MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/2025-12/2026-01-01_CRTA_Health-Alliance-Plan.json.gz

Size: 50.71 MB • MD5: 9038ccd13d3699f50f8a65261994495f

Payer: **Health Alliance Plan** • File Date: **2026-01-01** • Generated: **2026-04-24 20:51 EDT** • Tool Version: **1.0.0** • Elapsed: **571.40s**

83.6

Usable w/ Caution

Errors: 1 • Warnings: 6 • Info: 3

TOC Plan References

TOC: s3://talon-storage-private/mrf-feed-uploads/2025-12/2025-12_plan_ref_cff6f1a4b3fbd3810c4722edde594511_index.json • Talon - HAP

Canonical: s3://talon-storage-private/mrf-feed-uploads/2025-12/2026-01-01_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	70.5	40%	5

Schema Integrity — Findings

Score: 90.0

ERROR `file_freshness`

File is 113 days old (last_updated_on exceeds the 90-day threshold)

Provider Mapping — Findings

Score: 100.0

WARNING `ein_validity`

0.25% of EINs failed IRS prefix validation (63 of 24772)

- 001468376
- 071649500
- 076501987
- 077701981
- 080380433
- *… and 5 more*

INFO `duplicate_npis`

25369 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%)

Pricing Sanity — Findings

Score: 70.5

INFO `per_diem_rates`

3680 per-diem rates (0.1%) — not dollar amounts; excluded from spread analysis

INFO `percentage_rates`

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

WARNING `rate_spread_by_class`

billing_class='professional' / negotiated_type='fee schedule': P95/P50 spread is 11.8x (threshold: 5x, N=2,378,647 (1,000 sampled), high confidence)

WARNING `rate_spread_by_class`

billing_class='institutional' / negotiated_type='fee schedule': P95/P50 spread is 14.2x (threshold: 10x, N=175,006 (1,000 sampled), high confidence)

WARNING `per_code_rate_spread`

57363 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
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83
Q2042	HCPCS	fee schedule	professional	ffs	—	$2.06	$621312.51	$680022.23	$2071041.70	**1005360.0×**	83

Recommended Actions

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

File is 113 days old (last_updated_on exceeds the 90-day threshold)

2. **provider_mapping** `ein_validity` **P2**

0.25% of EINs failed IRS prefix validation (63 of 24772)

3. **pricing** `rate_spread_by_class` **P2**

billing_class='professional' / negotiated_type='fee schedule': P95/P50 spread is 11.8x (threshold: 5x, N=2,378,647 (1,000 sampled), high confidence)

4. **pricing** `rate_spread_by_class` **P2**

billing_class='institutional' / negotiated_type='fee schedule': P95/P50 spread is 14.2x (threshold: 10x, N=175,006 (1,000 sampled), high confidence)

5. **pricing** `per_code_rate_spread` **P2**

57363 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. **code_coverage** `billing_code_format` **P3**

