

MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2026-03/FP_in-network-rates_4.json

Size: 1015.37 MB • MD5: 1a4e7300c46d15f984c1497d5b93b526

Payer: **FrontPath** • File Date: **2026-02-25** • Generated: **2026-04-24 00:34 EDT** • Tool Version: **1.0.0** • Elapsed: **273.40s**

65.0

Limited Reliability

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

Errors: 2 • Warnings: 8 • Info: 2

CMS Official Schema Validation

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

Validator output:

```
info Running validator container...
info
```

Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	70.0	30%	2
Provider Mapping	99.1	15%	3
Code Coverage	70.0	15%	2
Pricing Sanity	69.3	40%	5

Schema Integrity — Findings

Score: 70.0

WARNING `file_freshness`

File is 58 days old (last_updated_on exceeds the 45-day threshold)

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: 99.1

WARNING `npi_validity`

0.62% of NPIs failed Luhn checksum validation (84 of 13608)

- 0
- 1003919619
- 1010783520
- 1033397618
- 1053778703
- ... and 5 more

WARNING `ein_validity`

0.10% of EINs failed IRS prefix validation (1 of 968)

- 28-2663275

INFO `duplicate_npis`

1961 NPIs appear in more than one provider group

Code Coverage — Findings

Score: 70.0

WARNING `billing_code_format`

76947 CPT codes do not match expected format

WARNING `duplicate_billing_codes`

17476 billing codes appear in more than one `in_network` item (96.1%)

Pricing Sanity — Findings

Score: 69.3

INFO `percentage_rates`

198 percentage rates (0.0%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

ERROR `zero_rates`

1870 zero-dollar rates (0.22%) — CMS schema requires `negotiated_rate > 0` (`exclusiveMinimum`)

WARNING `rate_spread_by_class`

`billing_class='institutional' / negotiated_type='negotiated'`: P95/P50 spread is 14.4x (threshold: 10x, N=66,725 (1,000 sampled), high confidence)

WARNING `rate_spread_by_class`

`billing_class='professional' / negotiated_type='fee schedule'`: P95/P50 spread is 14.7x (threshold: 5x, N=770,714 (1,000 sampled), high confidence)

WARNING `per_code_rate_spread`

183583 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
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4
J0461	CPT	fee schedule	professional	ffs	both	\$0.02	\$8883.19	\$9666.98	\$20901.50	1045075.0x	4

Recommended Actions

1. `schema` `cms_schema_validation`

P1

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

2. **pricing** zero_rates

P1

1870 zero-dollar rates (0.22%) — CMS schema requires negotiated_rate > 0 (exclusiveMinimum)

3. **provider_mapping** npi_validity

P2

0.62% of NPIs failed Luhn checksum validation (84 of 13608)

4. **provider_mapping** ein_validity

P2

0.10% of EINs failed IRS prefix validation (1 of 968)

5. **pricing** rate_spread_by_class

P2

billing_class='institutional' / negotiated_type='negotiated': P95/P50 spread is 14.4x (threshold: 10x, N=66,725 (1,000 sampled), high confidence)

6. **pricing** rate_spread_by_class

P2

billing_class='professional' / negotiated_type='fee schedule': P95/P50 spread is 14.7x (threshold: 5x, N=770,714 (1,000 sampled), high confidence)

7. **pricing** per_code_rate_spread

P2

183583 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.

8. **schema** file_freshness

P2

File is 58 days old (last_updated_on exceeds the 45-day threshold)

9. **code_coverage** billing_code_format

P3

76947 CPT codes do not match expected format

17476 billing codes appear in more than one in_network item (96.1%)

Provider Geographic Coverage

11036 unique NPIs found — 10763 geocoded (98%) — 1513 zip codes represented.



