Texas All-Payor Claims Database

Common Stage II Data Quality Checks

November 2024

Version 1.0

UTHealth
The University of Texas
Health Science Center at Houston

School of Public Health | Center for Health Care Data

1200 Pressler | Houston | TX 77030

Ph: 713.500.9000 | https://go.uth.edu/txapcd

UTHealth The University of Texas Health Science Center at Houston | School of Public Health

Check Number	Check Name	Description	Associated Files	Associated Fields	Rule	Unit	Output	Cutoff	Scope	Note
013	Subscriber has no record as member	Checks to see how many carrier_specific_unique_subscriber_id are also found in the carrier_specific_unique_member_id column	ME	CDLME061 - carrier_specific_unique_member_id CDLME062 - carrier_specific_unique_subscriber_id	Count how many unique subscriber_ids there are, and of those how many subscriber_ids are found in the member_id column	distinct carrier_specific_unique_subscriber_ids	percent	2%		We expect the majority of subscribers to also themselves be members.
018	Member existence in eligibility file (Medical)	Checks to see if the member listed in the medical file exists in the eligibility file	ME, MC	CDLME061 - carrier_specific_unique_member_id CDLMC161 - carrier_specific_unique_member_id	If carrier specific unique member ID in medical file does not exist in eligibilty file under the same payor code, flag as missing	distinct SSNs, distinct carrier specific unique member IDs	percent	1%		This check looks for whether or not a carrier specific unique member ID provided in the medical file is also listed in the eligibility file under the same payor code. We do expect some small percentage of claims to be from people without coverage, or for claims under adjudication for people who no longer have coverage, but high numbers of missing members warrant investigation.
019	Member existence in eligibility file (Pharmacy)	Checks to see if the member listed in the pharmacy file exists in the eligibility file	ME, PC	CDLME061 - carrier_specific_unique_member_id CDLMC161 - carrier_specific_unique_member_id	If carrier specific unique member ID in medical file does not exist in eligibilty file under the same payor code, flag as missing	distinct SSNs, distinct carrier specific unique member IDs	percent	1%		This check looks for whether or not a carrier specific unique member ID provided in the pharmacy file is also listed in the eligibility file under the same payor code. We do expect some small percentage of claims to be from people without coverage, or for claims under adjudication for people who no longer have coverage, but high numbers of missing members warrant investigation.
020	Member existence in eligibility file (Dental)	Checks to see if the member listed in the dental file exists in the eligibility file	ME, DC	CDLME061 - carrier_specific_unique_member_id CDLMC161 - carrier_specific_unique_member_id	If carrier specific unique member ID in medical file does not exist in eligibilty file under the same payor code, flag as missing	distinct SSNs, distinct carrier specific unique member IDs	percent	1%		This check looks for whether or not a carrier specific unique member ID provided in the dental file is also listed in the eligibility file under the same payor code. We do expect some small percentage of claims to be from people without coverage, or for claims under adjudication for people who no longer have coverage, but high numbers of missing members warrant investigation.
025	Consistency in claim line versions	Check to see that version_number starts from 0 and is not missing any versions	MC	CDLMC005 - payor_claim_control_number CDLMC006 - line_counter CDLMC007 - version_number	If the highest version is n, then versions 0 to n-1 should also be provided	payor_claim_control_number, line counter, version number	percent	20%		
028	Correct sequencing of claim line counters	Check to see line counters start at 1 and are not missing any counters	MC	CDLMC005 - payor_claim_control_number CDLMC006 - line_counter CDLMC007 - version_number	If the highest line counter is n, then counters 1 to n should also be provided	payor_claim_control_number, line counter, version number	percent	20%		
031	Invalid version numbers	Check to see if there are version numbers that are out of range	MC, PC, DC	CDLMC007, CDLPC007, CDLDC007 - version_nu	Percentage of claim lines where version_number is higher than 100		percent	0.1%		
032	Repetition in payor claim control number, line counters and version numbers	Check to see combination of payor claim control number, line counter and version number has been repeated	MC	CDLMC005 - payor_claim_control_number CDLMC006 - line_counter CDLMC007 - version_number	The combination of payor_claim_control_number, line_counter and version_number should not be repeated		percent	20%		
035	Consistency between claim line type and plan paid_amount	Check to see claim line type is aligned with plan paid amount	MC, PC, DC	CDLMC160, CDLPC066, CDLDC084 - claim_line_type CDLMC125, CDLPC037, CDLDC060 - paid_amount	If claim line type = "O", plan_paid_amount should be positive and if claim line type = ("D", "B", "R"), plan_paid_amount should be negative		percent	5%		
036	Unusually long inpatient stays	Check the percentage of claim lines with inpatient stays exceeding a month	MC	CDLMC121 - service quantity CDLMC029 - discharge date CDLMC025 - admission date	If service quantity (days) > 30 or discharge date - admission date > 30		percent	10%		It should be noted that there are scenarios where long term stays are the norm. This exclusion has not yet been accounted for in this rule, but will be done based on TYPE_OF_BILL.
039	Cross reference ID existence check	Checks to see if the cross reference ID provided refers back to a previously received claim's payor_claim_control_number	MC, PC, DC	CDLMC008, CDLPC008, CDLDC008 - cross_reference_claims_id CDLMC005, CDLPC005, CDLDC005 - payor_claim_control_number	If cross_reference_claims_id is not found in payor_claim_control_number for a given payor_code (in any submission) then flag	distinct cross_reference_claims_ids	percent	10%	Applies to payors who use cross reference ID system exclusively	Cross reference claims ID is only used when payors do not use a claim control number/version number system. This check is only applicable to payors who do NOT use the version number system and instead uses cross reference claims IDs. The cross reference claims ID is checked against the list of all previously received payor claim control numbers and a flag is raised if no previously received payor claim control number matches. Currently, a cross_reference_claim_id will pass if it matches any payor_claim_control_number regardless of when the payor_claim_control_number was received, even one that is set in the future relative to the cross_reference_claim_id, but as the dataset matures, this will be amended to only allow previously received payor_claim_control_numbers.

