

# MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/  
2025-10/2025-07-17\_Providence\_Health\_Plan\_e572c172672d3c4f8e527be3eb5f6ad2\_in-network-rates.json.gz

Size: 163.96 MB • MD5: 1edce631d221e1686886902ccc04ac2e

Payer: **Providence\_Health\_Plan** • File Date: **2025-07-17** • Generated: **2026-04-26 07:37 EDT** • Tool Version: **1.0.0** •

Elapsed: **1875.80s**

# 74.0

Limited Reliability

score capped at 74.0 — 1 native JSON parse error(s) require payer re-export (see meta for details)

Errors: 3 • Warnings: 10 • Info: 3 • **1 Raw JSON Error(s)**

## 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/2025-10/2025-07-17\_Providence\_Health\_Plan\_e572c172672d3c4f8e527be3eb5f6ad2\_in-network-rates.json.gz

Plan Name	Plan ID	Issuer / Sponsor	Market
HSA	941672743 (EIN)	—	group
OPTIONADV_PREM	941672743 (EIN)	—	group

## Raw JSON Parse Errors

### ResponseStreamingError

An error occurred while reading from response stream: ('Connection broken: IncompleteRead(171926412 bytes read, 6410327283 more expected)', IncompleteRead(171926412 bytes read, 6410327283 more expected))

Full message:

```
An error occurred while reading from response stream: ('Connection broken: IncompleteRead(171926412 bytes read, 6410327283 more expected)', IncompleteRead(171926412 bytes read, 6410327283 more expected))
```

# Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	88.2	30%	4
Provider Mapping	99.9	15%	4
Code Coverage	90.0	15%	1
Pricing Sanity	72.8	40%	7

## Schema Integrity — Findings

---

Score: 88.2

**ERROR** file\_freshness

File is 283 days old (last\_updated\_on exceeds the 90-day threshold)

---

**WARNING** expired\_prices

21836 negotiated\_prices have past expiration dates (3.7%)

---

**ERROR** raw\_json\_error

Native JSON parse error detected in unpatched source file: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(171926412 bytes read, 6410327283 more expected)', IncompleteRead(171926412 bytes read, 6410327283 more expected)). Scoring below reflects auto-patched data only. File must be re-exported by the payer to fully resolve.

---

**ERROR** parse\_crash

single pass crash after 40 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(9468334 bytes read, 6572785361 more expected)', IncompleteRead(9468334 bytes read, 6572785361 more expected))

---

## Provider Mapping — Findings

---

Score: 99.9

**WARNING** `npi_validity`

0.01% of NPIs failed Luhn checksum validation (4 of 69488)

- 1336588484
  - 1659650617
- 

**WARNING** `ein_validity`

0.45% of EINs failed IRS prefix validation (189 of 42343)

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

**WARNING** `empty_npi_groups`

43 provider groups contain no NPIs

---

**INFO** `duplicate_npis`

23626 NPIs appear in more than one provider group

---

# Code Coverage — Findings

---

Score: 90.0

**WARNING** `billing_code_format`

11 CPT codes do not match expected format

---

## Pricing Sanity — Findings

---

Score: 72.8

**INFO** `per_diem_rates`

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

---

**INFO** `percentage_rates`

889 percentage rates (0.1%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

---

**WARNING** `zero_rates`

3 zero-dollar rates (0.00%) — CMS schema requires `negotiated_rate > 0` (`exclusiveMinimum`); may represent unfiled or excluded services

---

**WARNING** `high_frequency_rate_value`

44 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': 113.1, 'count': 19624, 'pct': 3.31}`
  - `{'rate': 937.43, 'count': 19622, 'pct': 3.31}`
  - `{'rate': 2706.57, 'count': 19622, 'pct': 3.31}`
  - `{'rate': 74.39, 'count': 19622, 'pct': 3.31}`
  - `{'rate': 4524.0, 'count': 19622, 'pct': 3.31}`
  - ... and 5 more
- 

**WARNING** `rate_spread_by_class`

`billing_class='professional' / negotiated_type='fee schedule': P95/P50 spread is 18.0x (threshold: 5x, N=574,295 (1,000 sampled), high confidence)`

---

**WARNING** `rate_spread_by_class`

`billing_class='institutional' / negotiated_type='fee schedule': P95/P50 spread is 70.1x (threshold: 10x, N=18,320 (1,000 sampled), high confidence)`

