

# MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/  
2025-10/2025-06-30\_Providence\_Health\_Plan\_f15fd58926a54d09da48cdd927ff9abe\_in-network-rates.json.gz  
Size: 11.34 GB • MD5: 55f51bd2bc45053d7f7f38c5704597a1

Payer: **Providence\_Health\_Plan** • File Date: **2025-06-30** • Generated: **2026-04-25 16:54 EDT** • Tool Version: **1.0.0** • Elapsed: **27549.00s**

# 65.0

Limited Reliability

score capped at 65.0 — CMS official schema validation failed (see meta for details)

Errors: 3 • Warnings: 13 • Info: 3

## TOC Plan References

TOC: s3://talon-storage-private/mrf-feed-uploads/2025-10/2025-10-15\_Providence-Health-Plan-Master\_combine\_index.json • Providence Health Plan (Health Plan)

Canonical: s3://talon-storage-private/mrf-feed-uploads/2025-10/2025-06-30\_Providence\_Health\_Plan\_f15fd58926a54d09da48cdd927ff9abe\_in-network-rates.json.gz

Plan Name	Plan ID	Issuer / Sponsor	Market
HSA	201544716 (EIN)	—	group
HSA	208691906 (EIN)	—	group
HSA	260820665 (EIN)	—	group
HSA	421599749 (EIN)	—	group
HSA	451998958 (EIN)	—	group
HSA	464077672 (EIN)	—	group
HSA	843273798 (EIN)	—	group
HSA	852537298 (EIN)	—	group
HSA	910719737 (EIN)	—	group
HSA	910774723 (EIN)	—	group
HSA	910818628 (EIN)	—	group

Plan Name	Plan ID	Issuer / Sponsor	Market
HSA	910861800 (EIN)	—	group
HSA	910870695 (EIN)	—	group
HSA	910883323 (EIN)	—	group
HSA	910886063 (EIN)	—	group
HSA	910957371 (EIN)	—	group
HSA	911358195 (EIN)	—	group
HSA	911463668 (EIN)	—	group
HSA	911531721 (EIN)	—	group
HSA	930256820 (EIN)	—	group
HSA	930386915 (EIN)	—	group
HSA	930645870 (EIN)	—	group
HSA	931013415 (EIN)	—	group
HSA	931104323 (EIN)	—	group
HSA	931815916 (EIN)	—	group
HSA	954136880 (EIN)	—	group
HSA_E	201421296 (EIN)	—	group
OPEN_OPTION	936002286 (EIN)	—	group
OPTIONADV_BASE	201544716 (EIN)	—	group
OPTIONADV_BASE	208691906 (EIN)	—	group
OPTIONADV_BASE	237071436 (EIN)	—	group
OPTIONADV_BASE	451998958 (EIN)	—	group
OPTIONADV_BASE	464077672 (EIN)	—	group
OPTIONADV_BASE	811045408 (EIN)	—	group
OPTIONADV_BASE	843273798 (EIN)	—	group
OPTIONADV_BASE	910719737 (EIN)	—	group
OPTIONADV_BASE	910818628 (EIN)	—	group
OPTIONADV_BASE	910861800 (EIN)	—	group
OPTIONADV_BASE	910870695 (EIN)	—	group
OPTIONADV_BASE	910883323 (EIN)	—	group

