ProcedureRadar

Procedure-Pricing Data from Hospital Transparency Files: Methodology, Sources, and Limitations

Methodology version 1.0.0. Effective 2026-05-09. Authored by ProcedureRadar.

1. Executive Summary

1.1 Purpose and Scope

This white-paper is the actuarial communication for ProcedureRadar's v1 B2B API methodology. It is written to the standard that ASOP 41 (Actuarial Communications, Source 3) requires: a qualified actuary in the same practice area should be able to read this document and make an objective appraisal of the reasonableness of the work without access to the pipeline codebase. Where we reference implementation files, we do so to indicate where a reviewer can verify claims, not to substitute code references for methodological explanation.

ProcedureRadar produces procedure-pricing features from federally mandated hospital Machine-Readable Files published under 45 CFR Part 180. For every (hospital, procedure, payer, charge-type) combination with sufficient observations, we compute empirical severity distributions (percentile_25, percentile_75, percentile_95, percentile_99, mean, standard deviation) at credible sample sizes per ASOP 25 (Source 2), bias-corrected and accelerated bootstrap confidence intervals per the North American Actuarial Journal BCa standard (Source 15), BLS CUUR0000SAM time-series normalization with documented limitations (Sources 13, 14), covered-event episode mapping with fractional-contribution weights (Enterprise tier), and a six-component data quality score anchored to the AAA Big Data governance framework (Source 8). Fields that do not meet credibility thresholds are suppressed rather than approximated. The result is a tier-appropriate underwriting data product priced and scoped for specialty payers, MGAs, and regional TPAs that cannot clear the procurement hurdles of enterprise data vendors.

1.2 The Four Properties This Paper Claims

The four properties below are specific methodological claims, not marketing descriptors. A reader can check each one against the fields we publish and the source file named on every row:

First, severity distributions are computed only where the observation count is sufficient to produce statistically defensible empirical quantiles, with empirical quantiles computed at n = 30 and above and the credibility procedure documented per ASOP 25 (Source 2). Below that floor, we apply a documented small-sample t-interval fallback on the mean (10 to 29 observations) or suppress entirely (fewer than 10). No point estimate is presented as credible when the underlying sample does not support the claim.

Second, confidence intervals use the bias-corrected and accelerated bootstrap method (BCa) established in peer-reviewed actuarial literature as the standard for risk-measure confidence intervals at insurance sample sizes (Gruen and Miljkovic, NAAJ, 2022, Source 15). BCa corrects for the bias and skewness that plain percentile-bootstrap understates at the right-skewed price distributions typical in hospital charge data.

Third, time-series normalization applies a primary federal statistical series (BLS CUUR0000SAM, Sources 13, 14) with the specific limitation that the October 2024 BLS methodology change introduced a conceptual mismatch between the deflator (which now partly reflects adjudicated reimbursement prices) and our input data (which is charge-side). We name this limitation and explain why we retain CUUR0000SAM despite it.

Fourth, every pricing record carries a six-component data quality score governed by the ASOP 23 compliance layer (Source 1) and the AAA Big Data governance framework (Source 8), with published composite weights that a practitioner can audit. Records that fail quality gates are flagged, not silently passed through.

This paper does not claim that our data is the most comprehensive or largest dataset in the market. The enterprise data majors maintain larger record volumes and, in some cases, access to adjudicated claims data that hospital MRFs cannot provide. Our lane is documented, source-traceable data depth at a published price, accessible to organizations that cannot clear the procurement hurdles of enterprise data vendors.

1.3 What the Methodology Produces

The v1 methodology produces five categories of output, each documented in a dedicated section of this paper:

Severity distributions (Section 4). Empirical percentiles at the procedure-metro-payer-charge-type-setting grain for cohorts of 30 or more observations, with the percentile ladder withheld below that threshold and the severity fields suppressed entirely below 10. These give an underwriter the distributional shape of observed hospital charges, not just a midpoint.

Confidence intervals (Section 5). BCa bootstrap CIs at 95 percent confidence on the mean, 25th, 75th, 95th, and 99th percentiles for cohorts with 30 or more observations; a small-sample t-interval on the mean for cohorts with 10 to 29 observations; suppression for cohorts with fewer than 10 observations. These quantify sampling uncertainty around each severity estimate.

Time-series normalization (Section 6). CPI-adjusted and medical-CPI-adjusted price columns anchored to a stated reference month, with regional price indexing via BLS metro-area series. An append-only pricing-history table guarantees no look-ahead bias in historical comparisons.

Covered-event taxonomy (Section 7). A 10–category episode-of-care mapping with fractional-contribution weights, available at the Enterprise tier, that lets a practitioner query an episode (for example, a knee-replacement bundle) and receive aggregate pricing rather than individual line-item prices.

Data quality score (Section 8). A six-component composite score in [0, 1] with published weights, a three-band interpretation framework (above 0.85, 0.70 to 0.85, below 0.70), and a direct relationship to the categorical quality-flag enum for practitioners who need categorical filtering.

1.4 Data Source and Legal Basis

All pricing data in this methodology derives from hospital Machine-Readable Files published under 45 CFR Part 180 (the Hospital Price Transparency Rule, Source 10). Hospitals in the top 100 U.S. metropolitan markets are required to publish standard charges in machine-readable

format. We ingest each hospital's file directly, retain the source URL and publication date on every record, and treat each hospital's own published file as the authoritative source for that hospital's prices. Section 2 documents the legal basis in full.

TiC (Transparency in Coverage) payer-negotiated rates now ship through the B2B API on the Scale tier and above, which is a change from what this paper originally described. File scope varies by payer: some payer rates come from network-wide files, some from a prioritized slice of a payer's network files, and some from employer-group files, as Section 10.5 documents per payer along with the deterministic rule applied when a payer's own files disagree. Section 10.5 also states the measured size of the lane and names the payers whose rows exist but cannot be served. The methodology documented here was authored charge-side, and it has not yet been extended to cover the payer lane. Every payer-sourced row is distinguishable in an API response by `data_source = 'payer_tic'` and carries its own source file and date, and payer rows are excluded by construction from the severity distributions and confidence intervals this paper specifies. Section 10.5 states exactly what that exclusion covers.

1.5 Scope Boundary

This paper documents the methodology behind fields exposed through the v1 B2B API as of the methodology version 1.0.0 effective date (2026-05-09). Features documented in this paper correspond to what the API returns at each tier as of that date. Future methodology versions will carry a new `methodology_version_id` and a changelog entry. Section 11 documents the versioning and pre-notification commitments.

1.6 References

1. Source 1: ASOP No. 23, Data Quality (Revised Edition, Actuarial Standards Board, adopted December 2016, effective April 1, 2017). Operative sections: source identification, known limitations disclosure, adjustment documentation, extent-of-reliance disclosure for each upstream data layer.
2. Source 2: ASOP No. 25, Credibility Procedures (Revised Edition, Actuarial Standards Board, adopted December 2013). Operative sections: full-credibility threshold documentation, purpose-appropriate threshold selection, application of credibility procedures to external data.
3. Source 3: ASOP No. 41, Actuarial Communications (Revised Edition, effective May 1, 2011). Operative sections: clarity for peer actuary review; disclosure of methods, procedures, assumptions, models, and data; cautions regarding uncertainty.
4. Source 8: American Academy of Actuaries, "Big Data and Algorithms in Actuarial Modeling and Consumer Impacts," Data Science and Analytics Committee Issue Brief (October 2022). Operative sections: four-point governance framework for external data sources; validation, collection-procedure review, quality assurance guidelines, regulatory compliance analysis.

5. Source 13: Bureau of Labor Statistics, "How BLS Measures Price Change for Medical Care Services in the Consumer Price Index" (factsheet, current); "Incorporating Medical Claims Data in the CPI," Monthly Labor Review (2023). Operative sections: October 2024 CUUR0000SAM methodology change; claims-data switch for outpatient private-insurance component.
6. Source 14: Dunn, A., Hale, E., and Dauda, B., "Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States," Health Services Research, Vol. 53, No. 1, pp. 526–548 (2018). Operative sections: CPI medical care vs. PCE PHC deflator selection by expenditure type.
7. Source 15: Gruen, B. and Miljkovic, T., "The Automated Bias-Corrected and Accelerated Bootstrap Confidence Intervals for Risk Measures," North American Actuarial Journal, Vol. 27, No. 4, pp. 731–750 (2023, published online December 2, 2022). Operative sections: BCa as actuarial standard for risk-measure confidence intervals; 10,000–resample recommendation for tail stability.

2. Data Sources and Legal Basis

2.1 Federal Mandate: 45 CFR Part 180

The legal authority for the hospital charge-side data this paper documents is 45 CFR Part 180, "Hospital Price Transparency," published in the U.S. Code of Federal Regulations (Source 10). It is the authority for almost all of what the API returns; payer-published rates, described in Section 2.5, rest on the separate Transparency in Coverage rule. The rule requires hospitals with 25 or more beds to publish a machine-readable file of standard charges for all items and services they provide. The eCFR version of 45 CFR Part 180 is the authoritative current text; the rule has been amended multiple times since its original effective date.

A brief regulatory history: the rule became effective January 1, 2021. CMS implemented major enforcement updates through 2024, including civil monetary penalty escalations and mandatory attestation requirements. As of July 1, 2024, hospitals must publish MRFs that conform to a CMS-specified template format. As of January 1, 2025, MRFs must include additional payer-specific fields. As of January 1, 2026, hospitals must publish the 10th percentile, median, and 90th percentile of allowed amounts with a count of remittances for each shoppable service, representing the first statutory requirement for hospitals to publish empirical percentile data from their own adjudicated claims experience. This January 2026 expansion is documented separately in Section 10.11 and referenced again in Section 11 as a v1.1 integration target.

2.2 ASOP 23 Compliance Layer: Per-Source Documentation

ASOP 23 (Source 1) requires that for each upstream data source, the actuary document: (a) source identity and scope, (b) known biases or limitations, (c) adjustments applied, and (d) the extent of reliance placed on data supplied by others. We apply this requirement to each of our four upstream sources.

Hospital Machine-Readable Files. Source identity: per-hospital MRF files published at each hospital's own URL on each hospital's own cadence, ingested via the ProcedureRadar pipeline. Scope: standard charges as defined in 45 CFR Part 180, including gross charge, discounted cash price, payer-specific negotiated charges, and de-identified minimum and maximum negotiated charges. Known biases and limitations: the GAO found in October 2024 (GAO-25–106995, Source 11) that CMS cannot verify that hospital MRF data are sufficiently complete and accurate, and that health plans and employers have raised

concerns about data quality including prices that appear unusually high or low. The HHS Office of Inspector General found in November 2024 (Source 12) that approximately 46 percent of hospitals were non-compliant with one or more price transparency requirements as of early 2023. Adjustments applied: shoppable-service early filter at parse time, placeholder-value detection, outlier flagging against historical hospital pricing patterns, and the six-component data quality score (Section 8). Extent of reliance: high; the hospital MRF is the sole source of procedure-level pricing data in v1. We do not supplement with proprietary claims data, pharmacy benefit data, or third-party pricing databases. Where a hospital's MRF is incomplete, non-compliant, or unparseable, our data for that hospital is incomplete accordingly. The coverage sub-score in the data quality score reflects this directly.

BLS CUUR0000SAM (Medical Care Services CPI). Source identity: Bureau of Labor Statistics Consumer Price Index series CUUR0000SAM, published monthly as part of the CPI-U all-urban-consumers family. Scope: a national index of medical care services price changes, based on a household survey sample with an October 2024 methodology change for the outpatient private-insurance sub-component (Source 13). Known biases and limitations: as of October 2024, the outpatient private-insurance component of CUUR0000SAM switched from survey-based list-charge pricing to adjudicated claims data, creating a conceptual mismatch with our charge-side input data. This mismatch is documented in detail in Section 6.3 and Section 10.9. Additionally, CUUR0000SAM is a national series; it does not capture within-metro geographic price variation, which we address through regional indexing described in Section 6.4. Adjustments applied: none to the BLS series itself. We apply the series as published. Extent of reliance: the deflator is applied to every time-series-adjusted column. Practitioner actuaries should review Section 6.3 before relying on time-series-adjusted values for use cases sensitive to the charge-side versus reimbursement-side distinction.

NPPES (National Plan and Provider Enumeration System). Source identity: CMS National Plan and Provider Enumeration System NPI registry, used to match provider identifiers to hospital facility records and to support CCN-to-EIN bridging in the URL refresher. Scope: provider identification data, not pricing data. Known biases and limitations: NPPES records may be stale for providers that have changed address, ownership, or tax identification number. The pipeline uses NPPES as one of four URL-discovery systems, not as a sole source.

Adjustments applied: cross-validated against TPAFS and direct CMS hospital price transparency index records. Extent of reliance: supporting role in URL discovery and hospital identification; not load-bearing for pricing values.

TPAFS (Third-Party Administrator Filing System). Source identity: CMS TPAFS system, used as one component of the CCN-to-URL mapping for hospitals whose MRF URLs are not directly discoverable from the CMS price transparency index. Known biases and limitations: NPPES-to-TPAFS upstream mapping contains documented CCN-to-URL contamination for a subset of hospitals (the pipeline applies a canonicalized-equivalence guard to identify and exclude these contaminated mappings). Adjustments applied: contamination guard applied before any URL derived from TPAFS is used as an ingestion source. Extent of reliance: minor; TPAFS is a tertiary URL-discovery fallback. The primary URL source is the CMS hospital price transparency index; TPAFS is used only when the primary source does not return a valid URL.

2.3 Per-Hospital Ingestion Model

We treat each hospital's own published MRF as the authoritative source for that hospital's prices. We do not blend prices across hospitals or impute missing hospital data from peer-hospital distributions. The source URL and publication date are retained on every pricing record; practitioners can verify the provenance of any specific record by inspecting the `hospital_mrf_url` and `source_file_date` fields returned in API responses.

Hospitals publish MRFs on their own cadence, which varies from monthly to annually. The pipeline re-ingests each hospital's MRF on a monthly pipeline run and updates the `source_file_date` and freshness sub-score accordingly. Records from hospitals that have not published a new MRF since the previous run retain the prior `source_file_date`, which the freshness sub-score penalizes over time.

2.4 BLS Series Used for Time-Series Normalization

The primary deflator is BLS CUUR0000SAM (Medical Care Services), discussed above and in detail in Section 6. The secondary reference series is BLS CUUR0000SA0 (All Items CPI-U), included as a benchmark for expressing medical price changes in the context of broader consumer inflation. CUUR0000SA0 is not used as the operative deflator for any pricing-record adjustment.

Regional price indexing uses BLS metro-area-level CPI sub-indexes, mapped to our internal metro slugs via a mapping table maintained in the pipeline codebase. The October 2024 BLS methodology change to CUUR0000SAM (which switched the outpatient private-insurance component from survey-based list-charge pricing to claims-data-sourced adjudicated pricing, per Source 13) is a date marker for practitioners reviewing our time-series output: records with `reference_index_month` values from October 2024 onward are deflated using a series that has a different conceptual basis for its outpatient private-insurance component than records with earlier reference months.

2.5 TiC Integration: Shipped for the API, Methodology Update Owed

The Transparency in Coverage rule requires health insurers to publish payer-negotiated machine-readable files containing plan-specific allowed amounts for every service. TiC files represent the reimbursement side of the pricing picture: what the payer actually pays, after adjudication, as opposed to what the hospital charges. A complete underwriting data product would include both charge-side and reimbursement-side signals.

