained By: the National Council for Prescription Drug Programs - Example Form: - Used By: Retail pharmacies (e.g. Walgreens, CVS, Wal-Mart) to bill health insurers. In this section we'll cover the key data elements present in claims data.

Each claim form has two sections: a header section and a line section. Each data element is either entered on the header section or line section. Every data element in the header section may only be entered a specific number of times (typically one time, but not always). On the other hand, data elements in the line section may be entered an unlimited number of times.

See Knuth (1984) for additional discussion of literate programming.

2 Claims Data Elements

2.1 Administrative Codes

One of the great things about claims data (compared to clinical data) is that they contain a number of fields with standard terminology that are (usually) well-populated. We refer to these fields as Administrative Codes because they are used for billing (i.e. administrative) purposes as opposed for clinical purposes. These fields contain valuable information about the services performed, where they were performed, and where the patient came from and went to before/after the service.

In this section we review these codes, their analytic use cases, and how you can identify common data quality problems in them.

2.1.1 Admit Source

Admit source code is used in institutional claims to indicate where the patient was located prior to admission. The field does not exist in professional claims. The field exists at the header-level, meaning there should be only 1 distinct value for this field per claim.

Admit source, along with admit type, is the least reliable among the administrative codes because the accuracy of the code is not verified during the claims adjudication process (other than verifying that the code is in fact a valid code).

Despite this, it's possible to use admit source to help identify things like: - transfers from another hospital - inpatient stays that came through the emergency department

Admit source codes are maintained by the National Uniform Billing Committee (NUBC).

You can find a complete listing of admit source codes and their descriptions here.

2.1.2 Admit Type

Admit type code is used in institutional claims to indicate the priority of admission, e.g., urgent, emergent, elective, etc. The field does not exist in professional claims. The field exists at the header-level, meaning there should be only 1 distinct value for this field per claim.

Admit type along with admit source, is the least reliable among the administrative codes because the accuracy of the code is not verified during the claims adjudication process (other than verifying that the code is in fact a valid code).

Despite this, admit type is commonly used to identify things like elective procedures.

Admit type codes are maintained by the National Uniform Billing Committee (NUBC).

You can find a complete listing of admit type codes and their descriptions here.

2.1.3 Bill Type

Bill type code is by far the most complex of the administrative codes. Each digit in the bill type code has a distinct purpose and meaning:

- 1st digit: This is always "0" and often omitted.
- 2nd digit: Indicates the type of facility, e.g., skilled nursing facility
- 3rd digit: Indicates the type of care, e.g., inpatient part A
- 4th digit: Indicates the sequence of the bill (also referred to as the frequency code)

The thing that makes this code complex is that the possible values of the 3rd and 4th digits depend on the value of the 2nd digit. As a result, some claims datasets will separate out the digits of bill type code into distinct fields. However, we find it preferable to work with bill type code as a single field and the dictionary below lists all bill type codes this way.

Despite the complexity of this field, it's extremely useful. Bill type code is used extensively in the creation of service categories, including the identification of acute inpatient, outpatient, skilled nursing, and emergency department services, among many others. The field is generally considered reliable because the accuracy and suitability of the code is verified during the claims adjudication process, i.e., a claim may be denied if the code doesn't make sense.

Bill type codes are maintained by the National Uniform Billing Committee (NUBC).

You can find a complete listing of bill type codes and their descriptions here.

2.1.4 Discharge Disposition

Discharge disposition code indicates where the patient was discharged following a stay at a facility. The field only exists on institutional claims. The field is sometimes called discharge status or patient status. The field exists at the header-level, meaning there should be only 1 distinct value for this field per claim.

The code is commonly used to identify things like: - Patients that died during an institutional stay - Patients who were transferred - Patients who were discharged to home or home w/ home health services - Patients who left against medical advice (LAMA)

Discharge disposition codes are maintained by the National Uniform Billing Committee (NUBC).

You can find a complete listing of discharge disposition codes and their descriptions here.

2.1.5 HCPCS

HCPCS codes indicate the services and supplies rendered by providers to patients. These codes are used in both institutional and professional claims forms. These codes exist at the line-level, meaning there can be many HCPCS codes on a single claim. There are codes for many different types of supplies and services including: - physician visits - lab tests - imaging reads - durable medical equipment - remote patient monitoring devices

And many many other types of things. There are thousands of HCPCS codes spread across two levels. Level 1 codes, also called CPT codes, are maintained by the American Medical Association (AMA). Level 2 codes are maintained by CMS.

Professional contracted rates between payers and providers are established using HCPCS codes. These rates are referred to as a fee schedule. Conversely, institutional rates are often paid on a per encounter (e.g. DRG) or per diem basis.

You can find a complete listing of all level 2 HCPCS codes and their descriptions here.

2.1.6 Place of Service

Place of service codes indicate the type of care setting professional claim services were delivered in. This field only exists on professional claims. Place of service is coded at the line-level to reflect the fact that services during a particular encounter can occur in different locations. Because of this, a single professional claim can have multiple place of service codes.

Place of service codes are used to assign claims to services categories. For example, place of service code 11 indicates an office visit.

CMS maintains place of service codes.

You can find a complete listing of all place of service codes and their descriptions here.

2.1.7 Revenue Center Codes

Revenue center codes are used to account for the services and supplies rendered to patients in institutional care settings. These codes are only used in institutional claims. Typically these codes will correspond to a facility's chargemaster, which is a listing of all charges used by the institution in billing. Although a hospital will use these codes to "charge" the health insurer, they have no bearing on the contracted payment amount, i.e., the amount paid to the provider by the payer. The payment amount is entirely determined by MS-DRG for inpatient claims and often a per diem rate for skilled nursing.

Many different categories of revenue center codes exist including for example: - Room and Board - Emergency - IV Therapy

For a given institutional claim there may be dozens of revenue center codes used. These codes are submitted at the line-level of the claim, so there is no limit to the number of revenue center codes that may be used on a given claim.

Revenue center codes play an important role in identifying different types of insitutional claims, including acute inpatient, emergency department, and others.

Revenue center codes are maintained by the National Uniform Billing Committee (NUBC).

You can find a complete listing of revenue center codes and their descriptions here.

2.2 Dates

Claims data is longitudinal in nature i.e. it captures conditions, services and other healthcare events that occur over time to patients. This makes claims data extremely useful for analyzing sequences of events e.g. did patients who received drug X have better or worse outcomes? However the ability to reliably use claims data in this matter is predicated by the completeness and accuracy of a variety of key date fields found in claims data.

The date fields listed below are the names we give to these fields in the Tuva Project, but there can be all sorts of different names for these fields in different claims datasets. For example, in Medicare LDS the claim_end_date field is called clm_thru_dt.

2.2.1 Medical Claims

To understand the key date fields in medical claims, it's useful to consider an example of a patient who's been receiving care in a long-term care (i.e. skilled nursing) facility for 1 year, from January 1st to December 31st, and suppose the facility bills the insurer every month on the beginning of the month.

- **claim_start_date:** The start date of the billable period for the claim. In the example above this date would always be the first date of the month.
- **claim_end_date:** The end date of the billable period for the claim. In the example above this date would always be the last date of the month.
- **admission_date:** The date the patient was first admitted to the facility. In the example above this date would be January 1st. This field only exists on institutional claims, not professional.
- **discharge_date:** The date the patient was discharged from the facility. In the example above this date would be December 31st. This field only exists on institutional claims, not professional.
- **paid_date:** The date the claim was paid by the insurance company. This date could be any date after the claim_end_date. Often this date is within a couple weeks of claim_end_date.

There are 2 other date fields in medical claims. They are claim_line_start_date and claim_line_end_date. These date fields are less important - in fact we don't currently use them in any analytics in the Tuva Project.

2.2.2 Pharmacy Claims

- **dispensing_date:** The date when the prescription was filled by the pharmacy and given to the patient.
- **paid_date:** The date the claim was paid. Often this date lags the dispensing_date by days or weeks.

2.2.3 Eligibility

- **enrollment_start_date:** The date when a patient became enrolled in a health plan (i.e. insurance). Patients can gain and lose enrollment over time, so a given patient may have more than one enrollment_start_date.
- **enrollment_end_date:** The date when a patient loses enrollment in a health plan (i.e. insurance). Patients can gain and lose enrollment over time, so a given patient may have more than one enrollment_end_date. Patients who are currently enrolled will not have an enrollment_end_date, or they may have a long-dated enrollment_end_date e.g. 12/31/9999, which is meant to indicate they are still enrolled.
- **birth_date:** The date the patient was born.

- **death_date:** The date the patient died (if applicable). Just because a patient does not have a death date does not mean they aren't deceased! Many deaths do not occur in a healthcare facility and therefore are not captured in claims. Sometimes the death date is captured in eligibility data, but often it is inferred by discharge_disposition_code = 20 (this field is found in institutional claims).

2.2.4 Data Quality Issues

As you might expect, the date fields in claims often suffer from data quality issues. For example, date fields can be missing, or the dates can exist unnaturally far into the past or into the future.

Identifying these sorts of problems across all the key date fields can be challenging and require a lot of ad hoc querying. We've figured out a good way to look for these sorts of data quality problems and built it into the Tuva Project. Check out the code and video below for more info.

```sql
select *
from insights.count_claim_by_date_column
order by 1
```

3 Adjustments, Denials, and Reversals

3.1 Overview

One of the trickiest issues to deal with in claims data is adjustments, denials, and reversals (what we often refer to as "ADR"). There are three types of claim records (original, adjustment, and reversal) and three types of claim payment statuses (paid, denied, reversed). How you model claim record types and payment statuses will impact the analytics you're able to perform on your claims data.

Let's take a step back and think about the types of analytics we perform on claims data. At the highest level, we think there are two categories of claims analytics:

1. Cashflow analytics
2. Population health analytics

Cashflow analytics includes analyses like calculating Incurred But Not Reported (IBNR) claims. This sort of calculation is done by an actuary to measure and manage cash reserves for an insurance company or health plan. To perform this type of analysis you need to see every iteration of a claim that occurred. For example, if a claim was originally paid, then reversed and adjusted weeks later, you need to have full visibility into these payments and the dates when the payments occurred. Leveraging multiple iterations of a claim is also useful for the analytics team when it's necessary to align analytics with financial reporting, or when it's necessary to understand when a claim was first received or processed (the original claim received or paid date).

Population health analytics includes analyses around the cost of care, diagnosis and treatment, utilization, and risk of a patient population. This type of analysis is more concerned with the final amounts paid (rather than intermediate adjustments and reversals) and the dates when services were delivered (as opposed to paid dates).

The trick is to model your claims data in such a way that supports both types of analyses, satisfying your cashflow analysis folks (e.g. actuaries and financial analysts) and your population health analysis folks (e.g. quality measures folks, data scientists, epidemiologists, also actuaries here too, etc.). This involves keeping the multiple iterations of the claim available for cashflow analytics while allowing population health and other analytics use cases to only worry about the final disposition, which we'll detail below.

Let's quickly define the three different types of claim records:

- **Original:** This claim record type is the first (and sometimes only) claim submitted.
- **Adjustment:** This claim record type is submitted if the original claim was denied or if the provider found an issue with the original claim they needed to correct.
- **Reversals:** This claim record type is submitted if the original claim was paid, but then an adjustment was needed and the original claim needed to be backed out.

The three types of payment statuses are straightforward:

- **Paid:** Indicates the claim was paid (positive paid amount)
- **Denied:** Indicates the claim was denied (zero paid amount)
- **Reversed:** Indicates the claim is reversed (negative paid amount)

3.2 Modeling ADR

It's easiest to illustrate how adjustments, denials, and reversals manifest by looking at example scenarios. By looking at examples we can also see how to model ADR claims to support both cash flow and population health analytics.

Original Only

original_claim_number	claim_number	claim_sequence_number	claim_record_type	final_claim_flag	claim_status	claim_line_number	billed_amount	cumulative_billed_amount	allowed_amount	cumulative_allowed_amount	paid_amount	cumulative_paid_amount
C123001	C123001	1	Original	Y	Paid	1	100	100	50	50	25	25

Denied / Resubmitted

original_claim_number	claim_number	claim_sequence_number	claim_record_type	final_claim_flag	claim_status	claim_line_number	billed_amount	cumulative_billed_amount	allowed_amount	cumulative_allowed_amount	paid_amount	cumulative_paid_amount
C123001	C123001	1	Original	N	Denied	1	100	100	0	0	0	0
C123001	C123002	2	Adjustment	Y	Paid	1	75	75	25	25	10	10

Full Reversal / Adjustment

original_claim_number	claim_number	claim_sequence_number	claim_record_type	final_claim_flag	claim_status	claim_line_number	billed_amount	cumulative_billed_amount	allowed_amount	cumulative_allowed_amount	paid_amount	cumulative_paid_amount
C123001	C123001	1	Original	N	Paid	1	100	100	50	50	25	25
C123001	C123002	2	Reversal	N	Reversed	1	-100	0	-50	0	-25	0
C123001	C123003	3	Adjustment	Y	Paid	1	75	75	25	25	10	10

Full Reversal / Adjustment

original_claim_number	claim_number	claim_sequence_number	claim_record_type	final_claim_flag	claim_status	claim_line_number	billed_amount	cumulative_billed_amount	allowed_amount	cumulative_allowed_amount	paid_amount	cumulative_paid_amount
C123001	C123001	1	Original	N	Paid	1	100	100	50	50	25	25
C123001	C123001	1	Original	N	Paid	2	200	200	100	100	50	50
C123001	C123002	2	Reversal	N	Reversed	1	-100	0	-50	0	-25	0
C123001	C123002	2	Reversal	N	Reversed	2	-200	0	-100	0	-50	0
C123001	C123003	3	Adjustment	Y	Paid	1	75	75	25	25	10	10
C123001	C123003	3	Adjustment	Y	Paid	2	150	150	75	75	25	25

Partial Adjustment

original_claim_number	claim_number	claim_sequence_number	claim_record_type	final_claim_flag	claim_status	claim_line_number	billed_amount	cumulative_billed_amount	allowed_amount	cumulative_allowed_amount	paid_amount	cumulative_paid_amount
C123001	C123001	1	Original	N	Paid	1	100	100	50	50	25	25
C123001	C123002	2	Adjustment	Y	Paid	1	50	150	25	75	75	100

Figure 3.1: modeling_adr

Let's walk through each scenario in the image above. As we do, pay careful attention to how the cumulative amounts change from the original claim, to the reversal, to the adjustment claim. Modeling the amounts this way is what enables both cashflow and population health analytics.

