Medicare Advantage Encounter Data

Methodological Approach for Analyzing Institutional and Non-Institutional Services

Centers for Medicare and Medicaid Services
Office of Enterprise Data and Analytics

July 10, 2024

Introduction

Currently, nearly half of Medicare beneficiaries are covered by a Medicare Advantage plan, and this figure is expected to continue to grow. It's increasingly important that CMS develop and use available data sources to understand care, track disparities, and drive decisions for this population. The Agency has a wealth of internal analytic capability and public products that focus on fee-for-service; it's critical that we develop similar resources for Medicare Advantage encounter data.

OEDA has developed an analytical approach for different types of institutional and non-institutional providers in the Medicare Advantage (MA) encounter data – namely inpatient hospitals, skilled nursing facilities (SNF), outpatient facilities, professional or carrier services, and durable medical equipment (DME).[1] We are using this analytical approach to allow us to include metrics related to services delivered to MA beneficiaries that are consistent with metrics reported in our existing fee-for-service (FFS) data products. Currently, OEDA is focused on generating utilization metrics in the MA population for these service types.[2]

This document describes the methods used to analyze and expand on certain data elements included in the encounter data. Specifically, it covers the following topics:

- The methods used to edit and impute encounter National Provider Identifiers (NPI) variables which are used to identify duplicate encounter records as well as attach CMS Certification Number (CCN) provider IDs to inpatient hospital encounter records.
- The methods used to identify and address duplicate encounter records to produce more accurate MA encounter service counts.
- The algorithm generated to assign the CCN provider ID to inpatient hospital encounter records to categorize these records by inpatient hospital type.[3] This algorithm allows us to use the MA data in a way that is consistent with how FFS inpatient hospital records are classified, and will help ensure that comparisons (e.g., utilization metrics) between the FFS and MA populations are appropriate.

We are providing this document so that researchers with access to MA Research Identifiable Files (RIFs) may better understand our analytic approach. We expect to make enhancements to our approach, and we will continue to publish updates to allow collaboration with a wide range of encounter data users. If you have specific feedback on the information provided, please contact PDAG_Data_Products@cms.hhs.gov.

Section I. NPI Edit and Imputation Methods

The first major data issue to address in the development of analytical MA encounter data files includes editing the different NPI variables that exist on each of the file types. The NPI variable, along with other key service variables, is used to identify duplicate encounter records according to the Encounter Data Processing System (EDPS)

[1] All of these encounter data files are available through the Research Data Assistance Center (ResDAC). For more information on how to access these data, visit https://resdac.org/ (accessed 05/07/2024). OEDA does not include the home health encounter services in this paper because we have not evaluated these data yet.

[2] Information on the various encounter data files can be found here: https://resdac.org/cms-data?tid%5B6056%5D=6056 (accessed05/07/2024).

[3] The MA inpatient hospital encounter data file includes short-term hospital, long-term care hospital, inpatient rehabilitation facility, inpatient psychiatric facility, and other specialty hospital services.

manua.l.[4] These key variables, referred to in this paper as duplicate record identifying (DRI) key groups, are discussed in more detail in Section II. The NPI is also used in the inpatient hospital data file to attach the CCN codes to these records as described in Section III.

The NPI is a 10-digit unique provider identifier that health care institutional and individual providers covered by the 1996 Health Insurance Portability and Accountability Act (HIPPA) must use for administrative and financial transactions. CMS assigns NPIs to institutional and non-institutional providers when they register with the agency's National Plan and Provider Enumeration System (NPPES). The NPI is intelligence-free in that the identifier does not indicate any information on provider type (e.g., short stay hospital, critical access hospital, rehab).

NPI Edits

Each MA encounter data file includes different NPI variables that are used to identify duplicate records. The inpatient hospital, SNF, and outpatient facility encounter data files only use the organizational NPI variable. The carrier and DME files use a combination of NPI variables to identify duplicate records. Specifically, the EDPS manual states that the system creates a unified NPI variable that reflects the line-level rendering NPI variable if it is reported, the base file rendering NPI variable if the line-level NPI variable is missing, and the base file organizational NPI if the first two variables are missing.

The first step in editing the NPI variables involves checking the fidelity of NPI variables that are used in each of the encounter files. Specifically, we check that the:

- NPI has a non-missing value.
- NPI value has a valid length and structure, i.e., it has 10 characters and begins with the number 1.
- NPI exists in the NPPES-based provider reference table.
- NPI's provider type reflects an organization rather than an individual for the organizational NPI variables.[5]

Regarding the last check for organizational NPIs, the CMS submission rules for encounter records[6] do not require plans to provide the organizational NPI directly responsible for the service represented by the encounter record when the Medicare Advantage Organization (MAO) has multiple NPIs on file for a given institution. Rather, CMS rules allow the plans to submit any NPI that they have on record as merely associated with a facility's internal MAO provider identification number. This guidance may lead MAOs to report NPIs associated with individual providers in the organizational NPI field on the encounter record. As for the other rendering NPI variables used in the base and line carrier and DME encounter data files, there are no requirements that the NPI reported must be a particular type (i.e., individual, organizational).

[4] https://www.csscoperations.com/internet/csscw3_files.nsf/F2/2022ED_Submission_Processing_Guide_20221130.pdf/$FILE/2022ED_Submission_Processing_Guide_20221130.pdf (accessed, 07/10/2024)

[5] We identify an NPI as an organizational or individual NPI using the entity type field in the NPPES tables that contain all NPIs assigned by CMS to providers.

[6] For additional details, download the zip file found here: https://downloads.cms.gov/files/2017-HPMS-Q4.zip (accessed 10/31/2022). After downloading, please refer to the subfolder labeled "2017-12-21 Memo re Encounter Data Record Submissions - NPI Submission Guidance - Frequently Asked Questions".

Any NPI variable value that does not meet the above criteria is recoded to missing. For the carrier and DME files, we then create a unified NPI variable using the edited NPI variables by coalescing valid NPI values using the variable order described above.

NPI Imputation Methods

We attempt to impute missing NPI values within the same DRI key group, excluding the NPI variable, by checking if there are other records within the key group with a ***single*** valid NPI. Figure 1 demonstrates this imputation method.

Figure 1. Illustration of NPI Imputation Method

DRI Key Group Except NPI	Edited NPI	Imputed NPI	Imputated NPI Flag
Scenario A			
1234	.	7	1
1234	.	7	1
1234	7	7	0
Scenario B			
3456	.	7	1
3456	.	7	1
3456	7	7	0
3456	7	7	0
Scenario C			
4567	.	.	0
4567	.	.	0
4567	7	7	0
4567	8	8	0
Scenario D			
5678	.	.	0
5678	.	.	0
5678	.	.	0

In all the above scenarios, all the records have the same values for all the non-NPI DRI key variables as demonstrated by the first column. In Scenario A, there are two records with missing values for the edited NPI variable and one record with a valid value (7). Since there is only a single valid NPI value, we can use this ID to impute a value to the two records that are missing an NPI value. Similarly, Scenario B has two records that have a missing NPI value and two records with a valid NPI value and that valid value is the same. Even though there are two records with a valid NPI value, since the NPI value is the same, that value can be used to impute an ID to the records with a missing NPI value. In Scenario C, however, there are two records with two different valid NPI values (7 and 8). Since we do not know which value should be imputed to the records with a missing NPI value, those missing values persist in the data. Lastly, Scenario D shows that there are no valid NPI values in the key group, so all values remain missing.

This method would treat records where all the values for the DRI key group variables excluding NPI are the same, but one record has a missing NPI value and another has a valid NPI value, as belonging to the same key group to be processed in the deduping process. If there are multiple valid NPIs within the key group, then the record with the invalid NPI remains missing and is treated as a unique encounter record.