Check Number	Check Name	Description	Associated Files	Associated Fields	Rule	Unit	Output	Cutoff	Scope	Note
040	Sharp change in number of distinct claims	Looks for sharp month-over-month changes in the claim volume	MC	CDLMC005 - payor_claim_control_number	If absolute value of percent change of count(distinct payor_claim_control_number for current month) / count(distinct payor_claim_control_number for previous month) > 10%, flag as requiring more investigation **** The issue is flagged at 50% and 100%	distinct payor_claim_control_numbers	percent	20%		Finds the percent change of distinct payor_claim_control_numbers on a monthly basis (relative to previous month). If percent change exceeds 10%, flag as needing further investigation. We do expect seasonality to affect this number (for example, increased claims in December/January or August/September depending on if insurance is on a calendar year basis or fiscal year basis); as TX-APCD accumulates more data, this metric will be changed to year-over-year (each month compared to the same month in previous year).
041	Charges discrepancy - Medical	Verify that total charges are greater than or equal to the sum of plan payments and patient's share.	MC	CDLMC123 - charge_amount CDLMC125 - plan_paid_amount CDLMC126 - copay_amount CDLMC127 - coinsurance_amount CDLMC128 - deductible_amount CDLMC130 - cob_tpl_amount	charge_amount >= (plan_paid_amount + copay_amount + coinsurance_amount + deductible_amount + cob_tpl_amount)	distinct charges	percent	10%		In a very small number of cases, it is possible that charges are smaller than the sum of carrier and patient payments. An example is where a hospital might charge less than the contracted DRG amount while the carrier pays on the contracted amount.
042	Charges discrepancy - Pharmacy	Verify that total charges are greater than or equal to the sum of plan payments and patient's share.	PC	CDLPC036 - charge_amount CDLPC037 - plan_paid_amount CDLPC039 - sales_tax_amount CDLPC040 - ingredient_cost_list_price CDLPC043 - copay_amount CDLPC044 - coinsurance_amount CDLPC045 - deductible_amount CDLPC041 - postage_amount_claimed CDLPC042 - dispensing_fee CDLPC046 - cob_tpl_amount	charge_amount >= plan_paid_amount + sales_tax_amount + ingredient_cost_list_price + co_pay_amount + coinsurance_amount + deductible_amount + postage_amount_claimed + dispensing_fee + cob_tpl_amount	distinct charges	percent	10%		
045	Cost columns are in cents (Medical, distribution method)	Checks to make sure that cost columns are correctly interpreted	MC	CDLMC123 - charge_amount	Find the distributions of: costs < 10 ($0.10) costs < 1000 ($10.00) costs > 1000000 ($10,000.00)	rows	percent	-		Only significant deviations from expected distributions will be investigated.
046	Cost columns are in cents (Pharmacy, distribution method)	Checks to make sure that cost columns are correctly interpreted	PC	CDLPC036 - charge_amount	Find the distributions of: costs < 10 ($0.10) costs < 1000 ($10.00) costs > 1000000 ($10,000.00)	rows	percent	-		Only significant deviations from expected distributions will be investigated.
047	Cost columns are in cents (Dental, distribution method)	Checks to make sure that cost columns are correctly interpreted	DC	CDLDC059 - charge amount	Find the distributions of: costs < 10 ($0.10) costs < 1000 ($10.00) costs > 1000000 ($10,000.00)	rows	percent	-		Only significant deviations from expected distributions will be investigated.
