The Medicare Cost-Quality Matrix

Doing More With Less: which U.S. hospitals deliver high quality at low cost, and what their profile reveals

Author: Jay Callery | Big Jacket Research

May 2026

Data: HCRIS Form 2552-10 (FY 2024), CMS Care Compare quality measures (CY 2024), IPPS Final Rule Impact File (FY 2026), Provider of Services File (QIES Q1 2026), Medicare Spending Per Beneficiary (CY 2024)

Pipeline: Python 3.12 | dbt-fusion 2.0 | Google BigQuery | Dagster

Repository: github.com/jcal-2/CMS-HRRP-Analysis (Big Jacket Research data platform)

Executive Summary

This study compares every U.S. hospital's cost and quality against hospitals of similar size, type, and local labor market, so small rural hospitals aren't measured against large urban ones. Out of 3,020 hospitals analyzed across five public CMS datasets, 636 (21.1%) deliver above-average quality at below-average cost within their peer group. But the most extreme outperformers, the ten hospitals with top-tier quality and lowest cost, are almost exclusively specialty surgical hospitals (nine orthopedic/spine, one cardiac). None are regular community hospitals. This suggests the most extreme efficiency gains come from procedural concentration, not management practices any hospital could adopt. On the other end, 836 hospitals (27.7%) are both relatively expensive and relatively low-quality within their peer group, the largest single quadrant. Of those, 69 are flagged as outliers on all four cost drivers: length of stay, labor intensity, operating cost, and margin pressure. These are the hospitals where a deeper operational review would have the highest return.

Key Findings at a Glance

Finding	Value	Significance
Hospitals in analytical universe	3,020	Active short-term acute care hospitals, intersection of POS and IPPS rosters
Peer cohorts formed	28	Teaching status x bed-size band x wage-index band; smallest cohort = 27 hospitals
Doing More With Less list	10 hospitals	Top within-cohort quality AND bottom within-cohort cost
Specialty surgical concentration on the list	9 of 10	90.0% are orthopedic or spine; the tenth is cardiac. None are general acute-care community hospitals.
Low-quality / high-cost (Q4) population	836 hospitals	27.7% of universe; reserved as a commercial data product, not a published list
Quad-driver outliers	69 hospitals	Flagged top-quartile on all four cost drivers (LOS, labor intensity, operating cost, margin pressure) within their cohort
Quality composite coverage	96.9%	39 hospitals suppressed because fewer than 2 of 5 quality domains are populated

1. Background

The U.S. government tracks hospital performance in two separate ways. The first is cost: every Medicare-certified hospital files an annual cost report (HCRIS Form 2552-10) that details what it spends on operations, how many patients it treats, and how much revenue it generates. The second is quality: CMS publishes data on patient outcomes including mortality rates, readmission rates, surgical complications, hospital-acquired infections, and patient satisfaction scores.

Both datasets are public and are available at CMS.gov, and both use the same hospital ID number (CMS Certification Number) to identify each facility. However, they are often reported in isolation from one another. While hospital finance teams and vendors are looking at the cost data, public ranking systems like CMS Star Ratings, US News, and Leapfrog are looking at the quality data. The only place the two come together is in Medicare penalty programs like HRRP and VBP, which reduce hospital payments based on

quality performance. These penalty programs use the data to adjust payments, but none of them publish a ranking that compares hospitals on both cost and quality at the same time, relative to similar peers.

Public hospital rankings do exist. Leapfrog, US News, CMS Star Ratings, and Vizient all score hospitals, but they each have limitations. Some compare every hospital against every other hospital, which unfairly disadvantages teaching hospitals, safety-net hospitals, and rural facilities because they operate under fundamentally different conditions. Others use peer-based comparisons but charge expensive licensing fees for access.

