

Missing and Invalid Diagnosis Codes in 2017

October 2019

Brief #5132

2017 TAF version 2

TAF data quality brief – Service use information

This analysis focused on 46 states, the District of Columbia, and Puerto Rico. Mississippi, Missouri, Montana, and Nebraska were excluded from the analysis.

Key Findings

- In the T-MSIS Analytic File, diagnosis codes that denote beneficiaries' medical conditions are on the header records in the medical claims files (IP, LT, and OT). In this analysis, we evaluated the extent to which each state's claims were completely coded with at least one valid diagnosis code. We also calculated the mean number of valid diagnosis codes on claims in the IP and LT files to identify states with potentially incomplete diagnosis code data.
- The vast majority of IP claims had a valid ICD-10 primary diagnosis code. On average, states submitted six to eight diagnosis codes per IP claim. One state had unusable diagnosis code data in the IP file (Table 1).
- Although there were a handful of exceptions, most states had a valid ICD-10 primary diagnosis code on the majority of their LT claims. States generally submitted two to four diagnosis codes per LT claim. Three states had unusable or highly concerning diagnosis code data (Table 2).
- Valid ICD-10 primary diagnosis codes are available on most OT claims. It is noteworthy that valid ICD-10 primary diagnosis codes were frequently included on claims in the OT file submitted by providers who were not expected to have access to reliable diagnostic information about patients (for example, claims for medical supplies, prosthetic equipment, or non-emergency medical transportation services). Users of the T-MSIS Analytic File may want to exercise caution in using the diagnosis codes on these claims even if they appear to be valid ICD-10 codes. One state had unusable diagnosis code data in the OT file (Table 3).

Background

Providers submit diagnosis codes on medical claims that can be used to analyze the prevalence of different medical conditions. Both institutional providers (such as hospitals, nursing facilities, and clinics) and professional providers (such as physicians, other clinical professionals, and ambulances) submit claims on standardized forms that allow them to record multiple diagnosis codes.¹ Diagnosis codes are not, however, submitted on pharmacy claims. Before October 1, 2015, providers used diagnosis codes from the International Classification of

¹ Institutional claims are often referred to as "UB-04 claims" when submitted in paper form or as "837I claims" when submitted in electronic form. Professional claims are referred to as "CMS-1500 claims" when submitted in paper form or "837P" when submitted in electronic form. All of these formats allow providers to include multiple diagnosis codes on a given claim.

Diseases, 9th Revision, Clinical Modification (ICD-9-CM). For services provided on or after October 1, 2015, providers are expected to use diagnosis codes from the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM), which has nearly five times as many possible codes at a far greater level of detail than did the ICD-9-CM.

In the T-MSIS Analytic Files (TAF), all diagnosis codes are located on the header record, which contains information that relates to the claim as a whole. The maximum number of diagnosis codes for a single claim in the TAF varies by file. TAF inpatient (IP) records may have up to 12 diagnosis codes per claim, long-term care (LT) records may have up to 5 diagnosis codes per claim, and other services (OT) records may have up to 2 diagnosis codes per claim.² Complete and accurate diagnosis code information is essential, as these fields help TAF users to better understand the underlying health conditions and needs of the populations they are studying.³

For the 2017 TAF, we expect all claims in the IP and LT claims files to have a valid ICD-10 diagnosis code in the primary diagnosis code field (DGNS_1_CD).⁴ For the OT claims file, we do not expect that all claims will have diagnosis codes. For certain types of services (for example, medical supplies, prosthetic equipment, or non-emergency medical transportation [NEMT] services), states may not require providers to submit diagnosis codes on the claims because the provider billing for these services may not be in the best position to know and record an accurate diagnosis. Many states also do not require providers to submit diagnosis codes on dental claims. However, if the provider includes any diagnosis codes on those claims, CMS instructs states to pass through those diagnosis codes to T-MSIS as reported on the claims form even if the diagnosis codes are inaccurate or invalid. For this reason, TAF users may want to exercise caution in using diagnosis codes on certain types of OT claims even if they appear to be valid ICD-10 codes.

