

# MRF Quality Report

s3://talon-storage-private/mrf-feed-uploads/  
2025-10/2025-07-07\_Providence\_Health\_Plan\_975f5c8fe0167c36c5789f7ce9edc56b\_in-network-rates.json.gz  
Size: 11.49 GB • MD5: 63778af2517410952b8a9cb481de1781

Payer: **Providence\_Health\_Plan** • File Date: **2025-07-07** • Generated: **2026-04-25 18:29 EDT** • Tool Version: **1.0.0** • Elapsed: **2307.20s**

# 65.0

Limited Reliability

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

Errors: 4 • Warnings: 11 • Info: 3

## TOC Plan References

TOC: s3://talon-storage-private/mrf-feed-uploads/2025-10/2025-10-15\_Providence-Health-Plan-Master\_combine\_index.json • Providence Health Plan (Health Plan)

Canonical: s3://talon-storage-private/mrf-feed-uploads/2025-10/2025-07-07\_Providence\_Health\_Plan\_975f5c8fe0167c36c5789f7ce9edc56b\_in-network-rates.json.gz

Plan Name	Plan ID	Issuer / Sponsor	Market
HSA	930669184 (EIN)	—	group
HSA_E_XPPO	262863088 (EIN)	—	group
HSA_E_XPPO	930773498 (EIN)	—	group
HSA_E_XPPO	931329922 (EIN)	—	group
HSA_XPPO	262863088 (EIN)	—	group
HSA_XPPO	453966557 (EIN)	—	group
HSA_XPPO	814138803 (EIN)	—	group
HSA_XPPO	910866217 (EIN)	—	group
HSA_XPPO	930177773 (EIN)	—	group
HSA_XPPO	930586809 (EIN)	—	group
OPTIONADV_BASE	930669184 (EIN)	—	group

Plan Name	Plan ID	Issuer / Sponsor	Market
OPTIONADV_BASE_XPPO	453966557 (EIN)	—	group
OPTIONADV_BASE_XPPO	930177773 (EIN)	—	group
OPTIONADV_BASE_XPPO	930891433 (EIN)	—	group
OPTIONADV_BASE_XPPO	936090239 (EIN)	—	group
OPTIONADV_PLUS	930669184 (EIN)	—	group
OPTIONADV_PREM	262863088 (EIN)	—	group
OPTIONADV_PREM_XPPO	203988235 (EIN)	—	group
OPTIONADV_PREM_XPPO	814138803 (EIN)	—	group
OPTIONADV_PREM_XPPO	910866217 (EIN)	—	group
OPTIONADV_PREM_XPPO	930468612 (EIN)	—	group
OPTIONADV_PREM_XPPO	930682364 (EIN)	—	group
OPTIONADV_PREM_XPPO	930773498 (EIN)	—	group
OPTIONADV_PREM_XPPO	931329922 (EIN)	—	group
OPTIONADV_PREM_XPPO	936090239 (EIN)	—	group
PERSONAL_OPTION_XPPO	930996650 (EIN)	—	group
PERSONAL_OPTION_XPPO	936090239 (EIN)	—	group

## CMS Official Schema Validation

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

Validator output:

```
Timed out after 600 s
```

## Dimension Scores

Dimension	Score	Weight	Findings
Schema Integrity	<b>70.0</b>	30%	4
Provider Mapping	<b>100.0</b>	15%	4
Code Coverage	<b>92.8</b>	15%	1
Pricing Sanity	<b>58.0</b>	40%	9

## Schema Integrity — Findings

---

Score: 70.0

**ERROR** `file_freshness`

File is 292 days old (last\_updated\_on exceeds the 90-day threshold)

---

**WARNING** `expired_prices`

2341425 negotiated\_prices have past expiration dates (8.3%)

---

**ERROR** `cms_schema_validation`

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

---

**ERROR** `parse_crash`

single pass crash after 884 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(434250156 bytes read, 11897802067 more expected)', IncompleteRead(434250156 bytes read, 11897802067 more expected))

---

## Provider Mapping — Findings

---

Score: 100.0

**WARNING** `npi_validity`

0.01% of NPIs failed Luhn checksum validation (6 of 112525)