3.2.1 Scenario 1: Original Only

This first scenario is the simplest. There's only a single claim, with a single claim line, and it's the original claim (see claim_record_type) as you would expect. This claim has been paid (see claim_status). The billed, allowed, and paid amounts for this claim are equal to the cumulative amounts. You'll see how these individual and cumulative claims differ in subsequent scenarios.

3.2.2 Scenario 2: Denied / Re-submitted

In the second scenario, the original claim was denied (see claim_status), so an adjustment claim was submitted and adjudicated. The original claim could have been denied for any number of reasons. The billed amount on the original claim was $100 and the allowed and paid amounts were $0 (no payment was allowed since the claim was denied).

The billed, allowed, and paid amounts on the second claim (the adjustment claim) were $75, $25, and $10, respectively. Now pay attention to the cumulative billed, allowed, and paid amount columns. These columns sum up the current and previous claim billed, allowed, and paid amounts.

There's one unusual thing about cumulative_billed_amount in this scenario. For denied claims, you don't add the billed_amount to cumulative_billed_amount total. Doing so would screw up the billed-to-allowed ratio of the cumulative_billed_amount and cumulative_allowed_amount fields.

Note that the adjustment claim gets a new claim_number and we associate this claim ID with the original claim via original_claim_number.

3.2.3 Scenario 3: Full Reversal / Adjustment

In this scenario, the original claim was completely reversed and then an adjustment claim was subsequently submitted. The cumulative billed, allowed, and paid amounts match the billed, allowed, and paid amounts from the final adjustment claim because the reversal claim negated all the amounts from the original claim.

3.2.4 Scenario 4: Full Reversal / Adjustment (w/ multiple lines)

This scenario is similar to the third scenario but provides an example of how a full reversal and adjustment will look with multiple claim lines. Note that cumulative amounts are tracked at the line level.

3.2.5 Scenario 5: Partial Adjustment

This final example shows how a claim may be partially adjusted, with an additional positive or sometimes negative payment amount.

3.3 How to Identify ADRs

The number one thing you can do is ask the insurance company or health plan how to identify ADRs in the claims dataset. If they can't tell you, that's a bad sign (but unfortunately this happens a lot). In that case see below.

The exact manner in which claims adjustments and reversals manifest in your claims dataset may not be obvious or easy to identify. However, they typically show up in 1 of 3 ways, depending on how the health insurer adjudicates their claims.

- **Scenario 1:** Health insurer creates a new claim ID for each additional reversal and/or adjustment claim record
- **Scenario 2:** Health insurer uses the original claim ID for each reversal and/or adjustment claim, but includes an adjustment/reversal code on each claim to indicate whether each new record was a reversal or adjustment
- **Scenario 3:** Health insurer uses a combination of old and new claim IDs for reversals and/or adjustments (this is essentially a combination of scenario 1 and 2)

In our experience working across dozens of healthcare claims datasets, each scenario is equally common.

3.3.1 Scenario 1: New Claim IDs for Each Adjustment/Reversal

Dealing with new claim IDs is the most difficult scenario because you are in effect trying to determine a linkage across claim IDs when there isn't a piece of data that tells you this. That is, you need to identify new Claim IDs that are related to the original Claim ID. This can be tricky, but here are some steps you can follow to do this:

1. Start by looking for patients with multiple claims (i.e. multiple claim IDs).

2. To identify reversals, look for claims where all the data elements match, except the paid amount and the paid date.

3. To identify adjustments, look for claims that have a reversal, then look for subsequent claims where most of the information is the same but there may be minor differences (e.g. place of service code changed from 11 to 20).

3.3.2 Scenario 2: Original Claim IDs for Each Adjustment/Reversal

This is the simplest scenario. The payer / health plan has already done the bulk of the work for you - explicitly telling you which claims were adjusted / reversed and whether the claim was paid or denied. However in practice this information is usually not completely available, so like most things in healthcare data there tends to be a bit of detective work to fill in the gaps.

3.3.3 Scenario 3: Combination New and Original Claim IDs for Each Adjustment/Reversal

Sometimes adjustments and reversals will appear as new claim IDs, but these claim IDs will include the original claim ID plus some additional characters. For example:

- original claim ID: A1234
- adjustment claims ID: A12341
- reversal claim ID: A12342

In this example you can see the subsequent claims have an integer appended to them.

3.4 Real-world Impacts of ADRs

In an ideal world, all population health analytics should be based on a claims dataset that represents the true set of services rendered by the provider to the patient and payments rendered by the health insurer to the provider. However, unless you account and correct for claims adjustments and reversals in raw claims data, you are including records in your analysis for services that may never have been delivered. The problem with downstream analytics is much more on the utilization side than on the payments side.

Payments

Let's start by exploring the impact on payments first, since it is minimal and more straightforward. At the end of the day, claims adjustments and reversals will directly flow through to aggregate payment amounts (e.g. payments aggregated to PMPM level), generating the true/correct payment statistics. For example, if a claim was submitted in error and then later a reversal was submitted, the sum of the paid amount for these two claims will be zero, which is the true paid amount we would expect. As this example demonstrates, the aggregate payment amounts in any analysis (e.g. trending total medical PMPM by month) will be correct by default, without any changes made to the raw claims data. Therefore, it's not usually necessary to identify and correct adjustments, reversals, or denials before calculating payment statistics.

Utilization

However, the impact on utilization analytics is not as straightforward. In the simple example above we noticed that the place of service code changed from 11 (office visit) to 20 (urgent care). Place of service code is an important piece of information used to group claims into encounters. Without taking into account that a reversal and adjustment was made for this claim, it looks like two place of service codes exist for the same visit, and we are unsure which code to use to assign an encounter type to the claim (i.e. should the claim be labeled an office visit or urgent care visit).

This has serious consequences for utilization analytics and also impacts payment analytics if we want to analyze payments by care setting. For example, suppose we are interested in looking at spend across different care settings, e.g., acute inpatient, inpatient rehab, emergency department, urgent care, and office visit. Without properly identifying and correcting claims adjustments and reversals we won't be able to bucket spend or number of visits in the appropriate category.

4 Providers

This section describes provider information included in claims data - namely the National Provider Identity (NPI).

4.1 What provider data is included in claims?

Medical claims includes several fields containing information on providers. The fields vary based on the type of claim.

Facility Claims CMS-1450 or UB-04: Provider information in the header of facility claims. In addition to the facility billing the service, these claims contain several fields for NPIs from up to four individual providers involved in the care (e.g., Attending Physician). - Box 1 Billing Provider Name and Address - 2 Pay-to Proivder Name and Address - 5 Federal Tax ID - 76 Attending Physician - 56 Billing Provider NPI - 57 Other Provider ID - 77 Operating Physician - 78 Other Physician - 79 Other Physician

Professional Claims CMS-1500: Professional claims track the NPI of the provider who rendered each individual line item (i.e., CPT/HCPSCS code) in the claim. In addition, the claim header contains information on the organization submitting the claim. - Box 17 Referring Provider - 24J Rendering Provider - 25 Federal Tax ID - 32 Service Facility Location Information - 33 Billing Provider

4.2 What is an NPI?

Individual provider and facility information is encoded in claims data via National Provider Identity (NPI) codes. However, one needs to enhance individual provider NPI codes with specialty information and group facility provider NPI codes into distinct locations before this information is useful for analytics.

- An NPI is a unique 10-digit numeric identifier for covered healthcare providers and organizations.
  - It is a HIPAA standard created to help send health information electronically.

- An NPI won't change even if a provider's name, address, taxonomy (specialty), or other information changes. However, in some situations, an NPI may need to be deactivated or replaced, such as the retirement or death of an individual, disbandment of an organization, or fraudulent use of the NPI.

- Who must get an NPI?
  - All health care providers who are HIPAA-covered entities, individuals, or organizations.
  - It is required for enrollment in Medicare and submitting claims.
  - When a provider registers for an NPI, CMS attempts to verify only two things: (1) the provider's social security number and (2) that the provided business address is valid.
    * CMS does not verify whether the provider actually works at the submitted business address, and CMS does not attempt to verify the provider's self-reported specialty.
- NPI Entity Types
  - Entity Type 1: Individual
    * You may only get 1 NPI.
    * If you're an individual healthcare provider who's incorporated, you may need to get an NPI for yourself (Type 1) and an NPI for your corporation (Type 2).
  - Entity Type 2: Organization
    * The main difference with type 2 NPI numbers is that organizations can have several NPI numbers rather than just one.
    * Some organizations may have parts or locations that work independently from their parent organization, referred to as "subparts". Each subpart can get its own NPI.
  - Where are the NPIs found on claim forms?
    * On the Professional CMS-1500 form, insert the main or billing Entity Type 2 NPI in Box 33a (Billing Provider). Insert the service facility Entity Type 2 NPI (if different from main or billing NPI) in Box 32a (Service Facility). Insert Entity Type 1 NPIs for rendering providers in Box 24J (Rendering Provider).
    * On the Institutional UB-04 form (aka CMS-1450), insert the main Entity Type 2 NPI in Box 56 (Billing Provider); insert Entity Type 1 NPIs for rendering providers in boxes 78-79 (Other Provider).

4.3 What is NPPES?

- CMS developed the National Plan and Provider Enumeration System (NPPES) to assign NPIs.

- This information is publicly available and disclosed under FOIA.
- The data is distributed to the public via monthly data file downloads or via the API.
- The API has limitations. It's useful for single-provider lookups but not for getting batch information.
- Limitations with the NPPES data:
  - The monthly file is a very large full replacement file that must be unzipped.
  - Many fields are codes requiring a separate lookup file for human-readable descriptions. These code sets are not distributed via data files. They are instead in a PDF provided with the monthly download. The codes must be extracted and transformed before they are useful.
  - Many pieces of information are in several columns that require logic to get any meaningful value out of them.
  - Once a provider has an NPI, there are no scheduled requests for updated information; however, providers are instructed to update their information in NPPES within 30 days of a change of required data fields. The degree to which providers update their information is not fully known.
  - There is no explicit penalty for a provider having out-of-date information in NPPES.
  - This means the data in NPPES is not necessarily reliable but remains one of the few publicly available sources of provider data.
- Other sources of provider data:
  - PECOS (Provider Enrollment, Chain, and Ownership System)
    - This is a registry of providers eligible to bill Medicare.
    - PECOS gathers more detailed information than NPPES on the financial arrangement between an individual provider and a practice group or organization, as well as a number of other details about business ownership and history of adverse outcomes with malpractice claims.
    - Providers are required to update their information every five years or whenever changes occur.
    - Unfortunately, this is not publicly available for research.
  - AMA (American Medical Association) Masterfile
    - The Masterfile attempts to be a comprehensive registry of all physicians trained in the United States.
    - It captures information from institutions on individuals at the time that they enter medical school or a graduate medical program in the United States.
    - It relies primarily on physicians voluntarily completing questionnaires to update information about their practice and location.
    - This data is publicly available for research.

4.4 What are provider taxonomies?

- When providers register with NPPES they are required to provide one primary provider taxonomy code (and up to 14 additional taxonomies) which defines the health care service provider type, classification, and area of specialization.
- The Health Care Provider Taxonomy code set is a collection of unique alphanumeric codes (e.g. 207KA0200X), ten characters in length, maintained by the National Uniform Claim Committee (NUCC).
- Again, a separate terminology lookup data source is required to interpret this code which does not come with the NPPES data set.
- The taxonomy codes are updated twice a year (January and July).

4.5 Why are TAX IDs included in claims?

In addition to NPIs, federal tax IDs are required fields on both facility and professional claim forms. The tax ID can be either an employer identifier number (EIN), or an invidual's social security number.

Most provider analyses are conducted using the NPIs recorded in claims. Tax IDs are used for some financial use cases, since network contracts are written at the TIN level. For example, network discounts are often analyzed by tax ID, since network contracts are written at the tax ID level (Example).

5 Member Months

5.1 Overview

When trending population level statistics such as claims payments or utilization, it's often best to normalize for changes in patient enrollment i.e. eligibility. The common way to do this is by computing member months and using this as the denominator. Statistics that have been normalized changes in members months are often reported as "per-member-per-month" or "PMPM". For example, one would typically look at ED visits PMPM.

In case it isn't obvious, the reason it's a best practice to normalize for changes in enrollment when trending these sorts of statistics is because things like claims payments and utilization will change month-to-month simply because the eligible population changes as members gain/lose eligibility due to changes in employment, birth, death, etc.

The process of calculating PMPM requires assigning claims to a particular member month. The two date fields most commonly used to do this are claim_start_date and claim_end_date. Paid_date is less commonly used because doing so will include variation due to claims adjudication e.g. adjustments that occur for some claims over time. Sometimes this is desired, but for most analyses it's more common to take the date from when the healthcare encounter occurred.