One limitation of this NPI imputation method is that it does not resolve incorrect assignments of organizational NPIs to a particular encounter record by MAOs. For example, a MAO could erroneously assign the NPI of the

parent acute care hospital organization to the encounter record for an inpatient rehabilitation service instead of assigning the organizational NPI of the inpatient rehabilitation facility that performed the care.

Table 1 reports the distribution of MA encounter records by NPI value status for each of the encounter file types covered in this paper. The results show that for the institutional file types that only use organizational NPIs, i.e., inpatient hospital, SNF, and outpatient facility encounter files, the percentage of records with invalid NPIs is less than 0.5% between 2016 and 2021. By contrast, the carrier and DME encounter files that use a coalesced version of the rendering and organizational NPI variables have a significantly higher percentage of records with invalid NPI values, averaging 7.6% of records across the data years. However, the invalid NPI rate is declining over time going from 9.6% of records in 2016 to 5.9% of records in 2021. Across all the encounter files, the imputation process did not have a substantial impact on recoding missing NPI values with less than 0.2% of records with an imputed NPI value for all data years.

Table 1. Distribution of MA Encounter Records, by File Type and NPI Imputation Status

NPI Value Status	2016		2017		2018		2019		2020		2021	
	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%
Inpatient Hospitals												
Total Records	4.3	100.0%	4.7	100.0%	5.1	100.0%	5.7	100.0%	5.5	100.0%	6.3	100.0%
Invalid NPI Values	0.01	0.2%	0.01	0.1%	0.01	0.1%	0.01	0.2%	0.01	0.1%	0.01	0.2%
Valid NPI Values	4.3	99.8%	4.7	99.9%	5.1	99.9%	5.7	99.8%	5.5	99.9%	6.2	99.8%
Original	4.3	99.7%	4.7	99.8%	5.1	99.8%	5.6	99.7%	5.5	99.7%	6.2	99.6%
Imputed	0.003	0.1%	0.002	0.03%	0.002	0.03%	0.01	0.1%	0.01	0.1%	0.02	0.2%
Skilled Nursing Facilities												
Total Records	1.46	100.0%	1.62	100.0%	1.74	100.0%	1.85	100.0%	1.65	100.0%	1.95	100.0%
Invalid NPI Values	0.01	0.5%	0.01	0.5%	0.01	0.4%	0.01	0.5%	0.01	0.4%	0.01	0.4%
Valid NPI Values	1.45	99.5%	1.61	99.5%	1.74	99.6%	1.85	99.5%	1.64	99.6%	1.95	99.6%
Original	1.45	99.4%	1.61	99.5%	1.74	99.6%	1.84	99.5%	1.64	99.6%	1.95	99.6%
Imputed	0.001	0.1%	0.001	0.04%	0.001	0.07%	0.001	0.04%	0.0005	0.03%	0.001	0.04%
Outpatient Facilities												
Total Records	349.8	100.0%	393.8	100.0%	460.4	100.0%	537.5	100.0%	533.7	100.0%	688.7	100.0%
Invalid NPI Values	0.7	0.2%	0.3	0.1%	0.3	0.1%	0.7	0.1%	0.7	0.1%	1.0	0.2%
Valid NPI Values	349.1	99.8%	393.5	99.9%	460.1	99.9%	536.8	99.9%	533.0	99.9%	687.6	99.8%
Original	349.0	99.8%	393.4	99.9%	460.1	99.9%	536.2	99.8%	532.3	99.7%	686.4	99.7%
Imputed	0.1	0.03%	0.1	0.02%	0.0	0.01%	0.5	0.1%	0.7	0.1%	1.3	0.2%
Carrier and DME Services												
Total Records	972.6	100.0%	1,076.9	100.0%	1,221.9	100.0%	1,388.0	100.0%	1,375.9	100.0%	1,686.9	100.0%
Invalid NPI Values	93.8	9.6%	92.6	8.6%	96.6	7.9%	102.6	7.4%	87.0	6.3%	100.1	5.9%
Valid NPI Values	878.8	90.4%	984.3	91.4%	1,125.3	92.1%	1,285.4	92.6%	1,288.9	93.7%	1,586.8	94.1%
Original	877.2	90.2%	982.7	91.3%	1,123.5	91.9%	1,283.7	92.5%	1,287.2	93.6%	1,584.6	93.9%
Imputed	1.6	0.2%	1.6	0.1%	1.8	0.1%	1.7	0.1%	1.7	0.1%	2.2	0.1%

SOURCE: 2016-2021 Inpatient hospital, SNF, outpatient facility, carrier, and DME MA encounter base and line data files.

Section II. Unique Service Identification

Eliminating Duplicate Encounter Records

As previously mentioned, the EDPS manual specifies a DRI key group that is used to identify duplicate encounter records for each encounter file type. If the EDPS flags records with the same DRI key group values, those records

should be rejected by the processing system. This section discusses how we check for the existence of duplicate records based on the EDPS-defined DRI key group for the inpatient hospital and SNF encounter data and modify the DRI key groups for carrier, DME, and outpatient facility encounter data to include only variables that are made available to the external research community.

It should be noted that a unique encounter record does not necessarily correspond to a unique service but rather conveys distinct information about aspects of a service. Once the deduping process is applied, one encounter record may represent a unique service or event, but it does not necessarily have to. For example, there may be two encounter records for use of a drug where one line represents how much drug was used and another record represents how much drug was wasted. Both encounter records could represent a single event of using the drug depending on the specifications of a given analysis.

Lastly, eliminating duplicate encounter records only serves to ensure that utilization counts are more accurate. This data edit does do not resolve discrepancies across the duplicate service encounter records since we are not including any of these data elements in published reports of MA beneficiary utilization (e.g., two records with the same DRI key group value may contain different diagnosis code information). However, since OEDA does plan to expand the analytical files to use these additional fields, we want to take care to apply a reasonable strategy for how we select a single record from a DRI key group with multiple records.

Inpatient and SNF Duplicate Service Records

According to the Medicare Advantage Encounter Data User Guide[7], a unique encounter inpatient hospital and SNF base record is identified as the unique combination of beneficiary ID, encounter start date, encounter end date, type of bill, and organizational NPI. This DRI key is also known as the "5-part service key" in the encounter data user guide. To apply the 5-part service key, we use the imputed NPI variable discussed above that removes invalid and individual provider IDs from the organizational NPI column. Our analysis of the data identified a few thousand encounter records that had the same 5-part service key.

To address the issue of duplicate encounter records, OEDA selected the encounter record with the latest EDPS processing date to restore the one-to-one relationship between encounter record and the 5-part service key. If multiple records report the latest EDPS processing date within the DRI key group, then a record is selected at random among the subset of records with the latest processing date. Table 2 documents that we eliminated a small percentage of records (less than 0.2% of records) that indicated duplicate services based on the 5-part service key. The inpatient hospital and SNF encounter data have similar percentages for duplicate records removed from the data.

[7] The user guide can be found here: https://www2.ccwdata.org/documents/10280/19002246/ccw-medicare-encounter-data-user-guide.pdf (accessed 10/31/2022).