This study fills that gap: a free, open, peer-based ranking that evaluates both cost and quality together, with transparent methodology. The full dataset of 3,020 hospitals, the code, and the analytical models are openly available. This study is also upfront about what the data actually shows, including the finding that the top-performing hospitals on this matrix are specialty surgical facilities whose advantage comes from case selection, not transferable management practices. **This study connects both tracks directly to answer a different question: within groups of similar hospitals, which ones deliver high quality at low cost?**

2. Methodology

2.1 Data Sources and Pipeline

CMS Data Source	Coverage	Role
HCRIS Hospital Cost Reports (Form 2552-10)	FY 2024	Operating expenses, discharges, FTEs, patient days, net patient revenue, operating margin
IPPS Final Rule Impact File	FY 2026	Wage index, beds, case-mix index, DSH percentage, teaching intensity, urban/rural
Provider of Services (POS) File	QIES Q1 2026	Active certification, hospital type, ownership, teaching switches
Care Compare quality measures	CY 2024	30-day mortality (6 conditions), readmission ERR (6 conditions), complications, PSI safety composite + 11 individual PSIs, HAI infection ratios (6), HCAHPS star ratings (9)
Medicare Spending Per Beneficiary (MSPB)	CY 2024	National-relative spending-per-episode ratio (validator only; not in cost composite)

The pipeline runs Python ingestion into Google BigQuery raw datasets, transforms with dbt-fusion 2.0 (13 staging models, 6 mart models in this paper's analytical scope), and orchestrates with Dagster. 62 dbt tests enforce data-quality and methodology invariants on every refresh, including cohort minimum size, universe sanity bounds, and quality coverage floor.

Every measure is keyed to CCN (The CMS Certification Number mentioned above). Source-level field mapping is documented separately in the field-map appendix.

2.2 Cohort Definitions

The analytical universe for this study is the intersection of two filters:

- POS active short-term acute care hospitals (PRVDR_CTGRY_CD = '01', subcategory '01', termination code '00'): 3,097 hospitals.

- IPPS Final Rule Impact File: 3,103 hospitals.

The intersection is **3,020 hospitals**. The 77 POS-only and 83 IPPS-only differences reflect non-IPPS short-term acute care (e.g., Maryland's all-payer system) and recent IPPS terminations after the data freeze. Both are excluded from this edition.

Three peer-cohort axes are drawn from the IPPS Impact File:

- **Teaching status**: binary, derived from resident_to_bed_ratio > 0.
- **Bed-size band**: four levels, less than 100, 100 to 249, 250 to 499, 500 and above.
- **Wage-index band**: quartiles of the FY 2026 wage index across the universe.

Initial assignment is the Cartesian product: 4 x 2 x 4 = 32 cells. We require at least 20 hospitals per cohort. In the 2024 vintage, one cell, non-teaching hospitals with 500 or more beds, contained only 18 hospitals nationally, below the threshold. We applied a single merge rule: collapse the 500 or more bed band into the 250 to 499 band for non-teaching hospitals only.

After merge: **28 cohorts**, smallest contains 27 hospitals, largest 337. The methodology test assert_peer_cohort_minimum_size enforces the at-least-20 invariant at every refresh.

Cohort boundaries are analytical choices, not natural categories; findings are robust to reasonable boundary adjustments. The 500-bed non-teaching merge is data-dependent and reviewed at every vintage refresh. Hospitals that closed or merged between the IPPS data freeze and publication are not represented in the universe.

2.3 Statistical Methods and Measurement

Every cross-hospital comparison in this paper is computed as a within-cohort percentile rank: a hospital's rank position on a signal, normalized to a 0 to 1 scale, using only members of the hospital's own peer cohort. 0 corresponds to the lowest value in the cohort; 1 corresponds to the highest. Each individual measure is sign-corrected so that higher percentile always represents the same direction (higher cost = higher cost percentile; higher quality = higher quality percentile).

Cell-level numeric values from HCRIS are winsorized at the 1st and 99th percentile per (worksheet, line, column) in the staging layer. Approximately 2 to 3% of cells are clipped per field; clipped flags are preserved alongside values for auditability. Quality measures from Care Compare carry CMS's upstream risk adjustment; no additional risk adjustment is applied in this study.

