

MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11-24_encore-health_encore-prime_in-network-rates.json
Size: 493.81 MB • MD5: 27f1609c51b1f7636698e8cc4b54e3a1

Payer: **The Healthcare Group dba Encore Health Network** • File Date: **2022-08-31** • Generated: **2026-04-24 22:55 EDT**
• Tool Version: **1.0.0** • Elapsed: **131.60s**

65.0

Limited Reliability

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

Errors: 4 • Warnings: 9 • Info: 2

TOC Plan References

TOC: s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11_plan_ref_fd3325a3d9bd979a1d703e0d0ebb235e_index.json • Talon

Canonical: s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11-24_encore-health_encore-prime_in-network-rates.json

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

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%	5
Provider Mapping	99.9	15%	3

Dimension	Score	Weight	Findings
Code Coverage	90.0	15%	2
Pricing Sanity	69.0	40%	5

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.

Provider Mapping — Findings

Score: 99.9

WARNING `npi_validity`

0.07% of NPIs failed Luhn checksum validation (39 of 55362)

- 10027007
- 100267970
- 192971326
- 364797599
- 1033190625
- ... and 5 more

WARNING `ein_validity`

0.33% of EINs failed IRS prefix validation (18 of 5440)

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

INFO `duplicate_npis`

8149 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`

6396 HCPCS codes do not match expected format

Pricing Sanity — Findings

Score: 69.0

INFO percentage_rates

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

ERROR zero_rates

15454 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 ≥ 50 occurrences) — likely placeholder/sentinel values rather than real negotiated rates.

- {'rate': 80.0, 'count': 11624, 'pct': 0.81}
- {'rate': 85.0, 'count': 8162, 'pct': 0.57}

WARNING rate_spread_by_class

billing_class='professional' / negotiated_type='negotiated': P95/P50 spread is 10.5x (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. **schema** required_header_fields

P1

Missing required header fields: [version]

2. **schema** file_freshness

P1

File is 1332 days old (last_updated_on exceeds the 90-day threshold)

3. **schema** cms_schema_validation

P1

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

4. **pricing** zero_rates

P1

15454 zero-dollar rates (1.08%) — CMS schema requires negotiated_rate > 0 (exclusiveMinimum)

5. **provider_mapping** npi_validity

P2

0.07% of NPIs failed Luhn checksum validation (39 of 55362)

6. **provider_mapping** ein_validity

P2

0.33% of EINs failed IRS prefix validation (18 of 5440)

7. **pricing** high_frequency_rate_value

P2

2 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.

8. **pricing** rate_spread_by_class

P2

billing_class='professional' / negotiated_type='negotiated': P95/P50 spread is 10.5x (threshold: 5x, N=1,415,729 (1,000 sampled), high confidence)

9. **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.

10. **schema** header_conditional

P2

plan_sponsor_name is required when plan_id_type == 'ein'

