MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2025-10/2025-07-16_Providence_Health_Plan_642241d881deabc826d74e8d171e12f6_in-network-rates.json.gz

Size: 79.89 MB • MD5: 72085046971c4241e541788af2331dc1

Payer: **Providence_Health_Plan** • File Date: **2025-07-16** • Generated: **2026-04-26 06:57 EDT** • Tool Version: **1.0.0** • Elapsed: **3299.00s**

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/2025-10/2025-07-16_Providence_Health_Plan_642241d881deabc826d74e8d171e12f6_in-network-rates.json.gz

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

Raw JSON Parse Errors

ResponseStreamingError

An error occurred while reading from response stream: ('Connection broken: IncompleteRead(83768748 bytes read, 12708455064 more expected)', IncompleteRead(83768748 bytes read, 12708455064 more expected))

Full message:

```
An error occurred while reading from response stream: ('Connection broken: IncompleteRead(83768748 bytes read, 12708455064 more expected)', IncompleteRead(83768748 bytes read, 12708455064 more expected))
```

Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	85.7	30%	4
Provider Mapping	100.0	15%	4
Code Coverage	94.6	15%	1
Pricing Sanity	77.3	40%	7

Schema Integrity — Findings

Score: 85.7

ERROR `file_freshness`

File is 284 days old (last_updated_on exceeds the 90-day threshold)

WARNING `expired_prices`

58690 negotiated_prices have past expiration dates (8.7%)

ERROR `raw_json_error`

Native JSON parse error detected in unpatched source file: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(83768748 bytes read, 12708455064 more expected)', IncompleteRead(83768748 bytes read, 12708455064 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 28 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(13114834 bytes read, 12779108978 more expected)', IncompleteRead(13114834 bytes read, 12779108978 more expected))

Provider Mapping — Findings

Score: 100.0

WARNING `npi_validity`

0.00% of NPIs failed Luhn checksum validation (4 of 112070)

- 1336588484
- 1659650617

WARNING `ein_validity`

0.30% of EINs failed IRS prefix validation (191 of 63319)

- 000000001
- 001625898
- 003360711
- 003680161
- 004829723
- *… and 5 more*

WARNING `empty_npi_groups`

48 provider groups contain no NPIs

INFO `duplicate_npis`

36791 NPIs appear in more than one provider group

Code Coverage — Findings

Score: 94.6

WARNING `billing_code_format`

3 CPT codes do not match expected format

Pricing Sanity — Findings

Score: 77.3

INFO `per_diem_rates`

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

INFO `percentage_rates`

2085 percentage rates (0.3%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

WARNING `zero_rates`

261 zero-dollar rates (0.04%) — CMS schema requires negotiated_rate > 0 (exclusiveMinimum); may represent unfiled or excluded services

WARNING `extreme_rates`

0.96% of rates are extreme (6453 above class-specific high threshold, 0 below $0.01)

WARNING `high_frequency_rate_value`

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

- {'rate': 95.27, 'count': 19700, 'pct': 2.92}
- {'rate': 1649.52, 'count': 19700, 'pct': 2.92}
- {'rate': 1298.48, 'count': 19700, 'pct': 2.92}
- {'rate': 1105.34, 'count': 19700, 'pct': 2.92}
- {'rate': 504.6, 'count': 19700, 'pct': 2.92}
- *… and 5 more*

WARNING `rate_spread_by_class`

billing_class='institutional' / negotiated_type='negotiated': P95/P50 spread is 371.4x (threshold: 10x, N=192, moderate confidence)

WARNING `per_code_rate_spread`

3 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
C9145	HCPCS	fee schedule	professional	ffs	—	$1.71	$2.33	$4431.41	$2250000.00	**1315789.5×**	1016
82656	CPT	fee schedule	professional	ffs	—	$8.88	$12.68	$3738.66	$2250000.00	**253378.4×**	1208
92352	CPT	fee schedule	professional	ffs	—	$16.69	$43.46	$24393.76	$2550000.00	**152786.1×**	45256

Recommended Actions

1. **schema** `file_freshness` **P1**

File is 284 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(83768748 bytes read, 12708455064 more expected)', IncompleteRead(83768748 bytes read, 12708455064 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 28 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(13114834 bytes read, 12779108978 more expected)', IncompleteRead(13114834 bytes read, 12779108978 more expected))

