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

May 2026

Table of Contents

1. Introduction

CMS currently publishes utilization metrics for people enrolled in Medicare Advantage (MA). Unlike Original Medicare (OM), where claims data follow long established billing standards, MA encounter data come from various plans with different data practices and are collected by CMS for purposes that differ from OM. This difference in data standards and purpose of collection means the MA data can be less consistent and sometimes incomplete. As a result, CMS makes certain adjustments to the MA data before using them to produce summary statistics on MA enrollee utilization. In this document, CMS describes the methodological approach developed to support the construction of utilization metrics from the MA encounter data.[1] We use this approach to allow us to include metrics related to services delivered to MA beneficiaries that are consistent with metrics reported in our existing OM data products.

As described in the Medicare Advantage Encounter Data User Guide[2], there are two main limitations to analyzing MA encounter data. First is the potential for duplicate service events in MA data. In OM data, each service event has a unique, unduplicated identifier that makes it relatively simple to calculate utilization metrics. In contrast, Medicare Advantage encounter data files may contain multiple records of the same service event making counting utilization more complex.

The second limitation is the lack of the CMS Certification Number (CCN) in the MA data. The CCN is used in utilization reporting to identify a hospital type (e.g., acute care hospital, inpatient psychiatric hospital, inpatient rehab facility). As these hospitals provide different types of services, we utilize CCN information on OM claims to further differentiate types of utilization. Currently, the only unique provider identifier in the encounter data is the National Provider Identifier (NPI) which cannot be used to identify hospital type.

This document describes the methods we used to prepare the MA encounter data. Specifically, it covers:

- The methods used to edit and impute encounter NPI variables which are used to identify unique service events and link CCNs to inpatient hospital records.
- The methods used to identify and address records that appear to describe the same service event.
- The algorithm generated to assign the CCN to inpatient hospital encounter records to categorize these records by inpatient hospital type.[3]
- The reasoning used to arrive at the decision not to release home health Medicare Advantage encounter data due to our inability to uniquely identify home health service events and uniformly define home health episode metrics.

CMS is providing this document so that researchers with access to MA Research Identifiable Files (RIFs) may better understand our approach. We expect to make enhancements to this approach and will publish updates as methods evolve. We welcome feedback and collaboration from encounter data users. Comments or questions may be directed to PDAG_Data_Products@cms.hhs.gov.

---

[1] Researchers can request encounter data files by contacting the Research Data Assistance Center (ResDAC). For more information on how to request these data, visit https://resdac.org/ (accessed 09/30/2025).

[2] *Medicare Encounter Data File User Guide,* Chronic Conditions Warehouse Virtual Research Data Center, June 2025, Version 2.10 https://www2.ccwdata.org/documents/10280/19002246/ccw-medicare-encounter-data-user-guide.pdf, accessed 03/04/2026.

[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.

2. NPI Edit and Imputation Methods

One important step in preparing the MA encounter data files for public reporting is to ensure that each record contains the best available NPI. The accuracy of NPI is critical because this variable is used to identify duplicate encounter records, as discussed in more detail in Section 3, and to attach the CMS provider ID in the inpatient hospital data file, as described in Section 4. This section describes the methods we used to edit and impute NPIs.

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).

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 variables to identify the NPI value. CMS' Encounter Data System does not perform editing of NPI variables when processing records, therefore we perform additional edits and imputations as described below.

a. NPI Edits

The first step in editing the NPI variables involves checking the fidelity of the variables. Specifically, we check that the NPI:

- Has a non-missing value.
- Has a valid length and structure, i.e., it has 10 characters and begins with the number 1.
- Exists in the NPPES provider reference table.
- Provider type reflects an organization rather than an individual for the organizational NPI variables.[4]

Any NPI variable value that does not meet these criteria is recoded to missing. Then, for the Carrier and DME files, we create a unified NPI variable by coalescing three NPI variables used in those files.[5]

b. NPI Imputation Methods

To impute missing NPI values, we group records using a multi-variable key edit that is used to identify unique encounter records but exclude the NPI variable. Then, we check if there are other records within the key edit that have a *single* valid NPI. Figure 1 demonstrates this imputation method.

[4] The CMS submission rules for encounter records 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. For additional details, download the zip file found here: https://downloads.cms.gov/files/2017-HPMS-Q4.zip (accessed 03/04/2026).

[5] The three NPI variables used to create a unified NPI for the Carrier and DME files include the line level rendering NPI, the header level rendering NPI variable, and the header level organizational NPI variable.

Figure 1. Illustration of NPI Imputation Method

Multi-Variable Key Except NPI	Edited NPI	Imputed NPI	Imputed 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

As shown in Figure 1, all the records have the same values for all the 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.