² In the IP file, diagnosis codes are captured in the fields DGNS_1_CD through DGNS_12_CD. In the LT file, diagnosis codes are captured in the fields DGNS_1_CD through DGNS_5_CD. In the OT file, diagnosis codes are captured in the fields DGNS_1_CD and DGNS_2_CD. In all three claims files, the first diagnosis code field (DGNS_1_CD) contains the primary diagnosis code, and the additional fields contain other diagnosis codes submitted on the claim. The IP and LT files also include a separate field for the admitting diagnosis code (ADMTG_DGNS_CD), which captures the diagnosis code recorded by the provider when a patient is admitted to a facility. If a claim includes more diagnosis codes than are accommodated in T-MSIS, the source data for TAF, for its given file type, then some diagnosis code information may be missing from the T-MSIS and TAF data systems. This is most likely to be an issue in the OT file because states can only report up to two diagnosis codes per claim in their T-MSIS OT files.

³ Although diagnosis codes typically capture beneficiaries' medical conditions, diagnosis codes are sometimes used to provide additional information about the services received during a medical encounter. For example, on evaluation and management claims, diagnosis codes can capture the patient's age or whether he or she required an examination to participate in sports or to begin a new job.

⁴ Some states may not require nursing facilities to report diagnosis codes on their claims if the diagnoses are tracked separately through the federally required Preadmission Screening and Resident Review process or within a different clinical information system. In these cases, the data may be missing from the T-MSIS and TAF claims files.

Methods

This analysis is based on the 2017 TAF.⁵ We assessed the extent to which diagnosis codes are available on fee-for-service (FFS) claims and on managed care encounter records in the IP, LT, and OT claims files.⁶ We excluded crossover claims (those for which Medicare is the primary payer, and Medicaid is responsible only for covering the remaining cost-sharing on behalf of dually eligible beneficiaries). We did not examine the pharmacy claims file because diagnosis codes are not reported on pharmacy claims, so they are not on that file.

For 46 states, Puerto Rico, and the District of Columbia, we calculated the percentage of header records in each claims file that had a valid ICD-10 diagnosis code in the field for the primary diagnosis code (DGNS_1_CD).⁷ Mississippi, Missouri, Montana, and Nebraska were excluded from the analysis because of a very low volume of claims.⁸ If a valid ICD-10 diagnosis code was not available, we calculated the percentage of records for which the field (1) was missing, (2) had an ICD-9 diagnosis code,⁹ or (3) had another non-missing but invalid value. Because we examined claims for services provided in 2017, which is after the transition from ICD-9 to ICD-10, we considered only the ICD-10 diagnosis codes to be “valid” values. We classified all ICD-9 codes as “invalid,” although TAF users may be able to use those codes for certain analyses. For the OT claims file only, we calculated the same measure for just the subset of claims that included outpatient hospital services, physician services, or clinic

services.¹⁰ We grouped states into categories of low, medium, and high data quality concern based on the percentage of their records that had a valid ICD-10 primary diagnosis code.¹¹

⁵ This analysis used the same TAF data as the T-MSIS Substance Use Disorder Data Book, which is not the version of the data that will be released as TAF Research Identifiable Files (RIFs).

⁶ We included all claims with a claim type code for either FFS claims or managed care encounter records (CLM_TYPE_CD equal to “1,” “3,” “A,” or “C”). We did not include capitation payments, supplemental payments, or service tracking payments. We also did not include any records with a claim type of “other,” which are records that the state did not classify as Medicaid or CHIP; these records may capture services that do not qualify for federal matching funds under Title XIX or Title XXI.

⁷ The ICD-10-CM code set is updated each year. For this analysis, we considered codes to be valid ICD-10 diagnosis codes if they were included in either the 2017 ICD-10-CM (effective from October 1, 2016 through September 30, 2017) or in the 2018 ICD-10-CM (effective from October 1, 2017 through September 30, 2018). Between the 2017 and 2018 versions of the ICD-10-CM, 360 codes were added, 142 codes were deleted, and 226 codes were revised.

⁸ In addition, we did not include Puerto Rico in the LT analysis because the territory did not cover services in long-term care facilities (Government Accountability Office 2016).