11. **schema** empty_negotiated_rates

P2

2 in_network items have empty negotiated_rates arrays (0.0%) — dead records

12. **code_coverage** billing_code_format

P3

7874 CPT codes do not match expected format

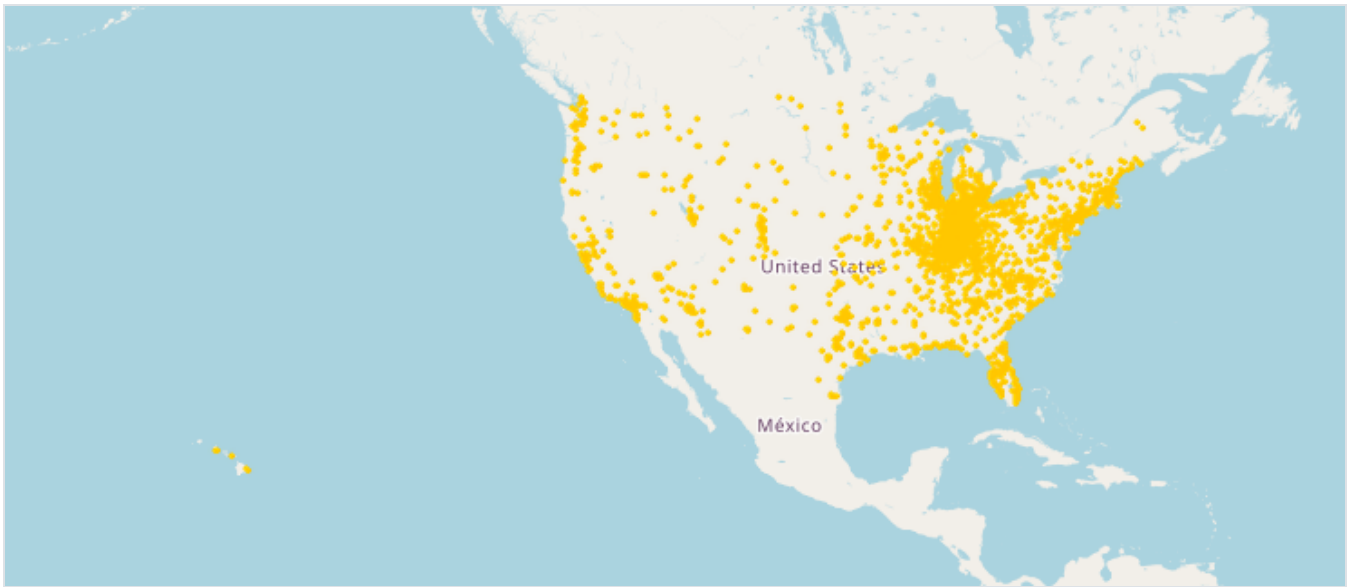
13. **code_coverage** billing_code_format

P3

6396 HCPCS codes do not match expected format

Provider Geographic Coverage

43497 unique NPIs found — 42882 geocoded (99%) — 3340 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	26304	
items_missing_required_pct	0.0	
items_empty_rates	2	
prices_total	1437466	
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	26304
billing_code_types	CPT	18784
	HCPCS	6565
	RC	164
	AP-DRG	790
	CSTM-ALL	1

expired_prices	0
invalid_expiration_format	0

Provider Mapping — Metrics

provider_references_in_file	5590
provider_group_ids_referenced	5440
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	55362
invalid_npi_count	39
npi_validity_rate_pct	99.93
invalid_npi_examples	10027007, 100267970, 192971326, 364797599, 1033190625, 1033190625, 1033190625, 1033190625, 1053561252, 1063131880
eins_validated	5440
invalid_ein_count	18
ein_validity_rate_pct	99.67
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	8149

Code Coverage — Metrics

unique_codes_total	26304										
duplicate_codes	0										
duplicate_pct	0.0										
by_code_type	<table border="1"> <tr> <td>CPT</td> <td>18784</td> </tr> <tr> <td>HCPCS</td> <td>6565</td> </tr> <tr> <td>RC</td> <td>164</td> </tr> <tr> <td>AP-DRG</td> <td>790</td> </tr> <tr> <td>CSTM-ALL</td> <td>1</td> </tr> </table>	CPT	18784	HCPCS	6565	RC	164	AP-DRG	790	CSTM-ALL	1
CPT	18784										
HCPCS	6565										
RC	164										
AP-DRG	790										
CSTM-ALL	1										

unknown_code_types			
format_invalid_by_type	CPT	7874	
	HCPCS	6396	
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	1437466
total_rates	1437423
per_diem_rates	0
percentage_rates	43
negative_rates	0
zero_rates	15454
extreme_high_rates	1356
extreme_low_rates	0

rate_distribution	sample_n	1437423
	sample_k	5000
	confidence	high
	p5	9.088000000000001
	p25	60.685
	p50	248.435
	p75	899.6625
	p95	3130.2935000000007
	p99	12422.940900000152

by_billing_class	Class / Type	Count	Median	p25	p75	p95	Confidence
	professional/ negotiated	1,415,729	260.7	66.5	873.1	2727.1	high
	institutional/ negotiated	21,694	12331.0	6580.0	18816.0	56776.6	high

negotiated_types	negotiated	1437423
-------------------------	-------------------	---------

unique_rate_contexts	148982
-----------------------------	--------

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%)	$\text{resolution_rate\%} \times 0.60$
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 = 47e15ab50b5a8d74 · 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.

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.