One limitation of this NPI imputation method is that it does not resolve incorrect assignments of organizational NPIs to a particular encounter record by MA Organizations (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 2022. 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 5.5% of records in 2023. 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, 2016 and 2022

NPI Value Status	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Inpatient Hospitals				
Total Records	4.3	100.0%	6.8	100.0%
Invalid NPI Values	0.01	0.2%	0.01	0.2%
Valid NPI Values	4.3	99.8%	6.8	99.8%
Original	4.3	99.7%	6.8	99.6%
Imputed	0.003	0.1%	0.01	0.2%
Skilled Nursing Facilities				
Total Records	1.5	100.0%	2.2	100.0%
Invalid NPI Values	0.01	0.5%	0.01	0.4%
Valid NPI Values	1.5	99.5%	2.2	99.7%
Original	1.4	99.4%	2.2	99.6%
Imputed	0.001	0.1%	0.001	0.04%
Outpatient Facilities				
Total Records	349.8	100.0%	780.6	100.0%
Invalid NPI Values	0.7	0.2%	1.1	0.1%
Valid NPI Values	349.1	99.8%	779.5	99.9%
Original	349.0	99.8%	778.2	99.7%
Imputed	0.1	0.03%	1.3	0.2%
Carrier and DME Services				
Total Records	972.6	100.0%	1,877.6	100.0%
Invalid NPI Values	93.8	9.6%	103.0	5.5%
Valid NPI Values	878.8	90.4%	1,774.6	94.5%
Original	877.2	90.2%	1,772.1	94.4%
Imputed	1.6	0.2%	2.5	0.1%

SOURCE: 2016 and 2022 Inpatient hospital, SNF, outpatient facility, carrier, and DME MA encounter base and line data files

3. Unique Service Identification

Accurate identification of unique encounter records is essential for producing valid utilization measures from MA encounter data that are comparable to OM utilization measures. CMS guidance in the Encounter Data Submission and Processing Guide describes CMS' approach to identifying and rejecting duplicate encounter records for each encounter file type. However, even after applying these system edits, encounter data may include duplicate encounter records that should have been eliminated. In addition, for some encounter file types, that guidance relies on multiple variables that are not available to the external research community. Therefore, we have developed an alternative approach to identifying and deleting duplicate encounter records using a more limited set of variables that can be reproduced externally. This section describes the methods we used to identify, evaluate, and resolve duplicate records to support consistent and reproducible public reporting.

It should be noted that eliminating duplicate encounter records only serves to ensure that utilization counts are more accurate, however the dataset may still include multiple records that relate to a single service. 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 records will remain in the data as they are not duplicates. Whether those two records are counted separately, or considered a single service, will be dependent on the purpose of a specific analysis. In addition, the methodology we developed to resolve duplicate records did not take into account variables such as diagnosis codes, and therefore deleted records may contain information that are of interest to researchers.

a. Inpatient and SNF Duplicate Service Records

CMS recommends using a 5-key edit (beneficiary ID, encounter start date, encounter end date, type of bill, and organizational NPI) to identify duplicate records in the inpatient hospital and SNF encounter data files (see the Chronic Conditions Warehouse Virtual Research Data Center Medicare Encounter Data File User Guide).

We apply the 5-key edit using the imputed NPI variable (discussed above) that removes invalid and individual provider IDs from the organizational NPI column. As shown in Table 2, our analysis identified 0.4% of encounter records that had the same 5-key edit between 2016 and 2022.

To select a single retained record, we identified the encounter record with the latest processing date. If multiple records report the latest processing date, then we selected a single retained record at random.

Table 2. Duplicate Inpatient Hospital and SNF MA Encounter Records Removed from Analytical Files, 2016 and 2022

Record Count Category	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Inpatient Hospitals				
Original Record Count	4.28	100.0%	6.84	100.0%
Single Records per 5-Key Edit	4.26	99.7%	6.81	99.6%
Multiple Records per 5-Key Edit	0.01	0.3%	0.03	0.4%
Encounter Records Dropped	0.01	0.2%	0.01	0.2%
Encounter Records Kept	0.01	0.2%	0.01	0.2%
Deduped Record Count	4.27	99.8%	6.83	99.8%
Skilled Nursing Facilities				
Original Record Count	1.46	100.0%	2.19	100.0%
Single Records per 5-Key Edit	1.45	99.7%	2.18	99.9%
Multiple Records per 5-Key Edit	0.005	0.3%	0.001	0.1%
Encounter Records Dropped	0.002	0.2%	0.001	0.0%
Encounter Records Kept	0.002	0.2%	0.001	0.0%
Deduped Record Count	1.46	99.8%	2.18	100.0%

SOURCE: 2016-2022 Inpatient hospital and SNF MA encounter base records.

b. Carrier, DME, and Outpatient Facility Duplicate Service Records

i. Identifying Duplicate Service Records

The process we use to identify duplicate encounter records for carrier, DME, and outpatient facility encounter data files is different than the process we used for inpatient hospital and SNF encounter data. As noted in the Encounter Data Submission and Processing Guide, CMS uses payment variables as part of the logic to flag and reject duplicate encounter records for carrier, DME, and outpatient service settings. However, encounter data files available to researchers do not include payment variables which will result in numerous records in the MA RIFs that look like duplicates, solely due to the payment fields. Therefore, we created an approach that external researchers can follow to resolve duplicate records without relying on the payment fields. Figure 3 presents the list of variables we used to develop a key to flag duplicate encounter records in the carrier/DME and outpatient facility encounter data files.