⁹ A small number of diagnosis codes are both an ICD-9 diagnosis code and an ICD-10 diagnosis code. For the purposes of our analysis, we classified those codes as valid ICD-10 diagnosis codes, not as ICD-9 codes.

¹⁰ We included OT claims in this subset if at least one of the claim lines had a type of service (TOS_CD) equal to “002” (outpatient hospital services), “012” (physician services), or “028” (clinic services). Claims with these types of services frequently appear to be the focus of in-depth research on service utilization patterns and are expected to have valid ICD-10 diagnosis codes. The subset of OT claims for outpatient hospital services, physician services, or clinic services captured 25 to 50 percent of OT claims in the majority of states.

¹¹ For the OT file, we used the measure calculated with just the subset of claims for outpatient hospital services, physician services, or clinic services to group states into categories of low, medium, and high data quality

The low-concern category includes states in which 90 percent or more of the records had a valid ICD-10 primary diagnosis code. The medium-concern category includes states in which 80 to 90 percent of the records had a valid ICD-10 primary diagnosis code, and the high-concern category includes states in which 50 to 80 percent of the records had a valid ICD-10 primary diagnosis code. In states in which less than 50 percent of the records had a valid ICD-10 primary diagnosis code, we considered the data to be unusable.

For the IP and LT files, we also calculated the mean number of valid ICD-10 diagnosis codes available per claim.¹² Although we did not use this measure to group states into categories of low, medium, or high data quality concern, a mean number of valid ICD-10 diagnosis codes that is substantially lower than that of other states may suggest that a state's diagnosis code information is incomplete. We did not assess the appropriateness of the available diagnosis codes against other information on the claims (such as procedure codes) because this complex undertaking is beyond the scope of this analysis.

Findings

Apart from a few exceptions, valid ICD-10 primary diagnosis codes were available on most IP, LT, and OT claims on which we would expect them to be available and reliable. It is noteworthy that valid ICD-10 primary diagnosis codes were also frequently included on OT claims on which we would not necessarily expect them to be reliable (for example, claims for medical supplies, prosthetic equipment, or NEMT services). TAF users may want to exercise caution in using the diagnosis codes on these claims even if they appear to be valid ICD-10 diagnosis codes.

We also analyzed the availability of diagnosis codes separately for FFS claims and managed care encounters (results not shown). We did not observe a clear pattern across states with respect to the availability of valid diagnosis codes on FFS claims compared with managed care encounters. However, there were substantial differences in the completeness of diagnosis codes by delivery system for certain states, which we note below.

IP file

- The vast majority of IP claims had valid ICD-10 primary diagnosis codes (Table 1). Forty-six states had valid ICD-10 primary diagnosis codes on 90 percent or more of their IP claims. Only one state (Maryland) had unusable primary diagnosis codes; 72 percent of its IP claims had an ICD-9 primary diagnosis code. We did not observe any substantial differences between FFS claims and managed care encounters with respect to the prevalence of valid ICD-10 primary diagnosis codes.

concern. For the IP and LT files, we used the measures calculated with all claims to assess the states' data quality.

¹² For the IP file, we counted the number of valid ICD-10 diagnosis codes in the fields DGNS_1_CD through DGNS_12_CD. For the LT file, we counted the number of valid ICD-10 diagnosis codes in the fields DGNS_1_CD through DGNS_5_CD. We then divided these counts by the total number of IP claims or LT claims, respectively.

- Most states submitted an average of six to eight valid ICD-10 diagnosis codes per IP claim. Although we have a low level of concern about the primary diagnosis codes in Georgia and Rhode Island, these two states submitted only about three diagnosis codes per IP claim, which is on the lower end of the range for all states. In Georgia, a large number of outpatient facility claims erroneously appear in the IP file instead of the OT file¹³; because outpatient facility claims tend to have fewer diagnosis codes per claim than do inpatient facility claims, the relatively low number of diagnosis codes in Georgia could be the result of claims that should not have been included in the IP file. In Rhode Island, this low number might be a result of the underlying variation in state billing practices, or it could suggest that the diagnosis code information was incomplete.