TiC ingestion has shipped and payer-negotiated rates are served on the Scale tier and above, with file scope that varies by payer; Section 10.5 states the scope per payer, including where a payer's rates come from a single large employer group's file. The computed fields this paper documents were specified against hospital charge-side data, and they remain charge-side: the cohort recompute excludes payer rows explicitly, so a payer rate neither enters a distribution nor receives one. What is still owed is a positive methodology for the payer lane itself, not a correction to the charge-side statistics. We disclose it in this form rather than quietly updating the text, because the charge-side versus reimbursement-side distinction is methodologically significant for rate-filing actuaries who work with adjudicated claims data in their primary models. Section 10.5 documents what a v1 user should assume in the meantime.

2.6 Future Integration: January 2026 Allowed-Amount Disclosure

As of January 1, 2026, 45 CFR Part 180 (Source 10) requires hospitals to publish the 10th percentile, median, and 90th percentile of allowed amounts with a count of remittances for each shoppable service. This is the first statutory requirement for hospital-published empirical percentile data from adjudicated claims. The v1 methodology does not yet incorporate these hospital-published percentiles in any computed output. Section 11 documents the v1.1 commitment to evaluate integration.

2.7 References

1. Source 1: ASOP No. 23, Data Quality (Revised Edition, Actuarial Standards Board, adopted December 2016, effective April 1, 2017). Operative sections: per-source documentation requirements; source identification, scope limitation, known bias description, adjustment documentation, extent-of-reliance disclosure.
2. Source 10: U.S. Code of Federal Regulations, 45 CFR Part 180, Hospital Price Transparency (eCFR current version; rule effective January 1, 2021, major updates through January 1, 2026). Operative sections: standard-charge publication requirements; shoppable-service definition; January 2026 allowed-amount percentile disclosure requirements; MRF publication requirements by hospital type.
3. Source 11: U.S. Government Accountability Office, GAO-25–106995, "Health Care Transparency: CMS Needs More Information on Hospital Pricing Data Completeness and Accuracy" (released October 2, 2024). Operative sections: CMS inability to verify MRF completeness and accuracy; health plan and employer data quality concerns; CMS Request for Information on enforcement.
4. Source 12: HHS Office of Inspector General, "Not All Selected Hospitals Complied With the Hospital Price Transparency Rule" (November 2024). Operative sections: approximately 46 percent non-compliance rate; deficient machine-readable files as primary compliance gap; small-hospital resource-constraint finding.
5. Source 13: Bureau of Labor Statistics, "How BLS Measures Price Change for Medical Care Services in the Consumer Price Index" (factsheet, current); "Incorporating Medical Claims Data in the CPI," Monthly Labor Review (2023). Operative sections: October 2024 CUUR0000SAM methodology change; switch to claims-data-sourced adjudicated pricing for outpatient private-insurance component; date marker for time-series interpretation.

4. Severity Distribution Methodology

4.1 Purpose

An underwriter evaluating a specialty-health rate filing or a TPA benchmarking inpatient surgery costs needs more than a median price. A median alone tells you the center of the distribution but nothing about its shape: whether prices cluster tightly around that midpoint or fan out across a wide range, whether the tail is fat enough to stress a stop-loss layer, and whether the 75th percentile is close to or far from the 95th. The severity distribution columns on every `pricing_record` give the underwriter that shape. Specifically, we compute and store `percentile_25`, `percentile_75`, `percentile_95`, `percentile_99`, `mean_price`, and `stdev_price` at the procedure-metro-payer-charge-type-setting grain. `percentile_10` and `percentile_90` are preserved in the consumer-facing surface, distinct from the B2B API, where a `low_sample` row returns both fields as `null`. This set of descriptors lets a practitioner actuary reconstruct the distributional structure of observed hospital charges for a given procedure in a given market without access to the underlying microdata.

4.2 Sample-Size Regimes and Credibility Thresholds

We divide every pricing cohort into one of three regimes based on the count of line-item observations at the procedure-metro-payer-charge-type-setting grain. Payer-published rows (data source `payer_tic`) are excluded from cohort statistics: a payer-published point rate is its own disclosure, not an observation of a hospital-published price distribution. The regime determines which computation method applies. This three-way structure is not arbitrary: it follows the framework ASOP 25 (Credibility Procedures) requires any actuary to apply when representing data as statistically credible for a specific actuarial purpose. ASOP 25 does not prescribe a universal observation count for full credibility; it requires the practitioner to select a procedure appropriate for the purpose, the data volume, and the intended use, and to document why the selection is appropriate. We do exactly that here.

Regime 1: n >= 30 (empirical quantiles). When a cohort contains 30 or more observations, we compute severity percentiles directly from the empirical distribution using `numpy.percentile`. This produces unbiased quantile estimates that do not depend on any distributional assumption. The choice of 30 as the floor for empirical quantile computation is grounded in two converging sources. To be precise about what we claim: n = 30 is the threshold at which empirical sample quantiles, including the tail percentiles, become stable enough for rate-filing use. It is not the classical full-credibility standard of ratemaking, which

for a frequency-based standard runs on the order of 1,082 claims. We make no claim that a 30-observation cohort carries full actuarial credibility in that sense; we claim only that the empirical quantiles are stable enough to report at and above this floor, and we report the sample size on every estimate so the practitioner can apply their own credibility weighting.

First, Werner and Modlin (Basic Ratemaking, Fifth Edition, CAS study note, Source 7) describe credibility as the degree of confidence an actuary places in a body of data. The text discusses the classical full-credibility threshold in the context of primary ratemaking data and notes that practitioners routinely anchor empirical distribution estimates at n = 30 as the minimum below which sample quantile stability deteriorates materially, particularly for tail percentiles such as the 95th and 99th. At n = 30, the empirical 25th percentile is the 7th or 8th order statistic and the empirical 75th percentile is the 22nd or 23rd. Both are reasonably stable. At n = 20, the tail order statistics begin to exhibit high variance across independent draws from the same population, producing estimates that can mislead a rate-filing actuary about the true distributional shape.

Second, ASOP 25 explicitly requires that the purpose and intended use of the data govern the threshold selection. Our data is used as input to underwriting rate calculations for specialty-health and MGA products where the 75th and 95th percentiles of procedure cost directly feed into credibility-blending worksheets, stop-loss attachment-point analysis, and corridor rating. At this intended use, a threshold below 30 would produce percentile estimates with standard errors large enough to contaminate the downstream rate calculation. We are not claiming that 30 is a universal standard; we are stating that 30 is the appropriate threshold for this purpose and this use case, which is exactly what ASOP 25 requires.

Regime 2: `10 ⩽ n < 30` (mean and interval only). When a cohort contains between 10 and 29 observations we cannot estimate quantiles from the empirical distribution, and as of the 2026-08-26 methodology revision we no longer estimate them any other way either. The cohort still carries actuarial information an underwriter should not discard, so we serve `cohort_median`, `sample_size`, `mean_price`, `stdev_price`, and a small-sample t-interval on the mean. We do not serve `percentile_10`, `percentile_25`, `percentile_75`, `percentile_90`, `percentile_95`, or `percentile_99`, and we do not serve percentile confidence intervals. All of those fields return `null` in this regime.

This is a withdrawal of a prior method, and we document it rather than quietly restate the current one. Through 2026-08-26 this regime derived each percentile parametrically from the standard normal z-score applied to the sample mean and standard deviation. The justification offered was the Central Limit Theorem, which does support treating the sample MEAN as approximately normal at n >= 10. It does not support extending that approximation to the shape of the distribution, which is what deriving a quartile from a z-score does. Doing so assumes the underlying price distribution is approximately symmetric in the observed range, and hospital procedure prices are characteristically right-skewed.

The failure was measurable, and we measured it across the full population rather than a sample. Of the 3,388,182 stored rows with 10 to 29 observations, 188,172 (5.55 percent) carried a `percentile_25` ABOVE their own `cohort_median`, and a further 139,623 (4.12 percent) carried a `percentile_75` BELOW it. Those are parametric quartiles landing on the wrong side of a measured median: not a rounding artifact, but a contradiction between two numbers served in the same row. The same two checks run against all 343,586 rows at n >= 30, where the ladder is empirical, found zero violations of either kind. A method that contradicts itself on roughly one row in ten does not meet the standard this document sets in Section 4.5, so we withdrew it rather than caveat it.

An underwriter working a thin cohort therefore gets a measured center, a measured spread, an honest interval around the mean, and the observation count with which to apply their own credibility weighting. What they no longer get is a distributional shape we cannot actually observe. The `sample_regime` field on every API row reports which regime produced it: `full` at n >= 30, `small_sample` at 10 to 29, `suppressed` below 10, so the absence of a percentile is always distinguishable from the absence of data.

Percentile estimates in the empirical regime remain clamped to a lower bound of zero. Price distributions are truncated at zero by construction: no price can be negative, and returning a negative lower price bound to an underwriter would be an absurdity regardless of what the arithmetic produced.

Regime 3: `n < 10` (suppression). When a cohort contains fewer than 10 observations, no distributional estimate we can produce is statistically defensible in a rate-filing context. At this observation count, even the sample mean carries a coefficient of variation large enough to render percentile estimates misleading. We suppress the severity fields entirely: `median_price`, `percentile_10`, `percentile_25`, `percentile_75`, `percentile_90`,

`percentile_95`, `percentile_99`, `mean_price`, and `stdev_price` are all returned as `null`, and the `quality_flag` field is set to `low_sample`. The confidence interval fields are likewise `null`.

This behavior is locked in the pipeline at `SUPPRESS_BELOW = 10` in `severity_compute.py`. It is also enforced by a check in `bca_bootstrap_ci` (Source: `packages/pipeline/scripts/bca_bootstrap.py`) that raises a `ValueError` if called on a sample of fewer than 10 observations, ensuring the bootstrap path cannot accidentally produce a CI for a suppressed cell.

The practitioner actuary encountering suppressed records should treat them as unquotable from this source and supplement with an appropriate complement of credibility (per Werner-Modlin, Source 7, Chapter 12) drawn from market-wide or regional pricing benchmarks. We do not provide the complement directly; we document the suppression so the actuary can apply their own credibility-weighting procedure.

4.3 Percentile Ordering Constraint

A severity distribution is only internally consistent if the percentiles are ordered: p10 is no larger than p25, which is no larger than p50, and so on through p99. Violations of this ordering constraint indicate a data-quality problem at the compute level and should never reach the API response layer.

We enforce this invariant at two independent points. First, the database layer enforces the constraint via the check condition added in migration 013: `p10 ⩽ p25 ⩽ p50 ⩽ p75 ⩽ p90 ⩽ p95 ⩽ p99`, with any position allowed to be `null`. This constraint is the model validation evidence required by ASOP 56 (Modeling, Section 4.4, Source 4), which requires that model output be validated against its intended purpose and that violations be detectable and logged. A pricing record that violates the ordering constraint cannot be written to the database, satisfying the ASOP 56 requirement that the model's valid input range be documented and that behavior outside that range be flagged.

Second, `severity_compute.py` applies non-negativity clamping after percentile computation (`p25 = max(p25, 0.0)`, etc.), which prevents any clamping operation from creating an out-of-order result when the pre-clamp value was negative. Any monotonicity

violation that somehow survived compute-level clamping would be caught at the write-time database constraint check. Pipeline logs record any constraint violation as an error-level event for operational review.

4.4 No Response-Layer Rounding

Every API key receives full-resolution `percentile_25` and `percentile_75` values from its first call. Earlier versions of the API rounded these two values to 50-dollar steps for new accounts; we retired that rule in October 2026. The stored values were always full-resolution, and nothing else about how percentiles are computed has changed.

4.5 Data Quality Score Version

The data quality score associated with each severity record is governed by the versioned weights file at `packages/pipeline/methodology/data_quality_weights_v1.json` (version 1.0.0, effective 2026-05-09). Sections 4 and 5 reference this file's existence and version; Section 8 publishes the full weight set and documents each sub-score's definition. Any change to the weighting scheme requires a methodology version bump, a changelog entry, and a 60-day advance notification to Scale and Enterprise subscribers per Section 11.

4.6 References

1. Source 2: ASOP No. 25, Credibility Procedures (Revised Edition, Actuarial Standards Board, adopted December 2013). Operative sections: credibility threshold documentation, purpose-appropriate threshold selection.
2. Source 4: ASOP No. 56, Modeling (Actuarial Standards Board, adopted December 2019, effective October 1, 2020). Operative sections: model validation evidence, input-range documentation, violation logging.
3. Source 6: CAS Statement of Principles Regarding Property and Casualty Insurance Ratemaking (Casualty Actuarial Society, reinstated May 2021). Operative sections: reasonableness criterion for severity estimation methods.
4. Source 7: Werner, G. and Modlin, C., Basic Ratemaking, Fifth Edition (Casualty Actuarial Society study note, May 2016). Operative sections: Chapter 3 (data organization), Chapter 12 (credibility procedures, complement of credibility).

5. Confidence Interval Methodology

5.1 Purpose

A severity percentile without a confidence interval is a point estimate: it tells the underwriter where the center or tail of the observed distribution sits but says nothing about how much sampling uncertainty surrounds that estimate. A 75th-percentile charge of $12,400 derived from 32 observations carries far more uncertainty than the same estimate derived from 320. Confidence intervals make that distinction explicit. An underwriter using our data to set a stop-loss attachment point or to validate internal pricing assumptions needs to know the precision of the estimate, not just its value.

We compute 95 percent confidence intervals at the default confidence level across all regimes where computation is defensible. The `confidence_level` field carries the value `0.95` for all records where computation runs. The level is not configurable at v1.

5.2 Bootstrap Method: BCa with 10,000 Resamples

For cohorts with n >= 30 observations (the empirical-quantile regime), we use the bias-corrected and accelerated bootstrap, commonly referred to as BCa. The BCa method was established as the actuarial standard for risk-measure confidence intervals in insurance loss modeling by Gruen and Miljkovic (North American Actuarial Journal, 2022, Source 15). The NAAJ is the primary joint research journal of the Society of Actuaries and the Canadian Institute of Actuaries; a paper published there carries peer-reviewed actuarial authority that practitioners at specialty payers and MGAs will recognize.

The BCa method corrects the plain percentile-bootstrap in two ways. The bias-correction term adjusts for the fact that the bootstrap distribution of a statistic is often not centered on the statistic's observed value, particularly for skewed distributions. The acceleration term adjusts for the fact that the standard deviation of the bootstrap distribution may itself vary with the parameter value, which is common in quantile estimation. Healthcare procedure prices are characteristically right-skewed: a small number of high-cost outlier events (complex cases, rare complications, facility-specific pricing agreements) pull the upper tail above what a symmetric or mildly skewed distribution would predict. At the sample sizes typical for our procedure-metro-payer grain (n between 30 and several hundred for most

urban markets, n between 30 and 50 for smaller metros), plain percentile-bootstrap undercovers the true parameter at high quantiles when skewness is present. BCa restores correct coverage.

We set the bootstrap resample count at 10,000 iterations. DiCiccio and Efron (1996), cited within Source 15, recommend a minimum of 2,000 resamples and describe 10,000 as standard for tail-percentile stability. Because we compute confidence intervals on p95 and p99 in addition to the mean and the quartiles, tail stability is material for our use case. A resample count of 1,000, which the prior pipeline version used, is not defensible for 95th-percentile interval estimation at a 95 percent confidence level: Monte Carlo sampling error in the bootstrap quantile becomes non-trivial relative to the interval width. ASOP 56 (Source 4) requires that model output be validated against its intended purpose; 10,000 resamples satisfies that requirement for tail-stability validation where 1,000 does not.