Using claim_start_date will often lead to slightly different results than claim_end_date, though the difference is often small. Although there is no hard rule, it's more common to use claim_start_date, the thinking being that if a patient loses eligibility during a long encounter, the insurer who covered the patient at the beginning of the encounter is more likely to pay. However we haven't seen much hard evidence supporting this hypothesis and our current use of claim_start_date is more out of convention than anything else.

5.2 Calculating Member Months

In this section we use an example to describe how to calculate member months. This is the same methodology we use in the Tuva Project.

To calculate member months, you need to convert each patient's eligibility record (with start and end dates) into multiple records, with one record for each month of eligibility. Let's take an example. Suppose member A1234 has coverage from Aetna from January 1st to June 15th of 2022. They lose coverage on June 16th and they regain coverage on August 10th. Further suppose member B2468 has coverage from January 1st through the entire year of 2022. These two members would have eligibility spans that look like the data below:

person_id	payer	enrollment_start_date	enrollment_end_date
A1234	Aetna	01-01-2022	06-15-2022
A1234	Aetna	08-10-2022	
B2468	Aetna	01-01-2022	12-31-2022

Finally, let's suppose the current date is January 31st 2023.

In this example, patient B2468 has an enrollment span with 12 months of continuous eligibility, and so should be counted as having 12 member months. And it turns out that A1234 should also be counted as having 12 member months, but the assignment isn't as straightforward. To unpack it, we need to take a brief detour into partial eligibility.

Partial eligibility occurs whenever a patient does not have eligibility for an entire month. A1234 has partial eligibility for the months of June 2022 and August 2022. There are multiple methods for handling partial eligibility when computing member months, but the most common method is to assume full eligibility for the entire month. Not every type of health insurance coverage works like this, but the majority do. In the example above we would give A1234 a full member month for both June 2022 and August 2022, following this method.

After converting the above enrollment spans to member months (e.g. by using the SQL at the end of this section), the data would look like this:

person_id	year_month	payer
A1234	2022-01	Aetna
A1234	2022-02	Aetna
A1234	2022-03	Aetna
A1234	2022-04	Aetna
A1234	2022-05	Aetna
A1234	2022-06	Aetna
A1234	2022-08	Aetna
A1234	2022-09	Aetna
A1234	2022-10	Aetna
A1234	2022-11	Aetna
A1234	2022-12	Aetna
A1234	2023-01	Aetna
B2468	2022-01	Aetna
B2468	2022-02	Aetna
B2468	2022-03	Aetna

person_id	year_month	payer
B2468	2022-04	Aetna
B2468	2022-05	Aetna
B2468	2022-06	Aetna
B2468	2022-07	Aetna
B2468	2022-08	Aetna
B2468	2022-09	Aetna
B2468	2022-10	Aetna
B2468	2022-11	Aetna
B2468	2022-12	Aetna

Notice that the last member month given to patient A1234 was for January 2023, since we supposed the present date was January 31st 2023 in our example.

5.3 Data Quality Problems

In order to correctly compute member months, it's important to take potential data quality issues into account, for example:

- Members with overlapping enrollment periods
- Enrollment end dates before enrollment start dates
- Duplicate enrollment records
- Null values in enrollment start date

6 Member Attribution

Provider panel attribution is the means of attributing individual patients to a provider, usually a primary care provider using medical claims or electronic medical records data. There are many reasons why one would like attribute patients to providers from evaluating performance or understanding healthcare outcomes.

There is not “one way” to approach provider attribution, and there are many valid approaches to how to go about attributing patients to providers. In this section, we will talk about provider attribution at a high level, and look at one of the attribution methods used by the Center of Medicare and Medicaid services (CMS) in the context of their Next Gen ACO models, Direct Contracting Entity models, and ACO REACH programs (the same attribution model being used in all these programs).

6.1 Before Reading Further

Ask yourself, how should patients be attributed to providers? What if you only had medical claims data available to you, and you need to attribute patients to primary care providers?

- How do you determine if a provider is a primary care provider?
- Should a blood draw or lab count in the attribution process?
- If a patient doesn’t visit a provider in a long time, should they be attributed to a provider?
- How do you handle if a patient is seeing multiple providers in a short period of time?
- What data fields are available to you from medical claims data?

6.2 Example attribution model from CMS:

This example will use a single attribution model from some of CMS’s CMMI programs. Full details about this attribution model can be found here in “Appendix B”. This attribution model from CMS uses exclusively Medicare claims data to attribute Medicare patients to providers. Specific data used from the claims data are:

1. **Rendering provider NPI** (used to determine if provider is primary care)
2. **Procedure code** (CPT / HCPCS) (used to determine if service is primary care service)

3. **Place of service code** (used to determine if service is primary care service)
4. **Date of service** (used for timing)
5. **Claim allowed amounts** (used for weighting)

Other sources of data are used as well for attribution in this model:

1. **NPPES – NPI registry data** (used to determine if a provider is a primary care provider via the provider's primary care taxonomy).
2. **Taxonomy crosswalk data** (Crosswalks Medicare specialty code to provider taxonomy code)

6.2.1 Attribution model at a high level

The attribution model CMS uses to attribute patients to providers is used in the context of primary care. First CMS only looks at a subset of providers it deems as "primary care." Then for a given patient, the attribution model looks at medical claims data for this subset of primary care providers and looks at which provider that patient has seen the most, with some weighting applied to give preference to more recently seen providers. Let's walk through an example patient.

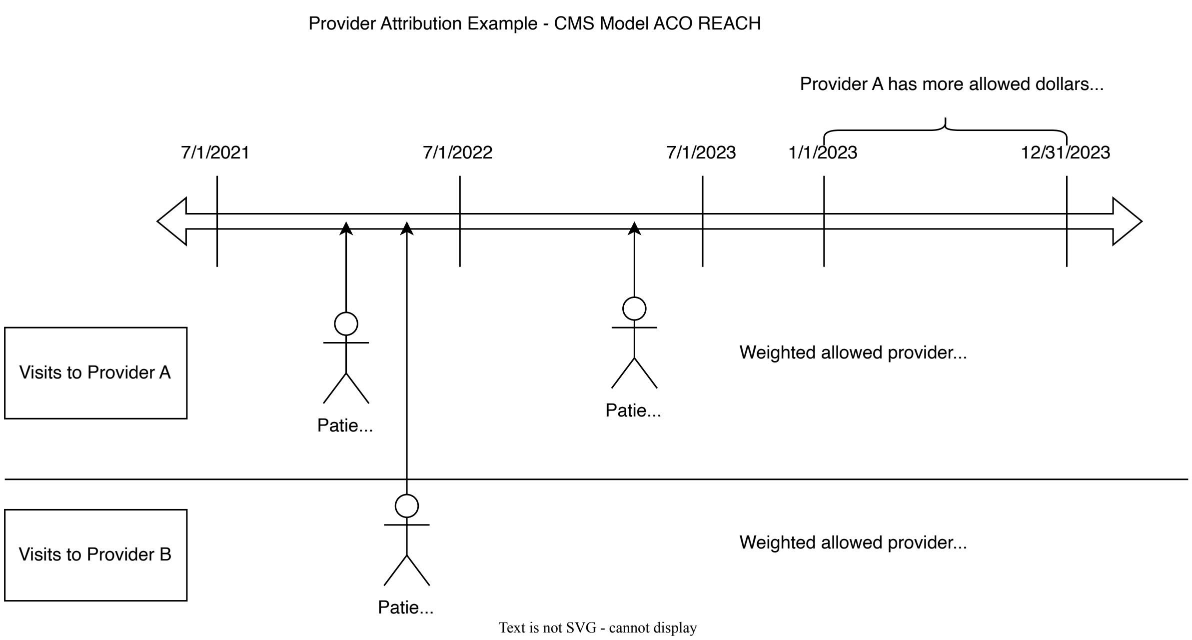


Figure 6.1: Example Patient Overview

6.2.2 Going into more detail

1. **Pull all medical claims data for a patient (professional claims only).**
2. **Filter the medical claims data based on the rendering provider NPI and cross-walks to filter to just primary care providers.** Based on the rendering provider NPI on the claim, lookup that provider's taxonomy on the npi registry database.

Table B.6.4. Specialty Codes Used to Identify Primary Care Specialists

Code[1]	Specialty
1	General Practice
8	Family Medicine
11	Internal Medicine
37	Pediatric Medicine
38	Geriatric Medicine
50	Nurse Practitioner
89	Clinical Nurse Specialist
97	Physician Assistant

[1] The Medicare Specialty Code. A cross-walk between Medicare Specialty Codes and the Healthcare Provider Taxonomy is published on the CMS website at https://www.cms.gov/Medicare/Provider-Enrollment-and-Certification/MedicareProviderSupEnroll/Downloads/TaxonomyCrosswalk.pdf

Figure 6.2: primary_care_specialty_codes

Here are the primary care specialty codes specified in this attribution model.

This crosswalk links the types of providers and suppliers who are eligible to apply for enrollment in the Medicare program with the appropriate Healthcare Provider Taxonomy Codes. This crosswalk includes the Medicare Specialty Codes for those provider/supplier types who have Medicare Specialty Codes. The Healthcare Provider Taxonomy Code Set is available from the Washington Publishing Company (www.wpc-edi.com) and is maintained by the National Uniform Claim Committee (www.nucc.org). The code set is updated twice a year, with the updates being effective April 1 and October 1 of each year. This document reflects Healthcare Provider Taxonomy Codes effective for use on October 1, 2017.

When changes are made to Medicare provider enrollment requirements, the Medicare Specialty Codes, or the Healthcare Provider Taxonomy Code Set, this document may need to be revised.

NOTE: This document does not alter existing Medicare claims preparation, processing, or payment instructions, nor does it alter existing Medicare provider enrollment requirements or policies.

MEDICARE SPECIALTY CODE	MEDICARE PROVIDER/SUPPLIER TYPE DESCRIPTION	PROVIDER TAXONOMY CODE	PROVIDER TAXONOMY DESCRIPTION: TYPE, CLASSIFICATION, SPECIALIZATION
01	Physician/General Practice	208D00000X	Allopathic & Osteopathic Physicians/General Practice
02	Physician/General Surgery	208600000X	Allopathic & Osteopathic Physicians/Surgery
		2086H0002X	Allopathic & Osteopathic Physicians/Surgery/Hospice and Palliative Medicine
		2086S0120X	Allopathic & Osteopathic Physicians/Surgery/Pediatric Surgery
		2086S0122X	Allopathic & Osteopathic Physicians/Surgery/Plastic and Reconstructive Surgery
		2086S0105X	Allopathic & Osteopathic Physicians/Surgery/Surgery of the Hand
		2086S0102X	Allopathic & Osteopathic Physicians/Surgery/Surgical Critical Care
		2086X0206X	Allopathic & Osteopathic Physicians/Surgery/Surgical Oncology
		2086S0127X	Allopathic & Osteopathic Physicians/Surgery/Trauma Surgery
		2086S0129X	Allopathic & Osteopathic Physicians/Surgery/Vascular Surgery
		208G00000X	Allopathic & Osteopathic Physicians/Thoracic Surgery (Cardiothoracic Vascular Surgery)
		204F00000X	Allopathic & Osteopathic Physicians/Transplant Surgery
		208C00000X	Allopathic & Osteopathic Physicians/Colon & Rectal Surgery
		207T00000X	Allopathic & Osteopathic Physicians/Neurological Surgery
		204E00000X	Allopathic & Osteopathic Physicians/Oral & Maxillofacial Surgery
		207X00000X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery
		207XS0114X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Adult Reconstructive Orthopaedic Surgery
		207XX0004X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Foot and Ankle Surgery
		207XS0106X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Hand Surgery
		207XS0117X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Orthopaedic Surgery of the Spine
		207XX0801X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Orthopaedic Trauma
		207XP3100X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Pediatric Orthopaedic Surgery
		207XX0005X	Allopathic & Osteopathic Physicians/Orthopaedic Surgery/Sports Medicine
		208200000X	Allopathic & Osteopathic Physicians/Plastic Surgery
		2082S0099X	Allopathic & Osteopathic Physicians/Plastic Surgery/Plastic Surgery Within the Head & Neck
		2082S0105X	Allopathic & Osteopathic Physicians/Plastic Surgery/Surgery of the Hand
03	Physician/Allergy/ Immunology	207K00000X	Allopathic & Osteopathic Physicians/Allergy and Immunology
		207KA0200X	Allopathic & Osteopathic Physicians/Allergy and Immunology/Allergy
		207KI0005X	Allopathic & Osteopathic Physicians/Allergy and Immunology/Clinical & Laboratory Immunology
04	Physician/Otolaryngology	207Y00000X	Allopathic & Osteopathic Physicians/ Otolaryngology
		207YS0123X	Allopathic & Osteopathic Physicians/ Otolaryngology/Facial Plastic Surgery
		207YX0602X	Allopathic & Osteopathic Physicians/Otolaryngology/Otolaryngic Allergy
		207YX0905X	Allopathic & Osteopathic Physicians/Otolaryngology/Otolaryngology/Facial Plastic Surgery
		207YX0901X	Allopathic & Osteopathic Physicians/Otolaryngology/Otology &Neurotology
		207YP0228X	Allopathic & Osteopathic Physicians/Otolaryngology/Pediatric Otolaryngology
		207YX0007X	Allopathic & Osteopathic Physicians/Otolaryngology/Plastic Surgery within the Head & Neck
		207YS0012X	Allopathic & Osteopathic Physicians/Otolaryngology/Sleep Medicine
05	Physician/Anesthesiology	207L00000X	Allopathic & Osteopathic Physicians/Anesthesiology
		207LA0401X	Allopathic & Osteopathic Physicians/Anesthesiology/Addiction Medicine
		207LC0200X	Allopathic & Osteopathic Physicians/Anesthesiology/Critical Care Medicine
		207LH0002X	Allopathic & Osteopathic Physicians/Anesthesiology/Hospice and Palliative Medicine
		207LP2900X	Allopathic & Osteopathic Physicians/Anesthesiology/Pain Medicine
		207LP3000X	Allopathic & Osteopathic Physicians/Anesthesiology/Pediatric Anesthesiology
06	Physician/Cardiovascular Disease (Cardiology)	207RC0000X	Allopathic & Osteopathic Physicians/Internal Medicine, Cardiovascular Disease

Figure 6.3: primary_care_specialty_code_taxonomy_crosswalk

They need to be taken in context of the crosswalk to get from specialty code to taxonomy code.