- 1336588484
  - 1659650617
- 

**WARNING** `ein_validity`

0.30% of EINs failed IRS prefix validation (191 of 62665)

- 000000001
  - 001625898
  - 003360711
  - 003680161
  - 004829723
  - ... and 5 more
- 

**WARNING** `empty_npi_groups`

52 provider groups contain no NPIs

---

**INFO** `duplicate_npis`

36425 NPIs appear in more than one provider group

---

# Code Coverage — Findings

---

Score: 92.8

**WARNING** `billing_code_format`

127 CPT codes do not match expected format

---

## Pricing Sanity — Findings

---

Score: 58.0

**INFO** `per_diem_rates`

657 per-diem rates (0.0%) — not dollar amounts; excluded from spread analysis

**INFO** `percentage_rates`

41965 percentage rates (0.1%) — values represent % of a reference rate, not dollar amounts; excluded from spread analysis

**ERROR** `zero_rates`

2010 zero-dollar rates (0.01%) — CMS schema requires `negotiated_rate > 0` (`exclusiveMinimum`)

**WARNING** `extreme_rates`

2.12% of rates are extreme (596576 above class-specific high threshold, 0 below \$0.01)

**WARNING** `high_frequency_rate_value`

1 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': 2250000.0, 'count': 401629, 'pct': 1.43}`

**WARNING** `rate_spread_by_class`

`billing_class='institutional' / negotiated_type='negotiated'`: P95/P50 spread is 1517.5x (threshold: 10x, N=2,568 (1,000 sampled), high confidence)

**WARNING** `rate_spread_by_class`

`billing_class='professional' / negotiated_type='fee schedule'`: P95/P50 spread is 5.3x (threshold: 5x, N=27,102,797 (1,000 sampled), high confidence)

**WARNING** `rate_spread_by_class`

`billing_class='professional' / negotiated_type='negotiated'`: P95/P50 spread is 2413.6x (threshold: 5x, N=7,222 (1,000 sampled), high confidence)

---

**WARNING** `per_code_rate_spread`

312 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
J1443	HCPCS	negotiated	institutional	ffs	—	\$0.03	\$600000.03	\$621562.52	\$1286250.00	42875000.0×	4
J0872	HCPCS	negotiated	institutional	ffs	—	\$0.04	\$600000.03	\$621562.53	\$1286250.00	32156250.0×	4
J1171	HCPCS	negotiated	institutional	ffs	—	\$0.09	\$600000.07	\$621562.56	\$1286250.00	14291666.7×	4
A6443	HCPCS	fee schedule	professional	ffs	—	\$0.30	\$0.37	\$8780.32	\$2400000.00	8000000.0×	25302
S5552	HCPCS	negotiated	institutional	ffs	—	\$0.22	\$600000.17	\$621562.64	\$1286250.00	5846590.9×	4
S5552	HCPCS	negotiated	institutional	ffs	—	\$0.22	\$600000.17	\$621562.64	\$1286250.00	5846590.9×	4
C9144	HCPCS	fee schedule	professional	ffs	—	\$0.44	\$0.60	\$4609.11	\$2400000.00	5454545.5×	1009
J7328	HCPCS	fee schedule	professional	ffs	—	\$0.55	\$0.70	\$21280.55	\$2400000.00	4363636.4×	1008
Q0164	HCPCS	negotiated	institutional	ffs	—	\$0.35	\$600000.26	\$621562.72	\$1286250.00	3675000.0×	4
Q0164	HCPCS	negotiated	institutional	ffs	—	\$0.35	\$600000.26	\$621562.72	\$1286250.00	3675000.0×	4

## Recommended Actions

1. **schema** `file_freshness`

P1

File is 292 days old (`last_updated_on` exceeds the 90-day threshold)

2. **schema** `cms_schema_validation`

P1

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

3. **schema** `parse_crash`

P1

single pass crash after 884 items: An error occurred while reading from response stream: ('Connection broken: IncompleteRead(434250156 bytes read, 11897802067 more expected)', IncompleteRead(434250156 bytes read, 11897802067 more expected))