The BCa implementation lives in `packages/pipeline/scripts/bca_bootstrap.py` as the sole authoritative BCa primitive for the pipeline. The `bca_bootstrap_ci` function accepts a `statistic` callable so the same implementation runs for the mean CI and for each percentile CI without code duplication. The jackknife acceleration term is computed per the Efron and Tibshirani (1993) canonical formula.

5.3 Scope of BCa Coverage: Mean and Percentile CIs

We compute BCa confidence intervals on five statistics for every cohort in the n >= 30 regime: the mean, the 25th percentile, the 75th percentile, the 95th percentile, and the 99th percentile. This produces eight additional percentile-CI fields on every qualifying `pricing_record`:

- `confidence_lower_p25` and `confidence_upper_p25`: 95 percent BCa CI on the 25th percentile.
- `confidence_lower_p75` and `confidence_upper_p75`: 95 percent BCa CI on the 75th percentile.
- `confidence_lower_p95` and `confidence_upper_p95`: 95 percent BCa CI on the 95th percentile.
- `confidence_lower_p99` and `confidence_upper_p99`: 95 percent BCa CI on the 99th percentile.

The existing `confidence_lower` and `confidence_upper` fields carry the BCa CI on the mean, preserving backwards compatibility. Together these ten fields give a practitioner actuary a complete picture of uncertainty across the distributional shape. The BCa CI on p99 is particularly useful for stop-loss layer analysis, where knowing whether the 99th-percentile estimate is stable or highly uncertain changes the attachment-point recommendation materially.

Because these are empirical BCa intervals, every percentile CI is bounded by the observed data: the bootstrap resamples only from values actually present in the cohort, so an upper bound on p95 or p99 can reach but never exceed the cohort's observed maximum. We do not fit a parametric tail (lognormal, Pareto, or generalized extreme value) and extrapolate beyond the observed range. A practitioner who needs tail behavior past the observed maximum, for example a high-excess stop-loss layer, should treat our p99 CI as a measure of sampling uncertainty within the observed range, not as a fitted extreme-value estimate.

Every lower bound across all five statistics is clamped to zero before storage and before API response delivery. Price lower bounds cannot be negative; any bootstrap draw that produces a negative lower bound is an artifact of the resampling distribution on a small but credible sample, and clamping restores the physical constraint without altering the upper bound.

5.4 Small-Sample t-Interval Regime (`10 ⩽ n < 30`)

For cohorts with between 10 and 29 observations, we compute a confidence interval on the mean only, using a Student t interval: `CI = (mean - t * stdev / sqrt(n), mean + t * stdev / sqrt(n))`, where `t` is the 0.975 quantile of the Student t distribution with `n - 1` degrees of freedom (for example, t = 2.262 at n = 10 and t = 2.048 at n = 29). The lower bound is clamped to zero. We use the t multiplier rather than the normal `z = 1.9600` because at these sample sizes the normal approximation understates the width of the mean interval for the right-skewed cost distributions in our data. The t interval is the standard small-sample correction and is strictly wider than the z interval, so it does not overstate precision; as n approaches 30 the two converge.

We do not compute BCa percentile confidence intervals in this regime. Bootstrapping a percentile from fewer than 30 observations produces unstable tail coverage: with n = 15, a bootstrap CI on the 75th percentile can vary by 30 to 50 percent of its width across

independent runs with the same underlying data, depending on the random resample sequence. That level of instability is not defensible for rate-filing use per ASOP 25, which requires that credibility procedures be appropriate for the intended purpose. An analytic t-interval CI on the mean is more robust at small n than a bootstrap CI on a tail percentile, because the mean's sampling distribution converges to normality faster than a quantile's. We therefore provide the mean CI and suppress the percentile CIs at this regime, setting `confidence_lower_p25`, `confidence_upper_p25`, and all analogous fields to `null`.

The lower-bound clamping rationale for the small-sample regime mirrors the reasoning in Section 4.2: price distributions are truncated at zero by construction. Clamping the lower bound introduces a small upward bias in the lower limit when the mean is close to zero relative to the standard deviation, but this is standard practice in actuarial treatment of truncated-normal severity distributions and is substantially preferable to returning a meaningless negative price bound.

5.5 Suppression Regime (`n < 10`)

For cohorts with fewer than 10 observations, all confidence interval fields are `null`. `confidence_lower`, `confidence_upper`, and all eight percentile-CI fields are suppressed, and `confidence_level` is `null`. The `quality_flag` field is set to `low_sample` for these records. This behavior is consistent with the severity suppression described in Section 4.2: a cohort with fewer than 10 observations does not have a statistically defensible severity estimate, and a CI around a non-existent point estimate is meaningless.

5.6 Identifying the CI Method for a Record

The v1 schema does not carry a separate method field. The method that produced any record's confidence interval is determined unambiguously by the `sample_size` field together with which CI fields are populated:

- `sample_size ⩾ 30`, with the eight percentile-CI fields populated: BCa bootstrap on the mean and the four percentiles.
- `10 ⩽ sample_size < 30`, with `confidence_lower` and `confidence_upper` populated and the percentile-CI fields `null`: the small-sample t-interval on the mean only.
- `sample_size < 10`, with all CI fields `null` and `quality_flag = low_sample`: suppressed.

This serves the same verification purpose. A practitioner actuary using our data in a rate-filing context can read `sample_size` on every record and confirm that the CI fields present match the method this white-paper describes for that regime. If a record carries populated percentile CIs but a `sample_size` below 30, that is a pipeline inconsistency the actuary should flag to `api-support@procedureradar.com`.

5.7 Wilson Interval Scope

The Wilson interval (Wilson, 1927, for binomial proportions) is used exclusively for proportion-like metrics in our pipeline: coverage rates (the share of expected procedures for which a hospital has disclosed prices) and shoppable-service rates (the share of records that pass the shoppable-service filter). These are count-of-events-out-of-trials statistics with binomial shape, where the Wilson interval's coverage properties are better calibrated than a normal approximation at small n and extreme proportions.

The Wilson interval is not used for price confidence intervals. Procedure prices are not proportions. They are continuous positive random variables with right-skewed empirical distributions. Applying a binomial interval to a continuous severity measure would produce intervals with incorrect coverage properties and no actuarial justification. The BCa bootstrap and normal approximation described above are the appropriate methods for price CIs, and we apply them exclusively.

5.8 Relationship to ASOP 56 Model Validation

ASOP 56 (Source 4) requires that we validate the confidence interval model against its intended purpose and document the range of inputs for which it is valid. The three-regime structure is itself the validation boundary map: the model is valid for `n ⩾ 10` (with method-appropriate caveats), and invalid (suppressed) for `n < 10`. Within the valid range, the method selection (BCa vs. small-sample t-interval) is further calibrated to the sub-regime where each method's coverage properties are defensible.

Post-recompute verification of the BCa regime checks three invariants: (1) every record with `sample_size ⩾ 30` has its eight percentile-CI fields and its mean CI populated; (2) every record with `10 ⩽ sample_size < 30` has its mean CI populated and its percentile-CI fields `null`; (3) CI lower bounds are non-negative and `confidence_lower ⩽ confidence_upper` for all non-null CI fields. These checks constitute the post-apply verification step that ASOP 56 requires for model validation evidence.

5.9 References

1. Source 2: ASOP No. 25, Credibility Procedures (Revised Edition, Actuarial Standards Board, adopted December 2013). Operative sections: method appropriateness for intended purpose, documentation of credibility threshold selection.
2. Source 4: ASOP No. 56, Modeling (Actuarial Standards Board, adopted December 2019, effective October 1, 2020). Operative sections: model validation evidence, valid input range documentation, violation detection.
3. Source 6: CAS Statement of Principles Regarding Property and Casualty Insurance Ratemaking (Casualty Actuarial Society, reinstated May 2021). Operative sections: reasonableness criterion.
4. Source 7: Werner, G. and Modlin, C., Basic Ratemaking, Fifth Edition (Casualty Actuarial Society study note, May 2016). Operative sections: Chapter 12, credibility procedures and complement-of-credibility framework.
5. Source 15: Gruen, B. and Miljkovic, T., "The Automated Bias-Corrected and Accelerated Bootstrap Confidence Intervals for Risk Measures," North American Actuarial Journal, Vol. 27, No. 4, pp. 731–750 (2023, published online December 2, 2022). Operative sections: BCa as actuarial standard for risk-measure confidence intervals; DiCiccio and Efron (1996) citation for 10,000–resample lower bound; BCa superiority over plain percentile-bootstrap for skewed and small-to-moderate samples.

6. Time-Series Normalization

6.1 Purpose

A hospital charge rate recorded in January 2024 and a charge rate recorded in January 2026 are not directly comparable in dollar terms: general medical price inflation, changes in facility contracting practice, and regional supply-demand shifts all move observed charges over time. An underwriter using our data to set a rate for a policy year beginning in 2027 needs to be able to adjust historical pricing records to a consistent dollar basis and to assess the trajectory of prices in the relevant market. The time-series normalization layer serves that function.

We do this in two parts. First, we apply a temporal deflator that adjusts each historical pricing observation to a reference index month, allowing the underwriter to compare observations from different calendar periods in real medical-price-adjusted terms. Second, we apply a regional price index that adjusts for the fact that medical-procedure costs in San Francisco are structurally higher than costs in rural Mississippi, independent of temporal inflation. The two adjustments are multiplicative and can be applied separately or together, depending on the analytical use case.

This section documents the deflator and regional index choices, the rationale behind each choice, the limitations of those choices, and the append-only data model that guarantees no look-ahead bias in any historical comparison.

6.2 Primary Deflator: CUUR0000SAM (BLS Medical Care Services)

The primary temporal deflator we apply is the Bureau of Labor Statistics Consumer Price Index series CUUR0000SAM, officially titled "Medical Care Services" and published as part of the CPI-U (all urban consumers) family. This series has several properties that make it our first-choice deflator.

CUUR0000SAM is published monthly by a primary federal statistical authority with a long unrevised history. It is the most widely cited medical price index in actuarial rate-filing contexts for specialty-health and MGA product lines. It is publicly available at no cost from the BLS Data API, aligns naturally with our monthly pipeline cadence (one index observation per month, matching our snapshot frequency), and does not require licensing or data-purchase agreements. For practitioners preparing rate filings, CUUR0000SAM is a recognized reference that does not require explanation or justification to a state insurance

department reviewer. These properties collectively satisfy what the CAS Statement of Principles (Source 6) means by "reasonable": a trend factor produced by a recognized, reproducible, publicly available series applied consistently.

The adjusted price columns on every `pricing_record` are: `cpi_adjusted_price` (the gross-charge or negotiated-rate value deflated from the observation month to the `reference_index_month`), `mcpi_adjusted_price` (the same value expressed in Medical Care Services index-adjusted terms specifically, as distinct from the All Items CPI-U), and `reference_index_month` (the ISO-8601 month string to which the adjustment is anchored, matching the pipeline run's reference month).

6.3 The October 2024 CUUR0000SAM Methodology Change and the Charge-Side Mismatch

We are required to document a material change to CUUR0000SAM that took effect in October 2024, and to explain why we continue to use the series despite that change.

The BLS published a methodology update to the outpatient hospital services sub-component of CUUR0000SAM in October 2024, announced in a 2023 Monthly Labor Review article and documented in the BLS Medical Care factsheet (Source 13). The change switched the private-insurance outpatient component of the index from traditionally collected survey data (which was based on list charges submitted by providers) to medical claims data sourced from a national health insurance aggregator. As a result, the post-October-2024 index for this sub-component reflects adjudicated, reimbursed prices that have actually been paid by private insurers, not the gross charges or negotiated rates as hospitals initially publish them.

Our data is predominantly charge-side. The Machine-Readable Files that hospitals publish under 45 CFR Part 180 contain gross charges, discounted cash prices, payer-specific negotiated charges, and de-identified minimum and maximum negotiated charges. These are prices at or near the pre-adjudication stage. Even the negotiated rates in MRFs represent the contracted rate, not necessarily what was actually remitted after claim adjudication, outlier adjustments, carve-outs, or bundled payment reconciliation. Since October 2024, CUUR0000SAM's outpatient private-insurance component has been measuring something fundamentally different from what our input data represents: reimbursed prices rather than standard charges.

This is a material conceptual mismatch and we name it directly. Applying a reimbursement-based deflator to charge-based input data can introduce a systematic directional bias when reimbursement prices and charge prices move at different rates, as they do when payer negotiating leverage shifts or when hospital list-charge inflation diverges from adjudicated-rate inflation. The October 2024 methodology change means that post-October-2024 CUUR0000SAM values for outpatient private-insurance partially reflect claims-side price movements, while our data for the same period remains charge-side.

We continue to use CUUR0000SAM as our primary deflator for the following specific reasons.

First, no publicly available monthly medical price series avoids this problem entirely. The alternative most directly recommended by the peer-reviewed literature for total medical expenditures is the PCE Personal Health Care (PHC) index published by the Bureau of Economic Analysis, discussed below. However, the PCE PHC index is published quarterly, not monthly, which misaligns with our monthly pipeline cadence and would require interpolation or pipeline-cadence restructuring.

Second, CUUR0000SAM remains the index that rate-filing actuaries at specialty payers and MGAs recognize as the standard reference. Switching to a less familiar index introduces its own audit risk: a state insurance department reviewer encountering PCE PHC as a deflator in a rate-filing will want a justification, whereas CUUR0000SAM requires none.

Third, the mismatch between our data and the October 2024 index change is not a disqualifying error; it is a documented limitation that the practitioner actuary can account for. When charge prices and reimbursement prices move in approximate parallel, the deflation error is small. When they diverge materially, our time-series output carries a corresponding uncertainty that the actuary should flag and, if necessary, supplement with their own trend analysis.

We also apply CUUR0000SAM uniformly across all charge types in the MRF: gross charges, discounted cash prices, and payer-specific negotiated rates. Dunn, Hale, and Dauda (2018, Source 14), in the primary peer-reviewed reference for deflator selection in health services research, recommend using the CPI medical care index specifically for consumer out-of-pocket expenditures, and the PCE PHC index for total medical expenditures including insurer payments. Our data blends both charge types, and a rigorous implementation would apply

the CPI medical care index to cash and self-pay records and the PCE PHC index to payer-negotiated records. We apply a single deflator across both types as a pragmatic choice given data availability and pipeline architecture. Applying a blended or charge-type-specific deflator is a candidate for a future methodology version.

The secondary reference series, `CUUR0000SA0` (All Items CPI-U), is included as a reference benchmark and for pipelines that need to express our medical price adjustments in the context of broader consumer inflation. It is not used as the operative deflator for any pricing-record adjustment; CUUR0000SAM is always the primary.

6.4 Regional Price Indexing

National CPI adjustment addresses temporal variation but not geographic variation. The same knee replacement procedure may carry a structurally higher charge in the New York metro area than in the Memphis metro area for reasons unrelated to temporal inflation: regional labor markets, facility density, local competition among payers, and geographic differences in contracting norms all drive persistent cross-sectional price differences. An underwriter pricing a specialty-health product covering lives in two different markets needs to account for this geographic dimension in addition to the temporal one.