Table 2. Duplicate Inpatient Hospital and SNF MA Encounter Records Removed from Analytical Files

Record Count Category	2016		2017		2018		2019		2020		2021	
	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%
Inpatient Hospitals												
Original Record Count	4.28	100.0%	4.73	100.0%	5.10	100.0%	5.66	100.0%	5.51	100.0%	6.26	100.0%
Single Records per DRI Key	4.26	99.7%	4.72	99.7%	5.09	99.9%	5.65	99.7%	5.49	99.7%	6.23	99.5%
Multiple Records per DRI Key	0.01	0.3%	0.01	0.3%	0.00	0.1%	0.02	0.3%	0.02	0.3%	0.03	0.5%
Encounter Records Dropped	0.01	0.2%	0.01	0.1%	0.00	0.0%	0.01	0.1%	0.01	0.1%	0.01	0.2%
Encounter Records Kept	0.01	0.2%	0.01	0.1%	0.00	0.0%	0.01	0.1%	0.01	0.1%	0.01	0.2%
Deduped Record Count	4.27	99.8%	4.73	99.9%	5.10	100.0%	5.65	99.9%	5.50	99.9%	6.24	99.8%
Skilled Nursing Facilities												
Original Record Count	1.46	100.0%	1.62	100.0%	1.74	100.0%	1.85	100.0%	1.65	100.0%	1.95	100.0%
Single Records per DRI Key	1.45	99.7%	1.62	99.7%	1.74	99.8%	1.85	99.9%	1.95	118.4%	1.95	99.9%
Multiple Records per DRI Key	0.005	0.3%	0.004	0.3%	0.003	0.2%	0.003	0.1%	0.001	0.1%	0.001	0.1%
Encounter Records Dropped	0.002	0.2%	0.002	0.1%	0.002	0.1%	0.001	0.1%	0.001	0.0%	0.001	0.0%
Encounter Records Kept	0.002	0.2%	0.002	0.1%	0.002	0.1%	0.001	0.1%	0.001	0.0%	0.001	0.0%
Deduped Record Count	1.46	99.8%	1.62	99.9%	1.74	99.9%	1.85	99.9%	1.65	100.0%	1.95	100.0%

SOURCE: 2016-2021 Inpatient hospital and SNF MA encounter base records.

Carrier, DME, and Outpatient Facility Duplicate Records

The process to identify duplicate encounter records for carrier, DME, and outpatient facility encounter data is more complicated than the process used for inpatient hospital and SNF encounter data for three reasons.

- First, there are instances where duplicate values for the DRI key can have the same values so we must identify records that meet criteria to bypass duplicate record checks.
- Second, the DRI key specified in the EDPS manual includes payment variables that are not publicly available. Therefore, our methods must modify the way we identify duplicates to exclude the payment-related fields so that the encounter records identify unique service elements for data available to external researchers.
- Third, given the number of duplicate encounter records that are identified due to this modification, we must use more than the latest EDPS processing date to resolve duplicate encounter records.

The following sections discuss each of these issues in more detail.

Duplicate Record Check Bypass

According to the EDPS manual, there are two instances where records are excluded from duplicate record checks. First, encounter records with the following HCPCS modifiers associated with the record are exempt from duplicate record checks:

Institutional- Outpatient	Professional
59 - Distinct Procedural Service	59 - Distinct Procedural Service
62 - Two Surgeons	(Not applicable)
66 - Surgical Team	(Not applicable)
76 - Repeat Procedure by Same Physician	76 - Repeat Procedure by Same Physician
77 - Repeat Procedure by Another Physician	77 - Repeat Procedure by Another Physician
91 - Repeat Clinical Diagnostic Laboratory Test	91 - Repeat Clinical Diagnostic Laboratory Test

Source: EDPS manual, Table 6.7
https://www.csscoperations.com/internet/cssc4.nsf/files/ED_Submission_Processing_Guide.pdf/$FIle/ED_Submission_Processing_Guide.pdf, accessed 06/04/2024.

Second, encounter records that represent ambulatory surgical center (ASC) services and meet the following criteria are exempt from duplicate record checks.

- For carrier and DME encounter data, records must contain the place of service (POS)='24' which represents ASCs, and the provider NPI assigned to the record must have provider specialty='49'.
- For outpatient facility encounter data, records must report type of bill (TOB)='83X' which represents ASC.
- For both data types, the HCPCS code on the encounter service line must be present in the ASC fee schedule and have a '1' for the Multiple Procedure Discount (MPD) Indicator.

The specialty values used to model the ASC bypass criteria in the carrier and DME encounter data are determined using the edited NPI variable defined in Section I. One limitation of this method for carrier and DME data is that it is unclear whether MAO plans require both the professional service and the facility service to be reported in the encounter data the way that FFS Part B non-institutional Medicare claims require both the professional and facility to submit claims independently for payment. If the MA encounter data do not reflect the FFS ASC billing standard, then the carrier and DME encounter data may not always capture the ASC facility NPI. In other words, we may not always see records with NPIs that have specialty code = "49". This limitation could lead us to underestimate the number of records that meet the ASC bypass criteria. Future versions of the MA analytical data will explore this issue further.

Defining Duplicate Record Identifying Keys

As previously mentioned, the EDPS uses payment variables as part of the DRI key group used to flag and reject duplicate encounter records. However, we modify the DRI key group definition to exclude payment variables so that the encounter records reflect unique service elements for data available to external researchers. Below is a list of the modified DRI key group definition used for the carrier, DME, and outpatient facility data files:

Data Field	Carrier and DME	Outpatient Facility
Beneficiary ID	x	x
Line From/Thru Date	x	x
HCPCS	x	x
HCPCS Modifier 1-4	x	x
Place of Service	x	
Rendering Provider NPI (1)	x	
Type of Bill		x
NPI [1]		x
Revenue Center Code		x

NOTES:[1] As described in the "NPI Edit and Imputation Methods" section, the carrier and DME files use an edited NPI variable that coalesces three NPI provider variables whereas the outpatient facility file only uses the organizational NPI variable.

Using this definition of the DRI key group, Table 3 documents the number of records that are bypassed for the duplicate record checks, that belong to a DRI key group with a single record, and that belong to a DRI key group with multiple records.

Table 3. Distribution of Carrier, DME, and Outpatient Facility MA Encounter Records by Dedupe Process Status

Dedupe Process Status	2016		2017		2018		2019		2020		2021	
	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%
Carrier and DME Services												
Total	972.57	100.0%	1076.91	100.0%	1221.84	100.0%	1387.96	100.0%	1375.88	100.0%	1686.94	100.0%
Bypass Records	21.92	2.3%	24.57	2.3%	29.04	2.4%	34.68	2.5%	33.93	2.5%	42.95	2.5%
Single Records per DRI Key	894.98	92.0%	992.94	92.2%	1121.70	91.8%	1273.20	91.7%	1259.80	91.6%	1542.80	91.5%
Multiple Records per DRI Key	55.67	5.7%	59.40	5.5%	71.11	5.8%	80.08	5.8%	82.15	6.0%	101.19	6.0%
Outpatient Facilities												
Total	349.81	100.0%	393.81	100.0%	460.43	100.0%	537.45	100.0%	533.70	100.0%	688.68	100.0%
Bypass Records	12.83	3.7%	14.57	3.7%	16.83	3.7%	20.63	3.8%	19.19	3.6%	24.18	3.5%
Single Records per DRI Key	307.68	88.0%	345.12	87.6%	405.07	88.0%	465.13	86.5%	459.69	86.1%	583.19	84.7%
Multiple Records per DRI Key	29.30	8.4%	34.12	8.7%	38.54	8.4%	51.68	9.6%	54.81	10.3%	81.32	11.8%

SOURCE: 2016-2021 Carrier, DME, and outpatient facility MA encounter base and line records.

As Table 3 illustrates, the carrier, DME, and outpatient facility encounter data have a significantly higher percentage of records that belong to DRI key groups with multiple records than the inpatient hospital and SNF encounter data. This result is likely because we are using modified DRI keys that exclude payment variables originally included by the EDPS for duplicate record checks.