---

**WARNING** `per_code_rate_spread`

15 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
K0041	HCPCS	fee schedule	professional	ffs	—	\$4.52	\$5.87	\$12392.07	\$2400000.00	530973.5×	11626
K0041	HCPCS	fee schedule	professional	ffs	—	\$5.28	\$6.62	\$12419.18	\$2400000.00	454545.5×	11601
K0041	HCPCS	fee schedule	professional	ffs	—	\$5.39	\$6.20	\$12418.95	\$2400000.00	445269.0×	11601
K0041	HCPCS	fee schedule	professional	ffs	—	\$33.90	\$46.50	\$12431.23	\$2400000.00	70796.5×	11626
K0041	HCPCS	fee schedule	professional	ffs	—	\$39.66	\$49.73	\$12460.93	\$2400000.00	60514.4×	11601
K0041	HCPCS	fee schedule	professional	ffs	—	\$40.41	\$46.50	\$12459.32	\$2400000.00	59391.2×	11601
K0041	HCPCS	fee schedule	professional	ffs	—	\$45.20	\$62.00	\$12446.31	\$2400000.00	53097.3×	11626
K0041	HCPCS	fee schedule	professional	ffs	—	\$45.20	\$62.00	\$12445.24	\$2400000.00	53097.3×	11627
K0041	HCPCS	fee schedule	institutional	ffs	—	\$56.57	\$56.57	\$3174.29	\$2700000.00	47728.5×	866
K0041	HCPCS	fee schedule	institutional	ffs	—	\$56.57	\$56.57	\$3174.29	\$2700000.00	47728.5×	866

## Recommended Actions

1. `schema` `file_freshness`

P1

File is 283 days old (`last_updated_on` exceeds the 90-day threshold)

2. `schema` `raw_json_error`

P1

Native JSON parse error detected in unpatched source file: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(171926412 bytes read, 6410327283 more expected)', IncompleteRead(171926412 bytes read, 6410327283 more expected)). Scoring below reflects auto-patched data only. File must be re-exported by the payer to fully resolve.

3. **schema** parse\_crash

P1

single pass crash after 40 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(9468334 bytes read, 6572785361 more expected)', IncompleteRead(9468334 bytes read, 6572785361 more expected))

4. **provider\_mapping** np\_i\_validity

P2

0.01% of NPIs failed Luhn checksum validation (4 of 69488)

5. **provider\_mapping** ein\_validity

P2

0.45% of EINs failed IRS prefix validation (189 of 42343)

6. **provider\_mapping** empty\_np\_i\_groups

P2

43 provider groups contain no NPIs

7. **pricing** zero\_rates

P2

3 zero-dollar rates (0.00%) — CMS schema requires negotiated\_rate > 0 (exclusiveMinimum); may represent unfiled or excluded services

8. **pricing** high\_frequency\_rate\_value

P2

44 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='professional' / negotiated\_type='fee schedule': P95/P50 spread is 18.0x (threshold: 5x, N=574,295 (1,000 sampled), high confidence)

10. **pricing** rate\_spread\_by\_class

P2

billing\_class='institutional' / negotiated\_type='fee schedule': P95/P50 spread is 70.1x (threshold: 10x, N=18,320 (1,000 sampled), high confidence)

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

P2

15 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.

12. **schema** expired\_prices

P2

21836 negotiated\_prices have past expiration dates (3.7%)

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

P3

11 CPT codes do not match expected format

## Provider Geographic Coverage

25259 unique NPIs found — 25127 geocoded (99%) — 3143 zip codes represented.



## Schema Integrity — Metrics

header\_missing\_fields

header\_conditional\_issues

file\_age\_days

283

items_total	40	
items_missing_required_pct	0.0	
items_empty_rates	0	
prices_total	593595	
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	40
billing_code_types	CPT	27
	CDT	1
	HCPCS	10
	APC	1
	RC	1
expired_prices	21836	
invalid_expiration_format	0	

## Provider Mapping — Metrics

---

provider_references_in_file	30148
provider_group_ids_referenced	27792
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	69488
invalid_npi_count	4
npi_validity_rate_pct	99.99
invalid_npi_examples	1336588484, 1336588484, 1659650617, 1659650617
eins_validated	42343
invalid_ein_count	189
ein_validity_rate_pct	99.55

<b>invalid_ein_examples</b>	000000001, 000000001, 000000001, 001625898, 001625898, 001625898, 003360711, 003360711, 003360711, 003680161
<b>empty_npi_groups</b>	43
<b>groups_without_tin</b>	0
<b>npis_in_multiple_groups</b>	23626