4. **provider_mapping** `npi_validity` **P2**

0.00% of NPIs failed Luhn checksum validation (4 of 112070)

5. **provider_mapping** `ein_validity` **P2**

0.30% of EINs failed IRS prefix validation (191 of 63319)

6. **provider_mapping** `empty_npi_groups` **P2**

48 provider groups contain no NPIs

7. **pricing** `zero_rates` **P2**

261 zero-dollar rates (0.04%) — CMS schema requires negotiated_rate > 0 (exclusiveMinimum); may represent unfiled or excluded services

8. **pricing** `extreme_rates` **P2**

0.96% of rates are extreme (6453 above class-specific high threshold, 0 below $0.01)

9. **pricing** `high_frequency_rate_value` **P2**

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

10. **pricing** `rate_spread_by_class` **P2**

billing_class='institutional' / negotiated_type='negotiated': P95/P50 spread is 371.4x (threshold: 10x, N=192, moderate confidence)

11. **pricing** `per_code_rate_spread` **P2**

3 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**

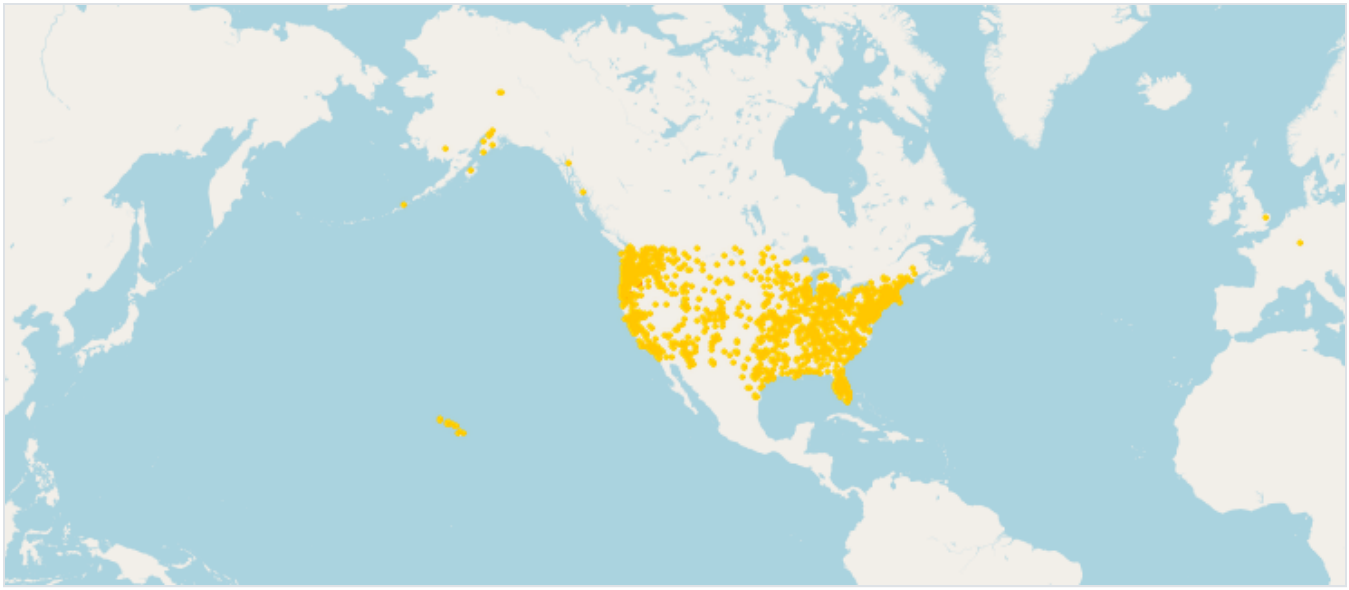
58690 negotiated_prices have past expiration dates (8.7%)

13. **code_coverage** `billing_code_format` **P3**

3 CPT codes do not match expected format

Provider Geographic Coverage

37788 unique NPIs found — 37623 geocoded (100%) — 3804 zip codes represented.