Figure 3. Data Fields Used in Key to Identify Duplicate Records

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.

In applying this key, we also wanted to align with the two instances where CMS applies criteria to bypass the duplicate records check. Specifically, CMS bypasses duplicate records with select HCPCS modifiers (see Encounter Data Submission and Processing Guide, Table 6.7). In addition, CMS bypasses duplicate records that represent ambulatory surgical center (ASC) services and meet the following criteria:

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

Table 3 documents the number of records that are bypassed for the duplicate record checks (bypass records), that are identified as single records and that are identified as duplicates. The carrier, DME, and outpatient facility encounter data have a significantly higher percentage of records that we identify duplicates relative to inpatient hospital and SNF encounter files. This result is likely because we are using modified keys that exclude payment variables.

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

Dedup Process Status	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Outpatient Facilities				
Total	349.8	100.0%	780.6	100.0%
Bypass Records	12.8	3.7%	26.1	3.3%
Single Records per Key	307.7	88.0%	657.4	84.2%
Multiple Records per Key	29.3	8.4%	97.0	12.4%
Carrier and DME Services				
Total	972.6	100.0%	1877.6	100.0%
Bypass Records	21.9	2.3%	50.9	2.7%
Single Records per Key	895.0	92.0%	1722.9	91.8%
Multiple Records per Key	55.7	5.7%	103.8	5.5%

SOURCE: 2016 and 2022 Carrier, DME, and outpatient facility MA encounter base and line records.

ii. Resolving Duplicate Service Records

Given that the data show a higher percentage of records that are duplicates relative to the inpatient hospital and SNF encounter data, we apply additional criteria to select a single retained record. Specifically, we use the following methods, applied hierarchically:

- Method 1: Latest Processing Date Resolution. This method determines if a single record contains the latest processing date within the 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 key). 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 key group, then one record is selected at random.

Table 4 shows the percentage of unduplicated records that fall into each deduping method category. For carrier and DME encounter data, over 85% of duplicate records across the data years are deduped using either the latest processing date or the claim frequency code methods, with the date method (Method 1) representing the largest proportion of records (approximately 72%). For the outpatient facility data, the predominant deduping methods (nearly 90% of records) use either the latest processing date (Method 1) or the same ENC_JOIN_KEY value (Method 3). Additionally, the percentage of duplicate records resolved using these methods are more evenly split, with the “latest date” method (Method 1) averaging 46% of records and the “same ENC_JOIN_KEY” method (Method 3) averaging about 42% between 2016 and 2022. Lastly, the outpatient facility data’s deduping algorithm does not use the “claim frequency code = 7” method (Method 2), which suggests that this file type submits multiple encounter lines as replacements for prior encounter records within the key groups.

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

Dedup Process Status	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Outpatient Facilities				
Total	140.6	100.0%	97.0	100.0%
Method 1: Last create date	13.0	9.2%	44.3	45.7%
Method 2: Claim Frequency Code=7	--	--	--	--
Method 3: Same ENC_JOIN_KEY	12.5	8.9%	43.1	44.5%
Method 4: Pure Duplicates	2.5	1.8%	5.1	5.2%
Method 5: Random Selection	1.3	0.9%	4.5	4.7%
Carrier and DME Services				
Total	55.7	100.0%	103.8	100.0%
Method 1: Last create date	40.4	72.6%	72.3	69.6%
Method 2: Claim Frequency Code=7	7.7	13.8%	16.8	16.1%
Method 3: Same ENC_JOIN_KEY	2.3	4.1%	4.1	3.9%
Method 4: Pure Duplicates	3.4	6.1%	7.0	6.8%
Method 5: Random Selection	1.9	3.5%	3.7	3.5%

SOURCE: 2016 and 2022 Carrier, DME, and outpatient facility MA encounter base and line records.

Table 5 shows the percentage of duplicate records that we deleted from the final carrier, DME, and outpatient facility data. 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 records that have the same ENC_JOIN_KEY value (Method 3).