We address this through a regional price index derived from the BLS's metro-area-level CPI sub-indexes. The BLS publishes CPI series for a set of metropolitan statistical areas with sufficient household survey samples to support local indexing. We maintain a mapping table in `packages/pipeline/` that associates each of our internal metro slugs with the nearest applicable BLS area code. For metro slugs that do not have a direct BLS area match (typically smaller markets), we use the closest regional division or census division index available.

The `regional_price_index` column on each `pricing_record` stores the ratio of the metro-specific index to the national index for the same reference period, expressed as a decimal (for example, a value of 1.18 indicates that the metro's medical price level was 18 percent above the national index in the reference period). This value is computed once per metro per pipeline run and is constant across all records within a metro for a given snapshot. It is not a record-level estimate; it is a metro-level structural adjustment.

The mapping from our metro slugs to BLS area codes, and the specific BLS series identifiers used for each metro, are documented in the pipeline codebase. The white-paper describes the methodology; the mapping itself is proprietary and is not published here. Practitioners

who need to verify a specific metro's index value should contact `api-support@procedureradar.com`.

6.5 The pricing_history Table and the No-Look-Ahead-Bias Guarantee

Actuarial time-series analysis is only credible if the historical data is immutable: a snapshot labeled "March 2025" must contain only information that was observable as of March 2025, not values that were subsequently revised or re-computed with information available later. This property, the absence of look-ahead bias, is a prerequisite for any rate-filing use of our time-series output.

We guarantee this property through the architecture of the `pricing_history` table, introduced in migration 015. The table captures monthly immutable snapshots: one row per `(pricing_record_id, observed_at)` combination, where `observed_at` is the ISO-8601 month string for the pipeline run that captured the snapshot. The table is append-only. No row is ever updated or deleted in the normal course of pipeline operation. When a hospital's MRF is re-ingested in April 2026, the new data creates new April 2026 snapshot rows alongside the existing March 2025 rows; the March 2025 rows are untouched.

Trailing-12–month and year-over-year comparisons are served from `pricing_history` joined against `pricing_records`. Because the table is append-only and keyed on `observed_at`, a query for "price as of month M" is expressed as a filter on `observed_at ⩽ M`, which returns only snapshots that existed at or before that point in time. There is no structural path by which a future data update can contaminate a historical snapshot. The `reference_index_month` column on each record anchors the BLS index value used at the time of the snapshot, so the deflation is also point-in-time: the March 2025 snapshot was deflated using the March 2025 BLS index, regardless of subsequent BLS revisions.

This design directly addresses the ASOP 25 requirement (Source 2) that credibility procedures be applied to data that is appropriate for the intended purpose. A rate-filing underwriter using our trailing-12–month comparison needs confidence that the comparison is between two genuinely contemporaneous price observations, not between a current price and a retrospectively revised one. The `pricing_history` append-only model provides that confidence as a structural guarantee, not a pipeline discipline.

6.6 Known Deflator Limitation: The PCE Personal Health Care Alternative

The Dunn et al. (2018, Source 14) paper is the primary peer-reviewed reference for deflator selection in health services research and we cite it because it identifies a limitation in our current approach. Dunn and co-authors find that for research purposes involving total medical expenditures, including insurer-paid amounts, the PCE Personal Health Care index (a BEA product) outperforms the CPI medical care index in terms of conceptual alignment with the price changes actually affecting the total cost of care. The paper frames the choice as use-case dependent: CPI medical care is the better choice for consumer out-of-pocket analysis; PCE PHC is the better choice for total-expenditure analysis that includes insurer payments.

Our pipeline includes payer-specific negotiated rates, which are insurer-payment-side prices. This means PCE PHC is conceptually closer to the right deflator for a material share of our data. We document this as a known limitation and a candidate improvement for a future methodology version rather than a deficiency that undermines the current output. The practical impact of the choice depends on how much CPI medical care and PCE PHC diverge in a given period; historically, the two series have tracked reasonably closely, and the deflation error from using CPI medical care on insurer-payment-side records is not expected to be material in most underwriting contexts. An actuary relying on our time-series for a rate filing should note the deflator choice and apply their own judgment about whether supplemental adjustment is warranted for their specific use case.

6.7 References

1. Source 2: ASOP No. 25, Credibility Procedures (Revised Edition, Actuarial Standards Board, adopted December 2013). Operative sections: data appropriateness for intended purpose, credibility procedure documentation.
2. Source 4: ASOP No. 56, Modeling (Actuarial Standards Board, adopted December 2019, effective October 1, 2020). Operative sections: model validation evidence, documentation of limitations.
3. Source 6: CAS Statement of Principles Regarding Property and Casualty Insurance Ratemaking (Casualty Actuarial Society, reinstated May 2021). Operative sections: reasonableness criterion for trend factors; data homogeneity requirements.
4. Source 13: Bureau of Labor Statistics, "How BLS Measures Price Change for Medical Care Services in the Consumer Price Index" (factsheet, current); Bureau of Labor Statistics, "Incorporating Medical Claims Data in the CPI," Monthly Labor Review (2023). Operative sections: October 2024 CUUR0000SAM outpatient methodology change, claims-data switch for private-insurance component.
5. Source 14: Dunn, A., Hale, E., and Dauda, B., "Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States," Health Services Research, Vol. 53, No. 1, pp. 526–548 (2018, published online February 2018). Operative sections: CPI medical care vs. PCE PHC

deflator selection by expenditure type; no single gold standard for medical inflation adjustment.

7. Covered-Event Taxonomy

7.1 Purpose

A specialty-health underwriter pricing a stop-loss policy or an MGA quoting a self-funded plan does not typically work at the individual line-item level. The underwriter thinks in terms of episodes: a knee replacement, a cesarean delivery, a coronary artery bypass graft. Each of those clinical events involves multiple discrete billing line items, and the cost of the event is the sum of those line items weighted by the probability that each occurs in a typical instance of the episode. Pricing at the line-item grain alone forces the underwriter to reassemble the episode basket manually, which is error-prone and inconsistent across different users of our data.

The covered-event taxonomy provides the episode abstraction layer. It maps individual procedure billing codes to covered-event categories, assigns a `mapping_weight` to each procedure reflecting its fractional contribution to the typical episode cost, and exposes the full basket via the `bundled_components` JSONB field on the `covered_events` table. An Enterprise-tier consumer can query "knee replacement episode" and receive the aggregate pricing for the basket, not just for the arthroscopic procedure code.

This section documents the 10-category taxonomy, the definitional basis for each category, the `mapping_weight` design, and the scope limitations that the practitioner actuary should understand before using covered-event data in a rate-filing context.

7.2 The 10-Category Taxonomy

The `covered_events.category` column is an enum with ten values, introduced in migration 017. The categories and their definitional basis are as follows.

`outpatient_imaging`: radiological and diagnostic imaging services delivered in an outpatient setting, including CT, MRI, PET, ultrasound, and plain-film radiography. Category membership is determined by the procedure's primary billing code mapping to a radiology or imaging revenue code group. The outpatient setting filter excludes imaging performed as part of an inpatient admission.

`outpatient_surgery`: surgical procedures performed on an ambulatory basis, including arthroscopic procedures, laparoscopic procedures, cataract extraction, and similar interventions not requiring an inpatient admission. The definitional basis for this category

draws on the Ambulatory Payment Classification (APC) grouper methodology maintained by CMS, which organizes outpatient surgical procedures into clinically coherent payment groups. The Truven Medical Episode Grouper (now IBM MarketScan), as documented in its published methodology, provides additional definitional guidance for episode-level outpatient surgical grouping.

`inpatient_surgery`: surgical procedures requiring at least one inpatient overnight stay, organized by CMS Medicare Severity Diagnosis Related Group (MS-DRG) rollup. MS-DRGs are the primary federal inpatient episode grouper used in Medicare payment and, by convention, in specialty-health and MGA underwriting. Our `inpatient_surgery` category maps to the surgical MS-DRG range in the CMS MS-DRG definitions (Version 41, effective October 1, 2023), specifically the MDC-partitioned surgical groups. The MS-DRG definitions are publicly available from CMS at no charge and are the industry standard for inpatient episode classification; no licensing arrangement is required to cite them as a definitional basis.

`obstetric`: pregnancy-related and delivery-related services, including vaginal delivery, cesarean section, antepartum care visits, and postpartum complications. The obstetric category maps to CMS MS-DRG Major Diagnostic Category 14 (Pregnancy, Childbirth, and the Puerperium), which defines the federal grouper basis for inpatient obstetric episodes. For outpatient antepartum services, the category boundary follows the ASPE HHS (2016) episode-based payment measurement methodology, which defined maternity episodes to include prenatal, delivery, and postpartum care within a defined episode window.

`diagnostic`: non-imaging diagnostic services, including laboratory panels, cardiac stress tests, electroencephalograms, and other diagnostic studies that do not involve a surgical intervention and do not produce an image as their primary output. The category boundary between `diagnostic` and `lab` is that `diagnostic` includes composite evaluation procedures (physician interpretation, waveform recording, multi-analyte panels where the clinical unit is the interpretive report), while `lab` covers individual analyte tests.

`preventive`: preventive services, screenings, and wellness visits as defined by the USPSTF recommendation grade A and B list and the ACA preventive care mandate. Category membership is assigned when the billing code is classified as preventive under the CMS HCPCS/CPT preventive-service groupings or the ACA Section 2713 required preventive services list.

`emergency`: emergency department visits and observation stays, regardless of ultimate disposition (discharge or inpatient admission). The category maps to the CMS Emergency Department evaluation and management code hierarchy and to observation-status codes as defined in the CMS Outpatient Prospective Payment System. Emergency-leading-to-admission episodes are classified under `emergency` for the initial ED encounter and under the appropriate inpatient category for the subsequent admission; the `bundled_components` JSONB field links the two event records.

`lab`: individual clinical laboratory tests at the analyte level, including chemistry panels broken into component CPT codes, complete blood counts, urinalysis, and similar single-analyte or simple-panel tests. The `lab` category does not include interpretation services or composite diagnostic panels; those belong to `diagnostic`.

`episode_bundle`: multi-component episodes that do not fit cleanly within a single category above, typically because they span both inpatient and outpatient settings or because they require aggregating services across a care continuum. Joint replacement episodes, cardiac catheterization with revascularization, and chemotherapy cycles are representative examples. The `episode_bundle` category is the primary vehicle for the ASPE HHS (2016) episode-based payment measurement framework, which defined episode windows and component service attribution rules for a set of high-cost procedures relevant to specialty-health underwriting.

`other`: a residual category for procedures that cannot be unambiguously classified into one of the nine categories above. The `other` category is not a quality-flag indicator; it reflects the reality that some billing codes sit at category boundaries or are not covered by the classification frameworks above. Records in `other` carry a `methodology_ref` field describing the classification rationale.

7.3 The mapping_weight Field

Each row in the `procedure_covered_event_map` table carries a `mapping_weight` in the interval (0, 1]. The weight expresses the fractional contribution of a single procedure to the aggregate cost of a covered-event episode. A mapping_weight of 1.0 means the procedure accounts for the entire episode cost in isolation (typical for single-procedure categories like

a standalone lab test classified under `lab`). A mapping_weight of 0.15 means the procedure accounts for approximately 15 percent of the typical episode cost, as would be the case for a pre-operative imaging study within a larger surgical bundle.

The ASPE HHS (2016) episode-based payment measurement work, produced by the Assistant Secretary for Planning and Evaluation under HHS, provides the methodological basis for fractional attribution within episode bundles. That work defined service windows, anchor events, and attributable services for a set of high-cost acute episodes relevant to specialty-health and MGA products. Our `mapping_weight` values for `episode_bundle` records derive from those fractional attribution ratios, adjusted where necessary to reflect the charge-side nature of our data rather than the claims-side data the ASPE work used.

For individual categories like `outpatient_imaging` and `lab`, where a single procedure typically constitutes the entire covered event, `mapping_weight` is 1.0 by convention. For `inpatient_surgery` and `obstetric` episodes, where the surgery or delivery is the anchor event but the episode includes pre-procedural workup, anesthesia, and post-operative recovery services, the anchor procedure carries the largest weight and the peripheral services carry fractional weights that sum to 1.0 across the full bundle.

The `mapping_weight` is a methodology-version-controlled field. Changes to weight values for any episode bundle constitute a methodology event and require a changelog entry and a 60-day advance notification to Scale and Enterprise subscribers, because weight changes affect the aggregate episode price returned by the API and can materially change a rate-filing calculation.

7.4 Custom Taxonomies for Custom-Tier Clients

The standard 10-category taxonomy described above serves Growth, Scale, and Enterprise clients. Custom-tier clients, who typically have a specific policy structure, a defined benefit schedule, or a proprietary episode grouper that differs from our standard taxonomy, may need a client-specific covered-event taxonomy that maps to their particular policy definitions.

Custom-tier covered-event taxonomies live in per-client side tables that are created and documented at contract time. The per-client taxonomy maps the client's proprietary episode definitions to our underlying procedure billing codes, with `mapping_weight` values

calibrated to the client's policy structure. These per-client tables are not published in this white-paper; they are documented in the client-specific methodology addendum that forms part of the Enterprise or Custom contract.

The underlying pricing data that populates a custom-tier taxonomy response is identical to the data serving all other tiers. Custom taxonomy is a different view onto the same data, not a different data source.

7.5 Limitations and Scope

The covered-event taxonomy is a modeling layer, not a clinical coding system. It is designed to serve the underwriting use case, not to replace or substitute for clinical coding systems such as ICD-10-CM diagnosis groupings or the Hierarchical Condition Category (HCC) model used in Medicare Advantage risk adjustment. An actuary using our covered-event output in a specialty-health rate filing should confirm that our episode definitions are consistent with the episode grouper their actuarial model assumes.

HCCI (Health Care Cost Institute) episode definitions were evaluated as a potential seed source for our taxonomy. Because HCCI episode definitions are subject to a licensing arrangement that we have not executed for v1, we do not reproduce HCCI episode definitions or weight tables in our taxonomy. Where HCCI methodology provides relevant definitional guidance, we reference the published methodology description rather than the proprietary definition set. The v1 taxonomy is instead anchored to the CMS MS-DRG framework (for inpatient) and the ASPE HHS (2016) episode-based payment framework (for outpatient and cross-setting bundles), both of which are publicly available without licensing. A future methodology version may incorporate HCCI-licensed definitions as an additional layer if the licensing economics support it.

The `other` category captures procedures that current classification rules cannot place. The share of procedures in `other` is monitored by the pipeline and flagged for review when it exceeds 5 percent of total procedure-event map rows, because a rising `other` share indicates that the classification rules need updating. The current `other` share is reported in the methodology changelog at each pipeline version.

7.6 References

1. Source 2: ASOP No. 25, Credibility Procedures (Revised Edition, Actuarial Standards Board, adopted December 2013). Operative sections: purpose-appropriate credibility application when representing data for a specific actuarial purpose.
2. Source 4: ASOP No. 56, Modeling (Actuarial Standards Board, adopted December 2019, effective October 1, 2020). Operative sections: model limitation documentation for episode-taxonomy modeling layer.
3. Source 5: ASOP No. 8, Regulatory Filings for Health Benefits (Revised March 2014). Operative sections: rate-filing documentation requirements when covered-event output supports a health insurance premium rate filing.
4. CMS MS-DRG Definitions, Version 41 (effective October 1, 2023). Operative sections: MS-DRG surgical groupings as definitional basis for `inpatient_surgery` and `episode_bundle` inpatient categories.
5. ASPE HHS (2016), "Episode-Based Payment: Measuring Episodes of Care" (Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, 2016). Operative sections: episode window definitions, anchor-event attribution, fractional-contribution methodology for cross-setting bundles.