Plan Name	Plan ID	Issuer / Sponsor	Market
OPTIONADV_BASE	910886063 (EIN)	—	group
OPTIONADV_BASE	910957371 (EIN)	—	group
OPTIONADV_BASE	911189908 (EIN)	—	group
OPTIONADV_BASE	911358195 (EIN)	—	group
OPTIONADV_BASE	911463668 (EIN)	—	group
OPTIONADV_BASE	911531721 (EIN)	—	group
OPTIONADV_BASE	931013415 (EIN)	—	group
OPTIONADV_BASE	931815916 (EIN)	—	group
OPTIONADV_PLUS	201421296 (EIN)	—	group
OPTIONADV_PLUS	421599749 (EIN)	—	group
OPTIONADV_PLUS	852537298 (EIN)	—	group
OPTIONADV_PLUS	931104323 (EIN)	—	group
OPTIONADV_PREM	204882257 (EIN)	—	group
OPTIONADV_PREM	260820665 (EIN)	—	group
OPTIONADV_PREM	910774723 (EIN)	—	group
OPTIONADV_PREM	930386915 (EIN)	—	group
OPTIONADV_PREM	930645870 (EIN)	—	group
OPTIONADV_PREM	930692796 (EIN)	—	group
OPTIONADV_PREM	942709226 (EIN)	—	group
OPTIONADV_PREM	954136880 (EIN)	—	group
PERSONAL_OPTION	930386915 (EIN)	—	group
PERSONAL_OPTION	936002286 (EIN)	—	group

## CMS Official Schema Validation

**FAILED** (exit code -1) — File does not conform to the CMS schema.

Validator output:

Timed out after 600 s

# Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	70.0	30%	4
Provider Mapping	100.0	15%	4
Code Coverage	89.5	15%	2
Pricing Sanity	58.9	40%	9

## Schema Integrity — Findings

---

Score: 70.0

**ERROR** `file_freshness`

File is 299 days old (last\_updated\_on exceeds the 90-day threshold)

---

**WARNING** `item_required_fields`

0.01% of in\_network items are missing required fields

---

**WARNING** `expired_prices`

66246002 negotiated\_prices have past expiration dates (8.3%)

---

**ERROR** `cms_schema_validation`

CMS official schema validator FAILED (exit code -1). File does not conform to the TIC in-network-rates schema.

---

## Provider Mapping — Findings

---

Score: 100.0

**WARNING** `npi_validity`

0.00% of NPIs failed Luhn checksum validation (4 of 111626)

- 1336588484
- 1659650617

---

**WARNING** `ein_validity`

0.31% of EINs failed IRS prefix validation (191 of 62130)

- 000000001
- 001625898
- 003360711
- 003680161
- 004829723
- ... and 5 more

---

**WARNING** `empty_npi_groups`

53 provider groups contain no NPIs

---

**INFO** `duplicate_npis`

36183 NPIs appear in more than one provider group

---

# Code Coverage — Findings

---

Score: 89.5

**WARNING** unknown\_billing\_code\_types

Unrecognized billing\_code\_type values: {'PROC': 1}

---

**WARNING** billing\_code\_format

3924 CPT codes do not match expected format

---

## Pricing Sanity — Findings

---

Score: 58.9

**INFO** `per_diem_rates`

16690 per-diem rates (0.0%) — not dollar amounts; excluded from spread analysis

**INFO** `percentage_rates`

1277514 percentage rates (0.2%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

**ERROR** `zero_rates`

63126 zero-dollar rates (0.01%) — CMS schema requires `negotiated_rate > 0` (`exclusiveMinimum`)

**WARNING** `extreme_rates`

1.95% of rates are extreme (15615235 above class-specific high threshold, 0 below \$0.01)

**WARNING** `high_frequency_rate_value`

1 rate value(s) appear with suspiciously high frequency ( $\geq 0.5\%$  of dollar rates and  $\geq 50$  occurrences) — likely placeholder/sentinel values rather than real negotiated rates.

- `{'rate': 2250000.0, 'count': 10924257, 'pct': 1.37}`

**WARNING** `rate_spread_by_class`

`billing_class='institutional' / negotiated_type='negotiated'`: P95/P50 spread is 2663.4x (threshold: 10x, N=79,962 (1,000 sampled), high confidence)

**WARNING** `rate_spread_by_class`

`billing_class='professional' / negotiated_type='fee schedule'`: P95/P50 spread is 7.7x (threshold: 5x, N=786,489,844 (1,000 sampled), high confidence)

**WARNING** `rate_spread_by_class`

`billing_class='professional' / negotiated_type='negotiated'`: P95/P50 spread is 3619.7x (threshold: 5x, N=162,125 (1,000 sampled), high confidence)

---

**WARNING** per\_code\_rate\_spread

9148 rate contexts have a max/min ratio exceeding the type-specific threshold (20x professional / 50x facility, min 3 occurrences required). Each context is a unique combination of all 10 rate-key dimensions. n= shows how many distinct provider rates exist for that exact context.