LT file

- Most states had valid ICD-10 primary diagnosis codes on the majority of their LT claims, with a handful of exceptions (Table 2). Two states (Maryland and South Carolina) had unusable primary diagnosis codes. In Maryland, 79 percent of LT claims had an ICD-9 primary diagnosis code. In South Carolina, 100 percent of LT claims were missing the primary diagnosis code.¹⁴ Another state (New Hampshire) had valid ICD-10 primary diagnosis codes on less than 80 percent of their LT claims because of high rates of missing primary diagnosis codes. We have a medium level of data quality concern for the primary diagnosis codes submitted by Illinois and Texas. In both states, this was driven entirely by missing primary diagnosis codes on FFS claims; all of the primary diagnosis codes on managed care encounters were valid ICD-10 diagnosis codes.
- Most states submitted an average of two to four valid ICD-10 diagnosis codes per LT claim. All of the states for which we have a low level of data quality concern also submitted an average of at least 1.5 diagnosis codes per LT claim.

OT file

- The vast majority of OT claims that we would expect to have reliable diagnostic information had a valid ICD-10 primary diagnosis code (Table 3). When we examined only the subset of OT claims that included outpatient hospital services, physician services, or clinic services, 44 states had valid ICD-10 primary diagnosis codes on 90 percent or more of these OT claims. Only one state (Maryland) had unusable data because of a high rate of ICD-9 primary diagnosis codes. We have a medium level of data quality concern for three additional states (Kansas, North Dakota, and Puerto Rico); 81 to 86 percent of their OT claims had a valid ICD-10 primary diagnosis code.¹⁵ In Puerto Rico, many of its other codes did not match to an ICD-9 or an ICD-10 code.

¹³ For more information, see TAF DQ brief #5042, “Using the Type of Bill to Classify Institutional Claims in 2017.”

¹⁴ In South Carolina, nursing facilities and intermediate care facilities use a non-standard claims form that does not include any fields for diagnosis codes.

¹⁵ In Kansas, the low percentage of valid ICD-10 diagnosis codes was driven entirely by the state’s claims for outpatient hospital services. All outpatient hospital claims were missing primary diagnosis codes; however, almost 100 percent of the state’s claims for physician services and clinic services had a valid ICD-10 primary diagnosis code.

- When we examined all OT claims without any type of service restriction, many states had missing primary diagnosis codes on between 5 to 20 percent of their OT claims (Table 3). A substantial percentage of these missing primary diagnosis codes appeared on dental claims (results not shown), which is expected if states do not require providers to submit diagnosis codes on these claims. It is noteworthy that valid ICD-10 primary diagnosis codes were frequently included on claims in the OT file submitted by providers who were not expected to have access to reliable diagnostic information about patients (for example, medical supplies, prosthetic equipment, and NEMT services; results not shown). TAF users may want to exercise caution in using the diagnosis codes on these claims even if they appear to be valid ICD-10 diagnosis codes.

All files

- Many states had a very small percentage of IP, LT, or OT claims with primary diagnosis codes that did not match to either an ICD-9 or ICD-10 diagnosis code (Tables 1 and 2). Of all of the states in our analysis, Puerto Rico submitted the highest proportion of its claims with this type of invalid value. For each claims file in all states, we examined the top 10 invalid primary diagnosis codes that occurred most often. In most cases, the invalid code would have been a valid ICD-10 diagnosis code if it had been submitted with an additional final character or without a decimal point.^{16,17} The invalid codes may reflect difficulties that the states had in extracting or reporting the data for these fields in their T-MSIS submissions. Alternatively, the invalid codes could reflect providers' errors in entering the diagnosis codes on claims.¹⁸

¹⁶ For example, the most frequently occurring invalid code in the IP file was "F1023." Although this code is not an ICD-9 or ICD-10 diagnosis code, adding a final character to the submitted code creates four possible valid ICD-10 diagnosis codes: "F10230" (alcohol dependence with withdrawal, uncomplicated), "F10231" (alcohol dependence with withdrawal delirium), "F10232" (alcohol dependence with withdrawal with perceptual disturbance), or "F10239" (alcohol dependence with withdrawal, unspecified).

¹⁷ States are instructed to report diagnosis codes to T-MSIS without a decimal point. For example, the code "210.5" should be submitted as "2105" without a decimal point.