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

Record Count Category	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Outpatient Facilities				
Original Record Count	349.8	100.0%	780.6	100.0%
Single Records per Key and Bypass Records	320.5	91.6%	683.6	87.6%
Multiple Records per Key	29.3	8.4%	97.0	12.4%
Encounter Records Dropped	8.9	2.5%	28.9	3.7%
Encounter Records Kept	20.4	5.8%	68.1	8.7%
Deduped Record Count	341.0	97.5%	751.6	96.3%
Carrier and DME Services				
Original Record Count	972.6	100.0%	1,877.6	100.0%
Single Records per Key and Bypass Records	916.9	94.3%	1,773.8	94.5%
Multiple Records per Key	55.7	5.7%	103.8	5.5%
Encounter Records Dropped	27.2	2.8%	51.0	2.7%
Encounter Records Kept	28.5	2.9%	52.8	2.8%
Deduped Record Count	945.4	97.2%	1,826.6	97.3%

SOURCE: 2016 and 2022 Carrier, DME, and outpatient facility MA encounter base and line records.

4. 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 we used to attach the CCN to the encounter data using the MA organizational NPI variable.

a. Background

The CCN is a provider identifier that CMS assigns to institutional providers to classify which OM 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 OM claims data also includes NPI like the MA encounter data, we use the CCN to identify hospital type for OM data products such as the CMS Program Statistics (CPS), since the NPI does not allow users to delineate facility type. 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 OM data analyses.

b. 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 created to associate an organizational NPI to a CCN. We use two data sources to construct this crosswalk:

- NPI-CCN relationships found in the OM institutional claims data (Part A and B); and
- Provider Master Index (PMI) NPI-CCN crosswalk that combines information from various CMS provider data systems. [6]

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

To construct the NPI-CCN crosswalk from the OM 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 2022. 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 relationships. 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 create annual data sets of NPI-CCN combinations that were in effect for a calendar year using NPI-CCN effective dates. Some PMI records did not have valid provider 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" and assign them to all data years. Finally, we merge the OM 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 OM and PMI longitudinal crosswalks, only on the OM longitudinal crosswalk, and only on the PMI longitudinal crosswalk. Across the data years, approximately 70% of NPI-CCN combinations exist in both the OM and PMI source data for at least one data year. Those combinations where no data years match between the data sources 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 for our CCN assignment algorithm. See Appendix B for more details on how to access the NPI-CCN crosswalk we developed to map CCN codes to MA encounter hospital NPIs and the crosswalk's file layout.

[6] 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.

Table 6. Distribution of NPI-CCN Combinations, by Crosswalk Type

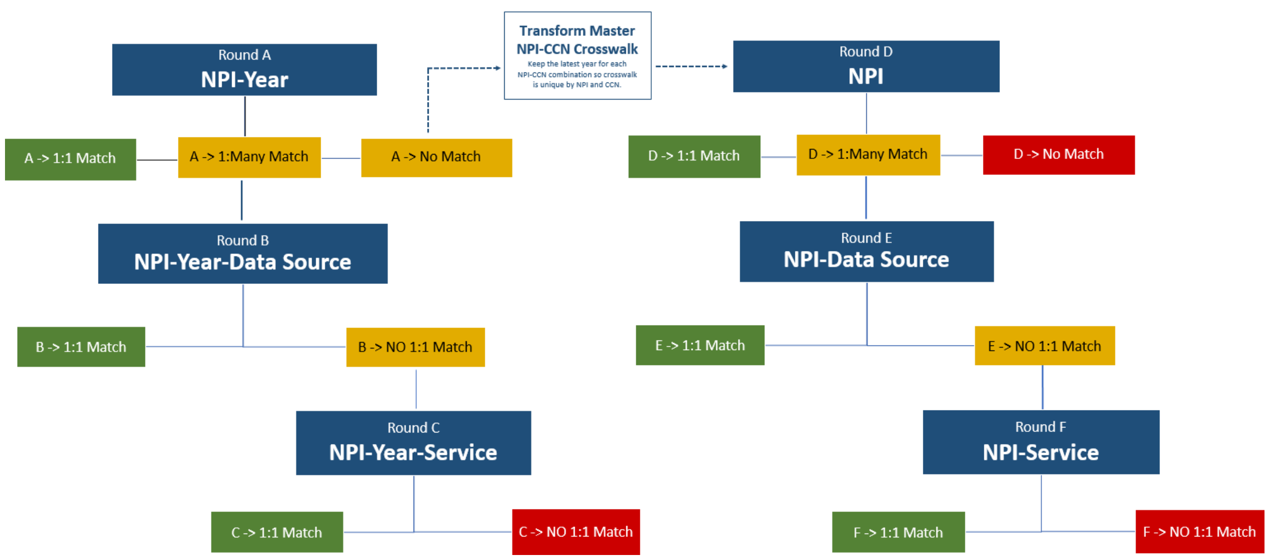
OM-PMI NPI-CCN Comparison Categories	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Total	106,744	100.0%	82,711	100.0%
All Years Match	49,111	46.0%	41,569	50.3%
Some Years Match	25,009	23.4%	19,795	23.9%
No Match = OM Only	476	0.4%	31	0.0%
No Match = PMI Only	24,513	23.0%	19,755	23.9%
No Match = Both	20	0.02%	9	0.01%
No Years Match	32,624	30.6%	21,347	25.8%
No Match = OM Only	1,529	1.4%	6,170	7.5%
No Match = PMI Only	31,071	29.1%	15,172	18.3%
No Match = Both	24	0.02%	5	0.01%

SOURCE: 2016 and 2022 OM Part A and Part B institutional claims and Provider Master Index NPI-Legacy ID Table.

c. NPI-CCN Linkage Algorithm

Figure 4 below is a diagram that illustrates the steps we used to assign a CCN to MA inpatient hospital encounter records.