3. **Filter out any claims that are not in the PQEM procedure code set provided.**

Table B.6.3: Evaluation & Management Services

Administration of HRA	
96160	Administration of patient-focused health risk assessment instrument
96161	Administration of caregiver-focused health risk assessment instrument
Office or Other Outpatient Services	
99201	New Patient, brief
99202	New Patient, limited
99203	New Patient, moderate
99204	New Patient, comprehensive
99205	New Patient, extensive
99211	Established Patient, brief
99212	Established Patient, limited

Global and Professional Direct Contracting Model
Financial Operating Guide: Overview Rev. 1.1

Appendix B: Beneficiary Alignment Procedures 43

99213	Established Patient, moderate
99214	Established Patient, comprehensive
99215	Established Patient, extensive
Domiciliary, Rest Home, or Custodial Care Services	
99324	New Patient, brief
99325	New Patient, limited
99326	New Patient, moderate
99327	New Patient, comprehensive
99328	New Patient, extensive
99334	Established Patient, brief
99335	Established Patient, moderate
99336	Established Patient, comprehensive
99337	Established Patient, extensive
Professional services provided in a non-skilled Nursing Facility[1]	
99304	Initial Nursing Facility Care
99305	Initial Nursing Facility Care

Figure 6.4: primary_care_procedure_codes

Here's a subset of some of the procedure codes that are considered primary care services. The full list can be found on page 40 of this pdf. Some have a place of service code pre-requisite to be considered valid.

4. **Filter out claims that do not meet the date range criteria for the alignment period.**
5. **Apply weighting to allowed amounts and sum by provider. Claims in the earlier alignment period should be weighted by 1/3, and claims in the more recent alignment period should be weighted 2/3.**

Table B.2.1 Alignment Years for each Performance Year and Base Year

Base/Performance Year	Period Covered	Alignment Year 1	Alignment Year 2
Base Year 1	CY2017	7/1/2014 – 6/30/2015	7/1/2015 – 6/30/2016
Base Year 2	CY2018	7/1/2015 – 6/30/2016	7/1/2016 – 6/30/2017
Base Year 3	CY2019	7/1/2016 – 6/30/2017	7/1/2017 – 6/30/2018
Performance Year 1	April 1, 2021 – December 31, 2021	7/1/2018 – 6/30/2019	7/1/2019 – 6/30/2020
Performance Year 2	CY2022	7/1/2019 – 6/30/2020	7/1/2020 – 6/30/2021
Performance Year 3	CY2023	7/1/2020 – 6/30/2021	7/1/2021 – 6/30/2022
Performance Year 4	CY2024	7/1/2021 – 6/30/2022	7/1/2022 – 6/30/2023
Performance Year 5	CY2025	7/1/2022 – 6/30/2023	7/1/2023 – 6/30/2024
Performance Year 6	CY2026	7/1/2023 – 6/30/2024	7/1/2024 – 6/30/2025

Figure 6.5: example_alignment_periods

6. **The provider with the most weighted dollars gets the patient attributed. In the case of a tie, choose the provider with the more recent claim.**

6.2.3 Issues that could be raised about the attribution model above.

- This model will not attribute patients who have not seen their primary care provider in the two-year observation period.
- This model will end up not weighting care provided by Advanced Practice Providers (APPs) less than care provided by MDs and DOs.
- This model might have specialty care services competing with primary care services due to the taxonomy classification system and crosswalk not reflecting APPs providing specialist care.
- This model might not be a timely reflection for attributing patients to providers. It may take up to two years for a patient to get appropriately attributed to a provider after a change in provider.

6.3 Things to consider when creating an attribution model

- How are you going to use the model? Is your use care specific to primary care?
- Does it make sense to use allowed amounts?
- How should timing of visits be taken into account?
- Is there other sources besides medical claims data that you have available to use?
- How should APP visits be handled?
- Does your model need to be straight forward and explainable to others?

6.4 References

This page just took a look at a single attribution model. More attribution models exist. * https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6549236/

7 Service Categories

In the Tuva Project, we've created a service category grouper to help us analyze payment and utilization metrics. We use it to categorize medical claim lines.

Data Elements

The data elements that we use to create this grouper are as follows:

- **bill_type_code:** Bill type code for the claim (institutional claims only).
- **revenue_center_code:** Revenue center code for the claim line (institutional only and typically multiple codes per claim).
- **ms_drg_code:** MS-DRG for the claim (inpatient claims only).
- **place_of_service_code:** Place of service for the claim (professional claims only).
- **hcpcs_code:** HCPCS level 1 or level 2 code for the claim line. Most definitions use the CCS groupings of codes instead of referencing codes individually.
- **icd_diagnosis_code:** Typically referenced through CCSR groupings instead of individual codes.
- **npi:** Used to reference the taxonomy code of a facility NPI and a provider's specialty.

The Tuva Project Service Category Grouper has three levels in a hierarchy with each subcategory rolling up to a high level category. Because all subcategories roll up to one and only one higher level category, the sum of all the logic for each subcategory in a category is the same as the logic for the category. As such, we'll describe the higher level categories conceptually without codes, and then we'll define each subcategory sharing the code sets. See table below for a quick view of the categories and subcategories:

SERVICE_CATEGORY_1	SERVICE_CATEGORY_2	SERVICE_CATEGORY_3
inpatient	acute inpatient	l/d - cesarean delivery
inpatient	acute inpatient	l/d - newborn
inpatient	acute inpatient	l/d - newborn nicu
inpatient	acute inpatient	l/d - other
inpatient	acute inpatient	l/d - vaginal delivery
inpatient	acute inpatient	medical
inpatient	acute inpatient	surgical
inpatient	acute inpatient	acute inpatient - other
inpatient	inpatient hospice	inpatient hospice
inpatient	inpatient psychiatric	inpatient psychiatric

SERVICE_CATEGORY_1	SERVICE_CATEGORY_2	SERVICE_CATEGORY_3
inpatient	inpatient rehabilitation	inpatient rehabilitation
inpatient	inpatient substance use	inpatient substance use
inpatient	skilled nursing	skilled nursing
office-based	office-based pt/ot/st	office-based pt/ot/st
office-based	office-based radiology	ct
office-based	office-based radiology	general
office-based	office-based radiology	mri
office-based	office-based radiology	pet
office-based	office-based surgery	office-based surgery
office-based	office-based visit	office-based visit
office-based	telehealth visit	telehealth visit
office-based	office-based other	office-based other
outpatient	ambulatory surgery center	ambulatory surgery center
outpatient	dialysis	dialysis
outpatient	emergency department	emergency department
outpatient	home health	home health
outpatient	observation	observation
outpatient	outpatient hospice	outpatient hospice
outpatient	outpatient hospital or clinic	outpatient hospital or clinic
outpatient	outpatient psychiatric	outpatient psychiatric
outpatient	outpatient pt/ot/st	outpatient pt/ot/st
outpatient	outpatient radiology	ct
outpatient	outpatient radiology	general
outpatient	outpatient radiology	mri
outpatient	outpatient radiology	pet
outpatient	outpatient rehabilitation	outpatient rehabilitation
outpatient	outpatient substance use	outpatient substance use
outpatient	outpatient surgery	outpatient surgery
outpatient	pharmacy	pharmacy
outpatient	urgent care	urgent care
ancillary	ambulance	ambulance
ancillary	durable medical equipment	durable medical equipment
ancillary	lab	lab
other	other	other

When developing the service category grouper we kept the following principles in mind:

- **Cardinality is Key:** If there were hundreds of categories, it would be too hard for a human to make sense of what was going on. But if you only had 2 categories for example, it wouldn't be enlightening. Almost all insights would come from breaking it down further.

- **Mutually Exclusive and Exhaustive:** Every healthcare claims can be grouped into one service category and only one service category. This implies that summing the total payments for all service categories would equal the sum of all payments for each individual claim.
- **The "Other" Category Isn't Too Large:** In order to make the grouper Exhaustive, we group everything we can into meaningful categories and then put everything else in the "other" category. If this "other" category is too large, that means we need to break it out into additional meaningful categories.
- **Hierarchical:** It's a balancing act to try to create groups with low cardinality but providing enough homogeneity inside each group for analysis to be actionable. This often leads us to create hierarchical groupers so that you can see high level groups first and then drill in to get more specific while still keeping the broader context simple.
- **Feasible:** Any categorization grouper is only useful if you're able to group things into the categories using data elements that are readily available and populated reasonably consistently.

The Tuva Project Service Category Grouper categorizes most institutional claims at the claim level using the bill type code for each claim. All inpatient institutional claims are defined at the claim level, while some outpatient institutional service categories are grouped at the line level (such as radiology which is defined using HCPCS codes). Professional claims are also defined at the claim line level.

8 Service Categories

8.1 Inpatient

Service Category 2 (Click to expand and see specific codes that make up each category. Service category 3 is listed where applicable.)

Acute Inpatient

Institutional Claims

- **DRG Codes**:
  - Any valid Diagnosis-Related Group (MS-DRG or APR-DRG) code: These classify hospital cases into groups expected to have similar hospital resource use.
- **Bill Type Codes**:
  - **11x**: General Inpatient
  - **12x**: Inpatient Psychiatric Services

Professional Claims

- **Place of Service Code**:
  - **21**: Inpatient Hospital

Service Category 3

- **Medical**:
  - DRGs designated as Medical per CMS DRG definition
- **Surgical**:
  - DRGs designated as Surgical per CMS DRG definition
- **Acute Inpatient**:

    - Any other acute inpatient claims that don't roll up to other service categories.

- **L/D Vaginal Delivery**:
    - **768**: Vaginal delivery with complicating diagnoses.
    - **796**: Vaginal delivery with other specified conditions.
    - **797**: Vaginal delivery with O.R. procedure except sterilization and/or D&C.
    - **798**: Vaginal delivery with sterilization and/or D&C.
    - **805**: Vaginal delivery without complicating diagnoses.
    - **806**: Vaginal delivery with tubal ligation/sterilization.
    - **807**: Vaginal delivery with antepartum conditions.
- **L/D Cesarean Delivery**:
    - **783**: Cesarean delivery with complicating diagnoses.
    - **784**: Cesarean delivery with sterilization and/or D&C.
    - **785**: Cesarean delivery with O.R. procedure except sterilization and/or D&C.
    - **786**: Cesarean delivery with other specified conditions.
    - **787**: Cesarean section without complicating diagnoses.
    - **788**: Cesarean section with tubal ligation/sterilization.

Inpatient Substance Use

Institutional Claims

- **Taxonomy Codes**:
    - **324500000X**: Substance Abuse Rehabilitation Facility
    - **261QR0405X**: Substance Use Disorder Rehabilitation Facility
    - **101YA0400X**: Addiction (Substance Use Disorder)
- **CCSR Category Codes**:
    - **MBD026**: Substance Use Disorders
    - **SYM008**: Mental Health and Substance Use Interventions
    - **MBD025**: Alcohol Use Disorders
    - **SYM009**: Mental Health and Substance Use Assessment
    - **MBD034**: Drug Use Disorders

Professional Claims

- None

Inpatient Hospice

Institutional Claims

- **Bill Type Codes**:
  - **82x**: Inpatient hospice services

Professional Claims

- **Place of Service Code**:
  - **34**: Hospice facility

8.2 Outpatient

Ambulatory Surgery Center

Institutional Claims

- **Revenue Codes**:
  - **0490**: Ambulatory Surgical Care - General classification
  - **0499**: Ambulatory Surgical Care - Other
- **Taxonomy Code**:
  - **261QA1903X**: Ambulatory Surgical Center

Professional Claims

- **Place of Service Code**:
  - **24**: Ambulatory Surgical Center

Dialysis

Institutional Claims

- **Bill Type Codes**:
  - **72**: Independent Renal Dialysis Center
- **Revenue Center Codes**:
  - **082x**: Hemodialysis
  - **083x**: Peritoneal Dialysis

Professional Claims

- **Place of Service Code**:
  - **65**: End-Stage Renal Disease Treatment Facility

9 Encounters

Part II

Advanced Topics

10 Acute Inpatient Stays

In this section we describe how to create acute inpatient encounters (i.e. hospital stays) from medical claims data. At the highest level this involves two steps:

1. Identifying claims that occurred during an acute inpatient hospital stay
2. Merging these claims into a single encounter

We go into significant detail about how this is done in the Tuva data model and how to use data tables in the data model to identify possible data quality problems and perform analytics.

First we need to connect to our database. We do this with the following connection.

```
%%capture

import snowflake.connector
import pandas as pd

# Connect to Snowflake with SSO
conn = snowflake.connector.connect(
    user="aaron@tuvahealth.com",
    account="ksa27360.us-east-1",
    authenticator="externalbrowser",
    warehouse="COMPUTE_WH",
    database="medicare_lds_five_percent",
    schema="PUBLIC",
    role="accountadmin"
)
```

10.1 Identifying Acute Inpatient Claims

The first step in building acute inpatient encounters is identifying claims that occurred during acute inpatient hospital stays. We need to do this for both institutional and professional claims.