Resolving Duplicate Encounter Records

Given that the data show a higher percentage of records belonging to multi-record DRI key groups, OEDA applies additional criteria to resolve duplicate encounter records beyond applying the latest EDPS processing date method that is used for the inpatient hospital and SNF encounter data. Specifically, we use the following methods, applied hierarchically, to select a single record from the duplicates within a DRI key group:

- **Method 1: Latest EDPS Processing Date Resolution.** This method determines if a single record contains the latest processing date within the DRI key group.
- **Method 2: Claim Frequency Code=7.** This method looks at the claim frequency code of records within the same key group and determines if there is a single record with claim frequency code ='7' which indicates that the record is a "replacement of a prior claim".
- **Method 3: Same ENC_JOIN_KEY.** The encounter join key, or ENC_JOIN_KEY, is the variable that identifies a unique base record submitted by a provider. If all the duplicate records appear with the same ENC_JOIN_KEY value, this method preserves those records as distinct encounter records. The logic behind this method is that providers are unlikely to submit duplicate service information on the same base encounter record, so these records likely represent distinct service line information.
- **Method 4: Eliminate Pure Duplicates**. This method identifies duplicate records that all have the same values for relevant variables (i.e., diagnosis, procedure, provider and claim administration variables not included in the DRI keys). It then selects one record within the group and eliminates the others from the data. See Appendix B for a complete list of variables used to identify pure duplicates.
- **Method 5: Random Selection**. If none of the other methods result in a resolution of multiple records within a DRI key group, then one record is selected at random.

Table 4 shows the percentage of unduplicated records that fall into each deduping method category.

Table 4. Distribution of Carrier, DME, and Outpatient Facility Duplicate MA Encounter Records by Dedupe Method Categories

Dedupe Method Category	2016		2017		2018		2019		2020		2021	
	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%
Carrier and DME Services												
Total	55.7	100.0%	59.40	100.0%	71.11	100.0%	80.08	100.0%	82.15	100.0%	101.19	100.0%
Last EDPS create date	40.4	72.6%	42.66	71.8%	52.69	74.1%	57.56	71.9%	59.20	72.1%	73.36	72.5%
Claim Frequency Code=7	7.7	13.8%	8.24	13.9%	8.93	12.6%	12.68	15.8%	12.58	15.3%	14.43	14.3%
Same ENC_JOIN_KEY	2.3	4.1%	2.48	4.2%	2.65	3.7%	2.82	3.5%	3.29	4.0%	4.75	4.7%
Pure Duplicates	3.4	6.1%	3.89	6.6%	4.28	6.0%	4.42	5.5%	4.37	5.3%	5.73	5.7%
Random Selection	1.9	3.5%	2.13	3.6%	2.56	3.6%	2.60	3.2%	2.70	3.3%	2.92	2.9%
Outpatient Facilities												
Total	29.3	100.0%	34.12	100.0%	38.54	100.0%	51.68	100.0%	54.81	100.0%	81.32	100.0%
Last EDPS create date	13.0	44.3%	16.10	47.2%	17.12	44.4%	23.83	46.1%	25.79	47.1%	39.15	48.2%
Claim Frequency Code=7	--	--	--	--	--	--	--	--	--	--	--	--
Same ENC_JOIN_KEY	12.5	42.7%	14.36	42.1%	17.13	44.4%	22.12	42.8%	23.49	42.9%	32.73	40.2%
Pure Duplicates	2.5	8.7%	2.20	6.4%	2.45	6.4%	3.30	6.4%	3.05	5.6%	4.95	6.1%
Random Selection	1.3	4.3%	1.47	4.3%	1.84	4.8%	2.43	4.7%	2.47	4.5%	4.49	5.5%

SOURCE: 2016-2021 Carrier, DME, and outpatient facility MA encounter base and line records.

For carrier and DME encounter data, over 85% of duplicate records across the data years are deduped using either the latest EDPS processing date or the claim frequency code methods, with the EDPS date method representing the largest proportion of records (approximately 72%). For the outpatient facility data, however, the predominant deduping methods (nearly 90% of records) use either the latest EDPS processing date or the same ENC_JOIN_KEY value. Additionally, the percentage of duplicate records resolved using these methods are more evenly split, with the "latest EDPS date" method averaging 46% of records and the "same ENC_JOIN_KEY" method averaging about 43% between 2016-2021. Lastly, the outpatient facility data's deduping algorithm does not use the "claim frequency code = 7" method, which suggests that this file type submits multiple encounter lines as replacements for prior encounter records within the DRI key groups.

Table 5 shows the percentage of duplicate records that were deleted from the final carrier, DME, and outpatient facility analytical data sets. For the carrier and DME encounter data files, the percentage of duplicate records that are dropped versus kept is roughly equal (approximately 3% each). However, for the outpatient facility encounter data files, the percentage of duplicate records that are kept is roughly twice as high as the percentage of duplicate records that are dropped largely due to the large proportion of duplicate records that have the same ENC_JOIN_KEY value.

Table 5. Duplicate Carrier, DME, and Outpatient Facility MA Encounter Records Removed from Analytical Files

Record Count Category	2016		2017		2018		2019		2020		2021	
	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%
Carrier and DME Services												
Original Record Count	972.6	100.0%	1,076.9	100.0%	1,221.8	100.0%	1,388.0	100.0%	1,375.9	100.0%	1,686.9	100.0%
Single Records per DRI Key and Bypass Records	916.9	94.3%	1,017.5	94.5%	1,150.7	94.2%	1,307.9	94.2%	1,293.7	94.0%	1,585.8	94.0%
Multiple Records per DRI Key	55.7	5.7%	59.4	5.5%	71.1	5.8%	80.1	5.8%	82.1	6.0%	101.2	6.0%
Encounter Records Dropped	27.2	2.8%	29.1	2.7%	34.9	2.9%	39.4	2.8%	40.3	2.9%	49.2	2.9%
Encounter Records Kept	28.5	2.9%	30.3	2.8%	36.2	3.0%	40.7	2.9%	41.8	3.0%	52.0	3.1%
Deduped Record Count	945.4	97.2%	1,047.8	97.3%	1,186.9	97.1%	1,348.6	97.2%	1,335.6	97.1%	1,637.7	97.1%
Outpatient Facilities												
Original Record Count	349.8	100.0%	393.8	100.0%	460.4	100.0%	537.4	100.0%	533.7	100.0%	688.7	100.0%
Single Records per DRI Key and Bypass Records	320.5	91.6%	359.7	91.3%	421.9	91.6%	485.8	90.4%	485.8	91.0%	607.4	88.2%
Multiple Records per DRI Key	29.3	8.4%	34.1	8.7%	38.5	8.4%	51.7	9.6%	54.8	10.3%	81.3	11.8%
Encounter Records Dropped	8.9	2.5%	10.4	2.6%	11.3	2.5%	15.6	2.9%	16.6	3.1%	25.9	3.8%
Encounter Records Kept	20.4	5.8%	23.7	6.0%	27.2	5.9%	36.1	6.7%	38.3	7.2%	55.5	8.1%
Deduped Record Count	341.0	97.5%	383.4	97.4%	449.1	97.5%	521.9	97.1%	517.1	96.9%	662.8	96.2%

SOURCE: 2016-2021 Carrier, DME, and outpatient facility MA encounter base and line records.

Section III. CCN Assignment Algorithm

The last issue to address in developing the MA encounter analytical files relates to attaching the CCN to the inpatient hospital encounter data. This section provides background information on the CCN, why it is needed to analyze MA inpatient hospital encounter data, and the algorithm used to attach the CCN to the encounter data using the MA organizational NPI variable.

