

# MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-12-01\_CRTA\_Health-Alliance-Plan.json.gz

Size: 50.78 MB • MD5: e43a028be82714e07a3bc50974a902a4

Payer: **Health Alliance Plan** • File Date: **2025-12-01** • Generated: **2026-04-24 23:08 EDT** • Tool Version: **1.0.0** •

Elapsed: **573.60s**

# 83.1

Usable w/ Caution

Errors: 1 • Warnings: 6 • Info: 3

## TOC Plan References

TOC: s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-11\_plan\_ref\_cff6f1a4b3fbd3810c4722edde594511\_index.json • Talon - HAP

Canonical: s3://talon-storage-private/mrf-feed-uploads/2025-11/2025-12-01\_CRTA\_Health-Alliance-Plan.json.gz

Plan Name	Plan ID	Issuer / Sponsor	Market
ELECTRICAL WORKERS INSURANCE FUND	381393235	—	group

## CMS Official Schema Validation

**PASSED** — File conforms to the CMS in-network-rates schema.

## Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	90.0	30%	1
Provider Mapping	100.0	15%	2
Code Coverage	89.2	15%	2
Pricing Sanity	69.2	40%	5

# Schema Integrity — Findings

---

Score: 90.0

**ERROR** file\_freshness

File is 144 days old (last\_updated\_on exceeds the 90-day threshold)

---

## Provider Mapping — Findings

---

Score: 100.0

**WARNING** ein\_validity

0.26% of EINs failed IRS prefix validation (65 of 24681)

- 001468376
- 071649500
- 076501987
- 077701981
- 080380433
- ... and 5 more

---

**INFO** duplicate\_npis

25194 NPIs appear in more than one provider group

---

# Code Coverage — Findings

---

Score: 89.2

**WARNING** `billing_code_format`

833 CPT codes do not match expected format

---

**WARNING** `duplicate_billing_codes`

776 billing codes appear in more than one `in_network` item (4.3%)