4. **pricing** `zero_rates`

P1

2010 zero-dollar rates (0.01%) — CMS schema requires `negotiated_rate > 0` (`exclusiveMinimum`)

5. **provider\_mapping** `npi_validity`

P2

0.01% of NPIs failed Luhn checksum validation (6 of 112525)

6. **provider\_mapping** `ein_validity`

P2

0.30% of EINs failed IRS prefix validation (191 of 62665)

7. **provider\_mapping** `empty_npi_groups`

P2

52 provider groups contain no NPIs

8. **pricing** `extreme_rates`

P2

2.12% of rates are extreme (596576 above class-specific high threshold, 0 below \$0.01)

9. **pricing** `high_frequency_rate_value`

P2

1 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.

10. **pricing** `rate_spread_by_class`

P2

billing\_class='institutional' / negotiated\_type='negotiated': P95/P50 spread is 1517.5x (threshold: 10x, N=2,568 (1,000 sampled), high confidence)

11. **pricing** `rate_spread_by_class`

P2

billing\_class='professional' / negotiated\_type='fee schedule': P95/P50 spread is 5.3x (threshold: 5x, N=27,102,797 (1,000 sampled), high confidence)

12. **pricing** `rate_spread_by_class`

P2

billing\_class='professional' / negotiated\_type='negotiated': P95/P50 spread is 2413.6x (threshold: 5x, N=7,222 (1,000 sampled), high confidence)

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

P2

312 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.

14. **schema** expired\_prices

P2

2341425 negotiated\_prices have past expiration dates (8.3%)

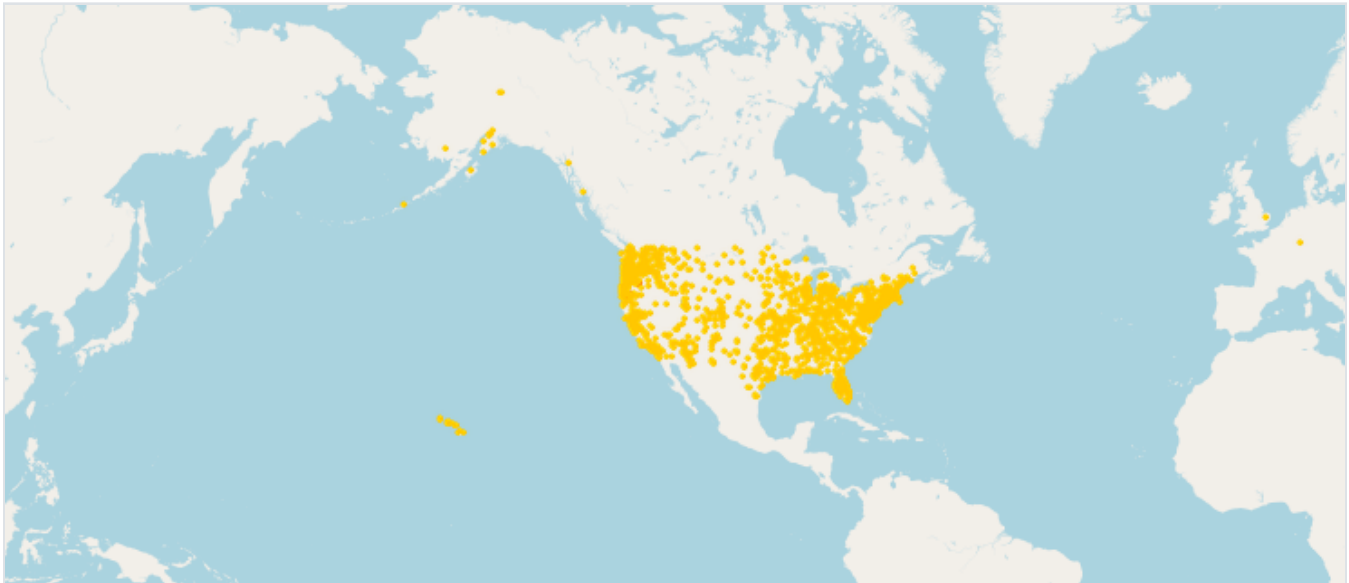
15. **code\_coverage** billing\_code\_format

P3

127 CPT codes do not match expected format

## Provider Geographic Coverage

37896 unique NPIs found — 37728 geocoded (100%) — 3809 zip codes represented.