Background

The CCN is a provider identifier that CMS assigns to institutional providers to classify which FFS payment system they fall under (e.g., critical access hospitals, inpatient rehab facilities). Specifically, the last four digits of the CCN indicate the facility type. For example:

0001-0879 = Short-Term (General and Specialty) Hospitals
2000-2299 = Long-Term Care Hospitals (Excluded from IPPS)
3025-3099 = Rehabilitation Hospitals (Excluded from IPPS)
4000-4499 = Psychiatric Hospitals (Excluded from IPPS)

While the FFS claims data also report NPI like the MA encounter data, OEDA uses the CCN to identify hospital type for FFS data products, such as the CMS Program Statistics (CPS), because the identifier is a CMS-generated identifier that categorizes hospitals in a consistent manner that is tied to FFS payment systems. The lack of CCN in MA data is one of the largest hurdles to evaluating MA encounter data inpatient hospitals by facility type in a manner consistent with FFS data analyses.

NPI-CCN Crosswalk

The first step to building the CCN assignment algorithm is to construct an NPI-CCN crosswalk. The NPI-CCN crosswalk is a data set we have created to associate an organizational NPI to a CCN. We use two data sources to construct this crosswalk:

- NPI-CCN relationships found in the FFS institutional claims data (Part A and B); and
- Provider Master Index (PMI) NPI-CCN crosswalk that combines information from various CMS provider data systems.[8]

The FFS crosswalk has the advantage of capturing providers that are currently active in Medicare fee-for-service. However, it may not represent the full universe of NPI-CCN relationships, particularly those that do not submit Medicare FFS claims. Using the PMI NPI-CCN crosswalk table allows us to supplement the NPI-CCN links we find in the FFS data.

To construct the NPI-CCN crosswalk from the FFS data, we collapse the Part A claims data and the Part B institutional claims data by NPI and CCN for each year of claims data between 2016 and 2021. We then concatenate the Part A and B annual data sets and further collapse the file for a given year so that there are unique combinations of NPI and CCN codes. Finally, we concatenate the annual files to create a longitudinal NPI-CCN crosswalk that contains year, NPI, and CCN data fields.

To construct the PMI-based NPI-CCN crosswalk, we subset the PMI Legacy ID table to records where the legacy ID type reflects IDs coming from the Online Survey Certification and Reporting (OSCAR) provider system to isolate the NPI-CCN combinations.[9] CMS does not check the quality of the provider information stored in the PMI tables. Therefore, we applied additional checks of the NPI and CCN values in the PMI table given that the PMI table may include erroneous or individual NPI data. Specifically, we checked that the:

- CCN value was valid, i.e., the value consisted of CCN with 6 characters, the first 2-digits were valid state codes, the last four digits fall within ranges for facility types, and the alpha characters used in the third character are valid letters used to indicate a subunit facility[10];
- NPI value was not linked to an individual provider; and

[8] The PMI is a suite of NPI-centered tables that combines information from various CMS provider systems. The PMI tables are only available for CMS data users. We used the PMI NPI-Legacy ID table to identify current and historical NPI-CCN links. The PMI legacy ID is sourced from the Provider Enrollment, Chain, and Ownership System (PECOS), which providers must enroll in to participate in Medicare; the National Plan and Provider Enumeration System (NPPES), which providers enroll in to receive an national provider identifier (NPI) as required by HIPPA of 1996 for use across public and private health care systems; and the National Provider Identifier Crosswalk System (NPICS), a temporary provider system that CMS implemented to ensure the continuity of claims processing during the implementation of the NPI. NPICS provides historical information NPI-legacy ID crosswalks and was ultimately replaced by PECOS.

[9] The OSCAR system is a legacy administrative database that was used to assign organizations a CCN identifier. This system was replaced by the Certification and Survey Provider Enhanced Reporting (CASPER) and the Quality Improvement Evaluation System (QIES) in 2012. However, internal CMS data systems still refer to elements from these data bases coming from the OSCAR system.

[10] These criteria are based on the CCN assignment rules published in the CMS State Operations Manual, Chapter 2, Certification Process, Section 2779, https://www.cms.gov/regulations-and-guidance/guidance/manuals/downloads/som107c02.pdf (accessed 10/31/2022).

- CCN values in the PMI matched to the Provider of Services..[11, 12]

We then create annual data sets of NPI-CCN combinations that were in effect for a calendar year using the legacy ID effective dates. Some PMI Legacy ID table records did not have valid effective and termination dates to determine whether the NPI-CCN relationship existed in a given data year. For these records, we only kept NPI-CCN combinations that had a record status set to "Current". Finally, we merge the FFS and PMI longitudinal NPI-CCN data sets by year, NPI, and CCN to create a table that we refer to as the Master NPI-CCN crosswalk.

Table 6 provides information on the number of NPI-CCN combinations that exist on both the FFS and PMI longitudinal crosswalks, only on the FFS longitudinal crosswalk, and only on the PMI longitudinal crosswalk.

Table 6. Distribution of NPI-CCN Combinations, by Source Data Comparisons

FFS-PMI NPI-CCN Comparison Categories	2016		2017		2018		2019		2020		2021	
	N	%	N	%	N	%	N	%	N	%	N	%
Total	106,744	100.0%	105,950	100.0%	105,143	100.0%	104,479	100.0%	103,948	100.0%	103,368	100.0%
All Years Match	49,111	46.0%	48,655	45.9%	48,139	45.8%	47,663	45.6%	47,161	45.4%	46,834	45.3%
Some Years Match	25,009	23.4%	25,069	23.7%	24,873	23.7%	24,717	23.7%	24,493	23.6%	24,299	23.5%
No Match = FFS Only	476	0.4%	477	0.5%	351	0.3%	264	0.3%	155	0.1%	71	0.1%
No Match = PMI Only	24,513	23.0%	24,572	23.2%	24,503	23.3%	24,436	23.4%	24,323	23.4%	24,220	23.4%
No Match = Both	20	0.02%	20	0.02%	19	0.02%	17	0.02%	15	0.01%	8	0.01%
No Years Match	32,624	30.6%	32,226	30.4%	32,131	30.6%	32,099	30.7%	32,294	31.1%	32,235	31.2%
No Match = FFS Only	1,529	1.4%	1,346	1.3%	1,323	1.3%	1,307	1.3%	1,428	1.4%	1,395	1.3%
No Match = PMI Only	31,071	29.1%	30,858	29.1%	30,788	29.3%	30,775	29.5%	30,849	29.7%	30,822	29.8%
No Match = Both	24	0.02%	22	0.02%	20	0.02%	17	0.02%	17	0.02%	18	0.02%

SOURCE: 2016-2021 FFS Part A and Part B institutional claims and Provider Master Index NPI-Legacy ID Table.

Approximately 70% of NPI-CCN combinations exist in both the FFS and PMI source data for at least one data year. Those combinations where no data years match between the data source are largely driven by those NPI-CCN relationships that only appear in the PMI data.

Since NPIs can link to multiple CCN codes, we create additional CCN-related fields that align with the MA data source field and the taxonomy fields in the MA encounter data. Specifically, we map the CCN codes to either an inpatient hospital, SNF, home health, or other institutional category to align with the MA data file types. We also map both CCN and the NPI taxonomy codes to common, broadly defined hospital type categories to see if the MA encounter taxonomy field can be used to resolve 1:Many NPI-CCN matches. See Appendix A for details on how CCNs and NPI taxonomy codes are mapped to common hospital type categories so that they can be used in the CCN assignment algorithm.

