

# MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11-24\_encore-health\_encore-combined\_in-network-rates.json

Size: 496.70 MB • MD5: e28633bd018e002f3dbe3ca162303370

Payer: **The Healthcare Group dba Encore Health Network** • File Date: **2022-08-31** • Generated: **2026-04-24 22:52 EDT**

• Tool Version: **1.0.0** • Elapsed: **168.50s**

# 59.0

Not Usable

score capped at 59.0 — 1 native JSON parse error(s) require payer re-export; CMS official schema validation failed (see meta for details)

Errors: 6 • Warnings: 9 • Info: 1 • **1 Raw JSON Error(s)**

## TOC Plan References

TOC: s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11\_plan\_ref\_5ccddb439b4bd7a45137b59cffdfb470\_index.json • Talon

Canonical: s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11-24\_encore-health\_encore-combined\_in-network-rates.json

Plan Name	Plan ID	Issuer / Sponsor	Market
Encore Combined	Encore Health (custom)	—	—

## Raw JSON Parse Errors

### IncompleteJSONError

parse error: premature EOF

Context:

```
(right here) -----^
```

Full message:

```
parse error: premature EOF
```

```
(right here) -----^
```

## CMS Official Schema Validation

---

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

Validator output:

```
error Data file does not contain version information. Please run again using the --schema-version option to specify a version.
```

## Dimension Scores

---

Dimension	Score	Weight	Findings
Schema Integrity	70.0	30%	7
Provider Mapping	99.9	15%	3
Code Coverage	90.0	15%	2
Pricing Sanity	70.3	40%	4

## Schema Integrity — Findings

---

Score: 70.0

**ERROR** `required_header_fields`

Missing required header fields: ['version']

---

**WARNING** `header_conditional`

plan\_sponsor\_name is required when plan\_id\_type == 'ein'

---

**ERROR** `file_freshness`

File is 1332 days old (last\_updated\_on exceeds the 90-day threshold)

---

**WARNING** `empty_negotiated_rates`

2 in\_network items have empty negotiated\_rates arrays (0.0%) — dead records

---

**ERROR** `cms_schema_validation`

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

---

**ERROR** `raw_json_error`

Native JSON parse error detected in unpatched source file: parse error: premature EOF. Scoring below reflects auto-patched data only. File must be re-exported by the payer to fully resolve.

(right here) -----^

---

**ERROR** `parse_crash`

single pass crash after 24,247 items: parse error: premature EOF (right here) -----^

---

## Provider Mapping — Findings

---

Score: 99.9

**WARNING** `npi_validity`

0.06% of NPIs failed Luhn checksum validation (30 of 51233)

- 10027007
- 364797599
- 1033190625
- 1053561252
- 1063131880
- ... and 5 more

---

**WARNING** `ein_validity`

0.38% of EINs failed IRS prefix validation (19 of 4995)

- 009466371
- 076587846
- 085425447
- 175426910
- 181347811
- ... and 5 more

---

**INFO** `duplicate_npis`

7622 NPIs appear in more than one provider group

---

# Code Coverage — Findings

---

Score: 90.0

**WARNING** `billing_code_format`

7874 CPT codes do not match expected format

---

**WARNING** `billing_code_format`

5463 HCPCS codes do not match expected format

---

## Pricing Sanity — Findings

Score: 70.3

### ERROR zero\_rates

15431 zero-dollar rates (1.08%) — CMS schema requires negotiated\_rate > 0 (exclusiveMinimum)

### WARNING high\_frequency\_rate\_value

2 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': 80.0, 'count': 11623, 'pct': 0.81}
- {'rate': 85.0, 'count': 8161, 'pct': 0.57}

### WARNING rate\_spread\_by\_class

billing\_class='professional' / negotiated\_type='negotiated': P95/P50 spread is 9.8x (threshold: 5x, N=1,415,729 (1,000 sampled), high confidence)

