MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2025-10/2025-07-07_Providence_Health_Plan_e6c380b99344bf4ce2b0761d0d338f79_in-network-rates.json.gz

Size: 11.99 GB • MD5: 1ab73231922e2ce5e7a36d9123d6e39f

Payer: **Providence_Health_Plan** • File Date: **2025-07-07** • Generated: **2026-04-26 01:36 EDT** • Tool Version: **1.0.0** • Elapsed: **9455.60s**

65.0

Limited Reliability

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

Errors: 4 • Warnings: 11 • 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/2025-10/2025-07-07_Providence_Health_Plan_e6c380b99344bf4ce2b0761d0d338f79_in-network-rates.json.gz

Plan Name	Plan ID	Issuer / Sponsor	Market
HSA_CONNECT	930586809 (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

Dimension	Score	Weight	Findings
Provider Mapping	100.0	15%	4
Code Coverage	92.0	15%	1
Pricing Sanity	59.6	40%	9

Schema Integrity — Findings

Score: 70.0

ERROR `file_freshness`

File is 293 days old (last_updated_on exceeds the 90-day threshold)

WARNING `expired_prices`

15643775 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.

ERROR `parse_crash`

single pass crash after 5,522 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(2748961666 bytes read, 10127545171 more expected)', IncompleteRead(2748961666 bytes read, 10127545171 more expected))

Provider Mapping — Findings

Score: 100.0

WARNING `npi_validity`

0.00% of NPIs failed Luhn checksum validation (6 of 120975)

- 1336588484
- 1659650617

WARNING `ein_validity`

0.30% of EINs failed IRS prefix validation (191 of 63759)

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

WARNING `empty_npi_groups`

52 provider groups contain no NPIs

INFO `duplicate_npis`

37071 NPIs appear in more than one provider group

Code Coverage — Findings

Score: 92.0

WARNING `billing_code_format`

861 CPT codes do not match expected format

Pricing Sanity — Findings

Score: 59.6

INFO `per_diem_rates`

4851 per-diem rates (0.0%) — not dollar amounts; excluded from spread analysis

INFO `percentage_rates`

285142 percentage rates (0.2%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

ERROR `zero_rates`

14492 zero-dollar rates (0.01%) — CMS schema requires negotiated_rate > 0 (exclusiveMinimum)

WARNING `extreme_rates`

1.83% of rates are extreme (3440762 above class-specific high threshold, 0 below $0.01)

WARNING `high_frequency_rate_value`

1 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': 2250000.0, 'count': 2343576, 'pct': 1.25}

WARNING `rate_spread_by_class`

billing_class='professional' / negotiated_type='fee schedule': P95/P50 spread is 7.2x (threshold: 5x, N=181,230,564 (1,000 sampled), high confidence)

WARNING `rate_spread_by_class`

billing_class='professional' / negotiated_type='negotiated': P95/P50 spread is 3922.6x (threshold: 5x, N=48,219 (1,000 sampled), high confidence)

WARNING `rate_spread_by_class`

billing_class='institutional' / negotiated_type='negotiated': P95/P50 spread is 1877.5x (threshold: 10x, N=17,375 (1,000 sampled), high confidence)

WARNING `per_code_rate_spread`

1814 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
90676	CPT	fee schedule	professional	ffs	—	$0.01	$366.32	$268295.73	$2400000.00	**240000000.0×**	89
90287	CPT	fee schedule	professional	ffs	—	$0.01	$2400000.00	$1834615.39	$2400000.00	**240000000.0×**	13
J2004	HCPCS	negotiated	institutional	ffs	—	$0.01	$1200000.00	$752250.01	$1286250.00	**128625000.0×**	5
J2003	HCPCS	negotiated	institutional	ffs	—	$0.01	$1200000.00	$752250.01	$1286250.00	**128625000.0×**	5
J1642	HCPCS	negotiated	institutional	ffs	—	$0.02	$1200000.00	$752250.01	$1286250.00	**64312500.0×**	5
J1836	HCPCS	negotiated	institutional	ffs	—	$0.03	$1200000.00	$752250.02	$1286250.00	**42875000.0×**	5
J1443	HCPCS	negotiated	institutional	ffs	—	$0.03	$1200000.00	$752250.02	$1286250.00	**42875000.0×**	5
J0641	HCPCS	negotiated	institutional	ffs	—	$0.04	$1200000.00	$752250.02	$1286250.00	**32156250.0×**	5
J0877	HCPCS	negotiated	institutional	ffs	—	$0.04	$1200000.00	$752250.02	$1286250.00	**32156250.0×**	5
J7614	HCPCS	fee schedule	professional	ffs	—	$0.08	$0.11	$7077.03	$2400000.00	**30000000.0×**	975