Schema Integrity — Metrics

header_missing_fields		
header_conditional_issues		
file_age_days	284	
items_total	28	
items_missing_required_pct	0.0	
items_empty_rates	0	
prices_total	675907	
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	28
billing_code_types	CPT	13
	HCPCS	10
	CDT	2
	MS-DRG	1
	APC	2

expired_prices	58690
invalid_expiration_format	0

Provider Mapping — Metrics

provider_references_in_file	50828
provider_group_ids_referenced	48105
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	112070
invalid_npi_count	4
npi_validity_rate_pct	100.0
invalid_npi_examples	1336588484, 1336588484, 1659650617, 1659650617
eins_validated	63319
invalid_ein_count	191
ein_validity_rate_pct	99.7
invalid_ein_examples	000000001, 000000001, 000000001, 001625898, 001625898, 001625898, 003360711, 003360711, 003360711, 003680161
empty_npi_groups	48
groups_without_tin	0
npis_in_multiple_groups	36791

Code Coverage — Metrics

unique_codes_total	28
duplicate_codes	0
duplicate_pct	0.0
by_code_type	CPT 13 HCPCS 10 CDT 2 MS-DRG 1 APC 2

unknown_code_types		
format_invalid_by_type	CPT	3
codes_not_in_reference	reference_not_loaded	

most_frequent_codes

Type	Code	Occurrences
CPT	31246	1
CPT	92352	1
HCPCS	M1159	1
CDT	D6722	1
HCPCS	G2014	1
CPT	35351	1
HCPCS	G4035	1
HCPCS	Q4352	1
CPT	1AHP5	1
HCPCS	G9065	1
MS-DRG	031	1
CPT	3EC21	1
CPT	29807	1
APC	9418	1
CPT	82656	1
CPT	01926	1
CPT	93454	1
CPT	15014	1
HCPCS	A0032	1
HCPCS	M1399	1

Pricing Sanity — Metrics

total_prices_checked	675907
total_rates	673809
per_diem_rates	13
percentage_rates	2085
negative_rates	0
zero_rates	261
extreme_high_rates	6453
extreme_low_rates	0

rate_distribution

sample_n	673809
sample_k	5000
confidence	high
p5	49.42
p25	199.47
p50	1585.41
p75	2606.52
p95	3677.56
p99	2250000.0

by_billing_class

Class / Type	Count	Median	p25	p75	p95	Confidence
`professional/ fee schedule`	655,006	1298.5	187.8	2503.3	3463.5	high
`institutional/ negotiated`	192	310.7	97.1	660.2	115400.1	moderate
`institutional/ fee schedule`	18,221	2326.3	190.5	3216.2	3498.5	high
`professional/ negotiated`	390	0.0	0.0	660.0	915.8	moderate

negotiated_types

fee schedule	673227
negotiated	582

unique_rate_contexts 87

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%)	`resolution_rate% × 0.60`
npi_validity (30%)	`(100 − invalid_npi_pct × 5) × 0.30`
ein_validity (10%)	`10 − (invalid_ein_pct × 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	`× 0.5 (cap 10)`
duplicate_code_pct	`× 2 (cap 20)`
codes_not_in_reference_pct	`× 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	`× 5 (cap 20)`
zero_rate_pct	`× 3 (cap 15) — warning (not error)`
extreme_rate_pct	`× 5 (cap 25) — ffs only`
class_spread_excess	`(spread − threshold) × 2, max across (billing_class, negotiated_type) buckets (cap 15)`
per_code_high_spread_count	`× 0.1 (cap 15)`
invalid_negotiated_type_pct	`× 3 (cap 10) — rates silently dropped`
invalid_setting_pct	`× 1 (cap 5) — silently defaults to wildcard`
invalid_severity_pct	`× 1 (cap 5) — silently normalised to ''`
institutional_with_service_codes_pct	`× 1 (cap 5) — extra key variation`
invalid_service_code_pct	`× 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 `reporting_entity_name` and `plan_id` are present. Key input: `plan|<entity>|<plan_id_type>|<plan_id>`. Stable across monthly re-exports.
Tier 2 — URL hash	Fallback when `plan_id` 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 `plan_id` 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 = `d62635c18a883335` · 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 ∝ 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.