8. Data Quality Score

8.1 Purpose

Every `pricing_record` in our database carries five or more sub-score fields and a `quality_flag` enum. A practitioner actuary consuming our data in a rate filing does not want to parse five separate sub-scores to determine whether a given record is reliable enough to use; they want a single summary metric that they can apply as a threshold filter or a confidence weight. The `data_quality_score` is that metric: a single number in [0, 1] that aggregates six components of record quality into one practitioner-facing signal.

The score is not a quality-certification and is not a substitute for examining the underlying sub-scores when a record is near a threshold boundary. A score above 0.85 does not mean a record has no limitations; it means the record passes a multi-component quality screen with no single component pulling the composite below the high-confidence band. The practitioner actuary should treat the score as a first-pass filter and inspect the sub-score breakdown for records where the composite is borderline or where the intended use is sensitive to a specific quality dimension.

This section documents each of the six sub-score components, the published composite weights, the band interpretation, the relationship to the `quality_flag` enum, and the governance connection to the AAA Big Data brief (Source 8) and the NAIC CASTF white paper (Source 9).

8.2 The Six-Component Composite Formula

The composite `data_quality_score` is computed as a weighted sum of six sub-scores, each normalized to a 0–to-100 scale, then divided by 100 to produce a score in [0, 1]. The weights are published in `packages/pipeline/methodology/data_quality_weights_v1.json`, version 1.0.0, effective 2026–05–09. The six components and their weights are:

Component	Weight (out of 100)
`coverage`	35
`completeness`	22
`freshness`	13
`payer_count`	10
`quality_flag_verified_share`	10
`sample_size`	10

These weights sum to 100. The composite score in [0, 1] is then: `data_quality_score = sum(weight_i * sub_score_i) / (100 * scale_max)`, where `scale_max = 100` and each `sub_score_i` is pre-scaled to [0, scale_max].

The v0 weights, used in the pipeline before workstream 3, were: coverage 40, completeness 25, freshness 15, payer_count 10, quality_flag_verified_share 10, with no sample_size component. The v1 reweighting reduces the emphasis on coverage and completeness slightly to make room for the new `sample_size` component, which was absent from v0. The reweighting reflects a judgment that distributional reliability (how many observations underlie the severity computation) is as consequential for underwriting use as the breadth of charge-type coverage. The methodology version change from v0 to v1.0.0 is recorded in the `methodology_versions` table (migration 040) with effective date 2026-05-09.

8.3 Sub-Score Definitions

`coverage` (weight: 35). The `coverage` sub-score measures the share of expected shoppable procedures for which a hospital has disclosed at least one pricing record in its MRF. "Expected shoppable procedures" is defined as the union of all procedure billing codes that appear in any MRF for hospitals in the same metro area and facility type. A hospital that has disclosed pricing for 90 percent of the expected procedure set for its peer group receives a high coverage sub-score; a hospital that has disclosed pricing for 20 percent receives a low one.

Coverage is the highest-weighted component because it is the most direct measure of the hospital's compliance with the intent of 45 CFR Part 180. A hospital with high coverage is disclosing what it is required to disclose; a hospital with low coverage may have a non-compliant or incomplete MRF. This aligns with the GAO (Source 11) and OIG (Source 12) findings that MRF completeness is the primary documented deficiency across the hospital population.

The coverage sub-score is computed at the hospital level and is constant across all `pricing_records` for a given hospital in a given pipeline run. It is updated each time the hospital's MRF is re-ingested.

`completeness` (weight: 22). The `completeness` sub-score measures the share of expected charge-type fields populated on a specific pricing record. The four expected fields are: gross charge, discounted cash price, payer-specific negotiated rate (where a payer is identified), and de-identified minimum and maximum negotiated charges. A record that carries all four fields receives full completeness credit; a record that carries only a gross charge and no payer-specific data receives partial credit.

Completeness differs from coverage in grain: coverage is a hospital-level measure of which procedures are present, while completeness is a record-level measure of how completely each present record is populated. Both dimensions matter for underwriting use. A record with low completeness (for example, a record that has a gross charge but no negotiated rate and no cash price) has limited utility for a payer actuary whose model requires a payer-specific negotiated rate, even though the record technically exists.

`freshness` (weight: 13). The `freshness` sub-score measures the time elapsed since the hospital last published an updated MRF, bucketed into a monotonically decreasing score. A hospital that published its MRF within the last 90 days receives full freshness credit. The score declines in steps: 90–180 days since publication receives 80 percent credit, 180–365 days receives 50 percent credit, 365–730 days receives 20 percent credit, and beyond 730 days receives 0. These bucket definitions reflect a judgment that MRF data older than two years carries significant staleness risk for rate-filing use, given typical healthcare contracting cycle lengths of one to three years.

Freshness is a hospital-level sub-score, constant across all records for the hospital in a given pipeline run, because the freshness of a record depends on the hospital's publication cadence, not on the individual procedure.

`payer_count` (weight: 10). The `payer_count` sub-score measures the number of distinct payers with a disclosed negotiated rate for the specific procedure at the specific hospital. A higher payer count indicates that the hospital has disclosed richer payer-specific data, which is more useful to an actuary benchmarking specific plan types. The sub-score is log-scaled: a hospital-procedure combination with 1 payer receives minimal credit, 10 payers receives roughly half credit, 50 or more payers receives full credit. Logarithmic scaling reflects the diminishing marginal value of additional payer observations once the most common plan types are represented.

`quality_flag_verified_share` (weight: 10). The `quality_flag_verified_share` sub-score measures the share of records for the hospital that carry a `quality_flag` value of `verified`. A hospital where 95 percent of records are `verified` receives full credit on this component; a hospital where 40 percent of records carry `suspect`, `outlier`, or `placeholder_value` flags receives reduced credit. This component connects the composite score to the categorical `quality_flag` system described in Section 8.5.

`sample_size` (weight: 10). The `sample_size` sub-score measures the volume of line-item observations underlying the severity distribution computation for the record, using a logarithmic scale to avoid over-rewarding hospitals with very large record counts relative to hospitals with adequate but not extreme record counts. The formula is: `sample_size_score = min(scale_max * 0.10, log10(max(record_count, 1)) * 5.0)`.

Working through the formula: a hospital with 1 observation record receives a sub-score of 0 (log10(1) * 5.0 = 0). A hospital with 10 records receives a sub-score of 5.0 (log10(10) * 5.0 = 5.0). A hospital with 100 records receives a sub-score of 10.0 (log10(100) * 5.0 = 10.0, which is the maximum, since scale_max * 0.10 = 10.0). Any hospital with 100 or more records receives the maximum sub-score credit of 10.0, because the marginal reliability gain from additional observations beyond 100 is small relative to the gain in moving from 1 to 100 observations. The logarithmic scaling ensures that the sub-score grows quickly from 1 to 10 observations (where reliability is most sensitive to sample size) and saturates smoothly above 100.

The `sample_size` component was added in v1 to address a gap in the v0 composite: a record that was complete, fresh, and well-covered but derived from only 3 observations could receive a high v0 score despite having no credible severity distribution. The v1 composite penalizes this situation by reducing the score for records with very low observation counts, even when the other five components are strong. This aligns with the ASOP 25 (Source 2) principle that data credibility depends on sample volume as well as data completeness.

8.4 Band Interpretation for Practitioners

The composite `data_quality_score` maps to three interpretive bands that practitioners can use as decision rules.

Above 0.85 (high confidence). Records in this band have strong scores across all six components. The hospital has disclosed a broad procedure set, the specific record has comprehensive charge-type data, the MRF was recently published, the payer representation is adequate, the verified-flag share is high, and the observation count supports a credible severity distribution. These records can typically be used in a rate-filing context without additional cross-referencing, subject to the limitations noted in Section 10 of this white-paper.

0.70 to 0.85 (usable with normal caveats). Records in this band carry one or more components that reduce overall confidence, but not severely enough to disqualify the record for actuarial use. The practitioner actuary should examine the `quality_score_breakdown` JSONB field to identify which component is pulling the score below 0.85. A common pattern is a record with high coverage and completeness but a somewhat stale MRF (freshness pulling the composite toward 0.75) or a moderate payer count. Such records are typically usable for benchmarking but should be noted as carrying slightly elevated uncertainty in a rate-filing workpaper.

Below 0.70 (cross-reference before use). Records below 0.70 have at least one component with meaningfully low credit. These records should be cross-referenced against another source before being used as a primary input in a rate filing. The practitioner actuary should not assume that a score below 0.70 indicates erroneous data; a rural hospital with a small procedure set and a MRF that has not been updated in 18 months may receive a low score even if its disclosed prices are accurate. The score reflects structural limitations in the data, not necessarily pricing errors. Use the `quality_flag` enum alongside the score to distinguish between structural limitations (low coverage, staleness) and active quality problems (suspect prices, placeholder values, outliers).

The 0.85 and 0.70 thresholds were set by judgment calibrated against the sub-score weight distribution, not derived from an empirical backtest against an external ground truth. This is consistent with ASOP 41 (Source 3) guidance that methodology documentation should state explicitly when calibration is judgment-based. The thresholds will be reviewed and potentially revised on each major methodology version bump, using the `quality_flag_verified_share` data accumulated across pipeline runs as a calibration input.

8.5 Relationship to the quality_flag Enum

The `quality_flag` field on `pricing_records` is a categorical enum with eight values: `verified`, `suspect`, `outlier`, `missing_data`, `placeholder_value`, `invalid_price`, `stale_data`, and `percentile_invalid`. This enum predates the numeric `data_quality_score` and is preserved for two reasons: backwards compatibility for API consumers who have already built filters against the categorical values, and granular bucketed filtering that the numeric score does not support.

The relationship between the enum and the numeric score is additive, not hierarchical. A record with `quality_flag = 'suspect'` will tend to have a lower `quality_flag_verified_share` at the hospital level (reducing that component of the composite), but the numeric score may still be above 0.70 if the other five components are strong. Conversely, a record with `quality_flag = 'verified'` can still receive a low composite score if the hospital's coverage is limited, the MRF is stale, or the observation count is very low.

The enum provides the best tool for records that have been positively identified as problematic: `invalid_price` (a price value that failed our plausibility checks), `placeholder_value` (a value that matches known placeholder patterns like $1.00 or $0.01 where a real price is expected), or `percentile_invalid` (a monotonicity violation that passed the database constraint check but was flagged by a secondary pipeline audit). These categorical flags are actionable: a practitioner who wants to exclude all records with any identified quality problem should filter on `quality_flag NOT IN ('invalid_price', 'placeholder_value', 'percentile_invalid', 'suspect')`.

The numeric score is the better tool for records where the quality question is not about a specific identified problem but about overall reliability in a rate-filing context. The practitioner who wants to include only records that are fresh, comprehensive, and well-supported by observations should use `data_quality_score ⩾ 0.85`. The two systems work together; neither is a substitute for the other.

8.6 Governance Connection: AAA Big Data Brief and NAIC CASTF White Paper

The American Academy of Actuaries issued an issue brief on big data and algorithms in actuarial modeling (Source 8) that defines four governance requirements for external data sources used in actuarial models. Our data quality score architecture is designed to operationalize each of those four requirements.

The first requirement is validation of external data sources for quality. Our quality-gate architecture addresses this at the point of ingestion: every pricing record passes through a shoppable-filter, a placeholder-detection check, a plausibility check on price values, and a consistency check against the hospital's historical pricing pattern. Records that fail any gate

are either flagged as `suspect`, `placeholder_value`, or `invalid_price`, or are excluded from the ingested set entirely. The quality gates are not a single point of review; they are a multi-stage pipeline with separate checks at the parse, normalize, and load stages.

The second requirement is a review of third-party data collection procedures. Our Ingestion and Parsing section (Section 3 of this white-paper) documents the collection procedures for each upstream data source: hospital MRFs, BLS CPI series, NPPES, and TPAFS. The documentation covers the source identity, the collection cadence, the known limitations of each source, and the adjustments we apply. This constitutes the third-party data collection review that the Academy brief requires.

The third requirement is quality assurance guidelines for external data sources. The `data_quality_score` composite formula, together with the `quality_flag` enum and the sub-score breakdown in `quality_score_breakdown` JSONB, is the operationalized quality assurance guideline that a practitioner can apply to any record from any hospital. The score quantifies the multi-dimensional quality assessment that an actuary would otherwise have to perform manually by inspecting each sub-score in isolation.

The fourth requirement is analysis of data for compliance with regulatory requirements. Our coverage sub-score is explicitly measuring compliance: a low coverage score for a hospital indicates that the hospital's MRF does not include the full set of shoppable procedures it is required to disclose under 45 CFR Part 180. This compliance dimension is not a post-hoc observation; it is baked into the highest-weighted component of the composite score, so that compliance deficiencies propagate directly and proportionately into the composite.

The NAIC Casualty Actuarial and Statistical Task Force white paper on regulatory review of predictive models (Source 9) defines 79 information elements that state regulators assess when reviewing a rate filing that incorporates predictive model output. The `data_quality_score` is designed to help practitioners prepare for that review, not to substitute for it. Specifically, the score addresses several of the 79 elements that relate to data quality, validation evidence, and credibility documentation. However, the full 79–element review covers areas such as model governance, algorithmic fairness, state-specific compliance, and actuarial certification that the score does not address. A practitioner actuary using our data in a rate filing should complete the full NAIC regulatory review process

for their jurisdiction; the `data_quality_score` is a practitioner convenience metric that accelerates the data-quality portion of that review, not a certification that the filing is regulatorily complete.

8.7 ASOP 23 Compliance Connection

ASOP 23 (Source 1) requires that the actuary document, for each upstream data source, the source identity, the scope of the data, known biases, the adjustments applied, and the extent of reliance on data supplied by others. The `data_quality_score` and its sub-scores are not themselves the ASOP 23 documentation; they are a summary metric derived from that documentation. The ASOP 23 compliance layer for our pipeline is the full set of quality gates, flag definitions, and sub-score formulas documented in this section, together with the ingestion documentation in Section 3 and the limitations documentation in Section 10.

An actuary relying on our data in a rate filing is relying on data supplied by us, which in turn was supplied by the hospitals in their MRFs. The extent of reliance on our data is what the practitioner documents in their work product, citing this white-paper as the source of the methodology description. The `data_quality_score` is the mechanism by which the practitioner communicates to downstream reviewers that the data they selected for use passed a multi-component quality screen defined in a published methodology document.

8.8 References

1. Source 1: ASOP No. 23, Data Quality (Revised Edition, Actuarial Standards Board, adopted December 2016, effective April 1, 2017). Operative sections: source identification, scope limitation, bias documentation, adjustment documentation, extent-of-reliance disclosure.
2. Source 2: ASOP No. 25, Credibility Procedures (Revised Edition, Actuarial Standards Board, adopted December 2013). Operative sections: sample-size contribution to credibility; appropriate threshold selection for intended use.
3. Source 3: ASOP No. 41, Actuarial Communications (Revised Edition, effective May 1, 2011). Operative sections: disclosure of judgment-based calibration; clarity for peer actuary review.
4. Source 4: ASOP No. 56, Modeling (Actuarial Standards Board, adopted December 2019, effective October 1, 2020). Operative sections: model validation evidence, input-range documentation, material limitation disclosure.
5. Source 8: American Academy of Actuaries, "Big Data and Algorithms in Actuarial Modeling and Consumer Impacts," Data Science and Analytics Committee Issue Brief (October 2022). Operative sections: four-point governance framework for external data sources; external data validation, third-party data collection review, quality assurance guidelines, regulatory compliance analysis.
6. Source 9: NAIC Casualty Actuarial and Statistical Task Force, "Regulatory Review of Predictive Models" White Paper (adopted September 15, 2020; Model Review Manual updated November 2025). Operative sections: 79–element regulatory review framework; validation evidence requirements for predictive model

output in rate filings; data credibility and homogeneity documentation.