## Schema Integrity — Metrics

header\_missing\_fields

header\_conditional\_issues

file\_age\_days

292

items_total	884	
items_missing_required_pct	0.0	
items_empty_rates	0	
prices_total	28214499	
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	884
billing_code_types	HCPCS	246
	CPT	529
	CDT	36
	APC	28
	RC	22
	MS-DRG	23
expired_prices	2341425	
invalid_expiration_format	0	

## Provider Mapping — Metrics

provider_references_in_file	50557
provider_group_ids_referenced	47473
unresolved_references	0
resolution_rate_pct	100.0
npis_validated	112525
invalid_npi_count	6
npi_validity_rate_pct	99.99
invalid_npi_examples	1336588484, 1336588484, 1336588484, 1659650617, 1659650617, 1659650617
eins_validated	62665
invalid_ein_count	191

ein_validity_rate_pct	99.7
invalid_ein_examples	000000001, 000000001, 000000001, 001625898, 001625898, 001625898, 003360711, 003360711, 003360711, 003680161
empty_npi_groups	52
groups_without_tin	0
npi_in_multiple_groups	36425

## Code Coverage — Metrics

unique_codes_total	884												
duplicate_codes	0												
duplicate_pct	0.0												
by_code_type	<table border="1"> <tr> <td>HCPCS</td> <td>246</td> </tr> <tr> <td>CPT</td> <td>529</td> </tr> <tr> <td>CDT</td> <td>36</td> </tr> <tr> <td>APC</td> <td>28</td> </tr> <tr> <td>RC</td> <td>22</td> </tr> <tr> <td>MS-DRG</td> <td>23</td> </tr> </table>	HCPCS	246	CPT	529	CDT	36	APC	28	RC	22	MS-DRG	23
HCPCS	246												
CPT	529												
CDT	36												
APC	28												
RC	22												
MS-DRG	23												
unknown_code_types													
format_invalid_by_type	<table border="1"> <tr> <td>CPT</td> <td>127</td> </tr> </table>	CPT	127										
CPT	127												
codes_not_in_reference	reference_not_loaded												

most_frequent_codes	Type	Code	Occurrences
	HCPCS	S2411	1
	CPT	27814	1
	CPT	81291	1
	CPT	19229	1
	CPT	27686	1
	CPT	0421T	1
	CPT	99467	1
	HCPCS	J0578	1
	CPT	0872T	1
	CPT	28124	1
	CPT	93641	1
	CPT	22853	1
	HCPCS	G9006	1
	CPT	0513U	1
	CDT	D4323	1
	HCPCS	G9497	1
	HCPCS	G9897	1
	HCPCS	B0106	1
	CPT	25676	1
	CPT	67221	1

## Pricing Sanity — Metrics

total_prices_checked	28214499
total_rates	28171877
per_diem_rates	657
percentage_rates	41965
negative_rates	0
zero_rates	2010
extreme_high_rates	596576
extreme_low_rates	0

<b>rate_distribution</b>	<b>sample_n</b>	28171877					
	<b>sample_k</b>	5000					
	<b>confidence</b>	high					
	<b>p5</b>	29.648999999999997					
	<b>p25</b>	253.22					
	<b>p50</b>	789.02					
	<b>p75</b>	1691.4175					
	<b>p95</b>	4611.1					
	<b>p99</b>	2250000.0					
<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>
	institutional/ negotiated	2,568	1084.0	7.4	33169.7	1644940.5	high
	professional/ fee schedule	27,102,797	881.7	267.0	1800.7	4710.8	high
	institutional/ fee schedule	1,059,290	701.1	153.9	2048.3	5264.9	high
	professional/ negotiated	7,222	932.2	362.2	1876.7	2250000.0	high
<b>negotiated_types</b>	<b>negotiated</b>	9790					
	<b>fee schedule</b>	28162087					
<b>unique_rate_contexts</b>	3491						
<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 = f49bccdcfc24da64 · entity = Providence_Health_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.