Recommended Actions

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

File is 293 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. **schema** `parse_crash` **P1**

single pass crash after 5,522 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(2748961666 bytes read, 10127545171 more expected)', IncompleteRead(2748961666 bytes read, 10127545171 more expected))

4. **pricing** `zero_rates` **P1**

14492 zero-dollar rates (0.01%) — CMS schema requires negotiated_rate > 0 (exclusiveMinimum)

5. provider_mapping npi_validity P2

0.00% of NPIs failed Luhn checksum validation (6 of 120975)

6. provider_mapping ein_validity P2

0.30% of EINs failed IRS prefix validation (191 of 63759)

7. provider_mapping empty_npi_groups P2

52 provider groups contain no NPIs

8. pricing extreme_rates P2

1.83% of rates are extreme (3440762 above class-specific high threshold, 0 below $0.01)

9. pricing high_frequency_rate_value P2

1 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='professional' / negotiated_type='fee schedule': P95/P50 spread is 7.2x (threshold: 5x, N=181,230,564 (1,000 sampled), high confidence)

11. pricing rate_spread_by_class P2

billing_class='professional' / negotiated_type='negotiated': P95/P50 spread is 3922.6x (threshold: 5x, N=48,219 (1,000 sampled), high confidence)

12. pricing rate_spread_by_class P2

billing_class='institutional' / negotiated_type='negotiated': P95/P50 spread is 1877.5x (threshold: 10x, N=17,375 (1,000 sampled), high confidence)

13. **pricing** `per_code_rate_spread` **P2**

1814 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.

14. **schema** `expired_prices` **P2**

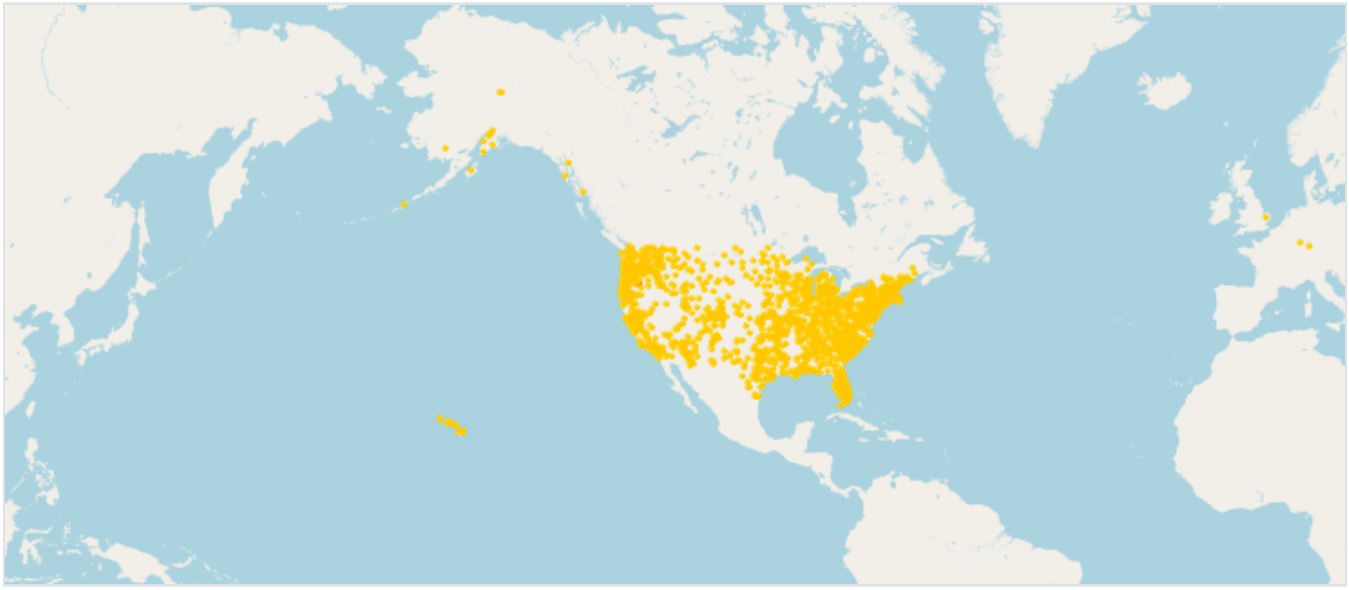
15643775 negotiated_prices have past expiration dates (8.3%)