051	Medical - date validity check	Checks to see if dates are within range of expectations	MC	CDLMC024 - paid_date CDLMC025 - admit_date CDLMC029 - discharge_date CDLMC119 - date_of_service_from CDLMC120 - date_of_service_through	If [date] is outside of the range [20170101, 20241231], flag as out of range		rows	1%		The data submission period for TX-APCD is 2019 forward. Dates occuring more than 2 years prior to the first data submission period are possible, but should be very rare as they should reflect only claims that are still being adjudicated 2 years after the claim was generated. Dates occuring in the future should not occur at all.
058	Medical - uniqueness of bill_type	Checks to see if a unique payor claim control number is associated with more than one type_of_bill	MC	CDLMC032 - type_of_bill_institutional	If a unique payor claim control number has > 1 distinct type_of_bill across submitted data per payor code, flag as an error	distinct payor_claim_control_numbers	percent	1%		Except for in the case of a typographical error, we expect that each PCCN is associated with only one type_of_bill.
059	Medical - uniqueness of claim_type	Checks to see if a unique payor claim control number is associated with more than one type_of_claim_institutional	MC	CDLMC156 - type_of_claim	If a unique payor claim control number has > 1 distinct type_of_claim_institutional across submitted data per payor code, flag as an error	distinct payor_claim_control_numbers	percent	1%		Except for in the case of a typographical error, we expect that each PCCN is associated with only one type_of_claim_institutional.
062	Version / cross-reference claim check	Checks to see if data uses versioning or cross_reference_claims_id	MC	CDLMC007 - version_number CDLMC008 - cross_reference_claims_id	Checks population rate of version_number, cross_reference_claims_id, and rate at which both columns are populated	distinct pccn/cross_reference version combination	-	-		This check is less error-checking and more informational, to allow TX-APCD to determine what primary method of versioning a payor code uses. However, payor codes that use both version_number and cross_reference_claims_id will warrant further scrutiny.
068	Each medical claim references only a single member	Checks to see if a given payor_claim_control_number for a given payor_code only has one unique carrier_specific_unique_member_id referenced	MC	CDLMC005 - payor_claim_control_number CDLMC161 - carrier_specific_unique_member_id	If a distinct payor_claim_control_number has count(distinct carrier_specific_unique_member_id) > 1, flag as an error	distinct payor_claim_control_number	percent	1%		Each unique payor_claim_control_number should only have 1 carrier_specific_unique_member_id across all claim lines. While it is possible that corrections to claims may include a member_id correction, we expect this rate to be very low (<1%).
075	Non-inpatient Claims - Admission date	admission_date should be null if claim is not an inpatient claim	MC	CDLMC025 - admission_date	Admit date should be null if claim is not an inpatient claim (for definition of inpatient claim see Appendix B)	Non-inpatient claim rows	percent	5%	Non-inpatient claims	If claim is not an inpatient claim, admit date is not expected to be populated.