### WARNING per\_code\_rate\_spread

2040 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
J1411	CPT	negotiated	professional	ffs	—	\$60.00	\$3570000.00	\$3109362.58	\$3570000.00	59500.0x	31
J3399	CPT	negotiated	professional	ffs	—	\$60.00	\$2299500.25	\$2007945.21	\$2299500.25	38325.0x	32
J7680	CPT	negotiated	professional	ffs	—	\$0.01	\$85.00	\$75.77	\$100.00	10000.0x	13
J7681	CPT	negotiated	professional	ffs	—	\$0.01	\$82.50	\$73.75	\$100.00	10000.0x	12
20697	CPT	negotiated	professional	ffs	—	\$0.28	\$0.32	\$807.40	\$2421.59	8648.5x	3
20697	CPT	negotiated	professional	ffs	—	\$0.28	\$0.32	\$807.40	\$2421.59	8648.5x	3
20697	CPT	negotiated	professional	ffs	—	\$0.28	\$0.32	\$807.40	\$2421.59	8648.5x	3
20697	CPT	negotiated	professional	ffs	—	\$0.28	\$0.32	\$807.40	\$2421.59	8648.5x	3
Q0171	CPT	negotiated	professional	ffs	—	\$0.01	\$0.91	\$7.39	\$85.00	8500.0x	36
J0665	CPT	negotiated	professional	ffs	—	\$0.01	\$0.02	\$4.31	\$60.00	6000.0x	56

## Recommended Actions

### 1. overall

P0

Score below minimum threshold — do not use for pricing without remediation.

2. **schema** required\_header\_fields

P1

Missing required header fields: [version]

3. **schema** file\_freshness

P1

File is 1332 days old (last\_updated\_on exceeds the 90-day threshold)

4. **schema** cms\_schema\_validation

P1

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

5. **schema** raw\_json\_error

P1

Native JSON parse error detected in unpatched source file: parse error: premature EOF. Scoring below reflects auto-patched data only. File must be re-exported by the payer to fully resolve.

6. **schema** parse\_crash

P1

single pass crash after 24,247 items: parse error: premature EOF (right here) -----^

7. **pricing** zero\_rates

P1

15431 zero-dollar rates (1.08%) — CMS schema requires negotiated\_rate > 0 (exclusiveMinimum)

8. **provider\_mapping** npi\_validity

P2

0.06% of NPIs failed Luhn checksum validation (30 of 51233)

9. **provider\_mapping** ein\_validity

P2

0.38% of EINs failed IRS prefix validation (19 of 4995)

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

P2

2 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.

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

P2

billing\_class='professional' / negotiated\_type='negotiated': P95/P50 spread is 9.8x (threshold: 5x, N=1,415,729 (1,000 sampled), high confidence)

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

P2

2040 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.

13. **schema** header\_conditional

P2

plan\_sponsor\_name is required when plan\_id\_type == 'ein'

14. **schema** empty\_negotiated\_rates

P2

2 in\_network items have empty negotiated\_rates arrays (0.0%) — dead records

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

P3

7874 CPT codes do not match expected format

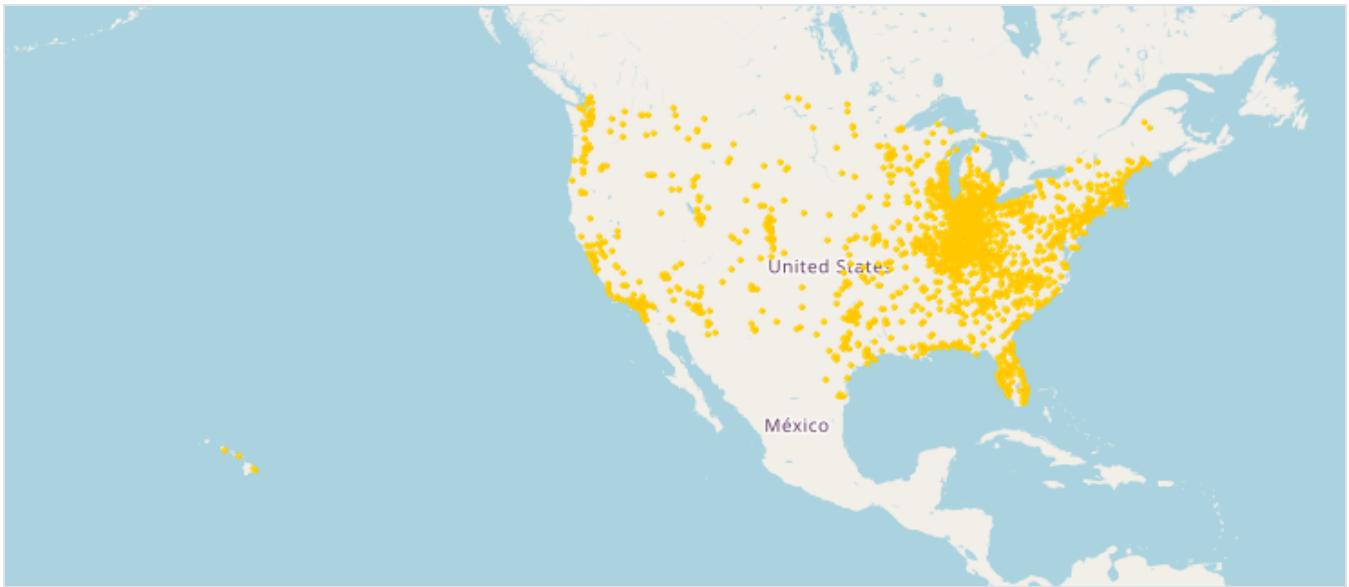
16. **code\_coverage** billing\_code\_format

