

MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/
2025-10/2025-07-11_Providence_Health_Plan_bc6ebe144c2cce3bf4a1b20a9d6fb122_in-network-rates.json.gz
Size: 23.23 MB • MD5: dbec848010b2f6b25ff958b776c6414b

Payer: **Providence_Health_Plan** • File Date: **2025-07-11** • Generated: **2026-04-26 03:37 EDT** • Tool Version: **1.0.0** • Elapsed: **2684.80s**

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/2025-10/2025-07-11_Providence_Health_Plan_bc6ebe144c2cce3bf4a1b20a9d6fb122_in-network-rates.json.gz

Plan Name	Plan ID	Issuer / Sponsor	Market
OPTIONADV_PREM_CHC	931244959 (EIN)	—	group

Raw JSON Parse Errors

ResponseStreamingError

An error occurred while reading from response stream: ('Connection broken: IncompleteRead(24357272 bytes read, 12503276889 more expected)', IncompleteRead(24357272 bytes read, 12503276889 more expected))

Full message:

```
An error occurred while reading from response stream: ('Connection broken: IncompleteRead(24357272 bytes read, 12503276889 more expected)', IncompleteRead(24357272 bytes read, 12503276889 more expected))
```

Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	86.1	30%	4
Provider Mapping	100.0	15%	4
Code Coverage	97.1	15%	1
Pricing Sanity	79.2	40%	7

Schema Integrity — Findings

Score: 86.1

ERROR file_freshness

File is 289 days old (last_updated_on exceeds the 90-day threshold)

WARNING expired_prices

145668 negotiated_prices have past expiration dates (7.8%)

ERROR raw_json_error

Native JSON parse error detected in unpatched source file: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(24357272 bytes read, 12503276889 more expected)', IncompleteRead(24357272 bytes read, 12503276889 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 35 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(28329050 bytes read, 12499305111 more expected)', IncompleteRead(28329050 bytes read, 12499305111 more expected))

Provider Mapping — Findings

Score: 100.0

WARNING `npi_validity`

0.00% of NPIs failed Luhn checksum validation (4 of 113824)

- 1336588484
- 1659650617

WARNING `ein_validity`

0.30% of EINs failed IRS prefix validation (192 of 63849)

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

WARNING `empty_npi_groups`

43 provider groups contain no NPIs

INFO `duplicate_npis`

36230 NPIs appear in more than one provider group

Code Coverage — Findings

Score: 97.1

WARNING `billing_code_format`

2 CPT codes do not match expected format

Pricing Sanity — Findings

Score: 79.2

INFO `per_diem_rates`

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

INFO `percentage_rates`

2405 percentage rates (0.1%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

WARNING `zero_rates`

300 zero-dollar rates (0.02%) — CMS schema requires `negotiated_rate > 0` (`exclusiveMinimum`); may represent unfiled or excluded services

WARNING `high_frequency_rate_value`

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

- `{'rate': 718.62, 'count': 19684, 'pct': 1.05}`
 - `{'rate': 1002.24, 'count': 19684, 'pct': 1.05}`
 - `{'rate': 662.07, 'count': 19684, 'pct': 1.05}`
 - `{'rate': 1961.85, 'count': 19684, 'pct': 1.05}`
 - `{'rate': 1286.3, 'count': 19684, 'pct': 1.05}`
 - ... and 5 more
-

WARNING `rate_spread_by_class`

`billing_class='institutional' / negotiated_type='negotiated'`: P95/P50 spread is 8271.3x (threshold: 10x, N=168, moderate confidence)

WARNING `rate_spread_by_class`

`billing_class='institutional' / negotiated_type='fee schedule'`: P95/P50 spread is 12.1x (threshold: 10x, N=31,891 (1,000 sampled), high confidence)

WARNING per_code_rate_spread

17 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
J0171	HCPCS	negotiated	institutional	ffs	—	\$0.76	\$600000.56	\$621562.97	\$1286250.00	1692434.2x	4
85292	CPT	fee schedule	professional	ffs	—	\$14.58	\$27.26	\$704.93	\$2250000.00	154321.0x	3311
E2366	HCPCS	fee schedule	professional	ffs	—	\$15.57	\$20.64	\$7918.10	\$2400000.00	154142.6x	28131
E2366	HCPCS	fee schedule	professional	ffs	—	\$18.18	\$22.40	\$7954.49	\$2400000.00	132013.2x	28005
E2366	HCPCS	fee schedule	professional	ffs	—	\$20.67	\$30.31	\$7963.51	\$2400000.00	116110.3x	28005
87487	CPT	fee schedule	professional	ffs	—	\$32.99	\$61.69	\$736.98	\$2250000.00	68202.5x	3311
A7048	HCPCS	fee schedule	professional	ffs	—	\$45.65	\$56.36	\$7878.84	\$2400000.00	52573.9x	28401
L2810	HCPCS	fee schedule	professional	ffs	—	\$66.06	\$85.93	\$7908.52	\$2400000.00	36330.6x	28402
J1000	HCPCS	negotiated	institutional	ffs	—	\$38.65	\$600028.60	\$621586.46	\$1286250.00	33279.4x	4
E2366	HCPCS	fee schedule	professional	ffs	—	\$116.80	\$154.83	\$8055.28	\$2400000.00	20547.9x	28131