The following fields are commonly used to identify acute inpatient institutional claims: - Bill type code in (11X, 12X) - Any valid MS- or APR-DRG - Room and board revenue center codes

Bill type codes equal to 11X or 12X should only be found on acute inpatient claims. The same is true for any valid MS- or APR-DRGs. However, room and board revenue center codes can be found on a wide variety of inpatient claims (e.g. SNF, inpatient rehab, etc.).

Given these fields are essential for this analysis, we need to quickly assess whether they have any data quality problems. For bill type and DRG fields we need to check the following issues: - Missing: Every claim line should have a value populated - Invalid: Every value should exist in the official terminology table - Duplicate: Every claim should have one and only one value

By comparison, revenue center codes should meet the missing and invalid criteria, however, each institutional claim can and typically does have more than one value.

The data quality summary in acute inpatient can quickly reveal whether these fields have any of these problems. Let's query this table and take a look.

```
%%capture
# Query the acute IP data quality summary table and print a dataframe with the relevent rows
query = "select * from input_layer.medical_claim limit 100000"
df = pd.read_sql(query, conn)
```

```
import pandas as pd
import plotly.express as px

# Normalize column names to upper case
df.columns = df.columns.str.upper()

# Ensure CLAIM_START_DATE is a datetime field
df['CLAIM_START_DATE'] = pd.to_datetime(df['CLAIM_START_DATE'])

# Extract month and year for grouping
df['YEAR'] = df['CLAIM_START_DATE'].dt.year
df['MONTH'] = df['CLAIM_START_DATE'].dt.strftime('%b')  # Short month name
df['YEAR_MONTH'] = df['CLAIM_START_DATE'].dt.to_period('M')

# Group by month-year and count unique claims
monthly_claims = df.groupby(['YEAR_MONTH', 'YEAR', 'MONTH'])['CLAIM_ID'].nunique().reset_inde
monthly_claims['YEAR_MONTH'] = monthly_claims['YEAR_MONTH'].dt.to_timestamp()

# Create the trend chart with Plotly
fig = px.line(
```

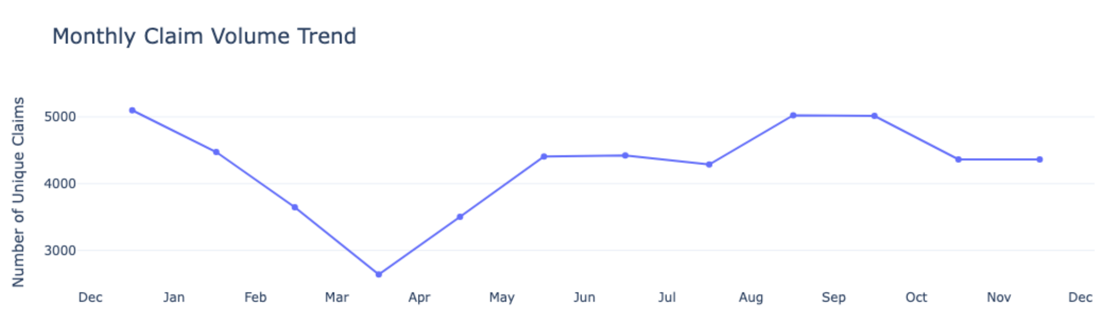
```python
    monthly_claims,
    x='YEAR_MONTH',
    y='CLAIM_ID',
    title='Monthly Claim Volume Trend',
    labels={'CLAIM_ID': 'Number of Unique Claims'},
    markers=True
)

# Customize the layout
fig.update_layout(
    title_font_size=20,
    xaxis_title='',
    yaxis_title='Number of Unique Claims',
    xaxis_tickangle=0,
    template='plotly_white',
)

# Update x-axis ticks to display months and years
fig.update_xaxes(
    tickformat='%b',  # Show abbreviated month names
    dtick='M1',       # Set ticks for every month
    ticklabelmode='period',  # Period-aligned ticks
    showgrid=False,
)

# Display the chart
fig.show()
```

```
Unable to display output for mime type(s): text/html
```



```
# Close the connection
conn.close()
```

```
# this is a test
```

11 Readmissions

Hospital readmissions are one of the most common healthcare concepts. They are also one of the most complicated concepts to define and implement as code. Here we provide a general overview of how to calculate a hospital readmission measure.

11.1 Overview

There are many different ways to define hospital readmission measures. However every readmission measure is built on two underlying concepts: the index admission and the readmission. The index admission is a hospitalization that qualifies to be included in the readmission measure. Not all hospitalizations will meet the criteria to be index admissions and will therefore not be included in the readmission measure. For example, if a patient dies during a hospitalization, that hospitalization will not be an index admission, and will not be included in the readmission measure. There are many more pieces of logic to define if a hospitalization counts as an index admission for each different readmission measure. For example, hospitalizations for medical treatment of cancer are not index admissions. We explain each piece of logic further below.

To better understand these fundamental concepts, it is helpful to think of building a readmission measure as follows:

1. Start with a data table of inpatient admissions, one record per admission.
2. Use logic (defined further below) to compute two additional columns which you append to this table:

    - The first column (index_admit_flag) is a binary variable that indicates whether each inpatient admission qualifies as an index admission.
    - The second column (readmit_flag) is a binary variable that indicates whether each inpatient admission had a subsequent admission that qualifies as a readmission.

encounter_id	patient_id	admit_date	discharge_date	ms_drg	facility	index_admit_flag	readmit_flag
A1234	0001	2019-01-04	2019-01-08	198	Good Samaritan	1	1
B1234	0001	2019-01-17	2019-01-28	197	Good Samaritan	1	0
C1234	0002	2019-02-01	2019-02-06	871	St. Joseph's	0	0

Figure 11.1: Readmission Table

The index admission (index_admit_flag) and readmission (readmit_flag) are then used to compute the readmission measure. The sum of the index_admit_flag column forms the denominator of the readmission measure while the sum of the readmit_flag column forms the numerator. In the example data table above there are 3 total inpatient admissions, 2 index admissions, and 1 readmission, resulting in a 50% readmission rate.

$$50\% = \frac{1 \text{ readmission}}{2 \text{ index admissions}}$$

Figure 11.2: Example Readmission Rate

The different definitions of readmission measures are simply variations in the inclusion and/or exclusion criteria that define the index_admit_flag and the readmit_flag. The most commonly used readmission measures are the CMS readmission measures.

There are 7 such measures grouped into 3 categories:

- All-cause Hospital-wide Readmissions (1)
- Condition-specific Readmissions (4)
    - Acute Myocardial Infarction
    - Congestive Obstructive Pulmonary Disease (COPD)
    - Pneumonia
    - Heart Failure

- Procedure-specific Readmissions (2)
  - Total Hip/Knee Arthroscopy
  - Coronary Artery Bypass Graft

11.2 CMS Readmissions Algorithm

The All-cause Hospital-wide Readmission Measure ("Hospital-wide Measure") is the most commonly used readmission measure definition of all. The sections that follow describe how to define and implement the Hospital-wide Measure on EHR data or claims data in your data warehouse. This measure has also been encoded into the Tuva Project, including all necessary terminology datasets and data quality tests.

11.2.1 Data Requirements

The Hospital-wide Measure can be implemented on either EHR data or claims data. Strictly speaking, CMS developed the Hospital-wide Measure to run on Medicare FFS claims data, but CMS explicitly states in their documentation that the measure may be adapted to run against an all-payer patient population.

The data elements needed to process the readmission measure are listed below.

- patient
  - person_id
  - gender
  - birth_date
- encounter
  - encounter_id
  - person_id
  - encounter_start_date
  - encounter_end_date
  - discharge_disposition_code
  - location
  - ms_drg
  - encounter_type
- condition
  - encounter_id
  - code
  - diagnosis_rank

  - code_type
  - condition_type

- procedure

  - encounter_id
  - procedure_code
  - code_type

11.2.2 Terminology Datasets

The Hospital-wide Measure requires several terminology datasets that are used as lookup tables to create the index admission and readmission. The following is a complete list of the terminology datasets that are needed:

#	Terminology Dataset	Description
1	HWR Specialty Cohort Inclusion	A list of CCS diagnosis and procedure categories based on ICD-10-CM and ICD-10-PCS codes that make up 4 specialty categories: medicine, cardiorespiratory, cardiovascular, and neurology.
2	HWR Surg-Gyn Cohort Inclusion	A list of ICD-10-PCS codes that make up the surgical-gynecological specialty category.
3	HWR Cohort Exclusion	A list of 47 CCS diagnosis categories based on ICD-10-CM codes that make up 3 categories: cancer, rehab, and mental health.
4	Planned Admissions: Always Planned Procedures	A list of CCS procedure codes that are considered always planned procedures.
5	Planned Admissions: Always Planned Diagnoses	A list of CCS diagnosis categories that are considered always planned diagnoses.
6	Planned Admissions: Potentially Planned Procedures	A list of CCS procedure categories and ICD-10-PCS codes that are considered potentially planned if the patient does not have an always acute diagnosis.
7	Planned Admissions: Always Acute Diagnoses	A list of CCS diagnosis categories and ICD-10-CM codes that are considered always acute diagnoses.
8	Custom CCS: ICD-10-CM Mapping	A modified version of the beta version 2019.1 CCS mapping from AHRQ for ICD-10-CM diagnosis codes.
9	Custom CCS: ICD-10-PCS Mapping	A modified version of the beta version 2019.1 CCS mapping from AHRQ for ICD-10-PCS procedure codes.

Figure 11.3: Terminology Datasets

CMS makes these terminology datasets available as files on the quality net website. The links below download two Excel spreadsheets that include all the terminology datasets listed above. - https://qualitynet.cms.gov/files/60943ca9fd340b002259fe16?filename=2021_HWR.xlsx - https://qualitynet.cms.gov/files/6092ab86fd340b002259fda0?filename=YaleMod_CCS_PCS_CM_Map_v202

These terminology datasets are already included in the Tuva Project.

11.2.3 Index Admission Algorithm

The index admission algorithm is the set of sub-algorithms (i.e. rules) that together determine whether an inpatient admission qualifies as an index admission (i.e. receives an index_admit_flag = 1 or 0). Not every inpatient admission qualifies as an index admission.

Here are the sub-algorithms used by the Hospital-wide Measure:

1. Cohort Inclusion Algorithm
2. Cohort Exclusion Algorithm
3. Discharged Alive and Not Against Medical Advice
4. Not a Transfer
5. Not a Same Day Readmission
6. Run-out

If an inpatient admission meets the criteria from all of these sub-algorithms then it qualifies as an index admission. Otherwise it does not. Below we walk through each sub-algorithm.

11.2.4 Cohort Inclusion Algorithm

The cohort inclusion algorithm is a set of rules that determine whether an inpatient admission belongs to 1 or more specialty categories. In order to qualify as an index admission, an inpatient admission must belong to 1 or more of these 5 specialty categories:

1. Medicine
2. Cardiorespiratory
3. Cardiovascular
4. Neurology
5. Surgery / Gynecology

Here are the steps to implement the cohort inclusion algorithm:

1. Each inpatient admission should have 1 primary ICD-10-CM diagnosis code and 0 or more ICD-10-PCS procedure codes associated with it.

2. Map each of these codes to the custom CCS Condition Categories (terminology datasets #8 and #9). Now each inpatient admission should have its corresponding CCS diagnosis and procedure categories assigned.

3. For each inpatient admission, map its assigned CCS diagnosis and procedure categories to those listed in the HWR Specialty Cohort Inclusion terminology dataset (terminology dataset #1) to determine if it should be assigned to any of the first 4 specialty categories (medicine, cardiorespiratory, cardiovascular, and neurology).

4. For each inpatient admission, map its ICD-10-PCS codes to those listed in the HWR Surg-Gyn Cohort Inclusion terminology dataset (terminology dataset #2) to determine if the readmission should be assigned to the Surgical / Gynecology specialty category.

5. Inpatient admissions are successfully assigned to one or more specialty categories based on steps 3 and 4 pass this algorithm. Otherwise the admission fails and is no longer a candidate for an index admission.