P3

5463 HCPCS codes do not match expected format

## Provider Geographic Coverage

40154 unique NPIs found — 39602 geocoded (99%) — 3269 zip codes represented.



## Schema Integrity — Metrics

header_missing_fields	version	
header_conditional_issues	plan_sponsor_name is required when plan_id_type == 'ein'	
file_age_days	1332	
items_total	24247	
items_missing_required_pct	0.0	
items_empty_rates	2	
prices_total	1428514	
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	24247
billing_code_types	CPT	18784
	HCPCS	5463
expired_prices	0	
invalid_expiration_format	0	

## Provider Mapping — Metrics

---

provider_references_in_file	5714
provider_group_ids_referenced	4995
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	51233
invalid_npi_count	30
npi_validity_rate_pct	99.94
invalid_npi_examples	10027007, 364797599, 1033190625, 1033190625, 1033190625, 1033190625, 1053561252, 1063131880, 1124163314, 1134378568
eins_validated	4995
invalid_ein_count	19
ein_validity_rate_pct	99.62
invalid_ein_examples	009466371, 076587846, 085425447, 175426910, 181347811, 184366929, 189663275, 196446983, 284500476, 287447787
empty_npi_groups	0
groups_without_tin	0
npis_in_multiple_groups	7622

## Code Coverage — Metrics

---

unique_codes_total	24247	
duplicate_codes	0	
duplicate_pct	0.0	
by_code_type	CPT	18784
	HCPCS	5463
unknown_code_types		
format_invalid_by_type	CPT	7874
	HCPCS	5463
codes_not_in_reference	reference_not_loaded	

most_frequent_codes	Type	Code	Occurrences
	CPT	0001A	1
	CPT	0001M	1
	CPT	0001U	1
	CPT	0002A	1
	CPT	0002M	1
	CPT	0002U	1
	CPT	0003A	1
	CPT	0003M	1
	CPT	0003T	1
	CPT	0003U	1
	CPT	0004A	1
	CPT	0004M	1
	CPT	0005U	1
	CPT	0006M	1
	CPT	0006U	1
	CPT	0007M	1
	CPT	0007U	1
	CPT	0008M	1
	CPT	0008T	1
	CPT	0008U	1

## Pricing Sanity — Metrics

total_prices_checked	1428514
total_rates	1428514
per_diem_rates	0
percentage_rates	0
negative_rates	0
zero_rates	15431
extreme_high_rates	1356
extreme_low_rates	0

<b>rate_distribution</b>	<b>sample_n</b>	1428514
	<b>sample_k</b>	5000
	<b>confidence</b>	high
	<b>p5</b>	7.7995
	<b>p25</b>	59.1775
	<b>p50</b>	242.815
	<b>p75</b>	888.6975
	<b>p95</b>	3178.78
	<b>p99</b>	8483.505500000025

<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>
	<b>professional/negotiated</b>	1,415,729	271.3	63.8	838.0	2662.9	high
	<b>institutional/negotiated</b>	12,785	10475.0	4593.0	12331.0	17264.0	high

<b>negotiated_types</b>	<b>negotiated</b>	1428514
-------------------------	-------------------	---------

<b>unique_rate_contexts</b>	146926
-----------------------------	--------

<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.

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).

<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/HCCPS/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)
<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 = 5183cd6ccda0d4e6 · entity = The Healthcare Group dba Encore Health Network · plan_id = 352067676 (EIN) · tier = 1 (entity + plan_id)`

## 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.