> **Methodology note**
> Findings in this paper describe relative position within peer cohort, not causation. Cost-driver flags are indicative of where a hospital diverges from its peers, not why. Suppression rules apply when input coverage is insufficient: the quality composite is suppressed for hospitals with fewer than two of five domains populated.

2.4 Cost Composite Construction

Two signals, each percentile-ranked within peer cohort, then simple-averaged:

1. Cost per CMI-adjusted discharge = HCRIS total operating expenses divided by HCRIS total discharges, then divided by IPPS case-mix index (Grouper v43).

2. Operating cost-to-charge ratio (CCR) = IPPS Operating CCR. Higher = costs are a larger fraction of charges.

MSPB is excluded from the composite (already national-relative; re-cohorting would over-shrink) and preserved as a row-level validator. Hospitals with neither cost signal available are suppressed and excluded from quadrant placement.

2.5 Quality Composite Construction

Five quality domains, each containing one or more measures from CMS Care Compare:

Domain	Measures (count)	Direction
Mortality	30-day risk-adjusted rates for AMI, CABG, COPD, HF, Pneumonia, Stroke (6)	Lower is better
Readmissions	30-day excess readmission ratio for AMI, CABG, COPD, HF, Pneumonia, Hip/Knee (6)	Lower is better
Complications	Hip/Knee complication rate, Hybrid hospital-wide mortality (2)	Lower is better
Safety	PSI_90 composite + 11 individual PSIs + 6 HAI standardized infection ratios (18)	Lower is better
HCAHPS	9 published star-rating composites (overall, hospital rating, recommend, communication, cleanliness, quietness)	Higher is better

For each measure, we compute a within-cohort percentile rank. For lower-is-better measures we use (1 minus percent_rank) so all measure percentiles are sign-corrected to 'higher = better quality.' We then compute one domain score per hospital as the simple average of its measures' sign-corrected percentiles. The overall quality composite is the simple average of available domain scores, **equal weight per domain**, so the safety domain (18 measures) does not dominate complications (2 measures).

We require **at least two domains** to compute the composite. 39 hospitals are suppressed by this rule; the composite has 96.9% coverage on the universe.

2.6 Cost-Driver Indicators

Four within-cohort percentile-ranked indicators, each top-quartile-flagged:

- **Length of stay** (LOS = total inpatient days divided by total discharges)
- **Labor intensity** (FTEs per 1,000 discharges)
- **Operating cost** (cost-per-CMI-adjusted-discharge, mirrors the cost composite signal)
- **Margin pressure** (1 minus operating margin)

Two additional cost-driver categories named in the strategic plan, supply chain (HCRIS Worksheet G-2) and administrative overhead (HCRIS Worksheet B-1 A&G cost center), are deferred to a future edition pending finalized CMS Form 2552-10 Rev 17 line-item mapping.

2.7 Matrix Quadrant and Named-List Cuts

We split each axis at the within-cohort median to form four quadrants. For the named 'Doing More With Less' list, we apply a tighter cut: top quartile of quality AND bottom quartile of cost, both within cohort. The

2024 vintage refresh cadence is annual, aligned to the IPPS Final Rule release each August and the Care Compare quarterly publication schedule.