Here is a diagram of the cohort inclusion algorithm:

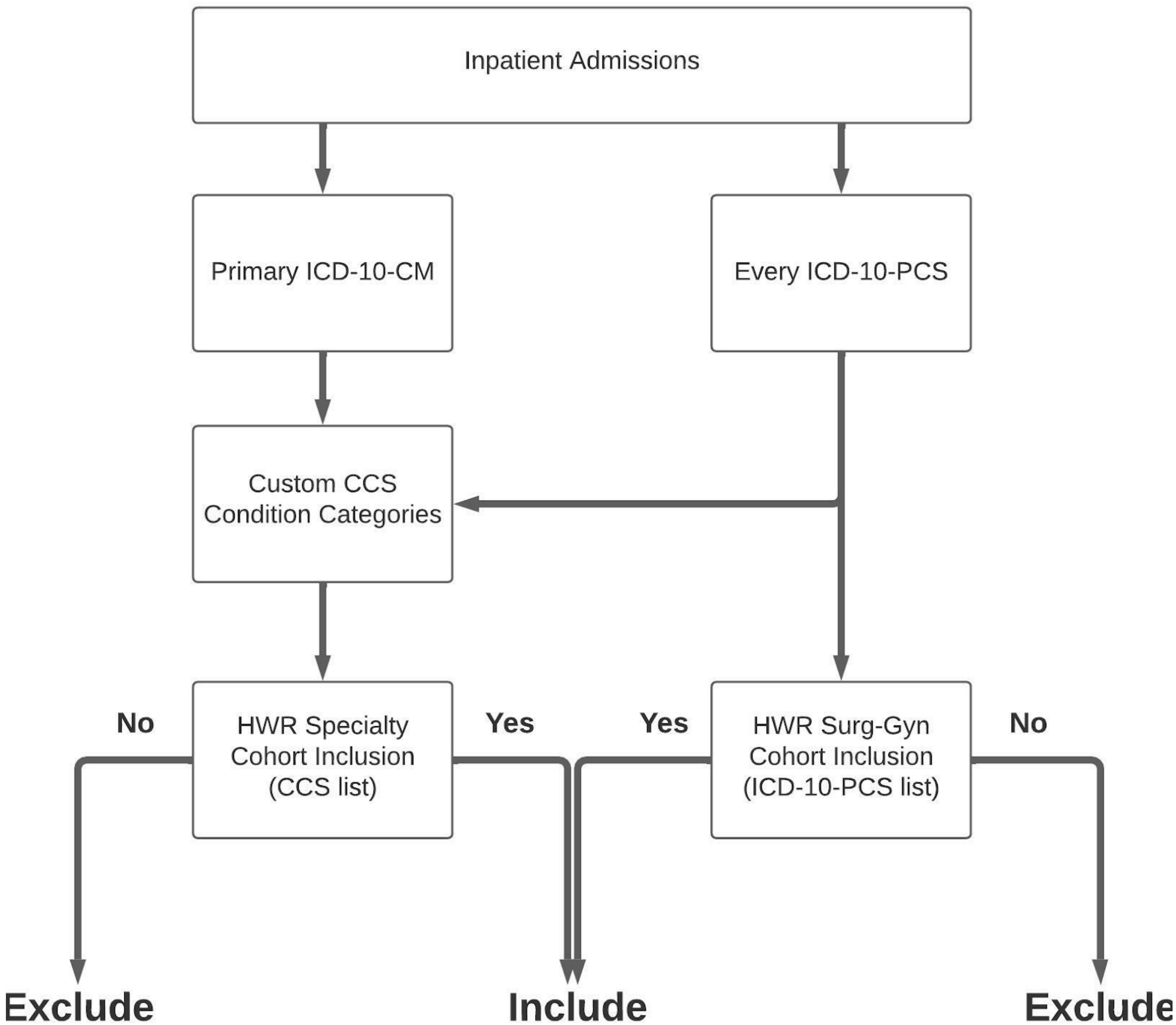


Figure 11.4: Cohort Inclusion Algorithm

11.2.5 Cohort Exclusion Algorithm

An inpatient admission does not qualify as an index admission if the primary reason for the index admission was related to cancer, rehabilitation, or mental health. These categories are defined in the HWR Cohort Exclusion terminology dataset (terminology dataset #3). In order to use this dataset, first map all primary ICD-10-CM codes to the Custom CCS mapping dataset and then lookup whether each inpatient admission had any of these CCS condition categories.

Inpatient admissions that do not map to any of these CCS condition categories pass this criteria for inclusion as an index admission.

11.2.6 Discharged Alive and Not Against Medical Advice

In order to qualify as an index admission the inpatient admission must not have a discharge status of 20 ('expired') or 07 ('left against medical advice').

11.2.7 Not a Transfer

Inpatient admissions that result in the patient being transferred to another acute care facility do not count as index admissions. However, the subsequent transfer does count as an index admission if:

- The patient is discharged alive and not against medical advice
- The patient is discharged to a non-acute care setting (e.g. home, SNF, etc.)

Importantly, the transfer does not necessarily have to meet the other index admission criteria. For example, the transfer does not have to pass the cohort inclusion algorithm as long as the initial admission does.

A transfer is defined as occurring if a patient is discharged from an acute care hospital and admitted to another acute care hospital on the same day or the following day. No other criteria such as discharge status or admit source is required to define a transfer.

11.2.8 Not a Same-day Readmission

An inpatient admission is not considered an index admission if the patient is readmitted to the same hospital on the same day they were discharged from a previous admission for the same condition. In this situation, the readmission itself qualifies as the index admission.

11.2.9 Run-out

In order to qualify as an index admission, the admission must occur at least 30 days before the last date of discharge in the dataset.

For example, suppose your dataset contains inpatient admissions that occurred in calendar year 2018. Without this rule, admissions in this dataset that occurred on the last date in the dataset (i.e. December 31, 2018) could be flagged as index admissions but would never have an associated readmission because the data does not exist. As a result the readmission measure for the month of December 2018 would be artificially low (there would be a typical number of index admissions but fewer than typical readmissions simply because the data does not exist).

11.2.10 Planned Admission Algorithm

The planned readmission algorithm is used to exclude planned admissions from being flagged as readmissions in the Hospital-wide Measure. Terminology datasets #4-7 are used in this algorithm.

Here are the steps to implement the algorithm:

1. This algorithm requires mapping the primary ICD-10-CM code to the custom CCS diagnosis categories and all the ICD-10-PCS codes to the custom CCS procedure categories for each inpatient admission (terminology datasets #8 and #9).
2. For each inpatient admission, check whether the CCS diagnosis and procedure categories fall under the "always planned" list of diagnoses and procedures (terminology datasets #4 and #5). If any match occurs, this is a planned admission.
3. For each inpatient admission, check whether the CCS procedure categories and ICD-10-PCS codes fall under the list of potentially planned procedures (terminology dataset #6). If any match occurs, check whether this patient had any primary ICD-10-CM code or CCS diagnosis category that is considered an "always acute" code (terminology dataset #7). If the answer is no, this is a planned admission.
4. Inpatient admissions that do not qualify as planned admissions based on steps #2 and #3 above are considered unplanned.

Here is a diagram of the algorithm:

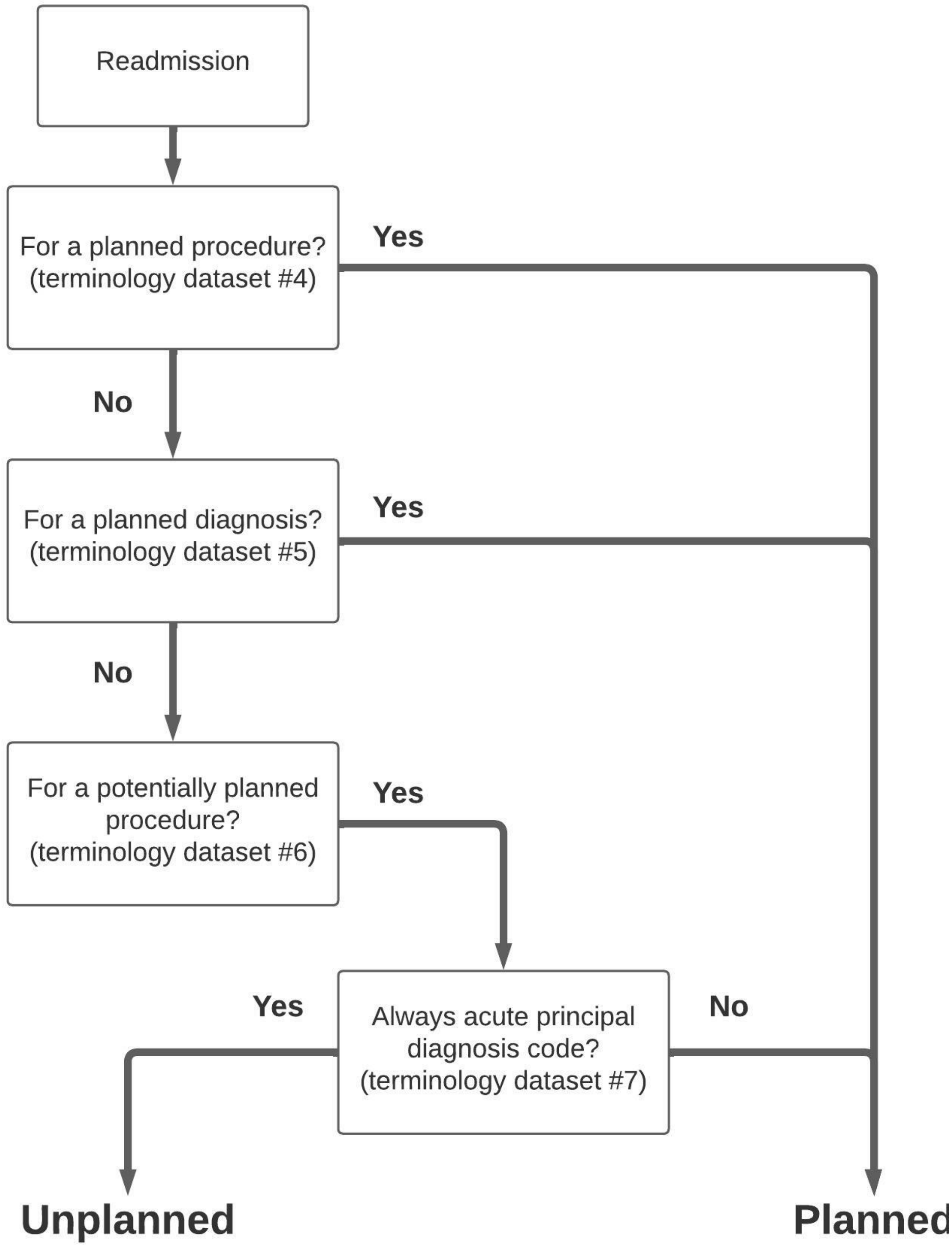


Figure 11.5: Planned Admission Algorithm

11.2.11 Unplanned Readmission Algorithm

An admission that occurs within 30 days of an index admission is considered a readmission if it meets the criteria below:

- Not a Planned Readmission
- Does Not Follow a Planned Readmission
- Not a Multiple Readmission

Planned readmissions are excluded from the readmission measure, as described in the previous section.

In a chain of readmissions, where the planned readmission occurs before an unplanned readmission, the unplanned readmission does not qualify as a readmission.

Figure 11.6: Unplanned Readmission 1

In a chain of readmissions, where two or more unplanned readmissions follow an index admission, only the first unplanned readmission qualifies as a readmission. This is because the readmission flag is binary.

Figure 11.7: Unplanned Readmission 2

12 Risk Adjustment

12.1 What is risk adjustment?

Risk adjustment was developed and created to dissuade payers from selectively enrolling patients who are healthy into health plans and make sure payers are compensated fairly for the illness burden of patients. Risk adjustment models are used to calculate payments to healthcare organizations for patients insured by: * Medicare Advantage * Accountable Care Organizations * Most Affordable Care Act plans. [1]

At a high level, if a patient is sicker or has higher predicted medical cost, the payer insuring that patient will be compensated more.

A Risk Adjustment Factor Score ("RAF score") is a measure of the estimated cost of an individual's care based on demographic factors such as age and gender and their associated disease factors (how costly/sick is the individual patient). This RAF score is then used to calculate payments or reimbursements to these organizations.

There are multiple risk adjustment models that exist for both Medicare patients and commercial patients. The most common risk adjustment models are:

- HHS-HCC
- CMS-HCC NON-ESRD
- CMS-HCC ESRD
- RxHCC
- PACE [2]

Each risk adjustment model has in common that it tries to predict medical expense using demographic information and disease factors for patients.

12.2 What are HCCs?

"HCC" stands for Hierarchical Condition Categories. They are sets of medical codes linked to clinical diagnoses that aim to predict cost for acute and chronic health conditions. HCC coding relies on ICD-10-CM coding to assign risk scores to patients. Not all risk adjustment models have to use HCCs, but almost all do.

12.3 Comparison between HHS-HCC and CMS-HCC

One way to separate HHS and CMS HCC models would be to call them respectively "commercial risk adjustment" (HHS) and "Medicare risk adjustment" (CMS). These two main models are similar but are separate:

CMS-HCC Model	**HHS-HCC Model**
Used for Medicare	Used for Affordable Healthcare Act (ACA) plans
Base year informs next year's payment	Current year diagnoses inform current year payment
Focused on patients 65 and older, but can include younger disabled population	Focused on all ages
Drug costs are carved out in separate model	Drug costs are included
Has distinct model and break out for ESRD patients	Does not carve out ESRD in separate risk adjustment model
Is not a zero-sum payment system	Is zero-sum between payers that participate in ACA
Has Normalization Factor and Coding Pattern Difference Adjustment	Has final adjustment for CSR benefit curve and utilization differences

12.4 FAQ

12.4.0.1 What does zero-sum mean?

Zero-sum in the context of commercial risk adjustment (HHS-HCC Model) means that there is no government subsidization for private companies in the risk adjustment process. If the illness burden of a state in total increases from the prior year, the government provides no additional funds. What happens during reconciliation is some health insurance companies will have to pay into a pool because they have healthier than average patients. Some organizations will get paid out from that pool because they have sicker than average patients. The total transfer of dollars adds up to $0.

12.4.0.2 If risk scores inform payments, how are risk scores checked to make sure they are not overinflated?

Both HHS-HCC and CMS-HCC models go through a validation process which is called "Risk Adjustment Data Validation" (aka RADV). The RADV processes and procedures are different for different risk adjustment models.

12.4.0.3 Can you compare risk scores between years to see the illness burden of a patient change over time?

This can be done, but is not recommended, especially if using different risk adjustment models between different years. There may be other models that can more accurately reflect illness burden, and risk adjustment is mainly focused on predicted medical expenditure. These two things are correlated, but not necessarily the same.

12.4.1 Footnotes

[1] ACA health plans that are not subject to risk adjustment are Grandfathered plans, Short-term, limited-duration plans, Health Reimbursement Arrangements (HRAs), and plans offered by states that have received a waiver from HHS.

[2] Program of All Inclusive Care for the Elderly

12.5 CMS Risk Score Models

CMS has implemented multiple models to address differences in program costs and the beneficiary population. For example, Medicare Part C versus Medicare Part D plans, the ESRD population (End-Stage Renal Disease) versus members without ESRD, or members enrolled in the Program of All Inclusive Care for the Elderly (PACE). CMS also segments each model, creating subpopulations with distinct cost patterns, such as dual enrollment in Medicaid or living in the community versus an institution.