¹⁸ If the invalid codes are the result of provider errors, and if those errors did not prevent the state or the managed care plan from adjudicating and paying the claim, then it is consistent with T-MSIS coding guidance for states to pass through paid claims with the invalid codes to T-MSIS.

Table 1. Valid, missing, and invalid diagnosis codes on IP FFS claims and managed care encounter records, 2017 TAF

State	Number of claims	Percentage with valid ICD-10 primary diagnosis code	Percentage with missing primary diagnosis code	Percentage with ICD-9 primary diagnosis code	Percentage with another invalid primary diagnosis code	Mean number of valid ICD-10 diagnosis codes per claim
Low concern about data quality (n = 46 states)						
Alaska	28,594	100.0	0.0	0.0	0.0	8
Connecticut	109,464	100.0	0.0	0.0	0.0	8
Georgia	2,175,923	100.0	0.0	0.0	0.0	3
Hawaii	52,059	100.0	0.0	0.0	0.0	8
Kentucky	197,732	100.0	0.0	0.0	0.0	8
Maine	18,348	100.0	0.0	0.0	0.0	7
Minnesota	198,434	100.0	0.0	0.0	0.0	8
Nevada	125,214	100.0	0.0	0.0	0.0	7
New Hampshire	12,079	100.0	0.0	0.0	0.0	8
New Jersey	294,673	100.0	0.0	0.0	0.0	7
New Mexico	98,298	100.0	0.0	0.0	0.0	7
North Dakota	10,771	100.0	0.0	0.0	0.0	7
Ohio	543,320	100.0	0.0	0.0	0.0	8
Oregon	122,687	100.0	0.0	0.0	0.0	7
Vermont	18,403	100.0	0.0	0.0	0.0	8
Washington	186,863	100.0	0.0	0.0	0.0	7
West Virginia	61,583	100.0	0.0	0.0	0.0	8
Wyoming	10,900	100.0	0.0	0.0	0.0	6
New York	1,642,058	100.0	0.0	0.0	0.0	7
Texas	846,386	100.0	0.0	0.0	0.0	6
Arizona	329,829	100.0	0.0	0.0	0.0	7
Colorado	116,378	100.0	0.0	0.0	0.0	8
Michigan	378,520	100.0	0.0	0.0	0.0	7
Louisiana	194,294	100.0	0.0	0.0	0.0	7
Arkansas	111,243	100.0	0.0	0.0	0.0	6
Kansas	71,694	100.0	0.0	0.0	0.0	7
Iowa	133,988	100.0	0.0	0.0	0.0	7
South Dakota	21,887	100.0	0.0	0.0	0.0	6
Delaware	32,241	100.0	0.0	0.0	0.0	7

Table 1 (continued)

State	Number of claims	Percentage with valid ICD-10 primary diagnosis code	Percentage with missing primary diagnosis code	Percentage with ICD-9 primary diagnosis code	Percentage with another invalid primary diagnosis code	Mean number of valid ICD-10 diagnosis codes per claim
North Carolina	340,586	100.0	0.0	0.0	0.0	7
Florida	711,187	100.0	0.0	0.0	0.0	6
Oklahoma	141,289	99.9	0.1	0.0	0.0	7
Pennsylvania	494,306	99.9	0.0	0.0	0.1	7
Idaho	34,731	99.9	0.1	0.0	0.0	6
Virginia	179,074	99.9	0.0	0.0	0.1	7
Alabama	163,616	99.8	0.2	0.0	0.0	7
Tennessee	234,108	99.7	0.0	0.0	0.3	7
Utah	69,513	99.7	0.3	0.0	0.0	7
Illinois	455,251	99.6	0.4	0.0	0.0	7
Massachusetts	1,451,390	99.2	0.0	0.0	0.8	8
Wisconsin	203,110	98.6	1.4	0.0	0.0	6
California	1,586,253	97.0	3.0	0.0	0.0	7
District of Columbia	36,151	96.9	3.1	0.0	0.0	7
Indiana	208,723	94.9	0.0	0.0	5.1	7
Rhode Island	23,377	94.4	5.6	0.0	0.0	3
South Carolina	166,073	91.8	8.2	0.0	0.0	6
Medium concern about data quality (n = 1 state)						
Puerto Rico	227,645	82.3	0.0	0.0	17.7	3
Unusable data (n = 1 state)						
Maryland	282,692	28.3	0.0	71.7	0.0	2
Not included in analysis (n = 4 states)						
Mississippi	DQ	DQ	DQ	DQ	DQ	DQ
Missouri	DQ	DQ	DQ	DQ	DQ	DQ
Montana	DQ	DQ	DQ	DQ	DQ	DQ
Nebraska	DQ	DQ	DQ	DQ	DQ	DQ