3. Findings

3.1 The Matrix at a Glance

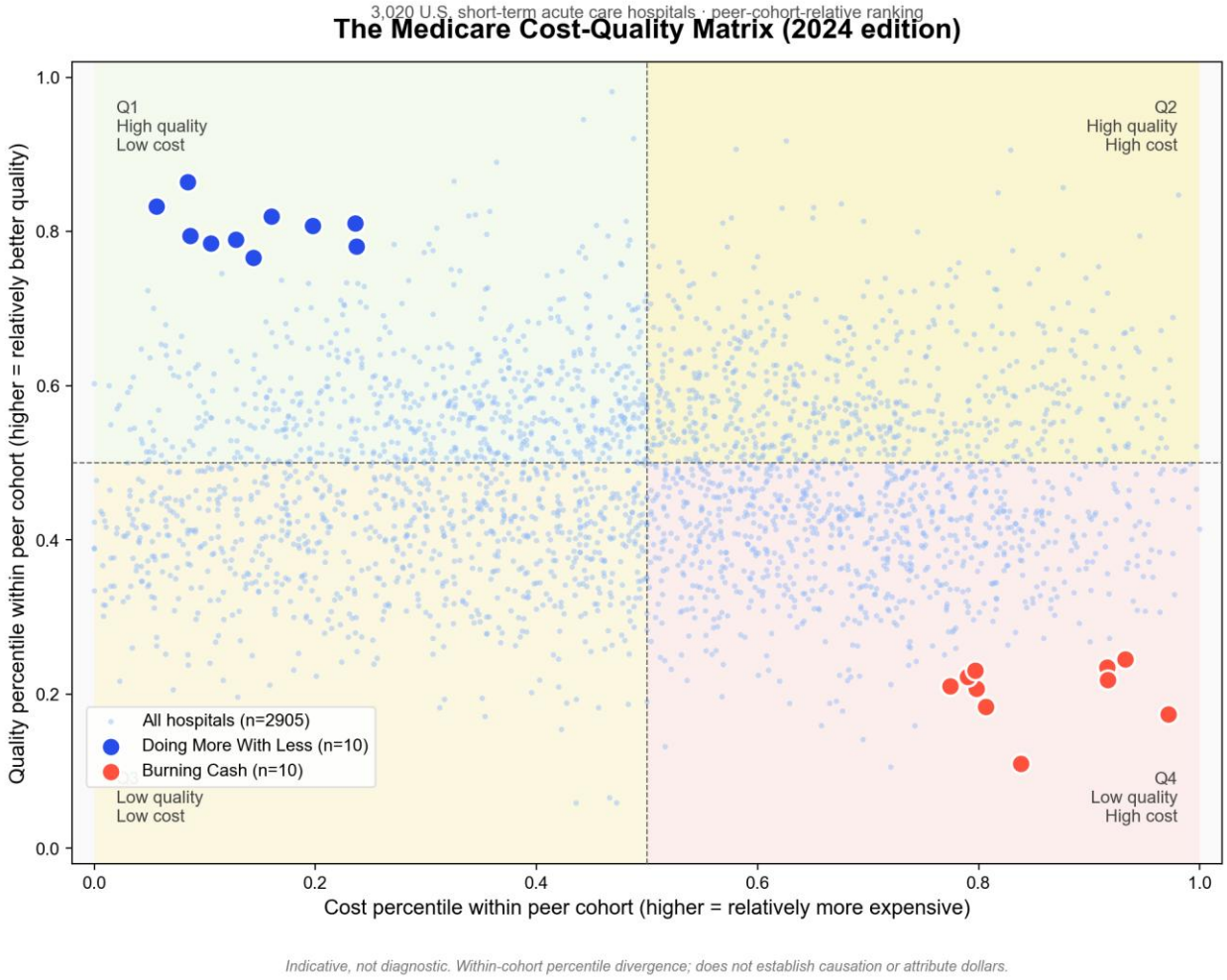

Figure 1. Each U.S. short-term acute care hospital plotted on within-peer-cohort cost and quality percentiles. Doing More With Less hospitals appear in the top-left corner; the opposite-corner cluster represents hospitals with relatively high cost and relatively low quality within their peer cohort.

Each axis is a within-peer-cohort percentile. The split is at the cohort median.