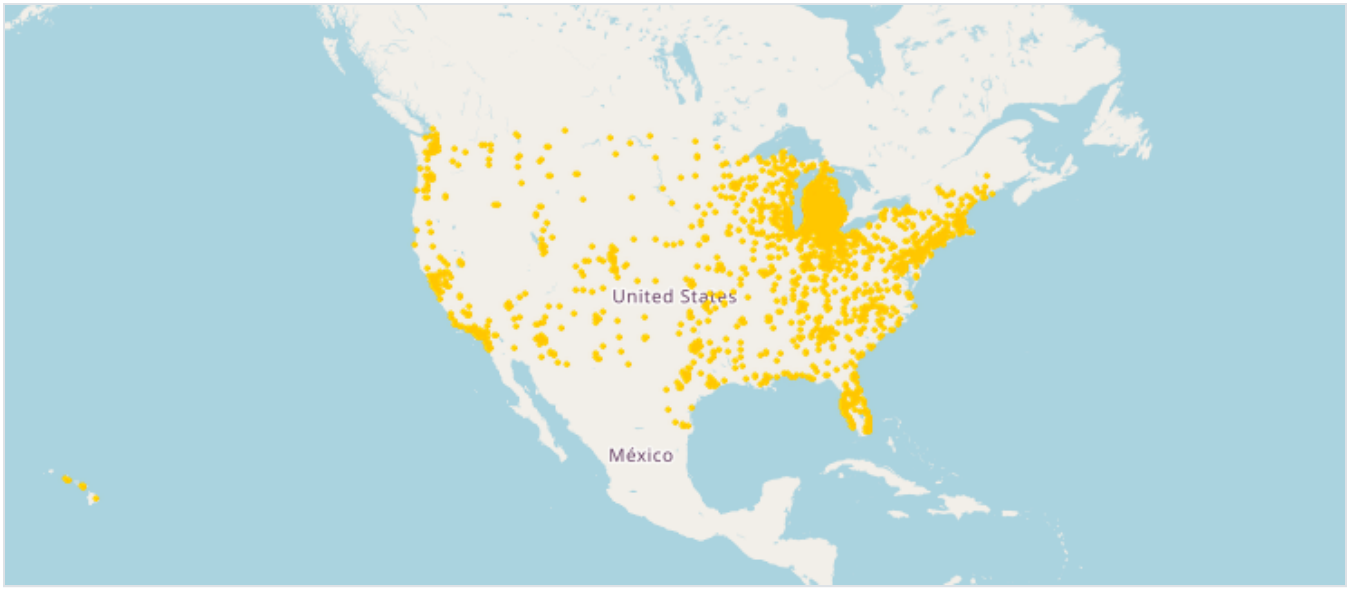
833 CPT codes do not match expected format

7. **code_coverage** `duplicate_billing_codes` **P3**

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

Provider Geographic Coverage

53113 unique NPIs found — 52968 geocoded (100%) — 3155 zip codes represented.



Schema Integrity — Metrics

header_missing_fields	
header_conditional_issues	
file_age_days	113
items_total	18991
items_missing_required_pct	0.0
items_empty_rates	0
prices_total	2599298
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** 17434
	bundle 1557

billing_code_types	CPT	10233
	HCPCS	5167
	ICD	1877
	RC	156
	CSTM-ALL	1
	MS-DRG	1557
expired_prices	0	
invalid_expiration_format	0	

Provider Mapping — Metrics

provider_references_in_file	1110
provider_group_ids_referenced	1110
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	100461
invalid_npi_count	0
npi_validity_rate_pct	100.0
invalid_npi_examples	
eins_validated	24772
invalid_ein_count	63
ein_validity_rate_pct	99.75
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	25369

Code Coverage — Metrics

unique_codes_total	18215
duplicate_codes	776
duplicate_pct	4.26

by_code_type	CPT	10233
	HCPCS	5167
	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

Type	Code	Occurrences
MS-DRG	667	2
MS-DRG	092	2
MS-DRG	064	2
MS-DRG	155	2
MS-DRG	145	2
MS-DRG	958	2
MS-DRG	488	2
MS-DRG	346	2
MS-DRG	165	2
MS-DRG	221	2
MS-DRG	441	2
MS-DRG	079	2
MS-DRG	469	2
MS-DRG	393	2
MS-DRG	579	2
MS-DRG	655	2
MS-DRG	480	2
MS-DRG	619	2
MS-DRG	368	2
MS-DRG	121	2

Pricing Sanity — Metrics

total_prices_checked	2599298
total_rates	2595369
per_diem_rates	3680
percentage_rates	249

negative_rates	0
zero_rates	0
extreme_high_rates	4147
extreme_low_rates	0

rate_distribution

sample_n	2595369
sample_k	5000
confidence	high
p5	4.7195
p25	40.0
p50	210.62
p75	831.6425
p95	3290.737500000003
p99	18015.272100000057

by_billing_class

Class / Type	Count	Median	p25	p75	p95	Confidence
`professional/ fee schedule`	2,378,647	207.1	47.0	744.4	2451.7	high
`institutional/ fee schedule`	175,006	180.3	35.0	693.5	2554.1	high
`institutional/ negotiated`	41,716	21090.6	13239.6	35057.0	78063.5	high

negotiated_types

fee schedule	2553653
negotiated	41716

unique_rate_contexts	1131726

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 = `8d99f48f3a9e5eeb` · 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.

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.