Recommended Actions

1. **schema** file_freshness

P1

File is 289 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(24357272 bytes read, 12503276889 more expected)', IncompleteRead(24357272 bytes read, 12503276889 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 35 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(28329050 bytes read, 12499305111 more expected)', IncompleteRead(28329050 bytes read, 12499305111 more expected))

4. **provider_mapping** `mpi_validity`

P2

0.00% of NPIs failed Luhn checksum validation (4 of 113824)

5. **provider_mapping** `ein_validity`

P2

0.30% of EINs failed IRS prefix validation (192 of 63849)

6. **provider_mapping** `empty_mpi_groups`

P2

43 provider groups contain no NPIs

7. **pricing** `zero_rates`

P2

300 zero-dollar rates (0.02%) — CMS schema requires `negotiated_rate > 0` (exclusiveMinimum); may represent unfiled or excluded services

8. **pricing** `high_frequency_rate_value`

P2

37 rate value(s) appear with suspiciously high frequency ($\geq 0.5\%$ of dollar rates and ≥ 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 8271.3x (threshold: 10x, N=168, moderate confidence)

10. **pricing** `rate_spread_by_class`

P2

`billing_class='institutional' / negotiated_type='fee schedule'`: P95/P50 spread is 12.1x (threshold: 10x, N=31,891 (1,000 sampled), high confidence)

11. **pricing** `per_code_rate_spread`

P2

17 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

145668 negotiated_prices have past expiration dates (7.8%)

13. **code_coverage** billing_code_format

P3

2 CPT codes do not match expected format

Schema Integrity — Metrics

header_missing_fields		
header_conditional_issues		
file_age_days		289
items_total		35
items_missing_required_pct		0.0
items_empty_rates		0
prices_total		1876096
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	35
billing_code_types		
	RC	2
	CPT	21
	HCPCS	8
	CDT	2
	APC	1
	MS-DRG	1
expired_prices		145668
invalid_expiration_format		0

Provider Mapping — Metrics

provider_references_in_file	50835
provider_group_ids_referenced	48575
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	113824
invalid_npi_count	4
npi_validity_rate_pct	100.0
invalid_npi_examples	1336588484, 1336588484, 1659650617, 1659650617
eins_validated	63849
invalid_ein_count	192
ein_validity_rate_pct	99.7
invalid_ein_examples	000000001, 000000001, 000000001, 001625898, 001625898, 001625898, 003360711, 003360711, 003360711, 003680161
empty_npi_groups	43
groups_without_tin	0
npis_in_multiple_groups	36230

Code Coverage — Metrics

unique_codes_total	35	
duplicate_codes	0	
duplicate_pct	0.0	
by_code_type	RC	2
	CPT	21
	HCPCS	8
	CDT	2
	APC	1
	MS-DRG	1
unknown_code_types		
format_invalid_by_type	CPT	2

codes_not_in_reference	reference_not_loaded																																																															
most_frequent_codes	<table border="1"> <thead> <tr> <th>Type</th> <th>Code</th> <th>Occurrences</th> </tr> </thead> <tbody> <tr><td>RC</td><td>0179</td><td>1</td></tr> <tr><td>CPT</td><td>45020</td><td>1</td></tr> <tr><td>CPT</td><td>24931</td><td>1</td></tr> <tr><td>CPT</td><td>27202</td><td>1</td></tr> <tr><td>HCPCS</td><td>J1000</td><td>1</td></tr> <tr><td>CPT</td><td>28344</td><td>1</td></tr> <tr><td>CPT</td><td>57420</td><td>1</td></tr> <tr><td>CDT</td><td>D1208</td><td>1</td></tr> <tr><td>HCPCS</td><td>A7048</td><td>1</td></tr> <tr><td>CPT</td><td>85292</td><td>1</td></tr> <tr><td>CPT</td><td>21267</td><td>1</td></tr> <tr><td>HCPCS</td><td>J3535</td><td>1</td></tr> <tr><td>HCPCS</td><td>M1060</td><td>1</td></tr> <tr><td>HCPCS</td><td>E0425</td><td>1</td></tr> <tr><td>CPT</td><td>44700</td><td>1</td></tr> <tr><td>CPT</td><td>67027</td><td>1</td></tr> <tr><td>CPT</td><td>69643</td><td>1</td></tr> <tr><td>CPT</td><td>36845</td><td>1</td></tr> <tr><td>CPT</td><td>11401</td><td>1</td></tr> <tr><td>CPT</td><td>87487</td><td>1</td></tr> </tbody> </table>	Type	Code	Occurrences	RC	0179	1	CPT	45020	1	CPT	24931	1	CPT	27202	1	HCPCS	J1000	1	CPT	28344	1	CPT	57420	1	CDT	D1208	1	HCPCS	A7048	1	CPT	85292	1	CPT	21267	1	HCPCS	J3535	1	HCPCS	M1060	1	HCPCS	E0425	1	CPT	44700	1	CPT	67027	1	CPT	69643	1	CPT	36845	1	CPT	11401	1	CPT	87487	1
Type	Code	Occurrences																																																														
RC	0179	1																																																														
CPT	45020	1																																																														
CPT	24931	1																																																														
CPT	27202	1																																																														
HCPCS	J1000	1																																																														
CPT	28344	1																																																														
CPT	57420	1																																																														
CDT	D1208	1																																																														
HCPCS	A7048	1																																																														
CPT	85292	1																																																														
CPT	21267	1																																																														
HCPCS	J3535	1																																																														
HCPCS	M1060	1																																																														
HCPCS	E0425	1																																																														
CPT	44700	1																																																														
CPT	67027	1																																																														
CPT	69643	1																																																														
CPT	36845	1																																																														
CPT	11401	1																																																														
CPT	87487	1																																																														