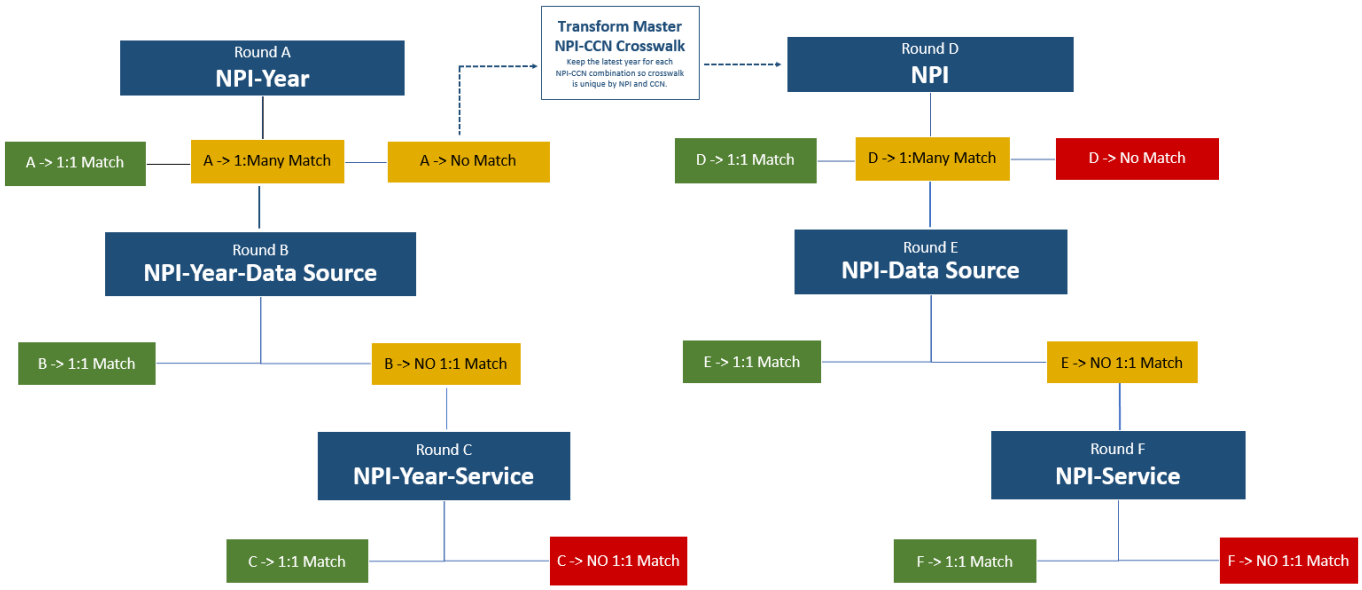
[11] Information on the POS file can be found here: https://data.cms.gov/provider-characteristics/hospitals-and-other-facilities/provider-of-services-file-hospital-non-hospital-facilities (accessed 10/31/2022).

[12] We performed similar checks on the FFS NPI-CCN combinations and found that all FFS combinations met these criteria.

NPI-CCN Linkage Algorithm

Figure 2 below is a diagram that illustrates the steps we used to assign a CCN to MA inpatient hospital encounter records.

Figure 2. Diagram of MA CCN Assignment Algorithm



The algorithm starts by merging on the CCN to the encounter data by using just the data year and the NPI (Round A). Matches from this step represent NPI-CCN relationships where there is a 1:1 relationship between NPI and CCN in the master crosswalk. We have the least uncertainty of these matches since they do not require information from the MA encounter record. Figure 3 contains an example of the Round A merge where each record is assigned a CCN by matching the crosswalk using just NPI and year.

Figure 3. Round "A" Example of CCN Assignment

MA Encounter Data

Round A

Encounter Join Key	Year	NPI	MA Data Source	Taxonomy Service Category	Taxonomy Code	Taxonomy Description	CCN
385565131	2016	1003000514	IP	ACUTE	282N00000X	General Acute Care Hospital	120123
357011136	2017	1003000514	IP	ACUTE	282N00000X	General Acute Care Hospital	120123
399394843	2018	1003000514	IP	ACUTE	282N00000X	General Acute Care Hospital	120123
380295462	2019	1003000514	IP	ACUTE	282N00000X	General Acute Care Hospital	120123

Master NPI-CCN Crosswalk

Round A

Year	NPI	CCN Data Source	CCN Service Category	CCN	CCN Description
2016	1003000514	IP	ACUTE	120123	Short-Term Hospitals
2017	1003000514	IP	ACUTE	120123	Short-Term Hospitals
2018	1003000514	IP	ACUTE	120123	Short-Term Hospitals
2019	1003000514	IP	ACUTE	120123	Short-Term Hospitals

MA encounter records that do not have a 1:1 match to the NPI-CCN crosswalk table are then divided into two groups: 1.) those that have a 1:Many relationship between NPI and CCN; and 2.) records that have no match to the master NPI-CCN crosswalk using NPI and year. The 1:Many group is then fed into the next round of merges where we use information on the encounter bill type to uniquely assign a CCN to the record (Round B). Figure 4 shows a Round B example where the NPI has two CCNs attached to it in the crosswalk. One CCN (370139) is mapped to the inpatient hospital bill type in the NPI-CCN crosswalk's "CCN Data Source" column; the other CCN (37U139) is mapped to the skilled nursing facility bill type. We then use the bill type mappings in the crosswalk to assign the inpatient hospital CCN to the encounter records with inpatient hospital bill type.

Figure 4. Round "B" Example of CCN Assignment

MA Encounter Data

Round B

Encounter Join Key	Year	NPI	MA Data Source	Taxonomy Service Category	Taxonomy Code	Taxonomy Description	CCN
337365886	2018	1003318692	IP	ACUTE	282N00000X	General Acute Care Hospital	370139
324453046	2019	1003318692	IP	ACUTE	282N00000X	General Acute Care Hospital	370139

Master NPI-CCN Crosswalk

Round B

Year	NPI	CCN Data Source	CCN Service Category	CCN	CCN Description
2018	1003318692	IP	ACUTE	370139	Short-Term Hospitals
2018	1003318692	SNF	SNF	37U139	Swing Beds
2019	1003318692	IP	ACUTE	370139	Short-Term Hospitals
2019	1003318692	SNF	SNF	37U139	Swing Beds

For those encounter records still left without a 1:1 CCN match after Round B, the algorithm attempts to match a CCN using broadly defined service categories (Round C). In this merge, we resolve 1:Many matches by assigning the CCN that matches the broadly defined service category mapped to the taxonomy reported on the encounter record. Figure 5 illustrates a Round C merge example. In this scenario, the NPI maps to CCNs that are both inpatient hospitals, but one CCN (01S007) reflects an inpatient psychiatric subunit facility (IPF) and the other CCN (010007) reflects a short-term acute care facility. The algorithm assigns the psychiatric hospital subunit CCN to the record that reports an encounter taxonomy code that maps to an IPF and the short-term acute care CCN to the record that reports an encounter taxonomy code that maps to an acute care hospital.