Source: 2017 TAF as of January 2019.

Note: States are ordered from highest to lowest with respect to the percentage of IP claims that had a valid ICD-10 primary diagnosis code. The low-concern category includes states in which 90 percent or more of the records had a valid ICD-10 primary diagnosis code. The medium-concern category includes states in which 80 to 90 percent of the records had a valid ICD-10 primary diagnosis code, and the high-concern category includes states in which 50 to 80 percent of the records had a valid ICD-10 primary diagnosis code. In states in which less than 50 percent of the records had a valid ICD-10 primary diagnosis code, we considered the data to be unusable.

DQ = Not reported because of concerns about the low volume of records in one or more claims files.

Table 2. Valid, missing, and invalid diagnosis codes on LT FFS claims and managed care encounter records, 2017 TAF

State	Number of claims	Percentage with valid ICD-10 primary diagnosis code	Percentage with missing primary diagnosis code	Percentage with ICD-9 primary diagnosis code	Percentage with another invalid primary diagnosis code	Mean number of valid ICD-10 diagnosis codes per claim
Low concern about data quality (n = 42 states)						
Alaska	18,200	100.0	0.0	0.0	0.0	3
Connecticut	108,750	100.0	0.0	0.0	0.0	3
Delaware	59,475	100.0	0.0	0.0	0.0	4
Maine	70,255	100.0	0.0	0.0	0.0	4
Michigan	415,224	100.0	0.0	0.0	0.0	4
Nevada	93,870	100.0	0.0	0.0	0.0	3
New Jersey	621,500	100.0	0.0	0.0	0.0	3
New Mexico	94,875	100.0	0.0	0.0	0.0	4
Pennsylvania	961,956	100.0	0.0	0.0	0.0	4
Tennessee	416,097	100.0	0.0	0.0	0.0	3
Washington	415,044	100.0	0.0	0.0	0.0	3
Wyoming	38,933	100.0	0.0	0.0	0.0	3
Ohio	638,837	100.0	0.0	0.0	0.0	4
New York	11,005,325	100.0	0.0	0.0	0.0	2
North Carolina	1,288,413	100.0	0.0	0.0	0.0	3
Massachusetts	710,879	100.0	0.0	0.0	0.0	4
Hawaii	39,838	100.0	0.0	0.0	0.0	5
Arizona	284,346	100.0	0.0	0.0	0.0	3
Idaho	80,711	100.0	0.0	0.0	0.0	3
Alabama	347,246	100.0	0.0	0.0	0.0	3
West Virginia	135,449	100.0	0.0	0.0	0.0	4
Vermont	43,007	99.9	0.0	0.0	0.1	4
Kentucky	343,384	99.9	0.0	0.0	0.1	4
Wisconsin	159,441	99.9	0.1	0.0	0.0	4
Oregon	84,977	99.8	0.0	0.0	0.2	3
Rhode Island	103,875	99.8	0.0	0.0	0.2	2
Arkansas	1,053,933	99.8	0.1	0.0	0.0	3
California	4,322,137	99.8	0.2	0.0	0.0	3
South Dakota	52,955	99.7	0.0	0.1	0.3	4
Kansas	339,443	99.6	0.4	0.0	0.0	3

Table 2 (continued)