Quadrant	Definition	Hospitals (n)	Share
Q1	High quality, low cost (within cohort)	636	21.1%
Q2	High quality, high cost	643	21.3%
Q3	Low quality, low cost	810	26.8%
Q4	Low quality, high cost	836	27.7%
Suppressed	Insufficient signal coverage	95	3.1%
Total		**3,020**	**100%**

3.2 The Doing More With Less List

Hospitals in the **top quartile of within-cohort quality AND the bottom quartile of within-cohort cost**:

Hospital	State	Cohort	Quality (pct)	Cost (pct)
Texas Orthopedic Hospital	TX	non-teaching, <100 beds, mid wage	86	8
Legent Orthopedic + Spine	TX	non-teaching, <100 beds, mid wage	83	6
Our Lady of the Lake Surgical Hospital	LA	non-teaching, <100 beds, low wage	82	16
Baylor Scott & White Texas Spine & Joint	TX	non-teaching, <100 beds, upper-mid wage	81	24
Pinnacle Hospital	IN	non-teaching, <100 beds, upper-mid wage	81	20
Oklahoma Spine Hospital	OK	non-teaching, <100 beds, low wage	79	9
Saint Thomas Hospital for Specialty Surgery	TN	non-teaching, <100 beds, low wage	79	13
Hoag Orthopedic Institute	CA	non-teaching, <100 beds, high wage	79	11
Arkansas Surgical Hospital	AR	non-teaching, <100 beds, low wage	78	24
The Heart Hospital Baylor Denton	TX	non-teaching, <100 beds, mid wage	77	14

Ten hospitals nationally satisfy this cut. **Nine of ten are orthopedic or spine surgical hospitals. The tenth is cardiac. None are general acute-care community hospitals.**

3.3 The Structural Pattern Behind the List

The contrarian quadrant pattern, that nearly all top-left-quadrant outliers are specialty surgical hospitals, is not noise. It is a structural observation about how the U.S. hospital industry is organized.

- **Procedural concentration drives both efficiency and outcome quality.** A hospital that performs orthopedic surgery and nothing else can build the surgical and post-op workflows, supply chain, and staffing model for that single procedure family. Its cost per case-mix-adjusted discharge is lower because its overhead is lower; its quality scores (especially on complications and HCAHPS) are higher because its volume per surgeon is higher.
- **Peer-cohort matching helps but does not fully neutralize this effect.** Bed-size, teaching status, and wage-index control for some of the structural difference between a 60-bed orthopedic specialty hospital and a 60-bed general community hospital, but they do not control for case-mix selection. Specialty hospitals select the cases they want; community hospitals do not.
- **The honest framing is: specialized scale produces operational efficiency.** It does not follow that any general acute-care hospital could close the gap by adopting these hospitals' practices.

> **Key Question:** Does the specialty surgical concentration in this list reflect a transferable operating model, or a structural case-mix selection effect that community acute-care hospitals cannot replicate? The data is consistent with the second interpretation; this paper does not claim the first.

3.4 The Opposite Corner

The opposite quadrant, within-cohort bottom quartile of quality and top quartile of cost, also contains ten hospitals at the tight cut. We do not name them in this artifact. The hospitals in that group are a mix of small rural facilities operating under labor and supply constraints they cannot solve internally, and large urban safety-net systems carrying structural payer-mix pressure their cost structure cannot escape. Those are operational and policy problems with different framings and different remediation paths from the public-research-artifact framing used here.

The broader Q4 quadrant (836 hospitals) is available as a commercial data product. Its purpose is to surface hospitals where cost-reduction conversations are most relevant, not to publish a worst-of list.

3.5 Cost-Driver Distribution

For every hospital we computed four cost-driver indicators (length of stay, labor intensity, operating cost, margin pressure), each percentile-ranked within peer cohort and flagged as an outlier if it falls in the top quartile.

Cost driver	Top-quartile flagged	Share of universe
Length of stay	763	25.3%
Labor intensity (FTE / discharges)	757	25.1%
Operating cost (CMI-adjusted)	763	25.3%
Margin pressure (1 minus operating margin)	744	24.6%

A hospital can be flagged on multiple drivers. **69 hospitals flagged on all four drivers** are the most severe candidates for diagnostic follow-up.