---

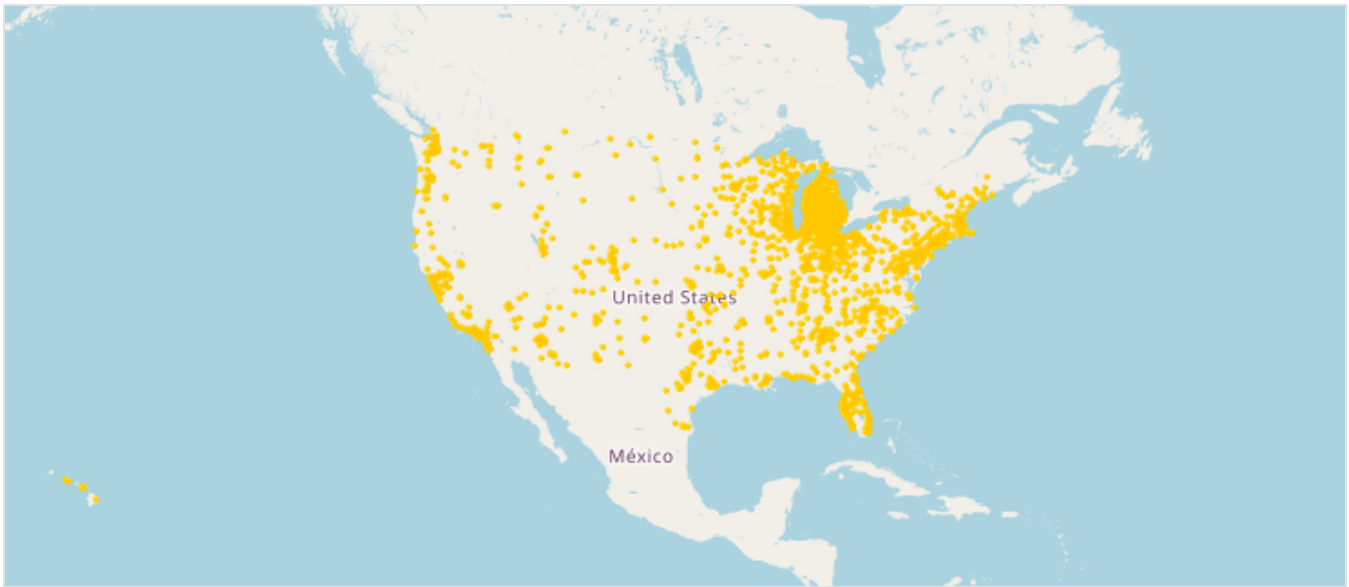


## Recommended Actions

1.	<b>schema</b>	file_freshness	P1
File is 144 days old (last_updated_on exceeds the 90-day threshold)			
2.	<b>provider_mapping</b>	ein_validity	P2
0.26% of EINs failed IRS prefix validation (65 of 24681)			
3.	<b>pricing</b>	rate_spread_by_class	P2
billing_class='professional' / negotiated_type='fee schedule': P95/P50 spread is 12.8x (threshold: 5x, N=2,372,359 (1,000 sampled), high confidence)			
4.	<b>pricing</b>	rate_spread_by_class	P2
billing_class='institutional' / negotiated_type='fee schedule': P95/P50 spread is 15.7x (threshold: 10x, N=174,413 (1,000 sampled), high confidence)			
5.	<b>pricing</b>	per_code_rate_spread	P2
80290 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.			
6.	<b>code_coverage</b>	billing_code_format	P3
833 CPT codes do not match expected format			
7.	<b>code_coverage</b>	duplicate_billing_codes	P3
776 billing codes appear in more than one in_network item (4.3%)			

## Provider Geographic Coverage

52853 unique NPIs found — 52699 geocoded (100%) — 3147 zip codes represented.



## Schema Integrity — Metrics

---

header_missing_fields		
header_conditional_issues		
file_age_days		144
items_total		18936
items_missing_required_pct		0.0
items_empty_rates		0
prices_total		2592423
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	17379
	bundle	1557

billing_code_types	CPT	10233
	HCPCS	5112
	ICD	1877
	RC	156
	CSTM-ALL	1
	MS-DRG	1557
expired_prices	0	
invalid_expiration_format	0	

## Provider Mapping — Metrics

provider_references_in_file	1109
provider_group_ids_referenced	1109
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	99972
invalid_npi_count	0
npi_validity_rate_pct	100.0
invalid_npi_examples	
eins_validated	24681
invalid_ein_count	65
ein_validity_rate_pct	99.74
invalid_ein_examples	001468376, 001468376, 071649500, 076501987, 077701981, 080380433, 080380433, 080385842, 080385842, 080660230
empty_npi_groups	0
groups_without_tin	0
npis_in_multiple_groups	25194

## Code Coverage — Metrics

unique_codes_total	18160
duplicate_codes	776
duplicate_pct	4.27

by_code_type	CPT	10233	
	HCPCS	5112	
	ICD	1877	
	RC	156	
	CSTM-ALL	1	
	MS-DRG	781	
unknown_code_types			
format_invalid_by_type	CPT	833	
codes_not_in_reference	reference_not_loaded		
most_frequent_codes	<b>Type</b>	<b>Code</b>	<b>Occurrences</b>
	MS-DRG	776	2
	MS-DRG	088	2
	MS-DRG	068	2
	MS-DRG	659	2
	MS-DRG	258	2
	MS-DRG	233	2
	MS-DRG	947	2
	MS-DRG	600	2
	MS-DRG	060	2
	MS-DRG	432	2
	MS-DRG	508	2
	MS-DRG	321	2
	MS-DRG	786	2
	MS-DRG	988	2
	MS-DRG	479	2
	MS-DRG	041	2
	MS-DRG	699	2
	MS-DRG	059	2
	MS-DRG	157	2
MS-DRG	862	2	

## Pricing Sanity — Metrics

total_prices_checked	2592423
total_rates	2588488
per_diem_rates	3686
percentage_rates	249

negative_rates	0
zero_rates	0
extreme_high_rates	3971
extreme_low_rates	0

rate_distribution	sample_n	2588488
	sample_k	5000
	confidence	high
	p5	4.918000000000001
	p25	41.205
	p50	208.02499999999998
	p75	804.1
	p95	2935.004500000002
	p99	19083.008300000016

by_billing_class	<b>Class / Type</b>	<b>Count</b>	<b>Median</b>	<b>p25</b>	<b>p75</b>	<b>p95</b>	<b>Confidence</b>
	professional/ fee schedule	2,372,359	203.3	41.5	798.9	2610.8	high
	institutional/ fee schedule	174,413	152.0	32.4	671.0	2391.2	high
	institutional/ negotiated	41,716	20739.1	12942.7	33863.4	77165.9	high

negotiated_types	fee schedule	2546772
	negotiated	41716

unique_rate_contexts	1128445
----------------------	---------

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).

<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 = 07cf69d157c63925 · entity = Health Alliance 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.

<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.