Check Number	Check Name	Description	Associated Files	Associated Fields	Rule	Unit	Output	Cutoff	Scope	Note
081	Inpatient Claims - admission date	admission date should be provided for inpatient claims	MC	CDLMC025 - admission_date	Admit date should be populated if claim is an inpatient claim (for definition of inpatient claim see Appendix B)	Inpatient claim rows	percent	5%	Inpatient claims	If claim is an inpatient claim, admit date is expected to be populated.
085	Inpatient Claims - discharge date before admit date	discharge dates should not precede admission dates	MC	CDLMC025 - admission_date CDLMC029 - discharge date	If discharge_date < admission_date, flag as an error	Inpatient claim rows	percent	1%	Inpatient claims	Discharge dates should be after admission dates.

UTHealth | School of Public Health
The University of Texas
Health Science Center at Houston

List of fields or values not explicitly defined in the Common Data Layout (CDL):

Field Name	Origin	Sample Value
source_file	Name of pipe-delimited text file as submitted to TX-APCD.	P_CHCDMDCR_20000206_201903_201903_ME_01.txt
data_period_start	Year/month of submission as indicated in file name.	201903
data_period_end	Year/month of submission as indicated in file name.	201903
submission_id	Unique numeric identifier for a submission to the TX-APCD. Note that all files submitted in the same .zip file will have the same submission ID.	6412
row_count	Count of rows in a given table.	

UTHealth | School of Public Health
The University of Texas
Health Science Center at Houston

SSN Validity Rules	
Reason	**Explanation**
NOT 9 DIGITS	Not 9 digits long
NON-NUMERIC CHARS	Contains non-numeric charcters
STARTS WITH 9	Starts with 9
AREA IS 666	Area = 666
AREA IS 000	Area = 000
GROUP IS 00	Group = 00
SERIAL IS 0000	Serial = 0000
REPEATED NUMBER	Social security number is a single number, repeated (such as 111111111)
OTHER INVALID	Social security number follows some other pattern that has been determined to be invalid (such as 123456789) *

*These “other” invalid numbers were identified by looking at how many distinct people share a given social security number. Distinct people were identified as people who have different first names (last names were not used as marriages/divorces would change a last name but not SSN). Of social security numbers shared by multiple people, numbers that followed some obvious pattern such as ‘121212121’, ‘234234234’, etc. were flagged as OTHER INVALID.

Definition of Inpatient

A claim is determined to be an inpatient claim if type_of_claim = 2 (Intituional/Facility)
and type_of_bill_institutional begins with:

Code	Description
11	Hospital Inpatient part A
12	Hospital Inpatient part B (includes HHA)
21	SNF Inpatient part A
22	SNF Inpatient part B
81	Special facility or hospital (Critical Access Hospitals (CAHs) or Ambulatory Surgical Centers (ASCs)) Inpatient part A
82	Special facility or hospital (Critical Access Hospitals (CAHs) or Ambulatory Surgical Centers (ASCs)) Inpatient part B

Explanation of Utilization Rate Checks

The purpose of these check are to determine if we are getting a complete eligibility table. We expect that there will always be some proportion of enrollees who have coverage but do not generate any claims.

To find the denominator for number of enrollees, first the enrollment table is filtered such that only rows where the plan_effective_date/plan_term_date pair includes the data_period are retained. From those rows, distinct year/member_id pairs are found. This does mean that if a member is enrolled for only one month of the year, they will still have a row for that year in the final table. Then, distinct pairs of year (from date of service) and member_id are generated from the medical table, and the two tables are compared.

The expectations for utilization rates vary according to insurance type. We expect some insurance types to have very low utilization rates, such as supplementary insurance. We also expect utilization rates to vary by population: insurances whose enrollees are primarily young adults (such as student insurance) will typically have lower utilization rates whereas older populations (e.g. Medicare Advantage) will have higher utilization rates. Additionally, dental insurance and medical insurance will have different rates as well.

Rates at the very low end and very high end will be examined to determine if numbers make sense in context.