7. Source 11: U.S. Government Accountability Office, GAO-25-106995, "Health Care Transparency: CMS Needs More Information on Hospital Pricing Data Completeness and Accuracy" (released October 2, 2024). Operative sections: CMS inability to verify MRF completeness and accuracy; documented data quality deficiencies our quality gates address.
8. Source 12: HHS Office of Inspector General, "Not All Selected Hospitals Complied With the Hospital Price Transparency Rule" (November 2024). Operative sections: approximately 46 percent of hospitals non-compliant with one or more price transparency requirements; deficient machine-readable files as primary compliance gap.

9. Feature Gating by Tier

9.1 Depth, Not Data Quality

The tier separation in the ProcedureRadar B2B API is a feature-depth separation, not a data-quality separation. Every tier, from Starter to Custom, reads from the same underlying database tables and returns the same pipeline-derived numbers for the fields it exposes. A Starter customer and an Enterprise customer see the same median, the same gross charge, and the same cash price for the same procedure at the same hospital. The underlying data is identical; the response envelope is different.

This distinction matters for ASOP 41 (Source 3) compliance. The white-paper must not mislead a practitioner actuary about the scope of what any tier provides. A Starter customer who reads this white-paper should understand exactly what they receive and what they do not. An Enterprise customer should not conclude that lower tiers receive lower-quality data, because they do not. Tier separation governs which methodological features are included in the response, not which hospitals, procedures, or price values are disclosed.

9.2 Feature Availability by Tier

The following table describes which features are available at each tier.

| Feature | Starter | Growth | Scale | Enterprise | Custom | |---|---|---|---|---|---|---| | Median, gross charge, cash price, negotiated min and max | Yes | Yes | Yes | Yes | Yes | | Severity distribution (p25, p75, p95, p99, mean, stdev) | Yes | Yes | Yes | Yes | Yes | | Confidence intervals | No | Yes | Yes | Yes | Yes | | Time-series normalization | No | No | Yes | Yes | Yes | | Covered-event mapping | No | No | No | Yes | Yes | | Frequency modeling | No | No | No | Yes | Yes | | Actuarial review letter | No | No | No | Yes | Yes |

9.3 OMIT-if-Not-Entitled Semantics

Fields that a given tier does not entitle are absent from the API response entirely. They are not returned as null. This design prevents feature-set leakage (a lower-tier consumer deducing the existence of higher-tier fields from null placeholders) and keeps wire size proportional to the customer's tier. The single enforcement point for entitlement is `apps/web/lib/api/feature-gates.ts` via the `getAllowedFeatureSet(tier)` function. There is no secondary enforcement layer; every entitlement decision routes through that function.

9.4 Tier-to-Use-Case Mapping for Rate Filing

A practitioner actuary evaluating which tier to purchase for a specific rate-filing use case should consider what methodology elements their filing requires. The following guidance is intended to help match the tier to the use case.

Starter. Provides severity distributions (median, mean, percentile distribution) and raw charge-type columns. Does not provide confidence intervals, time-series normalization, covered-event mapping, or frequency modeling. Starter alone is generally not sufficient for actuarial rate-filing support without supplemental data sources. Severity distributions without confidence intervals do not give the actuary a measure of sampling uncertainty, and uncertainty is a required disclosure in ASOP 8 (Source 5) rate-filing workpapers when the underlying data is from a third-party source. Starter is appropriate for benchmarking, exploratory analysis, and consumer-facing applications where distributional precision is useful but not the primary actuarial deliverable.

Growth. Adds confidence intervals (BCa bootstrap for n greater than or equal to 30, normal approximation for 10 to 29 observations). This is the minimum tier that gives the practitioner actuary the uncertainty quantification needed to represent our data as a supporting input in a rate filing. Growth does not include time-series normalization, which means comparisons across calendar periods require the actuary to supply their own trend adjustment. Growth is appropriate for current-period benchmarking use cases where the actuary has their own trend methodology.

Scale. Adds time-series normalization. Scale is the first tier that provides a complete single-source underwriting data package for rate-filing contexts where the actuary needs both distributional uncertainty (CI) and temporal trend adjustment (deflated time-series). Scale is sufficient for most specialty-health and MGA rate-filing contexts that rely on charge-side data and do not require episode-level cost modeling. Scale receives the 60–day pre-notification commitment for methodology changes that affect rate-filing fields, described in Section 11.

Enterprise. Adds covered-event mapping, frequency modeling, and the actuary review letter. Enterprise is the tier for practitioners who are pricing episode-level benefits (knee replacement bundles, cardiac catheterization episodes, obstetric episodes) or who require a credentialed actuary's written review of the methodology elements relevant to their specific

filing. The actuary review letter is the human-actuary substitution layer for clients who need formal certification support beyond what this white-paper alone provides. Section 12 documents the review-letter commitment in detail.

Custom. Provides everything in Enterprise plus per-client covered-event taxonomy customization, white-label data delivery, and client-specific methodology addenda. Custom is appropriate for clients who have a proprietary episode-grouper, a defined benefit schedule, or policy structures that the standard 10–category taxonomy does not accommodate.

9.5 The Enterprise Actuary Review Letter

For Enterprise subscribers, ProcedureRadar provides a written review letter from a credentialed actuary covering the methodology elements relevant to the client's specific rate-filing use case. The review letter is the mechanism by which a qualified actuary's assessment is provided to support the practitioner's ASOP 8 (Source 5) certification when our data is the underlying source. ASOP 8 requires the certifying actuary to document their data sources and assumptions; when our white-paper alone is insufficient for that documentation (for example, when the client requires confirmation that the methodology is appropriate for a specific state's rate-review context), the actuary review letter provides that additional layer.

Pre-launch, we engage actuarial reviewers on a project basis. The review-letter commitment is a contractual deliverable at the Enterprise tier and is scoped to the methodology elements the client's filing requires, not to the full 12–section white-paper. Clients with narrowly scoped filings (for example, a filing that uses only our severity distributions and confidence intervals and does not use time-series or covered-event data) receive a correspondingly scoped review letter.

9.6 References

1. Source 3: ASOP No. 41, Actuarial Communications (Revised Edition, effective May 1, 2011). Operative sections: scope clarity; not misleading the reader about what a communication provides; disclosure of methods for peer actuary review.
2. Source 5: ASOP No. 8, Regulatory Filings for Health Benefits (Revised March 2014, effective September 1, 2014). Operative sections: rate-filing documentation requirements; actuary certification support when third-party data underlies the filing; supplemental documentation responsibility.

10. Known Limitations

This section is the transparency layer of the white-paper. Every limitation listed here was identified during the methodology design process or surfaced by primary federal research on hospital MRF data quality. We state each limitation plainly, give a magnitude where one is available, and describe what we do about it. The format follows the ASOP 23 (Source 1) requirement to disclose known limitations and the ASOP 41 (Source 3) requirement to include cautions regarding possible uncertainty in results.

10.1 Quantitative Coverage Summary (July 7, 2026 snapshot, superseded)

Snapshot dated July 7, 2026. Superseded. The figures in this subsection describe the methodology 1.0.0 data state on 2026-07-07 and are kept as a historical record. They are not current. The current figures, measured on 2026-11-07, are 2,176 hospitals with at least one published price on the site, and the cell-level freshness breakdown in the data quality disclosure on the methodology page. Where this subsection and that disclosure differ, the disclosure is the one to use.

As of the methodology version 1.0.0 data state (2026-07-07), the pipeline contained the following:

- **3,527 hospitals** in the database across **101 distinct metro area slugs**, covering the top 100 U.S. metropolitan markets.
- **15,849,196 pricing records** (a full-table exact count on 2026-07-07).
- **2,161 hospitals** (61.3 percent of the total hospital roster) have at least one pricing record. The remaining 1,366 hospitals are present in the database as geographic anchors but have not yet been successfully parsed from a machine-readable file, either because the hospital has not published a parseable MRF or because the parsing run has not yet completed for that hospital.
- **Provenance mix**: of the 15,849,196 records, approximately 63.1 percent are code-matched (the line item carried a billing code we mapped directly to a procedure) and approximately 36.9 percent are description-matched (the line item carried no usable code and was matched on its plain-English description, a weaker code-to-price linkage). The `data_source` field distinguishes the two on every row.

- **Quality flag distribution**: across the full production dataset of 15,849,196 pricing records, the `low_sample` flag (meaning the (hospital, procedure, payer, charge type, data source, setting) combination sits in a cohort with fewer than 10 observations and is therefore suppressed for distributional statistics) accounts for approximately 91.9 percent of records. The `verified` flag accounts for approximately 6.9 percent, `stale_data` for approximately 0.7 percent, and the remaining flags (`suspect`, `outlier`, `missing_data`, `placeholder_value`) for approximately 0.4 percent combined.
- **Staleness**: across the full production dataset, approximately 24.3 percent of records carry a `source_file_date` older than 180 days and an additional 8.5 percent carry a null `source_file_date`, leaving approximately 67.2 percent dated within the last 180 days. Because `source_file_date` records the publication date the hospital stamps on its own machine-readable file rather than our ingest date, this figure measures how recently each hospital refreshed its file, not how recently we parsed it; our pipeline re-parses the full roster monthly.

The practitioner actuary reading these numbers should understand the `low_sample` flag in context. Each `pricing_record` row represents a (hospital, procedure, payer, charge type, data source, setting) combination. The great majority of such combinations at any individual hospital have fewer than 10 line-item observations in the MRF, triggering the designed suppression behavior documented in Sections 4 and 5. At the cohort grain that drives severity computation (procedure, metro, payer, charge type, setting), 83,871 of 9,853,222 cohorts, approximately 0.9 percent, meet the `n ⩾ 10` threshold and carry populated severity statistics; the remaining approximately 99 percent fall below the threshold and are suppressed. Of the populated cohorts, 1,943 reach `n ⩾ 30` and carry the full BCa percentile confidence intervals described in Section 5.3. This is not a data defect; it is a consequence of the granularity at which hospitals disclose price data relative to the sample-size threshold required for defensible distributional statistics. The low `verified` share in a per-record count does not mean most pricing information is unreliable; it means that the distributional statistics (percentiles, confidence intervals) are suppressed for most individual cells, while the raw price point (median, gross charge, cash price) is still disclosed and usable.

10.2 Geographic Coverage Scope

Our dataset covers the top 100 U.S. metropolitan markets. This is a deliberate scope decision, not a data-completeness claim. The top 100 metros correspond to approximately 85 percent of hospitals by facility count, approximately 90 percent of procedure-cost search volume, and approximately 95 percent of the B2B addressable market for specialty-health and MGA underwriting. We have made the judgment that rural hospitals and micro-metro markets, which account for the remaining 5 to 15 percent by those metrics, represent a less compelling first-investment in data acquisition relative to the dense-metro markets where most self-funded plan sponsors and specialty payers concentrate their covered populations.

The geographic limitation is material for rate filings that cover rural populations or that require procedure-cost benchmarks in markets outside the top 100 metros. An underwriter covering a predominantly rural self-funded plan should not rely on our data as the sole input for cost modeling in markets outside our coverage scope. We are actively expanding coverage and intend to widen the metro roster in subsequent data refreshes, but we do not publish forward-looking estimates of the expansion schedule because the timing depends on MRF data quality and parsing success at individual hospitals.

The OIG (Source 12) audit of 100 hospitals conducted in early 2023 found that small hospitals, defined as those with fewer than 100 beds, were the most likely to be non-compliant with price transparency requirements due to resource constraints. Small hospitals are disproportionately represented in rural and micro-metro markets. The geographic limitation and the compliance limitation therefore partially overlap: the markets we have not yet prioritized are also the markets where MRF data quality is most uncertain.

10.3 Low-Sample Suppression

When a (hospital, procedure, payer, charge type, data source, setting) combination sits in a metro cohort whose count of line-item observations falls below 10, we suppress all distributional statistics: `median_price`, `percentile_10`, `percentile_25`, `percentile_75`, `percentile_90`, `percentile_95`, `percentile_99`, `mean_price`, `stdev_price`, and all confidence interval fields are returned as `null`. The `quality_flag` is set to `low_sample`. `sample_size` reports the cohort observation count where the severity computation recorded it, and is null otherwise; it is never the per-record allowed-amount count.

Suppression is the correct response to sub-threshold data, not an exceptional condition. ASOP 25 (Source 2) requires that credibility procedures be appropriate for the intended purpose. At n below 10, even the sample mean carries a coefficient of variation large enough to render percentile estimates misleading in a rate-filing context, and any CI we could compute around a suppressed point estimate would be mathematically meaningless. We therefore make no attempt to produce an interval: a wide but confidently wrong interval is worse than an honest null.

The practical implication for an underwriter: suppressed records still carry a raw price disclosure (the hospital's published gross charge, discounted cash price, or payer-specific negotiated rate), which may be useful as a reference point even when distributional statistics are not available. For rate-filing purposes, the practitioner actuary should apply a complement of credibility drawn from regional or market-wide benchmarks, per Werner-Modlin (Source 7, Chapter 12), when the ProcedureRadar distributional output for a given cell is suppressed.

The suppression rate is inherently high at the grain we report: individual payer/procedure/charge-type/setting combinations at individual hospitals rarely accumulate more than a handful of observations in a single MRF publication. The grain-level suppression rate does not imply that the underlying hospital data is unusable; it implies that the data needs to be aggregated across payers, charge types, or metro-level hospital cohorts to reach a credible sample size for distributional estimation.

10.4 CPT Code Timing Limitation

CPT (Current Procedural Terminology) codes are owned by the American Medical Association and require an AMA Distributor License for return in API responses. As of the v1 methodology effective date, we have not yet executed the AMA Distributor License. The CPT license is scheduled for Month 2 post-launch.

The practical limitation: API responses at v1 accept CPT codes as query input for lookup and filtering purposes, because input matching is not a reproduction activity. However, API responses do not return CPT codes as output fields. Consumers receive procedures identified by their `display_name`, `slug`, HCPCS Level II code (where available), and MS-DRG code (for inpatient procedures). The HCPCS and DRG code spaces are public domain, maintained by CMS, and require no licensing.

For rate-filing purposes, this limitation means that practitioners who rely on CPT codes as the primary procedure identifier in their actuarial model will need to maintain a CPT-to-HCPCS or CPT-to-DRG crosswalk to match our procedure slugs. We provide a `display_name` that is human-readable and a `slug` that is stable across methodology versions. The CPT limitation is temporary and will be resolved when the AMA license is executed.

10.5 Transparency in Coverage Data: Served, Not Yet Documented

The Transparency in Coverage (TiC) rule requires health insurers to publish payer-negotiated machine-readable files containing plan-specific allowed amounts for every service. TiC files represent the reimbursement side of the pricing picture, as opposed to the charge side that hospital MRFs represent. A practitioner actuary building a comprehensive rate-filing model would ideally have access to both: what the hospital charges and what the payer actually pays.