Figure 4. Diagram of MA CCN Assignment Algorithm



The algorithm starts by merging 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 most certainty in these matches since they do not require information from the MA encounter record. Figure 5 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 5. 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 assign a unique CCN to the record (Round B). Figure 6 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 6. 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 7 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 7. 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 8 shows, the algorithm then addresses the records that did not find a match in previous rounds. 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 merges on the CCN to the MA encounter records only by NPI, excluding year, to check if the MA encounter NPI exists in any of the years in the NPI-CCN crosswalk. Figure 8 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 8. Round "D" Example of CCN Assignment

MA Encounter Data

Round D

Encounter Join Key	Year	NPI	MA Data Source	Taxonomy Service Category	Taxonomy Code	Taxonomy Description	CCN
332986107	2016	1013906221	IP	IPPS	282N00000X	General Acute Care Hospital	050305
337222864	2017	1013906221	IP	IPPS	282N00000X	General Acute Care Hospital	050305
328047172	2018	1013906221	IP	IPPS	282N00000X	General Acute Care Hospital	050305
337168102	2019	1013906221	IP	IPPS	282N00000X	General Acute Care Hospital	050305

Master NPI-CCN Crosswalk

Round D

Year	NPI	CCN Data Source	CCN Service Category	CCN	CCN Description
2016	1013906221	IP	IPPS	050305	Short-Term Hospitals
2017	1013906221	IP	IPPS	050305	Short-Term Hospitals
2018	**1013906221**	**IP**	**IPPS**	**050305**	**Short-Term Hospitals**

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 assigned CCN is consistent with the MA encounter inpatient hospital bill type.

Across all data years, 96% of records on average 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, 89% 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 the most certainty for its match assignment, i.e., we only use the reported organizational NPI. The more data fields 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 (less than 0.5% in 2016 and 2022) 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, a significant proportion 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. In any event, the non-match rate is declining over time, moving from 3.9% in 2016 to 2.3% in 2022.

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

CCN Merge Result	2016 N (Millions)	2016 % of Total	2022 N (Millions)	2022 % of Total
Total	4.3	100.0%	6.8	100.0%
Match	4.1	96.1%	6.7	97.7%
Match, Same	4.1	95.8%	6.7	97.6%
Match-Round A, Same	3.8	88.2%	6.1	90.0%
Match-Round B, Same	0.2	3.8%	0.2	3.0%
Match-Round C, Same	0.2	3.8%	0.3	4.6%
Match-Round D, Same	0.0	0.0%	0.0	0.1%
Match-Round E, Same	--	--	0.0	0.0%
Match-Round F, Same	--	--	0.0	0.0%
Match, Different	0.0	0.3%	0.0	0.1%
Match-Round A, Different	0.0	0.3%	0.0	0.1%
Match-Round C, Different	0.0	0.0%	0.0	0.0%
Match-Round D, Different	0.0	0.0%	0.0	0.0%
Match-Round F, Different	--	--	--	--
No Match	0.2	3.9%	0.2	2.3%
No Match, 1:many	0.1	2.7%	0.1	1.1%
No Match, Missing	0.0	1.1%	0.1	1.2%

SOURCE: 2016 and 2022 Inpatient hospital MA encounter base records.

5. Exclusion of Home Health Agency Medicare Advantage Encounter Records in Public Reporting

Home health (HH) encounter records come with unique challenges that are not found with other encounter record types. Our internal analysis determined that the unique structure of HH encounter records does not allow for accurate comparisons across MA plans, at the county or state levels, or to OM, and therefore HH encounter records are not included in our public release.[7] Our analysis focuses on the differences between home health claims and records collected for MA and OM beneficiaries in order to highlight the challenges in evaluating MA records. While many of the findings we share below compare MA to OM, these findings also persist across plans and geographic areas. A summary of our analysis is below.[8]

a. Background

Medicare Advantage home health encounter records are distinct from OM claims in two key ways that limit CMS's ability to publicly release aggregated data. First, there are different claim submission and processing requirements in OM and MA, which impact the quantity and type of records that we see in the two programs. Second, there are differences between the reimbursement methods to home health providers in the MA program and the OM program. In OM, there are strict rules around how the government pays home health providers that impact claim

[7] Please note that the underlying purpose behind collection of MA encounter data is for risk adjustment, rather than payment to providers as is done in fee-for-service. Therefore, it is not unexpected that the structure of MA encounter data would be different from FFS claims data.