Pricing Sanity — Metrics

total_prices_checked	1876096
total_rates	1873646
per_diem_rates	45
percentage_rates	2405
negative_rates	0
zero_rates	300
extreme_high_rates	5653
extreme_low_rates	0

rate_distribution	sample_n	1873646
	sample_k	5000
	confidence	high
	p5	48.298500000000004
	p25	232.55
	p50	854.995
	p75	1846.69
	p95	3792.92750000000025
	p99	4958.82

by_billing_class	Class / Type	Count	Median	p25	p75	p95	Confidence
	institutional/ negotiated	168	198.8	128.0	25637.9	1644414.3	moderate
	professional/ fee schedule	1,841,294	788.0	235.0	1782.6	3792.8	high
	institutional/ fee schedule	31,891	341.5	167.8	1478.5	4115.4	high
	professional/ negotiated	293	0.0	0.0	0.0	1289.3	moderate

negotiated_types	negotiated	461
	fee schedule	1873185

unique_rate_contexts	167
-----------------------------	-----

rate_key_dimension_validity	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. 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.

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)

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).

provider_resolution (60%)	$\text{resolution_rate\%} \times 0.60$
npi_validity (30%)	$(100 - \text{invalid_npi_pct} \times 5) \times 0.30$
ein_validity (10%)	$10 - (\text{invalid_ein_pct} \times 0.1)$ [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.

per_unknown_code_type	3 pts each (cap 20)
format_invalid_pct	$\times 0.5$ (cap 10)
duplicate_code_pct	$\times 2$ (cap 20)
codes_not_in_reference_pct	$\times 0.5$ (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.

negative_rate_pct	$\times 5$ (cap 20)
zero_rate_pct	$\times 3$ (cap 15) — warning (not error)
extreme_rate_pct	$\times 5$ (cap 25) — ffs only
class_spread_excess	$(\text{spread} - \text{threshold}) \times 2$, max across (billing_class, negotiated_type) buckets (cap 15)
per_code_high_spread_count	$\times 0.1$ (cap 15)
invalid_negotiated_type_pct	$\times 3$ (cap 10) — rates silently dropped
invalid_setting_pct	$\times 1$ (cap 5) — silently defaults to wildcard
invalid_severity_pct	$\times 1$ (cap 5) — silently normalised to ''
institutional_with_service_codes_pct	$\times 1$ (cap 5) — extra key variation
invalid_service_code_pct	$\times 2$ (cap 5) — encode raises ValueError

extreme high by billing class	professional: 25000.0, institutional: 2000000.0, both: 2000000.0, default: 500000.0
extreme low	0.01
spread warn p95 over p50 by class	professional: 5, institutional: 10, both: 10, default: 5
per rate context max min ratio	professional_codes: 20, facility_drg_codes: 50
flat rate iqr p75 threshold pct	5.0
flat rate min rates to check	100
spread min n to flag	50
per code min n to flag	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.

Tier 1 — entity + plan_id	Used when both <code>reporting_entity_name</code> and <code>plan_id</code> are present. Key input: <code>plan <entity> <plan_id_type> <plan_id></code> . Stable across monthly re-exports.
Tier 2 — URL hash	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 = 21d859147838b6ce · 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.

NPPES	CMS National Plan and Provider Enumeration System — monthly full-replacement CSV. Maps each NPI to its primary registered ZIP code.
ZCTA centroids	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.