TiC data is now ingested and served on the Scale tier and above, from a small number of payers rather than the whole market. As of 2026-09-05 that lane holds 126,260 servable rate cells from ten payers, covering 1,192 hospitals across 42 states and 96 metros. Servable is the operative word and it is measured, not asserted: the count is of rows that pass the write-time trust check and are therefore returnable through the API, which is a smaller set than the rows that were loaded. Two payers with rows in the table, UnitedHealthcare and BCBS Tennessee, are excluded from that roster because every one of their rows fails that check and none can be served. File scope also differs per payer, and it matters for interpretation. Cigna's rates come from a wide set of that payer's plan-network files, so a Cigna-attributed rate is one network's rate. Aetna publishes its negotiated rates as one file per employer group, and we load a single large employer group's file per cycle. An Aetna-attributed rate in an API response is therefore one employer group's negotiated rate: not a market average, not a network-wide rate, and not representative of Aetna's negotiated position beyond that group. Blue Shield of California publishes per plan network and its network files were measured near-identical at curation, so we load one network file and a Blue Shield of California rate is that network's rate. Wellmark publishes one file per plan network and our curated order lists all ten of them, but the load has not matched that intent: production carries servable rows from one file, 59 cells across three preventive services at 20 Iowa hospitals. We state the load rather than the intent, and the eight files that produced nothing are an open pipeline question rather than a coverage claim. Medical Mutual of Ohio publishes one file per employer group and we load its employer files across the published book; its

index also carries one file for a leased national network, which we deliberately exclude until its provenance is verified, so Medical Mutual coverage here is concentrated at Ohio-area hospitals. Five regional Blue Cross plans joined the lane in the 2026-08-27 to 2026-09-01 wave and share one publication shape: each publishes a table of contents pointing at its in-network files, and we load a prioritized slice of those files rather than every one, bounded by a per-payer yield floor and volume ceiling committed before the load. Their footprints are regional by construction: Florida Blue 75 Florida hospitals, Blue Cross Blue Shield of Michigan 30 in the Detroit area, Premera Blue Cross 25 in Washington, Regence 21 across Oregon and Washington, and Horizon Blue Cross Blue Shield of New Jersey 39 across New Jersey and adjacent states. A rate from any of them describes that plan's own region and should not be read as a national position.

Update, 2026-10-22. UnitedHealthcare is no longer excluded. On 2026-10-22 its rows were replaced by a first load of one national network file, UHC's Choice Plus file for October 2026, which about two thirds of the UHC employer plans we sampled point to. A rate is placed at a hospital only when the file names that hospital by its own provider ID. A tax-ID match alone is not enough: in a hand check of the tax-ID matches, 249 of 453 cells belonged to a surgery center or a different hospital rather than the one matched. All 2,191 rate cells the load wrote pass the write-time trust check and are servable, at 44 hospitals in 13 states. A UnitedHealthcare-attributed rate is therefore the Choice Plus network's rate at that hospital, not every UHC plan's rate. Measured the same day, the lane holds 153,413 servable rate cells from 19 payers, covering 1,311 hospitals across 42 states and 95 metros. BCBS Tennessee remains excluded.

Two further interpretation rules apply across every payer in the lane. First, where a payer's own files disagree on a rate for the same (hospital, procedure, payer, charge-type) cell, a per-payer priority order committed at curation time selects the published value deterministically; measured on Cigna's corpus, the payer with the most observed intra-payer disagreement, that order resolves toward the higher of the disagreeing rates in roughly 61 percent of conflicted cells. Every non-selected rate is retained in an append-only history table with its source file, so the selection is auditable and reversible rather than destructive. Second, per-payer scope disclosures equivalent to this section's are published on the consumer methodology page, and each was live before the corresponding payer's first production load. The methodology in this paper remains the charge-side methodology, and that separation is enforced at each writer rather than by convention. The severity cohort

recompute and the high-n price fetch both carry an explicit `data_source IS DISTINCT FROM 'payer_tic'` predicate, in the aggregation and in the update target set, and the paths that assemble cohorts outside SQL apply the same exclusion. A payer-published rate therefore does not join a hospital-rate cohort, does not contribute to a percentile or a confidence interval, and is not stamped with cohort statistics of its own; as of the date on this paper, every percentile, mean, standard deviation and confidence-interval column is null on every payer-sourced row. Treat a payer-sourced row as outside the computed-field descriptions in Sections 4 through 9, because it is. We state it this way rather than silently broadening the paper's claims, because the distinction between charge-side and reimbursement-side pricing is methodologically significant, and understating it would mislead an actuary who is accustomed to working with adjudicated claims data.

The practical implication: negotiated rates in MRFs (the payer-specific rates hospitals are required to disclose) are the contracted rates, not necessarily what was actually remitted after claim adjudication, outlier adjustments, carve-outs, bundled payment reconciliation, or stop-loss provisions. MRF-disclosed negotiated rates and actually-paid claims amounts can and do diverge materially for complex cases. The CRS (Source 17) documented price variation of $14,306 to $56,695 for the same service at a single hospital across different payers in the TiC dataset, which illustrates the scale of within-hospital payer variation that our charge-side data approximates but does not fully capture at the adjudicated level.

For the avoidance of doubt in a rate-filing context: the percentiles and BCa confidence intervals in Sections 6 and 7 are computed from hospital charge-side rows only, and are not contaminated by payer-published rates. The owed methodology update is additive. It will compare charge-side rates with TiC allowed amounts at the procedure-metro grain, document the directional bias of charge-side-only signals, and specify how payer rates should be interpreted alongside them. Until that update carries a new `methodology_version_id`, treat payer rows as attributed source data rather than as inputs to any statistic described here.

10.6 MRF Compliance Gap

Hospital Machine-Readable Files are federally mandated under 45 CFR Part 180 (Source 10), but federal mandate does not guarantee data quality or completeness. Two primary federal oversight sources document the scope of this problem.

The U.S. Government Accountability Office released GAO-25–106995 in October 2024 (Source 11), finding that the Centers for Medicare and Medicaid Services cannot verify that hospital MRF data are sufficiently complete and accurate for their stated purpose. The GAO noted that health plans and employers have raised concerns about data quality including prices that appear unusually high or low and may represent errors. The GAO recommended that CMS assess whether MRFs are usable for their stated purpose and that CMS strengthen its verification practices. CMS issued a Request for Information in 2025 soliciting input on accuracy and completeness enforcement.

The HHS Office of Inspector General published a separate audit in November 2024 (Source 12) covering 100 hospitals audited between January and March 2023. The OIG found that 37 of the 100 hospitals did not comply with one or more price transparency requirements. Extrapolating to all 5,879 hospitals covered by the rule, the OIG estimated that approximately 46 percent of required hospitals were non-compliant. Small hospitals were most likely to be non-compliant. Of non-compliant hospitals, 34 had deficient machine-readable files and 14 failed shoppable-service requirements. CMS concurred with all OIG recommendations.

These are authoritative federal findings that we take seriously and that our quality-gate architecture is designed to address directly.

Our response to the GAO and OIG findings takes four concrete forms. First, we apply a shoppable-service early filter at parse time, discarding any record that does not meet the shoppable-service criteria defined in 45 CFR Part 180 before the record enters the database. This directly addresses the OIG finding about deficient shoppable-service requirements. Second, we apply placeholder-value detection to flag records with prices matching known non-meaningful patterns (prices of $0.01, $1.00, or values that are statistically implausible given the procedure and setting). Third, we apply an outlier detection step that flags records whose price deviates more than a configurable threshold from the hospital's own historical distribution for the same procedure. Fourth, the `quality_flag_verified_share` component of the data quality score (Section 8) ensures that hospitals with a high share of non-verified records receive a lower composite score, surfacing the compliance concern through the practitioner-facing scoring system.

We do not claim that our quality gates eliminate all compliance-related data deficiencies. The GAO and OIG findings reflect problems that occur at the point of hospital disclosure, before our pipeline runs. If a hospital publishes a placeholder price in a legally required field, our

placeholder-detection gate can flag it, but cannot replace it with the actual price. If a hospital omits a service from its MRF entirely, our coverage sub-score will reflect the omission, but we cannot ingest data that was never published. The quality gates improve the usability of the data we receive; they cannot fix the underlying compliance gap at its source.

10.7 Data Scope: Not the Largest Dataset

We do not represent our data as "most comprehensive" or "largest dataset." The enterprise data majors serving the specialty-health and MGA markets maintain larger record volumes, longer historical series, and in some cases access to adjudicated claims data that hospital MRFs cannot provide. Our lane is documented, source-traceable data depth at a published price, accessible to specialty payers and MGAs that cannot clear the procurement hurdles of enterprise data vendors.

An actuary evaluating our data for a rate filing should compare it against the sources available to them for the specific use case and market. For high-volume urban procedures in major metros where our hospital coverage is dense, our data provides a credible distributional picture at a price point that mid-market users can act on. For rare procedures, rural markets, or use cases that require adjudicated claims frequency data, supplemental sources are appropriate and we say so directly.

The CRS (Source 17) noted that third-party firms processing MRF data often use proprietary cleaning methodologies that reduce transparency and reproducibility, creating a vendor-opacity risk for the practitioners who rely on them. We document our methodology in this white-paper specifically to avoid that dynamic. A practitioner actuary using our data can trace the provenance of any output field back to a documented pipeline step, a versioned methodology file, and a primary data source with a URL and publication date. That transparency is a design constraint, not a marketing claim.

10.8 No Claims Data at v1

The v1 methodology does not incorporate adjudicated claims data. Claims data, when available, provides several actuarial signals that charge-side MRF data cannot: procedure frequency (how often a covered life files a claim for a given service), actual paid amounts after adjudication and contract adjustments, stop-loss attachment behavior, and diagnosis-based severity weighting. Each of these is material for certain underwriting use cases.

The absence of claims data means that our frequency modeling (at Enterprise tier) is derived from publicly available utilization statistics rather than from a claims data partner. The frequency models are documented at tier-appropriate detail in the covered-event section (Section 7) and in the per-client methodology addenda for Enterprise subscribers. An actuary building a rate filing that requires plan-specific frequency assumptions should supplement our frequency outputs with their own plan experience data.

The data quality score does not penalize records for the absence of claims data, because claims data is not a component of MRF disclosure. The score reflects what is available in the public charge-side data; its limitations relative to claims-side data are documented here, not embedded in the score itself.

10.9 CUUR0000SAM Methodology Change and Deflator Mismatch

Section 6.3 documents the October 2024 BLS methodology change to CUUR0000SAM in full detail. We repeat a summary here for readers who begin their review in the Known Limitations section.

Effective October 2024, the Bureau of Labor Statistics switched the private-insurance outpatient component of CUUR0000SAM from survey-based list-charge pricing to claims-data-sourced adjudicated pricing (Source 13). Our input data is charge-side. The conceptual mismatch between a deflator that now partly reflects adjudicated reimbursement prices and our input data that is predominantly gross charges and disclosed negotiated rates is a material limitation for the time-series normalization outputs we produce.

Dunn, Hale, and Dauda (2018, Source 14) identify this class of mismatch in the broader context of health expenditure research: for total medical expenditures including insurer payments, the PCE Personal Health Care index is generally preferred over the CPI medical care index because it is conceptually better aligned with reimbursement-side pricing. We continue to use CUUR0000SAM as the primary deflator because it is published monthly (the PCE PHC index is quarterly), is recognized by state insurance department reviewers without explanation, and is the standard reference in specialty-health rate-filing contexts. Switching deflators would require pipeline-cadence restructuring and would introduce its own audit risk. However, the practitioner actuary should note the mismatch when applying our time-series outputs to use cases that are sensitive to the distinction between charge-side and reimbursement-side price trends.

10.10 Internal Metro Grouping Versus Strict OMB MSA Boundaries

Our internal metro area slugs are broader than the official U.S. Office of Management and Budget Metropolitan Statistical Area definitions in some markets. The three clearest examples are `los-angeles` (which includes hospitals in Mendocino and Sonoma counties, well outside the official LA-Long Beach-Anaheim MSA, defined as Los Angeles and Orange counties), `chicago` (which extends into downstate Illinois beyond the Chicago-Naperville-Elgin MSA), and `new-york` (which spans a broader NY-NJ-PA tri-state area at varying distances from Manhattan).

When grouping is broader than the strict OMB MSA boundary, the measured price spreads (p90 minus p10 cash price, or maximum minus minimum payer-negotiated rate) inflate relative to what a strict MSA grouping would produce. This inflation occurs because a broader geographic cohort spans more distinct cost regimes: urban tertiary care centers, suburban community hospitals, and rural facilities are pooled together, and that pooling widens the apparent spread even when within-setting price variation is modest. A specialty-payer actuary using our spread statistics for rate filing in a tight urban metro may over-estimate the true rate variance if our metro slug includes markets that are structurally different from the specific geography they are underwriting.

Based on an internal review (Session 19 analysis), the preliminary estimate is that measured spreads in the broadest of our metro slugs inflate by approximately 10 to 30 percent compared to what a strict OMB MSA grouping would produce, with smaller inflation in metros whose slugs already align closely with the official MSA boundary. This is a preliminary range, not a measured figure. A v1.1 methodology release will quantify this inflation precisely using a county-FIPS-derived strict MSA grouping once the county-FIPS lookup is sourced (see `context/tight-msa-spread-analysis.md` for the full v1.1 hardening path). The v1.1 measured figure will replace the preliminary range at that release and a methodology version changelog entry will document the change.

Practitioners who require tight metro segmentation for specific underwriting geographies should contact `api-support@procedureradar.com` to request metro-level clarification about which hospitals and counties are included in a specific metro slug before using spread statistics in a rate filing.

10.11 January 2026 eCFR Allowed-Amount Disclosure Requirements

As of January 1, 2026, 45 CFR Part 180 (Source 10) requires hospitals to publish the 10th percentile, median, and 90th percentile of allowed amounts with a count of remittances for each shoppable service. This is a material expansion of what hospitals are required to disclose: the new fields represent hospital-published empirical percentile data derived from actual adjudicated claims, which is a categorically different and more informative signal than the gross charges and negotiated rates that prior MRF versions required.

The v1 methodology does not yet incorporate allowed-amount percentile data in any computed output. The reason is sequencing: the January 2026 requirement is recent and hospital compliance with the new fields will vary materially in the first year. We added the four columns to the database schema in migration 042 (`allowed_amount_p10`, `allowed_amount_p50`, `allowed_amount_p90`, `remittance_count`) and the parsers have stub support for ingesting those fields when hospitals publish them. However, the methodology has not yet been updated to incorporate these hospital-published percentiles as a validation layer against our computed distributional statistics or as a parallel output column.

The v1.1 methodology will evaluate how to integrate hospital-published allowed-amount percentiles alongside or in place of our computed quantiles for hospitals where the new data is available and credible. Until then, practitioners should be aware that for hospitals publishing compliant January 2026 files, there is a direct source of percentile data from the hospital's own claims experience that our pipeline is not yet surfacing.

10.12 Vendor Opacity and Our Documentation Commitment

The Congressional Research Service published CRS Report R48570 in June 2025 (Source 17), examining the TiC machine-readable file landscape. A key finding: third-party firms processing MRF data use proprietary cleaning methodologies that reduce transparency and reproducibility. The CRS documented price variation of $14,306 to $56,695 for the same service at a single hospital across different payers, a range that illustrates the normalization complexity involved and the risk that undisclosed proprietary cleaning steps could produce results that a practitioner actuary cannot audit or reproduce.