Schema Integrity — Metrics

header_missing_fields	
header_conditional_issues	
file_age_days	58
items_total	802827
items_missing_required_pct	0.0
items_empty_rates	0
prices_total	837645
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	802827
billing_code_types	CPT	796617
	MS-DRG	6123
	RC	87
expired_prices	0	
invalid_expiration_format	0	

Provider Mapping — Metrics

provider_references_in_file	136
provider_group_ids_referenced	136
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	13608
invalid_npi_count	84
npi_validity_rate_pct	99.38
invalid_npi_examples	0, 0, 0, 0, 0, 0, 1003919619, 1010783520, 1033397618, 1053778703
eins_validated	968
invalid_ein_count	1
ein_validity_rate_pct	99.9
invalid_ein_examples	28-2663275
empty_npi_groups	0
groups_without_tin	0
npis_in_multiple_groups	1961

Code Coverage — Metrics

unique_codes_total	18180
duplicate_codes	17476
duplicate_pct	96.13

by_code_type	CPT	17387	
	MS-DRG	768	
	RC	25	
unknown_code_types			
format_invalid_by_type	CPT	76947	
codes_not_in_reference	reference_not_loaded		
most_frequent_codes	Type	Code	Occurrences
	CPT	85576	213
	CPT	95819	176
	CPT	95816	176
	CPT	95813	176
	CPT	73610	175
	CPT	73080	175
	CPT	73630	175
	CPT	72110	175
	CPT	72100	175
	CPT	72020	175
	CPT	73110	175
	CPT	72120	175
	CPT	72040	175
	CPT	73120	175
	CPT	73562	175
	CPT	72170	175
	CPT	72070	175
	CPT	73590	175
	CPT	76942	174
CPT	77003	174	

Pricing Sanity — Metrics

total_prices_checked	837645
total_rates	837447
per_diem_rates	0
percentage_rates	198
negative_rates	0
zero_rates	1870

extreme_high_rates	77																												
extreme_low_rates	0																												
rate_distribution	<table> <tr> <td>sample_n</td> <td>837447</td> </tr> <tr> <td>sample_k</td> <td>5000</td> </tr> <tr> <td>confidence</td> <td>high</td> </tr> <tr> <td>p5</td> <td>8.2195</td> </tr> <tr> <td>p25</td> <td>41.155</td> </tr> <tr> <td>p50</td> <td>237.08499999999998</td> </tr> <tr> <td>p75</td> <td>946.0749999999999</td> </tr> <tr> <td>p95</td> <td>3176.63900000000006</td> </tr> <tr> <td>p99</td> <td>9840.668700000113</td> </tr> </table>	sample_n	837447	sample_k	5000	confidence	high	p5	8.2195	p25	41.155	p50	237.08499999999998	p75	946.0749999999999	p95	3176.63900000000006	p99	9840.668700000113										
sample_n	837447																												
sample_k	5000																												
confidence	high																												
p5	8.2195																												
p25	41.155																												
p50	237.08499999999998																												
p75	946.0749999999999																												
p95	3176.63900000000006																												
p99	9840.668700000113																												
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>66,725</td> <td>1553.0</td> <td>1061.0</td> <td>3613.8</td> <td>22342.5</td> <td>high</td> </tr> <tr> <td>professional/ fee schedule</td> <td>770,714</td> <td>169.4</td> <td>34.2</td> <td>700.3</td> <td>2483.4</td> <td>high</td> </tr> <tr> <td>professional/ negotiated</td> <td>8</td> <td>90.0</td> <td>43.0</td> <td>90.0</td> <td>96.5</td> <td>low</td> </tr> </tbody> </table>	Class / Type	Count	Median	p25	p75	p95	Confidence	institutional/ negotiated	66,725	1553.0	1061.0	3613.8	22342.5	high	professional/ fee schedule	770,714	169.4	34.2	700.3	2483.4	high	professional/ negotiated	8	90.0	43.0	90.0	96.5	low
Class / Type	Count	Median	p25	p75	p95	Confidence																							
institutional/ negotiated	66,725	1553.0	1061.0	3613.8	22342.5	high																							
professional/ fee schedule	770,714	169.4	34.2	700.3	2483.4	high																							
professional/ negotiated	8	90.0	43.0	90.0	96.5	low																							
negotiated_types	<table> <tr> <td>negotiated</td> <td>66733</td> </tr> <tr> <td>fee schedule</td> <td>770714</td> </tr> </table>	negotiated	66733	fee schedule	770714																								
negotiated	66733																												
fee schedule	770714																												
unique_rate_contexts	1786981																												
rate_key_dimension_validity	<table> <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).

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/HCCPS/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)
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 = 90e1c74d0defdd47 · entity = FrontPath · 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.