15. **code_coverage** `billing_code_format` **P3**

861 CPT codes do not match expected format

Provider Geographic Coverage

41420 unique NPIs found — 41209 geocoded (99%) — 4874 zip codes represented.



Schema Integrity — Metrics

header_missing_fields	
header_conditional_issues	
file_age_days	293

items_total	5522
items_missing_required_pct	0.0
items_empty_rates	0
prices_total	187928967
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: 5496 capitation: 26
billing_code_types	CPT: 3273 HCPCS: 1581 MS-DRG: 174 CDT: 211 APC: 177 RC: 106
expired_prices	15643775
invalid_expiration_format	0

Provider Mapping — Metrics

provider_references_in_file	50980
provider_group_ids_referenced	48511
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	120975
invalid_npi_count	6
npi_validity_rate_pct	100.0
invalid_npi_examples	1336588484, 1336588484, 1336588484, 1659650617, 1659650617, 1659650617
eins_validated	63759

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	52
groups_without_tin	0
npis_in_multiple_groups	37071

Code Coverage — Metrics

unique_codes_total	5516
duplicate_codes	6
duplicate_pct	0.11
by_code_type	**CPT** 3269 **HCPCS** 1579 **MS-DRG** 174 **CDT** 211 **APC** 177 **RC** 106
unknown_code_types	
format_invalid_by_type	**CPT** 861
codes_not_in_reference	reference_not_loaded

most_frequent_codes	

Type	Code	Occurrences
CPT	99347	2
CPT	36415	2
CPT	17003	2
HCPCS	G0246	2
CPT	99215	2
HCPCS	G0180	2
CPT	57109	1
CPT	0391U	1
CPT	81504	1
CPT	3KB21	1
CPT	1EC31	1
CPT	26449	1
HCPCS	K0503	1
CPT	0541T	1
CPT	97113	1
CPT	15780	1
CPT	6100F	1
CPT	99342	1
HCPCS	M1244	1
HCPCS	Q4002	1

Pricing Sanity — Metrics

total_prices_checked	187928967
total_rates	187638974
per_diem_rates	4851
percentage_rates	285142
negative_rates	0
zero_rates	14492
extreme_high_rates	3440762
extreme_low_rates	0

rate_distribution

sample_n	187638974
sample_k	5000
confidence	high
p5	29.5195
p25	211.0675
p50	768.495
p75	1895.72
p95	5273.308500000001
p99	2250000.0

by_billing_class

Class / Type	Count	Median	p25	p75	p95	Confidence
professional/ fee schedule	181,230,564	704.2	203.8	1786.0	5048.0	high
institutional/ fee schedule	6,342,816	550.9	122.9	1870.4	4172.1	high
professional/ negotiated	48,219	573.6	290.0	1243.2	2250000.0	high
institutional/ negotiated	17,375	876.0	4.5	43184.2	1644677.5	high

negotiated_types

fee schedule	187573380
negotiated	65594

unique_rate_contexts 21301

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.

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).

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)`
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 = `05c0e1418c988fd4` · 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.