State	Number of claims	Percentage with valid ICD-10 primary diagnosis code	Percentage with missing primary diagnosis code	Percentage with ICD-9 primary diagnosis code	Percentage with another invalid primary diagnosis code	Mean number of valid ICD-10 diagnosis codes per claim
Colorado	315,587	99.4	0.0	0.0	0.6	4
Indiana	1,012,046	99.3	0.0	0.0	0.7	2
Utah	162,243	99.2	0.5	0.0	0.2	4
Georgia	1,377,399	99.2	0.0	0.0	0.8	4
Virginia	342,937	99.2	0.0	0.2	0.7	4
Iowa	231,257	98.9	1.1	0.0	0.0	3
District of Columbia	111,763	98.4	1.2	0.0	0.4	2
Louisiana	388,917	98.4	0.8	0.0	0.8	2
Florida	1,044,464	96.5	3.5	0.0	0.0	3
Minnesota	210,934	95.9	4.1	0.0	0.0	3
Oklahoma	857,279	95.7	4.1	0.0	0.2	3
North Dakota	56,259	95.1	4.9	0.0	0.0	3
Medium concern about data quality (n = 2 states)						
Illinois	503,783	89.1	10.9	0.0	0.0	3
Texas	3,386,189	81.0	19.0	0.0	0.0	2
High concern about data quality (n = 1 state)						
New Hampshire	154,303	51.3	48.7	0.0	0.0	1
Unusable data (n = 2 states)						
Maryland	285,097	20.6	0.0	79.4	0.0	1
South Carolina	193,166	0.0	100.0	0.0	0.0	0
Not included in analysis (n = 5 states)						
Mississippi	DQ	DQ	DQ	DQ	DQ	DQ
Missouri	DQ	DQ	DQ	DQ	DQ	DQ
Montana	DQ	DQ	DQ	DQ	DQ	DQ
Nebraska	DQ	DQ	DQ	DQ	DQ	DQ
Puerto Rico	0	—	—	—	—	—

Source: 2017 TAF as of January 2019.

Note: States are ordered from highest to lowest with respect to the percentage of LT claims that had a valid ICD-10 primary diagnosis code. The low-concern category includes states in which 90 percent or more of the records had a valid ICD-10 primary diagnosis code. The medium-concern category includes states in which 80 to 90 percent of the records had a valid ICD-10 primary diagnosis code, and the high-concern category includes states in which 50 to 80 percent of the records had a valid ICD-10 primary diagnosis code. In states in which less than 50 percent of the records had a valid ICD-10 primary diagnosis code, we considered the data to be unusable.

DQ = Not reported because of concerns about the low volume of records in one or more claims files.

— = Puerto Rico was not included in the analysis of the LT file because services in long-term care facilities were not a covered benefit in the territory, so there were no records in the LT file.

Table 3. Valid diagnosis codes on OT FFS claims and managed care encounter records, 2017 TAF

State	Subset of OT claims for outpatient hospital services, physician services, or clinic services		All OT claims	
	Number of claims	Percentage with valid ICD-10 primary diagnosis code	Number of claims	Percentage with valid ICD-10 primary diagnosis code
Low concern about data quality (n = 44 states)				
New Jersey	46,580,112	100.0	112,667,329	91.4
Tennessee	14,354,554	100.0	35,344,028	96.2
Pennsylvania	30,026,935	100.0	85,920,501	93.0
Washington	15,989,738	100.0	33,663,065	93.5
Vermont	944,774	100.0	3,681,596	98.6
Arkansas	6,457,486	100.0	21,860,795	99.2
Louisiana	27,183,017	100.0	70,893,573	94.7
Alaska	1,680,178	100.0	5,043,105	87.6
Connecticut	11,159,783	100.0	25,863,130	69.8
Minnesota	10,039,868	100.0	41,730,838	97.0
Wyoming	762,809	100.0	1,692,289	85.8
Delaware	2,409,376	100.0	5,554,889	96.7
Nevada	2,732,344	100.0	17,644,335	99.2
South Dakota	1,347,015	100.0	3,604,654	86.8
Maine	2,203,034	100.0	10,062,171	98.3
Oregon	7,754,309	100.0	23,769,198	89.9
Hawaii	3,105,679	100.0	5,770,319	94.8
Rhode Island	11,046,306	100.0	14,002,558	91.4
Oklahoma	10,894,968	100.0	18,732,243	94.4
Alabama	9,315,543	100.0	17,774,350	92.0
New York	97,564,139	100.0	237,533,417	97.4
Colorado	10,066,876	100.0	18,076,898	93.5
New Hampshire	1,137,824	100.0	6,696,069	96.9
Idaho	2,604,873	100.0	5,824,043	99.9
California	82,597,720	99.9	259,673,872	75.6
Wisconsin	8,926,611	99.8	41,042,333	78.4
Georgia	16,619,095	99.8	34,108,605	93.3
Arizona	25,316,741	99.7	78,177,180	95.7
North Carolina	19,604,200	99.6	44,738,916	93.2
District of Columbia	1,987,614	99.6	5,587,212	92.9
Virginia	12,287,714	99.6	29,431,398	95.6
Illinois	19,572,274	99.3	74,559,649	93.3
Ohio	41,893,250	99.1	106,274,252	96.4
Indiana	12,772,768	99.1	30,820,727	92.5
Florida	55,768,559	99.0	87,011,900	92.7
New Mexico	8,669,097	98.9	16,335,352	93.9
Texas	37,536,967	98.8	137,763,488	70.2
South Carolina	14,084,062	98.6	28,141,017	85.4
Iowa	7,605,034	97.5	15,410,499	94.9
West Virginia	4,810,476	96.7	12,293,016	92.9
Michigan	5,512,768	96.2	56,191,911	94.3
Utah	774,426	96.1	3,712,350	63.1