Figure 5. Round "C" Example of CCN Assignment

MA Encounter Data

Round C

Encounter Join Key	Year	NPI	MA Data Source	Taxonomy Service Category	Taxonomy Code	Taxonomy Description	CCN
359384435	2018	1013937705	IP	IPF	273R00000X	Psychiatric Unit	01S007
333037825	2018	1013937705	IP	ACUTE	282N00000X	General Acute Care Hospital	010007
394189577	2019	1013937705	IP	IPF	273R00000X	Psychiatric Unit	01S007
393068973	2019	1013937705	IP	ACUTE	282N00000X	General Acute Care Hospital	010007

Master NPI-CCN Crosswalk

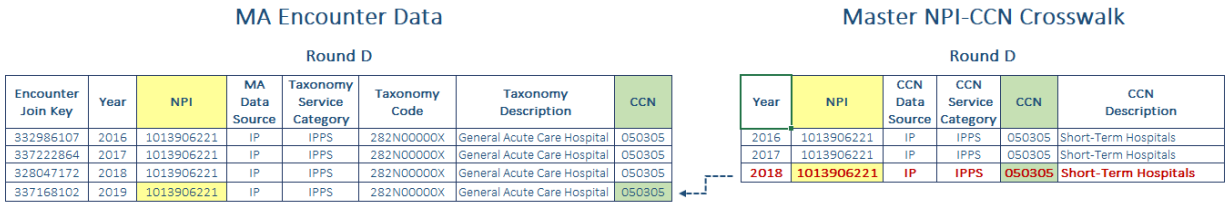
Round C

Year	NPI	CCN Data Source	CCN Service Category	CCN	CCN Description
2018	1013937705	IP	IPF	01S007	Psychiatric Hospital Subunits
2018	1013937705	IP	ACUTE	010007	Short-Term Hospitals
2019	1013937705	IP	IPF	01S007	Psychiatric Hospital Subunits
2019	1013937705	IP	ACUTE	010007	Short-Term Hospitals

The algorithm sets encounter records with no CCN match after Round C to have a missing value for the CCN variable since we cannot resolve 1:Many NPI-CCN relationships.

As Figure 2 shows, the algorithm then turns to addressing the records that did not find a match in Round A when it merges on the CCN by year and NPI. For Round D merges, we transform the NPI-CCN crosswalk by keeping the latest year for each NPI-CCN combination such that the crosswalk is unique by NPI and CCN. The algorithm then attempts to merge on the CCN to the MA encounter records by NPI only to see if the MA encounter NPI exists for any of the years in the NPI-CCN crosswalk. Figure 6 gives an example of Round D merges. In this figure, the MA encounter data for years 2016-2018 matches to the NPI-CCN crosswalk in Round A using NPI and year because there is a 1:1 NPI-CCN combination for those years in the crosswalk. However, the 2019 MA encounter record does not match to a CCN in Round A. In Round D, where the algorithm drops year and just merges by NPI using the latest NPI-CCN combination (2018), it can assign a CCN to the 2019 encounter record.

Figure 6. Round "D" Example of CCN Assignment



Rounds E (merge by NPI and data source) and F (merge by NPI and service category) of the algorithm essentially repeat Rounds B and C but without using the year variable to attach the CCN to the encounter data.

Table 7 shows the results of the CCN assignment algorithm in terms of the percentage of MA inpatient hospital encounter records that are assigned a CCN and the extent to which the CCN assigned is consistent with the MA encounter inpatient hospital bill type.

Table 7. Percentage of MA Inpatient Hospital Encounter Records with CCN Assignment, by Merge Result Category

CCN Merge Result	2016		2017		2018		2019		2020		2021	
	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%	N (Millions)	%
Total	4.3	100.0%	4.7	100.0%	5.1	100.0%	5.7	100.0%	5.5	100.0%	6.2	100.0%
Match	4.1	96.1%	4.6	96.6%	4.9	96.5%	5.5	96.7%	5.3	97.0%	6.1	97.7%
Match, Same	4.1	95.8%	4.5	95.1%	4.8	95.0%	5.4	95.6%	5.3	96.0%	6.1	97.5%
Match-Round A, Same	3.8	88.2%	4.1	87.4%	4.4	87.2%	5.0	87.8%	4.6	84.6%	5.4	86.5%
Match-Round B, Same	0.2	3.8%	0.2	3.6%	0.2	3.5%	0.2	3.5%	0.4	7.0%	0.4	6.1%
Match-Round C, Same	0.2	3.8%	0.2	4.0%	0.2	4.3%	0.2	4.3%	0.2	4.4%	0.3	4.8%
Match-Round D, Same	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%
Match-Round E, Same	--	--	--	--	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%
Match-Round F, Same	--	--	0.0	0.0%	0.0	0.0%	--	--	--	--	0.0	0.0%
Match, Different	0.0	0.3%	0.1	1.6%	0.1	1.5%	0.1	1.1%	0.1	1.0%	0.0	0.2%
Match-Round A, Different	0.0	0.3%	0.1	1.5%	0.1	1.5%	0.1	1.0%	0.0	0.9%	0.0	0.2%
Match-Round C, Different	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.1%	0.0	0.1%	0.0	0.0%
Match-Round D, Different	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%	0.0	0.0%
Match-Round F, Different	--	--	0.0	0.0%	--	--	--	--	--	--	--	--
No Match	0.2	3.9%	0.2	3.4%	0.2	3.5%	0.2	3.3%	0.2	3.0%	0.1	2.3%
No Match, 1:Many	0.1	2.7%	0.1	2.4%	0.1	2.4%	0.1	2.3%	0.1	1.8%	0.1	1.3%
No Match, Missing	0.0	1.1%	0.0	1.0%	0.1	1.1%	0.1	1.1%	0.1	1.1%	0.1	1.0%

SOURCE: 2016-2021 Inpatient hospital MA encounter base records.

Across all data years, approximately 96% of records are assigned a CCN. The algorithm assigns nearly all these inpatient hospital encounter records (approximately 95%) a CCN that is an inpatient hospital facility (i.e., the "Match, Same" merge result). On average, 87% of records are assigned a CCN in Round A where the algorithm uses NPI and year. This indicates that most records are matched using the method with the least amount of uncertainty in terms of the match assignment, i.e., we only use the reported organizational NPI. The more MA data we need to assign the CCN, the more uncertainty is introduced into the algorithm since there may be errors in the MAO-reported data elements. For example, we have not confirmed that MAOs are accurately reporting the bill type in the encounter data and bill type is used in Rounds B and E to resolve 1:Many CCN relationships to assign a CCN to the encounter record.

There is a small percentage of records (0.3% - 1.6%) that have a CCN assigned by the algorithm, but that CCN is not an inpatient hospital facility (i.e., the "Match, Different" merge result). Our analyses show that many of these

records are assigned home health agency CCNs. Future refinements to the algorithm will attempt to resolve this discrepancy by evaluating whether the bill type or the NPI was incorrectly assigned. For now, we set the CCN variable equal to missing.

Finally, among the encounter records with no match to the NPI-CCN crosswalk, the majority are not matched because we are unable to resolve the 1:Many relationships using the bill type or taxonomy encounter data elements. There may be opportunities in the future to improve the match rate by exploring other means for addressing 1:Many NPI-CCN relationships.

Appendix A: Mapping Inpatient Hospital CCN Facility and NPI Taxonomy Categories to Common Hospital Type Categories

Appendix Table A.1. Mapping Inpatient Hospital CCN Facility Categories to Common Hospital Type Categories

CCN Last 4 Digits	CCN Facility Categories	Common Hospital Type Categories
1300 - 1399	Critical Access Hospitals	Critical Access Hospital
4000 - 4499	Psychiatric Hospitals	Inpatient Psychiatric Facility
M*** [1]	Psychiatric Unit	Inpatient Psychiatric Facility
S*** [1]	Psychiatric Unit	Inpatient Psychiatric Facility
0001 - 0879	Short-Term Hospitals	Short-Term Care Hospital
0880 - 0899	ORD Demo Project Hospitals	Short-Term Care Hospital
0900 - 0999	Multiple Hospital Component-Medical Complex	Short-Term Care Hospital
3025 - 3099	Rehabilitation Hospitals	Inpatient Rehabilitation Facility
R*** [1]	Rehabilitation Unit	Inpatient Psychiatric Facility
T*** [1]	Rehabilitation Unit	Inpatient Psychiatric Facility
2000 - 2299	Long-Term Care Hospitals	Long-Term Care Hospital
1200 - 1224	Alcohol/Drug Hospitals	Other Hospital
1990 - 1999	Religious Non-Medical Hospitals	Other Hospital
3000 - 3024	Tuberculosis Hospitals	Other Hospital
3300 - 3399	Children's Hospitals	Other Hospital

SOURCE: CCN codes and facility descriptions come from the CMS State Operations Manual, Chapter 2, Certification Process, Section 2779 (https://www.cms.gov/regulations-and-guidance/guidance/manuals/downloads/som107c02.pdf, accessed 10/31/2022).
NOTE: [1] These CCN digits represent ID values where the third character in the 6-digit CCN indicates a particular hospital subunit.