The full list of models are:

- CMS-HCC
- CMS-HCC-ESRD
- PACE
- RxHCC

In addition to models for different populations, CMS has released versions of these models over the years, which include new HCC mappings and added or removed ICD-10-CM codes. Version 24 has been in use since 2020. Version 28 will be phased in over a three-year period starting in 2024.

- Payment year 2024 risk scores will be blended using 67% of the risk score calculated from v24 and 33% from v28.
- Payment year 2025 risk scores will be blended using 33% of the risk score calculated from v24 and 67% from v28.
- Beginning in payment year 2026 risk scores will be 100% from v28.

CMS also performs model calibration based on diagnostic and expenditure data. These changes can be found in the annual rate announcements on cms.gov.

These models generate risk scores by adding relative risk factors, demographics, and disease information. Additionally, they use hierarchies where the most severe manifestation of a condition is considered for risk scores.

12.6 Risk Score Calculation

Several resources are needed to calculate risk scores.

- Annual Rate Announcements for the applicable payment year (found on cms.gov)
- ICD-10-CM to HCC mapping (found on cms.gov)
- Risk adjustment model software (CMS makes a SAS program available on cms.gov)
- Model Output Report (MOR) or claims data
- Monthly Membership Detail Report or eligibility data

Once you have gathered the resources needed to calculate risk scores, you can begin identifying and calculating scores for your patients. A brief overview of the steps:

1. Identify demographic and enrollment information for each patient and cross-reference the risk factor value from the appropriate payment year's rate announcement document.
2. Identify disease information for each patient, apply the condition hierarchy, and cross-reference the risk factor value from the appropriate payment year's rate announcement document.
3. Identify additional relative and adjustment factors, such as disease and disabled interactions, and total HCC counts per patient, and cross-reference the risk factor value from the appropriate payment year's rate announcement document.
4. Calculate the raw risk scores for each patient, then apply the normalization factors and the MA coding pattern adjustment factors from the appropriate payment year's rate announcement document to calculate the normalized and payment risk scores.

12.7 CMS Risk Adjustment Files

The Monthly Membership Detail Report (MMR) and Model Output Report (MOR) are two types of files that CMS sends to Medicare Advantage organizations (MAOs). The files, in additional to other resources from CMS, are used to calculate the CMS HCC risk adjustment factor (RAF) scores.

CMS shares these files through the Medicare Advantage Prescription Drug (MARx) system and through the Health Plan Management System (HPMS).

MAOs can refer to the Plan Communications User Guide (PCUG) for additional details on the files exchanged through the MARx system.

12.7.1 Monthly Membership Detail Report (MMR)

The MMRs contain member eligibility, risk scores, and prospective payments the MAO receives for each member in the upcoming month. They also contain retroactive adjustments to prior months' records. This file contains the data for both Part C and Part D members. The key pieces of data to get from the MMR are:

- **Eligibility**: Understanding eligibility is a pre-requisite for modeling risk adjustment. The MMR has very detailed information regarding eligibility, including retrospective enrollments and disenrollment data. Just because a member exists in the MMR file, does not mean that they are eligible and should be included in risk adjustment.
- **Segment / Risk Model**: Depending on the risk adjustment model being used (most commonly v24 and v28), different co-efficients are used for the member based on if they are newly enrolled, are dual status, and their medicare status code. This information needs to be gathered from the MMR to calculate risk adjustment
- **Risk Scores**: The MMR will also disclose risk scores for a patient but only for certain time periods (start of the year, mid-year, and final). This can be used to understand funded premium, or validate calculated risk scores.

12.7.2 MAO-004

Not all medical claims are eligible for documenting HCCs for medicare risk adjustment. To address this, CMS provides a report called the MAO-004 which will inform Medicare Advantage organizations if a given diagnosis code submitted is eligible for risk adjustment. This file can be helpful for tying out risk score calculations to the MMR.

12.7.3 Model Output Report (MOR)

Similar to MMRs, the MORs contain a record for each member. That record shows the Hierarchical Condition Codes (HCCs) for each member used by the Risk Adjustment System (RAS) to calculate Part C or Part D risk adjustment factors for each beneficiary. There are two varieties of MORs, for Parts C and D respectively, as each uses different models.

In addition to these monthly files, CMS issues "final" MORs once per year with updated information after the year has ended, and planned runout data has been collected.

12.8 Sample Risk Score Calculation of a single patient

Let's walk through a single example with a single patient as to how the risk score is calculated in the context of medicare advantage for the year 2024.

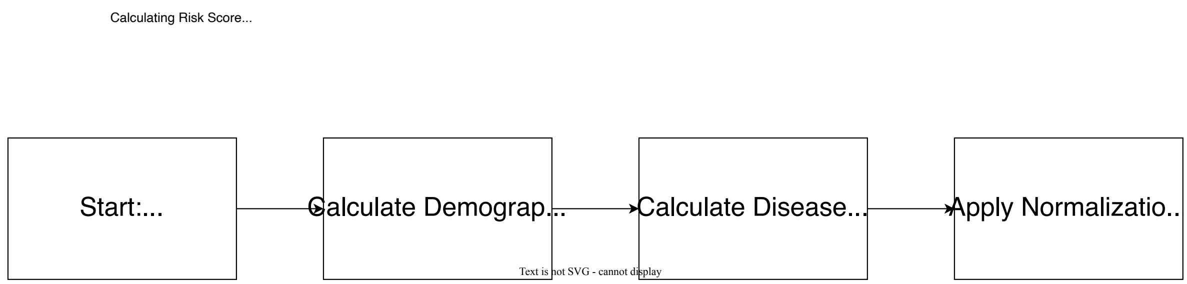


Figure 12.1: Overview

Here's some information about this patient.

- Patient is Female (BENE_SEX_CD = 2)
- Patient is 76 years old
- Patient is a partial dual patient (DUAL_ELGBL_CD_01 = 03)
- Patient is aged without ESRD (MDCR_STATUS_CODE_01 = 10)
- Patient was originally disabled (MDCR_OREC = 1)
- Patient has the following diagnosis documented in medical claims submitted to cms:

Diagnosis Code	Description
E10641	Type 1 diabetes mellitus with hypoglycemia with coma
E083293	Diabetes mellitus due to underlying condition with mild nonproliferative diabetic retinopathy without macular edema, bilateral
E139	Other specified diabetes mellitus without complications
E139	Other specified diabetes mellitus without complications

Walking through the steps listed above:

12.8.1 Calculate the demographic score

To calculate the demographics portion of the risk score, we need to look at the demographics information for the patient provided above. Let's take a look at a table from the 2024 final rule that contains the raw factors related to demographics.

Attachment VIII. CMS-HCC Risk Adjustment Factors & Predictive Ratio Tables

Table VIII-1. 2024 CMS-HCC Model Relative Factors for Continuing Enrollees

Variable	Description Label	Community, NonDual, Aged	Community, NonDual, Disabled	Community, FBDual, Aged	Community, FBDual, Disabled	Community, PBDual, Aged	Community, PBDual, Disabled	Institutional
Female								
0-34 Years		-	0.238	-	0.346	-	0.454	0.948
35-44 Years		-	0.288	-	0.332	-	0.420	0.810
45-54 Years		-	0.340	-	0.384	-	0.404	1.031
55-59 Years		-	0.385	-	0.421	-	0.424	0.949
60-64 Years		-	0.436	-	0.502	-	0.414	0.881
65-69 Years		0.330	-	0.435	-	0.365	-	1.188
70-74 Years		0.395	-	0.506	-	0.423	-	1.119
75-79 Years		0.465	-	0.596	-	0.485	-	0.965
80-84 Years		0.524	-	0.665	-	0.544	-	0.862
85-89 Years		0.624	-	0.775	-	0.618	-	0.750
90-94 Years		0.737	-	0.869	-	0.738	-	0.627
95 Years or Over		0.742	-	0.877	-	0.835	-	0.481
Male								
0-34 Years		-	0.106	-	0.191	-	0.306	0.826
35-44 Years		-	0.154	-	0.204	-	0.261	0.719
45-54 Years		-	0.215	-	0.293	-	0.300	0.991
55-59 Years		-	0.283	-	0.410	-	0.353	0.989
60-64 Years		-	0.345	-	0.504	-	0.374	0.917
65-69 Years		0.332	-	0.531	-	0.375	-	1.275
70-74 Years		0.396	-	0.626	-	0.417	-	1.224
75-79 Years		0.502	-	0.714	-	0.498	-	1.319
80-84 Years		0.571	-	0.789	-	0.565	-	1.238
85-89 Years		0.664	-	0.907	-	0.615	-	1.135
90-94 Years		0.800	-	0.993	-	0.712	-	0.946
95 Years or Over		0.896	-	1.058	-	0.904	-	0.825
Medicaid and Originally Disabled Interactions								
Originally Disabled, Female		0.228	-	0.160	-	0.103	-	-
Originally Disabled, Male		0.135	-	0.158	-	0.075	-	-
Medicaid		-	-	-	-	-	-	0.130

Figure 12.2: 2024_Final_Rule_Demographics_Table

- The demographic score for a female patient 76 years of age with partial dual status is 0.485.
- Given the patient has been originally disabled, they get an additional 0.103.

The final raw risk from demographics is $(0.485 + 0.103) = \mathbf{0.588}$

If the patient was an end stage renal disease (ESRD) patient, we would use a separate demographics table that uses the ESRD risk adjustment model.

12.8.2 Calculate the Disease Score

The disease score can be sourced from multiple places, either the MOR or claims data in combination with MAO-004 report. This example will be looking at calculating risk from claims data.

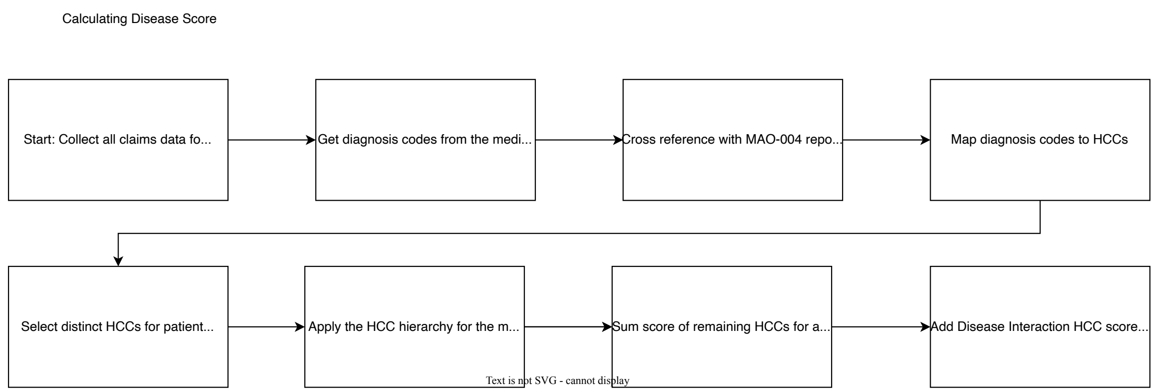
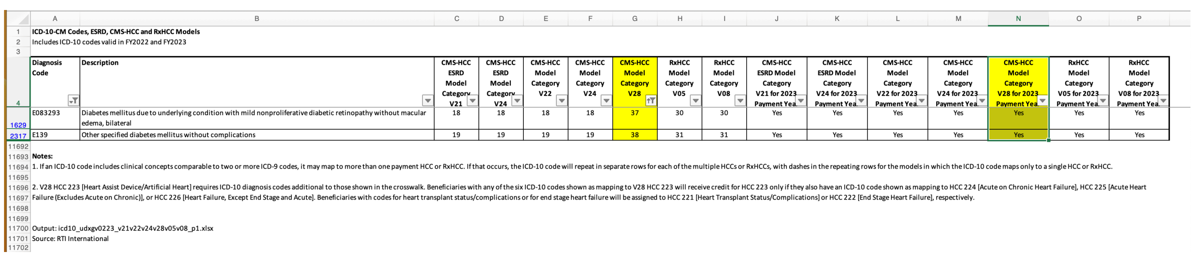


Figure 12.3: disease_score_calculation

In the sample patient, the diagnosis codes for that patient are provided above from claims data. Not all diagnosis are accepted for risk adjustment, so in this example we will say the diagnosis E10.641, (Type 1 diabetes mellitus with hypoglycemia with coma) is not accepted when checking the MAO-004. That leaves two diagnosis of E08.3293 and E13.9. Even though E13.9 is present twice in claims data, having more than one of the same accepted diagnosis code is the same as having a single instance of that diagnosis code being accepted.

Next we need to cross-reference the diagnosis codes to get the HCCs for the model. The crosswalk between diagnosis codes and HCCs can be found here under "2024 Initial ICD-10 Mappings".



ICD-10-CM Codes, ESRD, CMS-HCC and RxHCC Models

Includes ICD-10 codes valid in FY2022 and FY2023

Diagnosis Code	Description	CMS-HCC ESRD Model Category V21	CMS-HCC ESRD Model Category V24	CMS-HCC Model Category V22	CMS-HCC Model Category V24	CMS-HCC Model Category V28	RxHCC Model Category V05	RxHCC Model Category V08	CMS-HCC ESRD Model Category V21 for 2023 Payment Yea.	CMS-HCC ESRD Model Category V24 for 2023 Payment Yea.	CMS-HCC Model Category V22 for 2023 Payment Yea.	CMS-HCC Model Category V24 for 2023 Payment Yea.	CMS-HCC Model Category V28 for 2023 Payment Yea.	RxHCC Model Category V05 for 2023 Payment Yea.	RxHCC Model Category V08 for 2023 Payment Yea.
E083293	Diabetes mellitus due to underlying condition with mild nonproliferative diabetic retinopathy without macular edema, bilateral	18	18	18	18	37	30	30	Yes	Yes	Yes	Yes	Yes	Yes	Yes
E139	Other specified diabetes mellitus without complications	19	19	19	19	38	31	31	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes:

1. If an ICD-10 code includes clinical concepts comparable to two or more ICD-9 codes, it may map to more than one payment HCC or RxHCC. If that occurs, the ICD-10 code will repeat in separate rows for each of the multiple HCCs or RxHCCs, with dashes in the repeating rows for the models in which the ICD-10 code maps only to a single HCC or RxHCC.