Code	Code Type	Neg. Type	Billing Class	Arrangement	Setting	Min	Median	Mean	Max	Ratio	n
90399	CPT	fee schedule	professional	ffs	—	\$0.01	\$0.01	\$930000.01	\$2400000.00	240000000.0x	5
90758	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90667	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90676	CPT	fee schedule	professional	ffs	—	\$0.01	\$367.34	\$284243.54	\$2400000.00	240000000.0x	84
90666	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90664	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90668	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90296	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90287	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13
90393	CPT	fee schedule	professional	ffs	—	\$0.01	\$2400000.00	\$1834615.39	\$2400000.00	240000000.0x	13

## Recommended Actions

1. **schema** file\_freshness P1

File is 299 days old (last\_updated\_on exceeds the 90-day threshold)

2. **schema** cms\_schema\_validation P1

CMS official schema validator FAILED (exit code -1). File does not conform to the TIC in-network-rates schema.

3. **pricing** zero\_rates P1

63126 zero-dollar rates (0.01%) — CMS schema requires negotiated\_rate > 0 (exclusiveMinimum)

4. **provider\_mapping** `npi_validity`

P2

0.00% of NPIs failed Luhn checksum validation (4 of 111626)

5. **provider\_mapping** `ein_validity`

P2

0.31% of EINs failed IRS prefix validation (191 of 62130)

6. **provider\_mapping** `empty_npi_groups`

P2

53 provider groups contain no NPIs

7. **pricing** `extreme_rates`

P2

1.95% of rates are extreme (15615235 above class-specific high threshold, 0 below \$0.01)

8. **pricing** `high_frequency_rate_value`

P2

1 rate value(s) appear with suspiciously high frequency ( $\geq 0.5\%$  of dollar rates and  $\geq 50$  occurrences) — likely placeholder/sentinel values rather than real negotiated rates.

9. **pricing** `rate_spread_by_class`

P2

billing\_class='institutional' / negotiated\_type='negotiated': P95/P50 spread is 2663.4x (threshold: 10x, N=79,962 (1,000 sampled), high confidence)

10. **pricing** `rate_spread_by_class`

P2

billing\_class='professional' / negotiated\_type='fee schedule': P95/P50 spread is 7.7x (threshold: 5x, N=786,489,844 (1,000 sampled), high confidence)

11. **pricing** `rate_spread_by_class`

P2

billing\_class='professional' / negotiated\_type='negotiated': P95/P50 spread is 3619.7x (threshold: 5x, N=162,125 (1,000 sampled), high confidence)

12. **pricing** per\_code\_rate\_spread

P2

9148 rate contexts have a max/min ratio exceeding the type-specific threshold (20x professional / 50x facility, min 3 occurrences required). Each context is a unique combination of all 10 rate-key dimensions. n= shows how many distinct provider rates exist for that exact context.

13. **schema** item\_required\_fields

P2

0.01% of in\_network items are missing required fields

14. **schema** expired\_prices

P2

66246002 negotiated\_prices have past expiration dates (8.3%)