[8] We did not evaluate the data for completeness or quality, but rather focused our analysis on the structural differences between FFS claims and MA encounter records. For a discussion of the completeness of MA encounter HH data, please see https://www.medpac.gov/wp-content/uploads/2025/06/Jun25_Ch3_MedPAC_Report_To_Congress_SEC.pdf

submissions. In comparison, under MA, health plans have the flexibility to reimburse home health providers based on differing contractual agreements, which impacts encounter record submissions such that they may not be comparable to OM claims.

To the first point above, MA plans are required to submit a record to CMS for each encounter that a beneficiary has had with a provider. CMS has a set of criteria for which encounter records will be accepted or rejected based on defined data elements. CMS does not assess whether each encounter is the final action of record for the plan as the encounter can be paid, denied, or still pending adjudication by the plan; rather, all of these records are accepted by CMS. In the case of MA carrier, outpatient, DME, and home health encounter records, these criteria enable records for duplicative services to pass through the duplicate checking criteria. As described previously in this paper, we have been able to address issues of duplicative services in the MA carrier, outpatient, and DME settings by using a methodology that relies on the principle that same-day services should not repeatedly occur. However, home health is unique in that “exact” same-day services are not unusual for all Medicare beneficiaries. For example, a patient may receive two skilled nursing or home health aide visits in one day and this scenario occurs in both the OM claims and MA encounter data; however due to the unique nature of how MA encounter records are collected, we are unable to determine whether an MA encounter HH record is a true duplicate or multiple distinct services.

Second, because the contractual relationship between each home health provider and plan is unique, and the plans are not required to describe these contractual arrangements to CMS, it makes it difficult to standardize the definition of episodes of care in a home health setting for MA similar to OM. In OM, an episode of care for home health is currently defined as a 30-day episode under the Patient-Driven Groupings Model (PDGM) beginning with the start of care date.[9] Home health agencies (HHAs) are paid one lump sum for all services provided during the episode, regardless of quantity.[10] In contrast, under MA, each contractual arrangement between a plan and a home health agency may be different, therefore an episode of care cannot be standardized across plans and providers. For example, MA plans may pay for each service individually (such as a nursing or physical therapy visit) or they may pay for an episode similar to OM in that all services are covered under a lump sum for a specific time period.[11]

b. Analysis

We performed two analyses related to home health records to evaluate whether we could publicly release MA home health data – 1) checking for duplicate records, and 2) assessing the episode structure.

First, we analyzed home health records from both MA and OM to assess the presence of duplicates at the header level, line level, and visit/service level. Duplicate header records present instances where an MA plan may be submitting new standalone records, amended records, or duplicated records. We identified duplicate header records by creating a 5-key edit to detect identical record submissions for beneficiaries. The 5-key edit includes Bene ID, From Date, Through Date, Bill Type, and Organizational NPI (MA)/ CCN (OM).[12] To determine whether a duplicate header record was an exact duplicate, we extended the 5-key edit to include line-level details, specifically HCPCS code, revenue center codes, revenue center from date, revenue center through date, and revenue center unit count.

---

[9] Prior to 2020, HH services were paid under the Home Health Prospective Payment system and episodes were 60 days in length.

[10] In FFS, in certain cases, HHAs may be paid less than the episode amount if fewer than 5 services were provided. This occurs in less than 10% of episodes. HHAs may also qualify for an additional outlier payment if the cost of the episode rises above a certain level.

[11] FFS has a limit of 8 hours of skilled nursing and home health aide services, combined, per day, while MA does not. MA plans may choose to provide care beyond 8 hours on a given day; therefore we are unable to accurately identify a duplicate record versus multiple legitimate services provided in one day for MA beneficiaries.

[12] Only final action encounter records are included in this analysis.

Figure 9 provides an example of a header duplicate and an exact line-level duplicate (i.e., all values at the line level match perfectly):

Figure 9. Example of Header and Exact Line-Level Duplicate Records

5 Part Key: Header Duplicate Example

Bene ID	Clm Id	Bill Type	From Date	Thru Date	Org NPI
123	456	32x	01/01/2020	01/05/2020	123456789
123	789	32x	01/01/2020	01/05/2020	123456789

Exact Duplicate Example:

Bene ID	Clm Id	Bill Type	From Date	Thru Date	Org NPI	Line Number	HCPCS	Rev Cd	Rev From Date	Rev Through Date	Rev Unit Count
123	456	32x	01/01/2020	01/04/2020	123456789	1	G0151	0421	01/02/2025	01/02/2025	1
123	456	32x	01/01/2020	01/04/2020	123456789	2	G0299	0551	01/03/2025	01/03/2025	1
123	456	32x	01/01/2020	01/04/2020	123456789	3	G0153	0441	01/04/2025	01/04/2025	1
123	789	32x	01/01/2020	01/04/2020	123456789	1	G0151	0421	01/02/2025	01/02/2025	1
123	789	32x	01/01/2020	01/04/2020	123456789	2	G0299	0551	01/03/2025	01/03/2025	1
123	789	32x	01/01/2020	01/04/2020	123456789	3	G0153	0441	01/04/2025	01/04/2025	1

i. 5-Key Edit Header and Exact Duplicate Line-Level Analysis Results

Table 8 shows the results of our analysis to identify header and line-level exact duplicates for MA. In the first step we applied the 5-Key Edit at the header level and we observed that MA plans submitted duplicate MA encounter records on average 11.1% of the time for the period of 2016-2022. In contrast, OM had zero header level claims with the same 5-Key Edit. As described above, the original 5-Key Edit was then expanded for the MA encounter records to check the line-level information for exact duplicates. With the line-level details included, we found that the exact duplicate rate was on average 3.1% for the same period, 2016-2022.

Table 8. Duplicate Header and Exact Duplicate Line-Level Results (Medicare Advantage only)

Year	Total Record Counts	5-Key Edit Header Duplicate Counts	5-Key Edit Header Duplicate Percent	Exact Duplicate Line-Level Counts	Exact Duplicate Line-Level Percent
2016	5,694,748	694,226	12.2%	202,725	3.6%
2017	6,601,036	737,770	11.2%	217,649	3.3%
2018	7,530,250	713,718	9.5%	200,422	2.7%
2019	9,541,462	1,079,927	11.3%	323,425	3.4%
2020	11,392,985	1,459,329	12.8%	346,772	3.0%
2021	13,713,616	1,324,274	9.7%	314,096	2.3%
2022	14,135,638	1,526,882	10.8%	446,318	3.2%

Note: OM results were omitted from Table 8 since there were no duplicate header level claims.

ii. Visit/Service Level Results

CMS analyzed visit/service level home health records from both MA and OM to check for same-day visits/services. Table 9 shows that, on average, from 2016-2022, OM had a 1.6% duplication rate of same-day visits/services, whereas MA had a 6.9% duplication rate. A large portion of the same-day services were concentrated in revenue center code 0551 (skilled nursing) in both MA and OM, but the reasons behind this may differ. In OM, for most service types other than home health, exact duplicate records would typically be denied by the Medicare Administrative Contractors (MACs) that process claims due to Medically Unlikely Edits or similar enforcement mechanisms. However, since it's possible for a beneficiary to receive multiple visits by the same service discipline on the same day under the home health benefit, OM home health claims that have duplicate information can be considered distinct services rather than true duplicates. In MA however, as discussed previously, CMS allows for duplicate encounters to be submitted, regardless of whether the encounter was paid or denied by the MA plan. Therefore, the OM results indicate that same-day services can happen under OM, whereas the MA results are likely indicative of a combination of same-day services and encounter header record duplication.

Table 9. Visit/Service Duplicate Results

Year	OM Total Line Record Counts	OM Visit/Service Duplicate Counts	OM Visit/Service Duplicate Percent	MA Total Line Record Counts	MA Visit/Service Duplicate Counts	MA Visit/Service Duplicate Percent
2016	124,967,176	2,530,459	2.0%	32,035,776	2,158,672	6.7%
2017	120,632,032	2,058,344	1.7%	39,208,227	2,601,492	6.6%
2018	119,467,484	1,891,772	1.6%	47,053,405	2,632,316	5.6%
2019	115,178,553	1,679,296	1.5%	56,985,051	4,473,987	7.9%
2020	105,659,751	1,544,157	1.5%	60,264,544	4,781,432	7.9%
2021	98,364,879	1,328,309	1.4%	68,613,882	4,260,432	6.2%
2022	89,427,951	1,251,126	1.4%	71,603,086	5,021,048	7.0%

iii. Episode Comparison

We then analyzed home health records from both MA and OM to examine episodes of care as shown in Tables 10 and 11. On average, OM episodes tended to fall within defined time frames by the payment system - typically between 0-30 days (under the current PDGM system) and 0-60 days (under the former HH Prospective Payment System). In contrast, MA encounter records were more likely to reflect only a single day of care, show greater variability, and not follow a consistent episode billing pattern.

For example, on average, 32% of MA encounter records had a day range for the service of 0 days (i.e., same- day service) and another 34% of records had a day range for services between 1-10 days. In contrast this only occurs at a rate of 1% and 6% respectively in OM.

Table 10. Distribution of Length of Home Health Care in Medicare Advantage Home Health Header Claims

Day Range	2016	2017	2018	2019	2020	2021	2022
0*	29%	29%	31%	32%	32%	35%	33%
1-10	35%	36%	35%	35%	33%	32%	34%
11-20	12%	12%	12%	12%	13%	12%	12%
21-30	13%	12%	12%	12%	19%	20%	20%
31-40	2%	2%	2%	2%	0%	0%	0%
41-50	1%	1%	1%	1%	0%	0%	0%
51-60	7%	7%	7%	6%	1%	0%	0%
60+	0%	0%	0%	0%	0%	0%	0%
Other**	0%	0%	0%	0%	0%	0%	0%

NOTE: * ”0” Denotes same-day services. ** ”Other” denotes a few claims where the claim through date occurred before the claim from date.