2. V28 HCC 223 [Heart Assist Device/Artificial Heart] requires ICD-10 diagnosis codes additional to those shown in the crosswalk. Beneficiaries with any of the six ICD-10 codes shown as mapping to V28 HCC 223 will receive credit for HCC 223 only if they also have an ICD-10 code shown as mapping to HCC 224 [Acute on Chronic Heart Failure], HCC 225 [Acute Heart Failure (Excludes Acute on Chronic)], or HCC 226 [Heart Failure, Except End Stage and Acute]. Beneficiaries with codes for heart transplant status/complications or for end stage heart failure will be assigned to HCC 221 [Heart Transplant Status/Complications] or HCC 222 [End Stage Heart Failure], respectively.

Output: icd10_udxgv0223_v21v22v24v28v05v08_p1.xlsx

Source: RTI International

Figure 12.4: diagnosis_to_hcc_crosswalk_example

When looking up the two diagnosis codes in the crosswalk, we see they are both valid and map to two HCCs for the 2024 v28 risk adjustment model (37 and 38).

Next we need to check the hierarchy to drop HCCs that exist within the hierarchy. This hierarchy exists within the announcement document.

Table VIII-4. 2024 CMS-HCC Model with Disease Hierarchies

CMS-HCC	If the Disease Group is listed in this column…	…Then drop the CMS-HCC listed in this column
	CMS-HCC Hierarchical Condition Category Label	
17	Cancer Metastatic to Lung, Liver, Brain, and Other Organs; Acute Myeloid Leukemia Except Promyelocytic	18, 19, 20, 21, 22, 23
18	Cancer Metastatic to Bone, Other and Unspecified Metastatic Cancer; Acute Leukemia Except Myeloid	19, 20, 21, 22, 23
19	Myelodysplastic Syndromes, Multiple Myeloma, and Other Cancers	20, 21, 22, 23
20	Lung and Other Severe Cancers	21, 22, 23
21	Lymphoma and Other Cancers	22, 23
22	Bladder, Colorectal, and Other Cancers	23
35	Pancreas Transplant Status	36, 37, 38
36	Diabetes with Severe Acute Complications	37, 38
37	Diabetes with Chronic Complications	38
62	Liver Transplant Status/Complications	63, 64, 65, 68
63	Chronic Liver Failure/End-Stage Liver Disorders	64, 65, 68, 202
64	Cirrhosis of Liver	65, 68
77	Intestine Transplant Status/Complications	78, 80, 81
80	Crohn's Disease (Regional Enteritis)	81
93	Rheumatoid Arthritis and Other Specified Inflammatory Rheumatic Disorders	94
107	Sickle Cell Anemia (Hb-SS) and Thalassemia Beta Zero	108
111	Hemophilia, Male	112
114	Common Variable and Combined Immunodeficiencies	115
125	Dementia, Severe	126, 127
126	Dementia, Moderate	127
135	Drug Use with Psychotic Complications	136, 137, 138, 139
136	Alcohol Use with Psychotic Complications	137, 138, 139
137	Drug Use Disorder, Moderate/Severe, or Drug Use with Non-Psychotic Complications	138, 139
138	Drug Use Disorder, Mild, Uncomplicated, Except Cannabis	139
151	Schizophrenia	152, 153, 154, 155
152	Psychosis, Except Schizophrenia	153, 154, 155
153	Personality Disorders; Anorexia/Bulimia Nervosa	154, 155
154	Bipolar Disorders without Psychosis	155
180	Quadriplegia	181, 182, 253, 254
181	Paraplegia	182, 254
191	Quadriplegic Cerebral Palsy	180, 181, 182, 192, 253, 254
192	Cerebral Palsy, Except Quadriplegic	180, 181, 182, 253, 254
195	Myasthenia Gravis with (Acute) Exacerbation	196
211	Respirator Dependence/Tracheostomy Status/Complications	212, 213
212	Respiratory Arrest	213

Figure 12.5: hcc_hierarchy_2024_final_rule

Based on this table, we see that HCC-37 is on the left hand side "If the Disease Group is listed in this column…" and on the right hand side "…Then drop the CMS-HCC listed in this column" there is a match on HCC-38. This means that we drop the HCC-38 and are left with a single remaining HCC (HCC-37) for this patient.

Once we have our remaining HCCs after the hierarchy is applied, we need to find the score related to HCC-37 for this single patient example. If there was more than one HCC remaining, the values would be summed,

Variable	Description Label	Community, NonDual, Aged	Community, NonDual, Disabled	Community, FBDual, Aged	Community, FBDual, Disabled	Community, PBDual, Aged	Community, PBDual, Disabled	Institutional
Disease Coefficients								
HCC1	HIV/AIDS	0.301	0.213	0.397	0.237	0.196	0.109	1.322
HCC2	Septicemia, Sepsis, Systemic Inflammatory Response Syndrome/Shock	0.500	0.598	0.649	0.780	0.447	0.591	0.605
HCC6	Opportunistic Infections	0.381	0.763	0.588	0.833	0.518	0.685	0.728
HCC17	Cancer Metastatic to Lung, Liver, Brain, and Other Organs; Acute Myeloid Leukemia Except Promyelocytic	4.209	3.995	3.896	4.235	3.946	4.103	1.952
HCC18	Cancer Metastatic to Bone, Other and Unspecified Metastatic Cancer; Acute Leukemia Except Myeloid	2.341	2.486	2.277	2.537	2.166	2.403	1.110
HCC19	Myelodysplastic Syndromes, Multiple Myeloma, and Other Cancers	1.798	1.989	1.563	1.661	1.520	1.554	0.957
HCC20	Lung and Other Severe Cancers	1.136	0.978	1.166	1.173	1.214	1.067	0.672
HCC21	Lymphoma and Other Cancers	0.671	0.540	0.654	0.739	0.627	0.618	0.493
HCC22	Bladder, Colorectal, and Other Cancers	0.363	0.366	0.382	0.409	0.410	0.351	0.314
HCC23	Prostate, Breast, and Other Cancers and Tumors	0.186	0.233	0.196	0.218	0.203	0.237	0.197
HCC35	Pancreas Transplant Status	0.949	1.393	1.117	0.573	1.117	2.740	1.106
HCC36	Diabetes with Severe Acute Complications	0.166	0.191	0.186	0.235	0.166	0.210	0.280
HCC37	Diabetes with Chronic Complications	0.166	0.191	0.186	0.235	0.166	0.210	0.280
HCC38	Diabetes with Glycemic, Unspecified, or No Complications	0.166	0.191	0.186	0.235	0.166	0.210	0.280
HCC48	Morbid Obesity	0.186	0.144	0.300	0.178	0.164	0.118	0.442
HCC49	Specified Lysosomal Storage Disorders	9.256	13.778	2.833	6.399	3.269	7.771	1.528

Figure 12.6: disease_coefficients_v28_2024_announcement

For this patient the score is **0.166**.

Next we need to evaluate disease interactions. Since we are left with only a single HCC, disease interactions don't apply for this specific example patient. However, please see below for the disease interactions that exist within the v28 model. This is also in the announcement document.

Variable	Description Label	Community, NonDual, Aged	Community, NonDual, Disabled	Community, FBDual, Aged	Community, FBDual, Disabled	Community, PBDual, Aged	Community, PBDual, Disabled	Institutional
HCC454	Stem Cell, Including Bone Marrow, Transplant Status/Complications	1.068	0.452	1.326	0.608	1.338	0.416	1.596
HCC463	Artificial Openings for Feeding or Elimination	0.673	0.914	0.891	0.947	0.526	0.853	0.634
Disease Interactions								
DIABETES_HF	Diabetes*Heart Failure	0.112	0.023	0.183	0.041	0.164	0.053	0.209
HF_CHR_ LUNG	Heart Failure*Chronic Lung Disorder	0.078	0.062	0.109	0.097	0.140	0.108	0.145
HF_KIDNEY	Heart Failure*Kidney	0.176	0.314	0.194	0.420	0.140	0.328	-
CHR_LUNG_CARD _RESP_FAIL	Chronic Lung Disorder*Cardiorespiratory Failure	0.254	0.242	0.340	0.275	0.329	0.270	0.331
HF_HCC238	Heart Failure*Specified Heart Arrhythmias	0.077	0.257	0.140	0.372	0.135	0.314	-
gSubUseDisorder_ gPsych_	Substance Use Disorder*Psychiatric	-	0.087	-	0.152	-	0.149	-
Disabled/Disease Interactions								
DISABLED_ HF	Disabled, Heart Failure	-	-	-	-	-	-	0.488
DISABLED_ ULCER_	Disabled, Skin Ulcer	-	-	-	-	-	-	0.537
DISABLED_ CANCER	Disabled, Cancer	-	-	-	-	-	-	0.367
DISABLED_ NEURO_	Disabled, Neurological	-	-	-	-	-	-	0.154
DISABLED_ CHR_LUNG	Disabled, Chronic Lung Disorder	-	-	-	-	-	-	0.278

NOTES:

1. The denominator used is $10,402.34.
2. In the "disease interactions" and "disabled interactions," the variables are defined as follows:
   Cancer = HCCs 17-23
   Cardiorespiratory Failure = HCCs 211-213
   Chronic Lung Disorder = HCCs 276-280
   Diabetes = HCCs 35-38
   Heart Failure = HCCs 221-226
   Kidney = HCCs 326-329
   Neurological = HCCs 180-192, 195, 196, 198, 199
   Psychiatric = HCCs 151-155
   Skin Ulcer = HCCs 379-382
   Specified Heart Arrhythmias = HCC 238
   Substance Use = HCCs 135-139

SOURCE: 2018-2019 100% Medicare data.

Finally, we need to count the number of HCCs remaining after the application of the hierarchy. In this example, we only have a single HCC, so there is no additional score applied.

Variable	Description Label	Community, NonDual, Aged	Community, NonDual, Disabled	Community, FBDual, Aged	Community, FBDual, Disabled	Community, PBDual, Aged	Community, PBDual, Disabled	Institutional
Payment HCC Counts								
D1	1 payment HCCs	-	-	-	-	-	-	-
D2	2 payment HCCs	-	-	-	-	-	-	-
D3	3 payment HCCs	-	-	-	-	-	-	-
D4	4 payment HCCs	-	-	-	-	-	-	-
D5	5 payment HCCs	0.050	0.088	0.049	0.095	0.016	0.105	-
D6	6 payment HCCs	0.102	0.223	0.071	0.245	0.096	0.191	-
D7	7 payment HCCs	0.188	0.380	0.160	0.472	0.207	0.435	-
D8	8 payment HCCs	0.316	0.440	0.267	0.607	0.345	0.581	-
D9	9 payment HCCs	0.444	0.750	0.353	0.841	0.345	0.823	-
D10P	10 or more payment HCCs	0.728	1.431	0.746	1.471	0.901	1.268	0.373

192

Figure 12.7: hcc_counts_2024_announcement

12.8.3 Bringing it all together

We sum both the demographic score and the disease score to get the final raw raf for the patient. (0.588 + 0.166) = **0.754**. This score is the raw risk score for the patient. To get the final risk score for a patient, the formula is (raw_risk_score / normalization_factor).

For 2024, the CMS-HCC risk adjustment model normalization factor is **1.015** meaning. For **medicare advantage** organizations, another Coding Pattern Difference Adjustment (aka Coding Intensity Factor CIF) of 5.9% should be applied on top of the normalization factor.

So the final risk score for this single patient would be (0.754 / 1.015) * (1 - 0.059) = **0.699**

12.8.4 But wait! There's more

In the above example, it only looked at the scores and weights for a single risk adjustment model, v28. However, for medicare advantage organizations in the year 2024, final funded risk is not based solely on the outputs of the v28 risk adjustment model. There is a transition period where risk will be determined with a blended model, where 33% of the risk score will be weighted with the v28 model and 67% of the risk score will be weighted with the v24 risk adjustment model.

What does this mean? This means we have to go back and repeat the steps prior to "Bringing it all together" for the v24 risk adjustment model, then apply the 33% and 67% weighting for v28 and v24 risk scores respectively, then apply normalization factor and CIF to get the final risk score for the patient.

12.8.5 Additional notes

- In this specific example, this was looking at the risk adjustment model for Medicare Advantage. Different programs and different use cases can use different risk adjustment models.
- Different years going forward (2025 and 2026) have different weighting of the v24 vs v28 risk adjustment models.
- This can all be subject to change if there is new legislation or final rules for 2025 and 2026.
- In the context of some CMMI programs, the terms of "coding intensity factor" and "normalization factor" can seem to be the same as the Medicare Advantage definitions, but can be derived in different ways specific to that program.

12.8.6 References

- https://www.milliman.com/en/insight/medicare-advantage-and-the-encounter-data-processing-system-be-prepared
- https://www.cms.gov/files/document/2024-advance-notice-pdf.pdf

Knuth, Donald E. 1984. "Literate Programming." *Comput. J.* 27 (2): 97–111. https://doi.org/10.1093/comjnl/27.2.97.