15. **code\_coverage** unknown\_billing\_code\_types

P3

Unrecognized billing\_code\_type values: {'PROC': 1}

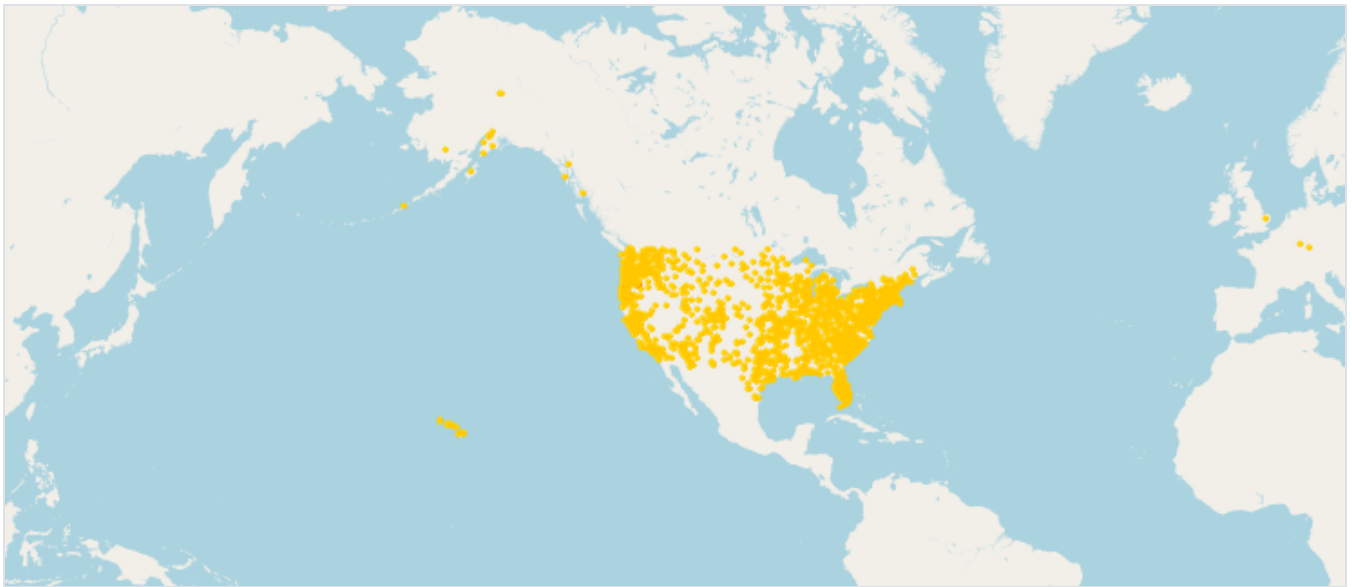
16. **code\_coverage** billing\_code\_format

P3

3924 CPT codes do not match expected format

## Provider Geographic Coverage

41630 unique NPIs found — 41413 geocoded (99%) — 4889 zip codes represented.



## Schema Integrity — Metrics

header_missing_fields		
header_conditional_issues		
file_age_days	299	
items_total	26125	
items_missing_required_pct	0.008	
items_empty_rates	0	
prices_total	801474358	
prices_missing_required_pct	0.0	
prices_missing_field_breakdown		
prices_missing_service_code	0	
prices_invalid_billing_class	0	
rates_without_providers	0	
negotiation_arrangements	ffs	26125

billing_code_types	MS-DRG	818
	HCPCS	7812
	CPT	15112
	CDT	939
	RC	566
	APC	877
	PROC	1
expired_prices	66246002	
invalid_expiration_format	0	

## Provider Mapping — Metrics

---

provider_references_in_file	50784
provider_group_ids_referenced	46897
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	111626
invalid_npi_count	4
npi_validity_rate_pct	100.0
invalid_npi_examples	1336588484, 1336588484, 1659650617, 1659650617
eins_validated	62130
invalid_ein_count	191
ein_validity_rate_pct	99.69
invalid_ein_examples	000000001, 000000001, 000000001, 001625898, 001625898, 001625898, 003360711, 003360711, 003360711, 003680161
empty_npi_groups	53
groups_without_tin	0
npis_in_multiple_groups	36183