Table 11. Distribution of Length of Home Health Care in Original Medicare Home Health Header Claims

Day Range	2016	2017	2018	2019	2020	2021	2022
0*	1%	1%	1%	0%	1%	1%	1%
1-10	5%	5%	5%	5%	6%	7%	6%
11-20	10%	10%	10%	9%	13%	14%	13%
21-30	11%	12%	12%	12%	77%	83%	83%
31-40	8%	8%	9%	9%	1%	0%	0%
41-50	7%	7%	7%	7%	1%	0%	0%
51-60	58%	57%	57%	57%	5%	0%	0%

NOTE:* “0” denotes same-day services.

iv. Summary of Duplicate and Episode Analyses

The results of these analyses hinder our ability to publish MA home health data at an aggregated level either for comparison to OM or for comparison across MA plans or geographic levels. First, a high rate of duplication at the header level of 11.1% creates uncertainty around whether these duplicate header records in MA represent adjudicated encounters with modified service lines (additions or deletions) or entirely distinct encounters. In contrast, OM has no duplication at the header level. Second, the presence of exact duplicates at the line level reveals that MA plans may be submitting identical records for, on average, 3.1% of all entries. Third, both programs permit the submission of duplicated same-day visit/service lines. Therefore, CMS cannot assume that an MA encounter record that looks like an exact duplicate is in fact a duplicate and not a unique visit/service. Finally, the results of the episode analysis highlight the differences among contractual arrangements between providers and MA plans, thereby precluding accurate comparisons of episode/utilization counts across MA plans, geographic levels, or between MA and OM.

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

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 03/04/2025).

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 9/30/2025).

Appendix B: NPI-CCN Crosswalk File Layout and Access Instructions

File Layout

The NPI-CCN crosswalk is constructed using the 2016-2022 OM Part A and Part B institutional claims data and the Provider Master Index (PMI) provider ID data. These crosswalk data are used to assign CCN codes to NPI values in the MA inpatient hospital encounter data as described in Section 4. The data contain NPI-CCN combinations for each year between 2016 and 2022 and are restricted to NPIs that appear in the inpatient hospital Medicare Advantage encounter data files during that time period.

Appendix B.1. NPI-CCN Crosswalk File Layout

Variable Name	Variable Type	Variable Length	Variable Description
year	Num	8	Year in which NPI-CCN combination appears
npi	Char	10	National Provider Identifier
ccn	Char	20	CMS Certification Number
ccn_desc	Char	50	CMS Certification Number description
ccn_datasrc_code	Char	3	CCN MA encounter data source category code (used for Round B/E matching)
ccn_datasrc_desc	Char	25	CCN MA encounter data source category description (used for Round B/E matching)
ccn_chtsrvc_code	Char	4	CCN MA encounter common hospital service category code (used for Round C/F matching)
ccn_chtsrvc_desc	Char	35	CCN MA encounter common hospital service category description (used for Round C/F matching)
mst_flg_ffs	Num	8	Master crosswalk flag | NPI-CCN combination found in OM
mst_flg_pmi	Num	8	Master crosswalk flag | NPI-CCN combination found in PMI
mst_yrs_total	Num	8	Master crosswalk years count | Total number of years NPI-CCN combination appears in crosswalk
mst_yrs_ffs	Num	8	Master crosswalk years count | Number of years NPI-CCN combination appears in OM data only
mst_yrs_pmi	Num	8	Master crosswalk years count | Number of years NPI-CCN combination appears in PMI data only
mst_yrs_both	Num	8	Master crosswalk years count | Number of years NPI-CCN combination appears in both OM and PMI
mst_yrs_comp_code	Num	8	Master crosswalk data source comparison code
mst_yrs_comp_desc	Char	40	Master crosswalk data source comparison description

Access Instructions

For users with access to the Medicare Advantage encounter data research identifiable files (RIF) wishing to use our generated NPI-CCN crosswalk to reproduce the NPI-CCN linkage algorithm, the crosswalk can be downloaded here: https://download.cms.gov/encounter_data/maed_xnc_ry25_p01_v10_dyt22_xwlk.zip

Appendix C: Variables Used to Identify Pure Duplicates When Resolving 5-Key Edit Duplicate Carrier, DME, and Outpatient Records

The variables listed below are used in Method 4 of the algorithm used to dedup encounter records that belong to 5-key edit group with multiple records. See Section 3 for more details about the methods used to dedup encounter records.

Appendix Table C.1. Carrier and DME Variables Used to Identify Pure Duplicates When Resolving Service-Event 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 C.2. Outpatient Variables Used to Identify Pure Duplicates When Resolving 5-Key Edit 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