Appendix Table A.2. Mapping Inpatient Hospital NPI Specialty Taxonomy Categories to Common Hospital Type Categories

NPI Taxonomy Codes	NPI Taxonomy Descriptions	Common Hospital Type Categories
282NC0060X	General Acute Care Hospital Critical Access	Critical Access Hospital
283Q00000X	Psychiatric Hospital	Inpatient Psychiatric Facility
273R00000X	Psychiatric Unit	Inpatient Psychiatric Facility
281P00000X	Chronic Disease Hospital	Short-Term Care Hospital
282N00000X	General Acute Care Hospital	Short-Term Care Hospital
282NR1301X	General Acute Care Hospital Rural	Short-Term Care Hospital
282NW0100X	General Acute Care Hospital Women	Short-Term Care Hospital
286500000X	Military Hospital	Short-Term Care Hospital
2865M2000X	Military Hospital Military General Acute Care Hospital	Short-Term Care Hospital
2865X1600X	Military Hospital Military General Acute Care Hospital Operational (Transportable)	Short-Term Care Hospital
276400000X	Rehabilitation, Substance Use Disorder Unit	Short-Term Care Hospital
283X00000X	Rehabilitation Hospital	Inpatient Rehabilitation Facility
283XC2000X	Rehabilitation Hospital Children	Inpatient Rehabilitation Facility
273Y00000X	Rehabilitation Unit	Inpatient Rehabilitation Facility
282E00000X	Long Term Care Hospital	Long-Term Care Hospital
281PC2000X	Chronic Disease Hospital Children	Other Hospital
282NC2000X	General Acute Care Hospital Children	Other Hospital
282J00000X	Religious Nonmedical Health Care Institution	Other Hospital
284300000X	Special Hospital	Other Hospital

SOURCE: The NPI specialty taxonomy codes and descriptions come from the National Uniform Claim Committee (NUCC) Provider Taxonomy code set, https://www.nucc.org/images/stories/CSV/nucc_taxonomy_210.csv (accessed 10/31/2022).

Appendix B: Variables Used to Identify Pure Duplicates When Resolving DRI Key Duplicate Carrier, DME, and Outpatient Records

The variables listed below are used in Method 4 of the algorithm used to dedupe encounter records that belong to an DRI key group with multiple records. See Section II for more details about the methods used to dedupe encounter records.

Appendix Table B.1. Carrier and DME Variables Used to Identify Pure Duplicates When Resolving DRI Key Group Duplicates

Encounter Data File Type	Variable Name	Variable Label
Base	BENE_ID	CCW Beneficiary ID
Base	CLM_PLACE_OF_SRVC_CD	Claim Place of Service Code
Base	CLM_FREQ_CD	Claim Frequency Code
Base	CLM_FROM_DT	Claim From Date
Base	CLM_THRU_DT	Claim Through Date
Base	CLM_TYPE_CD	Claim Type Code
Base	EDPS_CREATE_DT	Encounter Data Processing System (EDPS) Create Date
Base	ICD_DGNS_CD1-13	Claim Diagnosis Code 1-13
Base	ICD_DGNS_VRSN_CD1-13	Claim Diagnosis Code 1-13 Diagnosis Version Code (ICD-9 or ICD-10)
Base	PRNCPAL_DGNS_CD	Claim Principal Diagnosis Code
Base	PRNCPAL_DGNS_VRSN_CD	Claim Principal Diagnosis Code Diagnosis Version Code (ICD-9 or ICD-10)
Base	RFRG_PHYSN_NPI	Claim Referring Physician NPI Number
Base	TAX_NUM	Provider Tax Number
Line	HCPCS_1ST_MDFR_CD	HCPCS Initial Modifier Code
Line	HCPCS_2ND_MDFR_CD	HCPCS Second Modifier Code
Line	HCPCS_3RD_MDFR_CD	HCPCS Third Modifier Code
Line	HCPCS_4TH_MDFR_CD	HCPCS Fourth Modifier Code
Line	HCPCS_CD	HCPCS Code
Line	LINE_1ST_EXPNS_DT	Line First Expense Date
Line	LINE_LAST_EXPNS_DT	Line Last Expense Date
Line	LINE_NDC_CD	Line National Drug Code (NDC)
Line	LINE_RX_NUM	Line RX Number
Line	LINE_SRVC_CNT	Line Service Count

Appendix Table B.2. Outpatient Variables Used to Identify Pure Duplicates When Resolving DRI Key Group Duplicates

Encounter Data File Type	Variable Name	Variable Label
Base	BENE_ID	CCW Beneficiary ID
Base	CLM_FAC_TYPE_CD	Claim Facility Type Code
Base	CLM_FREQ_CD	Claim Frequency Code
Base	CLM_SRVC_CLSFCTN_TYPE_CD	Claim Service classification Type Code
Base	ORG_NPI	Organization NPI Number
Base	CLM_THRU_DT	Claim Through Date
Base	CLM_FROM_DT	Claim From Date
Base	CLM_TYPE_CD	Claim Type Code
Base	EDPS_CREATE_DT	Encounter Data Processing System (EDPS) Create Date
Base	CLM_1ST_DGNS_E_CD	First Claim Diagnosis E Code
Base	ICD_DGNS_CD1-25	Claim Diagnosis Code 1-25
Base	ICD_DGNS_E_CD1-10	Claim Diagnosis E Code 1-10
Base	ICD_PRCDR_CD1-13	Claim Procedure Code 1-13
Base	PRCDR_DT1-13	Claim Procedure Code 1-13 Date
Base	PRNCPAL_DGNS_CD	Claim Principal Diagnosis Code
Base	PTNT_DSCHRG_STUS_CD	Patient Discharge Status Code
Base	AT_PHYSN_NPI	Claim Attending Physician NPI Number
Base	OP_PHYSN_NPI	Claim Operating Physician NPI Number
Base	OT_PHYSN_NPI	Claim Other Physician NPI Number
Base	RFRG_PHYSN_NPI	Claim Referring Physician NPI Number
Base	RNDRNG_PHYSN_NPI	Claim Rendering Physician NPI Number
Base	TAX_NUM	Provider Tax Number
Line	HCPCS_1ST_MDFR_CD	HCPCS Initial Modifier Code
Line	HCPCS_2ND_MDFR_CD	HCPCS Second Modifier Code
Line	HCPCS_3RD_MDFR_CD	HCPCS Third Modifier Code
Line	HCPCS_4TH_MDFR_CD	HCPCS Fourth Modifier Code
Line	HCPCS_CD	HCPCS Code
Line	REV_CNTR	Revenue Center Code
Line	REV_CNTR_FROM_DT	Revenue Center From Date
Line	REV_CNTR_THRU_DT	Revenue Center Thru Date
Line	REV_CNTR_RNDRNG_PHYSN_NPI	Revenue Center Rendering Physician NPI
Line	REV_CNTR_IDE_NDC_UPC_NUM	Revenue Center IDE, NDC, or UPC Number
Line	REV_CNTR_NDC_QTY	Revenue Center National Drug Code (NDC) Quantity
Line	REV_CNTR_NDC_QTY_QLFR_CD	Revenue Center NDC Quantity Qualifier Code
Line	REV_CNTR_UNIT_CNT	Revenue Center Unit Count