## Code Coverage — Metrics

---

unique_codes_total	26125
duplicate_codes	0

<b>duplicate_pct</b>	0.0		
<b>by_code_type</b>	<b>MS-DRG</b>	818	
	<b>HCPCS</b>	7812	
	<b>CPT</b>	15112	
	<b>CDT</b>	939	
	<b>RC</b>	566	
	<b>APC</b>	877	
	<b>PROC</b>	1	
<b>unknown_code_types</b>	<b>PROC</b>	1	
<b>format_invalid_by_type</b>	<b>CPT</b>	3924	
<b>codes_not_in_reference</b>	reference_not_loaded		
<b>most_frequent_codes</b>	<b>Type</b>	<b>Code</b>	<b>Occurrences</b>
	<b>MS-DRG</b>	827	1
	<b>HCPCS</b>	J9400	1
	<b>CPT</b>	3CHPX	1
	<b>HCPCS</b>	G2118	1
	<b>CPT</b>	26650	1
	<b>HCPCS</b>	J7178	1
	<b>HCPCS</b>	E1015	1
	<b>CPT</b>	32488	1
	<b>CPT</b>	33732	1
	<b>CDT</b>	D0476	1
	<b>CPT</b>	27275	1
	<b>HCPCS</b>	S3652	1
	<b>CPT</b>	24615	1
	<b>CPT</b>	43250	1
	<b>HCPCS</b>	A7526	1
	<b>RC</b>	0133	1
	<b>CPT</b>	00222	1
	<b>CPT</b>	32667	1
	<b>CPT</b>	99000	1
<b>CPT</b>	3493F	1	