## Code Coverage — Metrics

---

<b>unique_codes_total</b>	40										
<b>duplicate_codes</b>	0										
<b>duplicate_pct</b>	0.0										
<b>by_code_type</b>	<table border="1"> <tr> <td>CPT</td> <td>27</td> </tr> <tr> <td>CDT</td> <td>1</td> </tr> <tr> <td>HCPCS</td> <td>10</td> </tr> <tr> <td>APC</td> <td>1</td> </tr> <tr> <td>RC</td> <td>1</td> </tr> </table>	CPT	27	CDT	1	HCPCS	10	APC	1	RC	1
CPT	27										
CDT	1										
HCPCS	10										
APC	1										
RC	1										
<b>unknown_code_types</b>											
<b>format_invalid_by_type</b>	<table border="1"> <tr> <td>CPT</td> <td>11</td> </tr> </table>	CPT	11								
CPT	11										
<b>codes_not_in_reference</b>	reference_not_loaded										

most_frequent_codes	Type	Code	Occurrences
	CPT	86042	1
	CPT	3315F	1
	CPT	1BFKW	1
	CPT	4CHL4	1
	CPT	1CFM4	1
	CPT	81334	1
	CPT	88188	1
	CPT	00942	1
	CPT	1BFMX	1
	CPT	82495	1
	CDT	D5915	1
	HCPCS	E0486	1
	APC	706	1
	CPT	2BA11	1
	CPT	3BFMT	1
	CPT	54304	1
	CPT	4013F	1
	HCPCS	E1084	1
	CPT	00124	1
	HCPCS	L5859	1

## Pricing Sanity — Metrics

total_prices_checked	593595
total_rates	592698
per_diem_rates	8
percentage_rates	889
negative_rates	0
zero_rates	3
extreme_high_rates	782
extreme_low_rates	0

<b>rate_distribution</b>	<b>sample_n</b>	592698
	<b>sample_k</b>	5000
	<b>confidence</b>	high
	<b>p5</b>	6.79
	<b>p25</b>	74.39
	<b>p50</b>	413.76
	<b>p75</b>	1131.32
	<b>p95</b>	9398.31
	<b>p99</b>	19178.3

  

<b>by_billing_class</b>	<b>Class / Type</b>	<b>Count</b>	<b>Median</b>	<b>p25</b>	<b>p75</b>	<b>p95</b>	<b>Confidence</b>
	professional/ fee schedule	574,295	521.6	113.1	1595.3	9398.3	high
	institutional/ fee schedule	18,320	128.0	44.7	884.7	8972.3	high
	institutional/ negotiated	31	178.2	150.4	228.6	600226.6	low
	professional/ negotiated	52	306.1	300.0	325.5	391.1	moderate

  

<b>negotiated_types</b>	<b>fee schedule</b>	592615
	<b>negotiated</b>	83

  

<b>unique_rate_contexts</b>	77
-----------------------------	----

  

<b>rate_key_dimension_validity</b>	<b>invalid_negotiated_type</b>	0
	<b>invalid_negotiated_types_seen</b>	{}
	<b>invalid_setting</b>	0
	<b>invalid_settings_seen</b>	{}
	<b>invalid_severity_of_illness</b>	0
	<b>severity_on_non_apr_drg</b>	0
	<b>institutional_with_service_codes</b>	0
	<b>invalid_service_code_format</b>	0
	<b>billing_code_modifier_too_long</b>	0

## Scoring Methodology

Embedded in this report at generation time.

### Overall Score

Weighted sum of four structural dimensions, normalized to a 0–100 scale. Pricing sanity is excluded when the file contains only non-dollar rate types (capitation, per-diem, or percentage) — dollar-amount spread analysis is not applicable in that case.

<b>Normalized Weights</b>	<b>Schema Integrity</b>	30%
	<b>Provider Mapping</b>	15%
	<b>Code Coverage</b>	15%
	<b>Pricing Sanity</b>	40%
<b>Confidence Bands</b>	<b>High</b>	≥90
	<b>Usable With Caution</b>	≥75
	<b>Limited Reliability</b>	≥60
	<b>Not Usable</b>	<60
<b>Score Caps</b>	<b>Raw Json Errors Only → 74.0</b>	Native JSON syntax errors in the unpatched source file. File must be re-exported by the payer; scoring reflects auto-patched data only.
	<b>Cms Validation Failure Only → 65.0</b>	CMS official schema validator reports the file does not conform to the TIC spec.
	<b>Both Raw Json Errors And Cms Failure → 59.0</b>	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 <b>billing_code_type</b>	Schema (required field) + Code Coverage (enum + format validation)
2 <b>billing_code_type_version</b>	Schema (required field)
3 <b>billing_code</b>	Schema (required field) + Code Coverage (format, duplicates, reference lookup)
4 <b>billing_code_modifier</b>	Pricing (modifier length, key normalization)
5 <b>service_code</b>	Pricing (POS format, normalization, institutional-class check)
6 <b>negotiated_type</b>	Pricing (CMS TIC enum validation)
7 <b>billing_class</b>	Schema (CMS TIC enum validation) + Pricing (spread thresholds)

Field	Validated by
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)
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/HCPCS/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) — warning (not error)
<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 = af936453c2c74426 · 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  $\propto$  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.