Table 3 (continued)

State	Subset of OT claims for outpatient hospital services, physician services, or clinic services		All OT claims	
	Number of claims	Percentage with valid ICD-10 primary diagnosis code	Number of claims	Percentage with valid ICD-10 primary diagnosis code
Massachusetts	9,423,849	95.1	97,689,561	91.6
Kentucky	12,941,353	92.4	37,893,322	89.8
Medium concern about data quality (n = 3 states)				
North Dakota	437,045	85.5	1,432,216	68.8
Kansas	4,770,398	84.0	10,380,640	90.3
Puerto Rico	8,246,779	80.7	22,074,087	68.8
Unusable data (n = 1 state)				
Maryland	14,065,653	17.9	68,386,993	15.7
Not included in analysis (n = 4 states)				
Mississippi	DQ	DQ	DQ	DQ
Missouri	DQ	DQ	DQ	DQ
Montana	DQ	DQ	DQ	DQ
Nebraska	DQ	DQ	DQ	DQ

Source: 2017 TAF as of January 2019.

Note: States are ordered from highest to lowest with respect to the percentage of OT claims for outpatient hospital services, physician services, or clinic services that had a valid ICD-10 primary diagnosis code. We included OT claims in this subset if at least one of the claim lines had a type of service (TOS_CD) equal to “002” (outpatient hospital services), “012” (physician services), or “028” (clinic services). The low-concern category includes states in which 90 percent or more of these records had a valid ICD-10 primary diagnosis code. The medium-concern category includes states in which 80 to 90 percent of these records had a valid ICD-10 primary diagnosis code, and the high-concern category includes states in which 50 to 80 percent of these records had a valid ICD-10 primary diagnosis code. In states in which less than 50 percent of these records had a valid ICD-10 primary diagnosis code, we considered the data to be unusable.

DQ = Not reported because of concerns about the low volume of records in one or more claims files.

References

Government Accountability Office. "Increased Funding in U.S. Territories Merits Improved Program Integrity Efforts." April 2016. Available at <https://www.gao.gov/assets/680/676682.pdf>. Accessed on March 13, 2019.

Amy Mangum¹, Kimberly Proctor², and Jessie Parker². "Missing and Invalid Diagnosis Codes in 2017." TAF DQ Brief #5132. Baltimore, MD: CMS, 2019.

Reviewers: Keith Branham²; Jeff Galecki²; Allison Barrett¹; Carol V. Irvin¹; Brian Johnston¹; Laura Nolan¹

¹Mathematica, ²Centers for Medicare & Medicaid Services, Center for Medicaid and CHIP Services, Data and Systems Group, Division of Business and Data Analytics

Correspondence should be addressed to MACBISData@cms.hhs.gov