## Pricing Sanity — Metrics

<b>total_prices_checked</b>	801474358
<b>total_rates</b>	800180154

per_diem_rates	16690																																			
percentage_rates	1277514																																			
negative_rates	0																																			
zero_rates	63126																																			
extreme_high_rates	15615235																																			
extreme_low_rates	0																																			
rate_distribution	<table border="1"> <tr> <td>sample_n</td> <td>800180154</td> </tr> <tr> <td>sample_k</td> <td>5000</td> </tr> <tr> <td>confidence</td> <td>high</td> </tr> <tr> <td>p5</td> <td>28.494500000000002</td> </tr> <tr> <td>p25</td> <td>174.4525</td> </tr> <tr> <td>p50</td> <td>678.26</td> </tr> <tr> <td>p75</td> <td>1885.31</td> </tr> <tr> <td>p95</td> <td>5215.44200000000055</td> </tr> <tr> <td>p99</td> <td>2250000.0</td> </tr> </table>	sample_n	800180154	sample_k	5000	confidence	high	p5	28.494500000000002	p25	174.4525	p50	678.26	p75	1885.31	p95	5215.44200000000055	p99	2250000.0																	
sample_n	800180154																																			
sample_k	5000																																			
confidence	high																																			
p5	28.494500000000002																																			
p25	174.4525																																			
p50	678.26																																			
p75	1885.31																																			
p95	5215.44200000000055																																			
p99	2250000.0																																			
by_billing_class	<table border="1"> <thead> <tr> <th>Class / Type</th> <th>Count</th> <th>Median</th> <th>p25</th> <th>p75</th> <th>p95</th> <th>Confidence</th> </tr> </thead> <tbody> <tr> <td>institutional/ negotiated</td> <td>79,962</td> <td>617.4</td> <td>6.0</td> <td>33836.4</td> <td>1644300.4</td> <td>high</td> </tr> <tr> <td>professional/ fee schedule</td> <td>786,489,844</td> <td>782.4</td> <td>208.1</td> <td>2019.4</td> <td>6008.1</td> <td>high</td> </tr> <tr> <td>institutional/ fee schedule</td> <td>13,448,223</td> <td>596.9</td> <td>135.0</td> <td>1825.8</td> <td>5563.4</td> <td>high</td> </tr> <tr> <td>professional/ negotiated</td> <td>162,125</td> <td>621.6</td> <td>240.0</td> <td>2129.0</td> <td>2250000.0</td> <td>high</td> </tr> </tbody> </table>	Class / Type	Count	Median	p25	p75	p95	Confidence	institutional/ negotiated	79,962	617.4	6.0	33836.4	1644300.4	high	professional/ fee schedule	786,489,844	782.4	208.1	2019.4	6008.1	high	institutional/ fee schedule	13,448,223	596.9	135.0	1825.8	5563.4	high	professional/ negotiated	162,125	621.6	240.0	2129.0	2250000.0	high
Class / Type	Count	Median	p25	p75	p95	Confidence																														
institutional/ negotiated	79,962	617.4	6.0	33836.4	1644300.4	high																														
professional/ fee schedule	786,489,844	782.4	208.1	2019.4	6008.1	high																														
institutional/ fee schedule	13,448,223	596.9	135.0	1825.8	5563.4	high																														
professional/ negotiated	162,125	621.6	240.0	2129.0	2250000.0	high																														
negotiated_types	<table border="1"> <tr> <td>negotiated</td> <td>242087</td> </tr> <tr> <td>fee schedule</td> <td>799938067</td> </tr> </table>	negotiated	242087	fee schedule	799938067																															
negotiated	242087																																			
fee schedule	799938067																																			
unique_rate_contexts	101074																																			
rate_key_dimension_validity	<table border="1"> <tr> <td>invalid_negotiated_type</td> <td>0</td> </tr> <tr> <td>invalid_negotiated_types_seen</td> <td>{}</td> </tr> <tr> <td>invalid_setting</td> <td>0</td> </tr> <tr> <td>invalid_settings_seen</td> <td>{}</td> </tr> <tr> <td>invalid_severity_of_illness</td> <td>0</td> </tr> <tr> <td>severity_on_non_apr_drg</td> <td>0</td> </tr> <tr> <td>institutional_with_service_codes</td> <td>0</td> </tr> <tr> <td>invalid_service_code_format</td> <td>0</td> </tr> <tr> <td>billing_code_modifier_too_long</td> <td>0</td> </tr> </table>	invalid_negotiated_type	0	invalid_negotiated_types_seen	{}	invalid_setting	0	invalid_settings_seen	{}	invalid_severity_of_illness	0	severity_on_non_apr_drg	0	institutional_with_service_codes	0	invalid_service_code_format	0	billing_code_modifier_too_long	0																	
invalid_negotiated_type	0																																			
invalid_negotiated_types_seen	{}																																			
invalid_setting	0																																			
invalid_settings_seen	{}																																			
invalid_severity_of_illness	0																																			
severity_on_non_apr_drg	0																																			
institutional_with_service_codes	0																																			
invalid_service_code_format	0																																			
billing_code_modifier_too_long	0																																			

# Scoring Methodology

Embedded in this report at generation time.

## Overall Score

Weighted sum of four structural dimensions, normalized to a 0–100 scale.

Normalized Weights		
	Schema Integrity	30%
	Provider Mapping	15%
	Code Coverage	15%
	Pricing Sanity	40%

Confidence Bands		
	High	≥90
	Usable With Caution	≥75
	Limited Reliability	≥60
	Not Usable	<60

Score Caps		
	Raw Json Errors Only → 74.0	Native JSON syntax errors in the unpatched source file. File must be re-exported by the payer; scoring reflects auto-patched data only.
	Cms Validation Failure Only → 65.0	CMS official schema validator reports the file does not conform to the TIC spec.
	Both Raw Json Errors And Cms Failure → 59.0	Both native JSON syntax errors and CMS schema validation failure present.

## Rate Context Key — 14-Tuple Field Coverage

Every rate in a CMS TIC MRF file is described by a 14-field tuple. Fields 1–10 form the rate-context key used to group and compare rates across the system. Fields 11–12 (provider, expiration date) are validated separately and excluded from the grouping key for analytical reasons. Each of the four scoring dimensions validates a distinct slice of this tuple — together they cover all 14 fields.