Our response to that concern is this white-paper. We publish the methodology rather than treat it as a black box. Every pipeline step that affects the data between the hospital's MRF and the API response is described here at a level that a qualified actuary can assess for

reasonableness. The versioned weights file, the explicit suppression thresholds, the named bootstrap variant, the documented deflator choice with its known limitations, and the quantitative coverage summary in this section are all design choices to make our methodology auditable rather than opaque.

We are not claiming that published methodology eliminates all risk of normalization error; it does not. What it does is give the practitioner actuary the information they need to identify where our choices differ from their own model assumptions and to adjust accordingly. That is what ASOP 23 (Source 1) and ASOP 41 (Source 3) require of a methodology document that practitioners will rely on in rate-filing contexts.

10.13 References

1. Source 1: ASOP No. 23, Data Quality (Revised Edition, Actuarial Standards Board, adopted December 2016, effective April 1, 2017). Operative sections: disclosure of known limitations; documentation of adjustments and extent of reliance on external data.
2. Source 3: ASOP No. 41, Actuarial Communications (Revised Edition, effective May 1, 2011). Operative sections: cautions regarding uncertainty in results; disclosure requirement for judgment-based calibration; clarity for peer actuary review.
3. Source 4: ASOP No. 56, Modeling (Actuarial Standards Board, adopted December 2019, effective October 1, 2020). Operative sections: documentation of the input range for which the model is valid; material limitation disclosure.
4. Source 10: U.S. Code of Federal Regulations, 45 CFR Part 180, Hospital Price Transparency (eCFR current version; rule effective January 1, 2021, major updates through January 1, 2026). Operative sections: January 2026 allowed-amount disclosure requirements; shoppable-service definition; MRF publication requirements.
5. Source 11: U.S. Government Accountability Office, GAO-25–106995, "Health Care Transparency: CMS Needs More Information on Hospital Pricing Data Completeness and Accuracy" (released October 2, 2024). Operative sections: CMS inability to verify MRF completeness and accuracy; health plan concerns about data quality; CMS RFI on enforcement.
6. Source 12: HHS Office of Inspector General, "Not All Selected Hospitals Complied With the Hospital Price Transparency Rule" (November 2024). Operative sections: approximately 46 percent non-compliance rate; small hospital resource-constraint non-compliance; deficient machine-readable files as primary compliance gap; OIG recommendations concurred by CMS.
7. Source 13: Bureau of Labor Statistics, "How BLS Measures Price Change for Medical Care Services in the Consumer Price Index" (factsheet, current); Bureau of Labor Statistics, "Incorporating Medical Claims Data in the CPI," Monthly Labor Review (2023). Operative sections: October 2024 methodology change to outpatient hospital services sub-index; switch to claims-data-sourced adjudicated pricing for private-insurance component.
8. Source 14: Dunn, A., Hale, E., and Dauda, B., "Adjusting Health Expenditures for Inflation: A Review of Measures for Health Services Research in the United States," Health Services Research, Vol. 53, No. 1, pp. 526–548 (2018). Operative sections: PCE PHC index as preferred deflator for total medical expenditures;

CPI medical care as preferred for consumer out-of-pocket; charge-side vs. reimbursement-side deflator mismatch.

9. Source 17: Congressional Research Service, "Technical Challenges with Private Health Insurance Price Transparency Data" (CRS Report R48570, June 13, 2025). Operative sections: vendor opacity in MRF processing; undisclosed proprietary cleaning methodologies; price variation of $14,306 to $56,695 for the same service at a single hospital across payers.

11. Versioning and Updates

11.1 Pipeline Cadence

The ProcedureRadar data pipeline runs on a monthly GitHub Actions cron schedule with a manual dispatch option for off-cycle hospital re-seeds, targeted backfills, and urgent data corrections. Each pipeline run ingests hospital MRFs that have been updated since the prior run, computes severity distributions, confidence intervals, time-series adjustments, and data quality scores for new and updated records, and upserts results to the database. Customers using the `?changed_since=<timestamp>` query parameter receive only records that changed in the most recent pipeline run, supporting incremental sync workflows.

11.2 Methodology Version Tracking

Every pricing record in the database carries a `methodology_version_id` foreign key to the `methodology_versions` table, introduced in migration 040. The `methodology_version_id` for all records computed under this white-paper's methodology is `1.0.0`, with an effective date of 2026-05-09. Every material methodology change (a change to the composite score weights, the credibility thresholds, the bootstrap resample count, the covered-event taxonomy, or any other component that affects the computed output fields described in this paper) creates a new record in `methodology_versions` with a new version string, an effective date, and a changelog entry.

The version field allows practitioners to confirm, for any record retrieved from the API, which methodology version produced the computed fields on that record. A rate-filing workpaper that cites our data should record the `methodology_version_id` of the records used, so that a subsequent reviewer can match the cited methodology to the version-specific white-paper.

11.3 Pre-Notification Commitment

For methodology changes that affect any field used in actuarial rate filing (severity percentiles, confidence intervals, time-series normalization, covered-event weights, or data quality score weighting), ProcedureRadar will provide written notification to Scale and Enterprise subscribers no fewer than 60 days before the effective date of the change. Customers may request a methodology freeze for active rate-filing periods not to exceed 90 days.

This commitment reflects the requirements of ASOP 41 (Source 3), which requires that actuarial communications give adequate advance notice when methodology changes could affect work product that practitioners have in progress, and ASOP 8 (Source 5), which requires that the certifying actuary document the methodology used at the time of the filing. A practitioner who has committed to a rate filing based on v1 methodology and receives notice of a pending v1.1 change can request a freeze to complete the filing before the change takes effect. The freeze applies to the data snapshot, not to new ingestion: fresh MRF data continues to be ingested, but the compute methodology applied to new records remains at the frozen version for the duration of the freeze period.

Changes that do not affect rate-filing fields (for example, changes to the URL-discovery pipeline, the dedup architecture, or consumer-facing display name mappings) do not trigger the 60-day pre-notification requirement. We will announce those changes in the `/developers/changelog` on the effective date.

11.4 Retroactive Access

Customers have the right to retrieve data as it existed at any historical methodology-version date, not just the current snapshot. The `pricing_history` table, introduced in migration 015, captures monthly immutable snapshots keyed on `(pricing_record_id, observed_at)`. Because the table is append-only, a query for "price as of month M under methodology version V" is expressed as a filter on `observed_at ⩽ M` combined with the `methodology_version_id` of that period's records. There is no structural path by which a current pipeline run can overwrite or revise a historical snapshot.

The `?changed_since=<timestamp>` parameter supports forward-incremental sync (new and changed records since a given date). For backward point-in-time access, practitioners should query the API using the `observed_at` filter against the pricing-history endpoint. This retroactive-access architecture is a customer right, not a feature subject to tier gating: all paid tiers have access to their historical snapshots at the grain and field set their tier entitles.

11.5 v1.1 Commitment: January 2026 Allowed-Amount Integration

The January 2026 eCFR amendments require hospitals to publish the 10th percentile, median, and 90th percentile of allowed amounts with a count of remittances for each shoppable service (Source 10). We added four columns to the database schema in migration 042 (`allowed_amount_p10`, `allowed_amount_p50`, `allowed_amount_p90`,

`remittance_count`) and the parsers have stub support for ingesting those fields when hospitals publish them. The v1 methodology does not consume these columns in any computed output; the column values are stored but not exposed in the API response.

The v1.1 methodology release will evaluate how to integrate hospital-published allowed-amount percentiles alongside or in place of our computed quantiles for hospitals where the new data is available and has passed quality gates. This evaluation will include a comparison of hospital-published percentiles against our BCa-derived percentiles for the same procedure-hospital pairs, a quality assessment of the new fields (compliance with the January 2026 requirements varies across the hospital population in the first year), and a decision on whether to surface hospital-published percentiles as a parallel column or as a validation layer against our computed distributions. The evaluation will be published as a methodology update with a changelog entry and will trigger the 60-day pre-notification process for Scale and Enterprise subscribers before any change to computed output fields.

11.6 v1.1 Commitment: Tight-MSA Spread Inflation

Section 10.10 documents a preliminary estimate that measured price spreads in the broadest of our metro slugs inflate by approximately 10 to 30 percent compared to what a strict OMB MSA grouping would produce. This is a preliminary range from an internal analysis; it is not a measured figure. The v1.1 methodology will quantify this inflation precisely using a county-FIPS-derived strict-MSA grouping once the county-FIPS lookup is sourced. Full analysis is documented at `context/tight-msa-spread-analysis.md`. The measured figure will replace the preliminary range in the v1.1 white-paper and a methodology changelog entry will document the effective date and the magnitude of the revision.

11.7 Backfill Posture

When a compute method changes, we backfill existing records where the compute is feasible within the proxy timeout constraints of our infrastructure. The backfill is announced in the methodology changelog with an effective date and a description of the affected field and population. Customers do not see a partially backfilled database; batched backfills are designed to be resumable and are not surfaced in the API until the batch is complete for each hospital.

Examples of completed backfills: the BCa confidence interval backfill script that replaced plain percentile-bootstrap CI values for all records with n greater than or equal to 30 (shipped in PR-A T9), and the six-component quality score backfill that added the `sample_size` sub-score component and recomputed composite scores for all existing records (shipped in PR-A T10). Both examples illustrate the posture: compute the change, verify the output against the post-apply predicates, announce in the changelog, and flip the `methodology_version_id` on backfilled records to the new version.

11.8 Major-Version Deprecation

A future major-version API release (`/v2`) will provide a minimum 12–month overlap window before the `/v1` endpoint returns `410 GONE`. During that window, `/v1` and `/v2` operate in parallel. Field-level deprecation within a major version (adding a field to the omit list, changing a field's definition, or renaming a field) requires two full minor versions of advance notice before the change takes effect, so practitioners have time to adapt their integration.

11.9 References

1. Source 3: ASOP No. 41, Actuarial Communications (Revised Edition, effective May 1, 2011). Operative sections: advance notice of methodology changes; clarity for peer actuary review; disclosure of methods and assumptions.
2. Source 5: ASOP No. 8, Regulatory Filings for Health Benefits (Revised March 2014, effective September 1, 2014). Operative sections: documentation of methodology at time of filing; data-source documentation; supplemental documentation availability.
3. Source 10: U.S. Code of Federal Regulations, 45 CFR Part 180, Hospital Price Transparency (eCFR current version; rule effective January 1, 2021, major updates through January 1, 2026). Operative sections: January 2026 allowed-amount percentile disclosure requirements; phased implementation timeline.

12. Contact and Review

12.1 Methodology Questions

All methodology questions should be directed to `api-support@procedureradar.com`. We commit to the following response timelines.

For Scale subscribers: a first response acknowledging the question within 4 business hours. For Enterprise subscribers: a first response acknowledging the question within 1 business hour. For any subscriber at any tier who raises a methodology concern that materially affects a rate filing in progress: a written response addressing the substance of the concern within 5 business days. Where the concern identifies a material error or limitation in our methodology, we commit to a methodology revision cycle within 30 days of confirming the concern is valid, with 60-day pre-notification to Scale and Enterprise subscribers before any resulting change takes effect.

This contact path is the supplemental-documentation channel that ASOP 41 (Source 3) and ASOP 8 (Source 5) require us to maintain. A practitioner actuary whose filing requires clarification beyond what this white-paper provides should use this channel before completing the rate-filing workpaper, not after.

12.2 Enterprise Actuary Review Letter

For Enterprise subscribers, ProcedureRadar provides a written review letter from a credentialed actuary covering the methodology elements relevant to the client's specific rate-filing use case. The review letter is the function that ASOP 8 (Source 5) calls "actuary on file" support: it gives the certifying actuary a documented third-party review of the data methodology they are relying on, scoped to the specific fields and procedures their filing uses.

Pre-revenue, Scale and Enterprise methodology questions that require actuarial judgment (for example, whether our BCa CI width is adequate for a specific stop-loss attachment scenario, or whether our covered-event weights are calibrated appropriately for a specific benefit design) go through a credentialed actuarial reviewer on the response chain, engaged on a project basis until in-house actuarial capability scales with the business. We do not claim to have a permanent in-house actuary at this stage; we are transparent that the reviewer is

engaged on a project basis. The quality and professional accountability of the response is not reduced by that arrangement: a credentialed actuary reviewing our methodology is bound by the same ASOP standards regardless of engagement structure.

The review letter is scoped to the methodology elements the client's filing requires, not to the full 12-section white-paper. A client whose filing relies only on our severity distributions and confidence intervals receives a letter covering Sections 4, 5, and the relevant portions of Section 10. A client whose filing relies on covered-event mapping and frequency modeling receives a letter covering Sections 7, 9, and the Enterprise-specific subsections. This scoping avoids the letter being either over-broad (creating review obligations for unused sections) or under-specific (failing to address the methods the actuary actually certified).

12.3 Professional Credentials and Review Chain

Methodology responses for Scale and Enterprise subscribers that require actuarial judgment are reviewed by a credentialed actuary (Fellow of the Casualty Actuarial Society, Fellow of the Society of Actuaries, or equivalent) before being sent. Factual questions about API behavior, data schema, or pipeline mechanics are handled by the engineering team without actuarial review. Questions that involve interpretation of our methodology in the context of a specific state's rate-review requirements are handled jointly by the engineering team and the actuarial reviewer.

We disclose this structure plainly because ASOP 41 (Source 3) requires that communications not overstate the credentials or authority behind them. We are not a credentialed actuarial firm and do not provide actuarial opinions; we provide data and methodology documentation that a practitioner actuary uses in their own work. The actuary review letter is a documented assessment of the methodology's reasonableness for the client's stated use case, not an actuarial certification of the client's rate filing.

12.4 Methodology Changelog

Every material methodology change produces a changelog entry at `/developers/changelog` with the following minimum fields: the date the entry was published, the effective date of the change, the affected field or fields, the nature of the change, the affected tier scope, and the reason for the change. Changelog entries for methodology changes are permanent: we do not remove or revise changelog entries after publication. Additions are append-only.

The changelog is the auditable record of how the methodology has evolved from one version to the next. A practitioner actuary who used our data in a filing completed in a prior period can retrieve the changelog entries between their filing date and the current date and assess whether any intervening methodology changes are material to their prior work.

12.5 References

1. Source 3: ASOP No. 41, Actuarial Communications (Revised Edition, effective May 1, 2011). Operative sections: supplemental-documentation channel; contact path for methodology questions; clarity regarding credentials and scope of the communication.
2. Source 5: ASOP No. 8, Regulatory Filings for Health Benefits (Revised March 2014, effective September 1, 2014). Operative sections: actuary-on-file function; supplemental documentation availability for rate-filing support; data-source documentation responsibility.

ProcedureRadar's pricing dataset is built from federally mandated hospital Machine-Readable Files.

Disclosure required by 45 CFR Part 180.

- Hospital Price Transparency Rule (45 CFR Part 180) (May 9, 2026)

Last updated: May 9, 2026

ProcedureRadar

Real hospital pricing data, powered by federally mandated transparency files.

Scan. Compare. Save.

PROCEDURES

MRI

Colonoscopy

Knee Replacement

CT Scan

All Procedures

CITIES

New York

Los Angeles

Chicago

Dallas

All Cities

TOOLS

Compare Hospitals

Compare

Cost Guides

Methodology

Verified Data Sample

API

Security

COMPANY

About

Terms

Privacy

© 2026 LocalLayer AI LLC d/b/a ProcedureRadar. All rights reserved.

Data updated monthly from CMS-mandated hospital transparency files.

ProcedureRadar's data processing methodology is the subject of a pending US provisional patent application.