Field	Validated by
1 billing_code_type	Schema (required field) + Code Coverage (enum + format validation)
2 billing_code_type_version	Schema (required field)
3 billing_code	Schema (required field) + Code Coverage (format, duplicates, reference lookup)
4 billing_code_modifier	Pricing (modifier length, key normalization)
5 service_code	Pricing (POS format, normalization, institutional-class check)
6 negotiated_type	Pricing (CMS TIC enum validation)
7 billing_class	Schema (CMS TIC enum validation) + Pricing (spread thresholds)
8 negotiation_arrangement	Schema (CMS TIC enum validation) + Pricing (FFS vs bundle/capitation gating)
9 severity_of_illness	Pricing (APR-DRG only, valid values 1–4)

Field	Validated by
10 setting	Pricing (CMS TIC enum validation)
11 provider (NPI/EIN)	Provider Mapping (Luhn checksum, IRS prefix, group resolution) — excluded from grouping key
12 expiration_date	Schema (date validity, far-future sanity) — excluded from grouping key
13 additional_generic_notes	not validated (free-text)
14 negotiated_rate	Pricing (negative/zero/extreme-value checks, spread analysis)

- Fields 1–10 are the grouping key. Each unique combination is a distinct rate context — rates with different modifiers, POS codes, or arrangements land in separate buckets and are never compared against each other.
- Provider (field 11) is excluded from the key: the spread check is cross-provider by design. Partitioning by provider produces singleton buckets and eliminates the spread signal.
- Expiration date (field 12) is excluded because it is a contract lifecycle attribute, not a clinical context. Rates for the same service should be comparable regardless of when they expire.
- service\_code (field 5) arrays are flattened and normalized before keying: '1' → '01', and a rate with ['11','22'] contributes to both the '11' and '22' buckets so rates are compared apples-to-apples by place of service.

## Schema Integrity

Validates required fields, enum values, conditional requirements, and date validity per the CMS TIC in-network-rates schema. Also checks file freshness and expiration date sanity.

**Method:** Penalty-based deductions from 100, capped per category.

per_missing_required_header_field	5
per_header_conditional_issue	2
freshness_warn	5
freshness_error	10
item_missing_fields_pct	×5 (cap 30)
empty_rates_pct	×0.5 (cap 5)
price_missing_fields_pct	×10 (cap 30)
rates_without_providers_rate	×200 (cap 20)
expired_prices_pct	×0.5 (cap 5)
file freshness warn days	45
file freshness error days	90
expiry far future years	3

## Provider Mapping

Verifies that all provider\_group\_id references in in\_network items resolve to an entry in the provider\_references array. Validates NPI integrity via Luhn checksum and EIN integrity via IRS-issued 2-digit prefix.

**Method:** Weighted component sum (not purely penalty-based).

<b>provider_resolution (60%)</b>	$\text{resolution\_rate\%} \times 0.60$
<b>npi_validity (30%)</b>	$(100 - \text{invalid\_npi\_pct} \times 5) \times 0.30$
<b>ein_validity (10%)</b>	$10 - (\text{invalid\_ein\_pct} \times 0.1)$ [0% invalid → 10 pts, 100% invalid → 0 pts, linear]

## Code Coverage

Tracks every (billing\_code\_type, billing\_code) pair and flags unrecognized CMS TIC code types, format violations for CPT/HCCPS/NDC, and duplicates (same code appearing in multiple in\_network items).

**Method:** Penalty-based deductions from 100.

<b>per_unknown_code_type</b>	3 pts each (cap 20)
<b>format_invalid_pct</b>	$\times 0.5$ (cap 10)
<b>duplicate_code_pct</b>	$\times 2$ (cap 20)
<b>codes_not_in_reference_pct</b>	$\times 0.5$ (cap 30) — only when reference set is loaded

## Pricing Sanity

Detects invalid rates (negative, zero, extreme-value) and distribution anomalies (per-class P95/P50 spread, per-code max/min ratio, flat-rate distributions). Exact counts are used for all validity checks (negative, zero, extreme, dimension validity). Percentile-based checks (spread, IQR) use reservoir sampling —  $k=5\,000$  global,  $k=1\,000$  per (billing\_class, negotiated\_type) bucket — so memory stays bounded on large files. Per-code max/min spread is exact (all rates seen, no sampling).

**Method:** Penalty-based deductions from 100.

<b>negative_rate_pct</b>	$\times 5$ (cap 20)
<b>zero_rate_pct</b>	$\times 3$ (cap 15)
<b>extreme_rate_pct</b>	$\times 5$ (cap 25) — ffs only
<b>class_spread_excess</b>	$(\text{spread} - \text{threshold}) \times 2$ , max across (billing_class, negotiated_type) buckets (cap 15)
<b>per_code_high_spread_count</b>	$\times 0.1$ (cap 15)
<b>invalid_negotiated_type_pct</b>	$\times 3$ (cap 10) — rates silently dropped
<b>invalid_setting_pct</b>	$\times 1$ (cap 5) — silently defaults to wildcard
<b>invalid_severity_pct</b>	$\times 1$ (cap 5) — silently normalised to ''
<b>institutional_with_service_codes_pct</b>	$\times 1$ (cap 5) — extra key variation
<b>invalid_service_code_pct</b>	$\times 2$ (cap 5) — encode raises ValueError
<b>extreme high by billing class</b>	professional: 25000.0, institutional: 2000000.0, both: 2000000.0, default: 500000.0

<b>extreme low</b>	0.01
<b>spread warn p95 over p50 by class</b>	professional: 5, institutional: 10, both: 10, default: 5
<b>per rate context max min ratio</b>	professional_codes: 20, facility_drg_codes: 50
<b>flat rate iqr p75 threshold pct</b>	5.0
<b>flat rate min rates to check</b>	100
<b>spread min n to flag</b>	50
<b>per code min n to flag</b>	3

## Dashboard: MRF Identity Key

(ingest-time — not stored in report JSON)

The dashboard assigns a persistent `mrf_key` to each MRF so that all validation runs of the same file are grouped together in the score-history view, even if the payer re-exports the file at a new URL.

<b>Tier 1 — entity + plan_id</b>	Used when both <code>reporting_entity_name</code> and <code>plan_id</code> are present. Key input: <code>plan &lt;entity&gt; &lt;plan_id_type&gt; &lt;plan_id&gt;</code> . Stable across monthly re-exports.
<b>Tier 2 — URL hash</b>	Fallback when <code>plan_id</code> is absent. Key input: the raw file location URL/path. Entity name alone is not used — a payer publishes multiple distinct plans under the same entity name and without <code>plan_id</code> they cannot be safely distinguished. A URL change produces a different key.

The key is a 16-character MD5 hex digest of the input string (case-insensitive, whitespace-stripped). **This run:** `mrf_key = c65eaec782fec5eb · entity = Providence_Health_Plan · tier = 2 (URL hash)`

## Provider Geographic Coverage

(supplemental — does not affect score)

Geographic analysis is a supplemental feature computed on demand after scoring completes. It does not affect any scoring dimension — it is an observational overlay to assess the breadth and distribution of in-network providers.

<b>NPPES</b>	CMS National Plan and Provider Enumeration System — monthly full-replacement CSV. Maps each NPI to its primary registered ZIP code.
<b>ZCTA centroids</b>	GeoNames US postal code file. Maps each 5-digit ZIP to a (latitude, longitude) centroid for map placement.

**Process:** Extract all NPIs from the MRF file → resolve each NPI to its primary practice ZIP via NPPES → aggregate provider count per ZIP → map each ZIP to a lat/lon centroid via ZCTA → render as a weighted heatmap (intensity ∝ provider count per ZIP).

**Limitations:** NPIs absent from NPPES (recently issued, test NPIs, EINs) are excluded and reduce the geocoding match rate. Location reflects the provider's NPPES-registered primary address, not necessarily where they accept this specific plan. Map viewport covers the bounding box of ZIP codes representing 90% of total provider count, dropping sparse geographic outliers.