Each driver reports relative position, not causation. A high LOS percentile does not establish that a hospital's discharge planning is broken; it reports that this hospital, compared with its peer cohort, keeps patients longer per admission. The discrimination of cause from observation is the work of a future custom-diagnostic engagement, not this paper.

4. Limitations

These are observations that should travel forward with the data, not blockers to releasing this edition.

- **Observational study: all findings are correlational.** Nothing in this paper attributes dollars, quantifies waste, or establishes causation. Cost-driver outlier flags are indicative, not diagnostic.
- **Single-vintage snapshot.** The 2024 edition uses one fiscal year of HCRIS, one calendar year of Care Compare. Vintage drift is expected; future editions will refresh annually.
- **CMS data quality.** HCRIS line items are self-reported and adjusted via cost report settlement; the inputs are not error-free. Care Compare measures are risk-adjusted by CMS using methodologies subject to ongoing review and revision.
- **Cohort minimums.** A 27-hospital cohort is a statistically thin reference set, and within-cohort percentiles in those cells should be treated with corresponding skepticism. Cohort sizes are documented alongside every result.

- **Specialty hospital effect.** The Doing More With Less list is dominated by specialty surgical hospitals; we have not adjusted for the case-mix selection that those facilities perform within the bounds of their CCN. Section 3.3 frames this transparently.
- **PSI reporting lag.** Patient Safety Indicators on Care Compare typically lag admission dates by 1 to 2 fiscal years due to claim-runout and risk-adjustment windows.
- **HCAHPS response-rate variation.** Star ratings are computed off completed surveys; response rates vary widely (often 15 to 40%) and interact with the specialty-hospital pattern. A future edition sensitivity analysis is planned for hospitals with response rates below 20%.
- **Survivorship.** Hospitals that closed or merged between the data freeze and the present are not represented. The universe is the active short-term acute care universe at the IPPS data freeze.

5. Conclusion

Which U.S. hospitals deliver high quality at low cost? Within peer cohorts, 636 hospitals (21.1%) fall in the high-quality, low-cost quadrant. But the tighter cut, top quartile quality and bottom quartile cost, reveals a structural truth about the U.S. hospital system: the most extreme cost-quality outperformers are almost exclusively specialty surgical hospitals. They achieve efficiency through procedural concentration and case-mix selection, not through practices that generalize to community hospitals serving undifferentiated patient populations.

The opposite corner (836 hospitals, 27.7%) represents the highest-need segment: facilities that are both relatively expensive and relatively low-quality within their peer group. The 69 hospitals flagged on all four cost drivers are the most acute candidates for further diagnostic work.

Appendix: Glossary

Terms and acronyms used in this paper, organized by category.

A. Acronyms

Acronym	Definition
CBSA	Core-Based Statistical Area. A U.S. metropolitan or micropolitan area used by CMS to define labor markets for the wage index.
CCN	CMS Certification Number. The six-digit identifier CMS assigns to every certified Medicare provider. Used as the universal hospital key across every data source.
CCR	Cost-to-Charge Ratio. The ratio of a hospital's Medicare-allowable costs to its Medicare-covered charges. Higher CCR = costs are a larger fraction of charges.
CMI	Case-Mix Index. A weighted average of DRG relative weights for the cases a hospital billed, indicating the relative severity / resource intensity of its patient mix.
DSH	Disproportionate Share Hospital. A CMS designation and payment adjustment for hospitals serving a high proportion of low-income patients.
ERR	Excess Readmission Ratio. Predicted divided by expected 30-day readmission rate from CMS's readmission measure. ERR > 1.0 = worse than the national peer-comparison group; lower is better.
HAI	Healthcare-Associated Infection. CMS publishes 6 standardized infection ratios (CLABSI, CAUTI, SSI colon, SSI hysterectomy, MRSA, C. diff).
HCAHPS	Hospital Consumer Assessment of Healthcare Providers and Systems. The CMS patient experience survey. Star ratings (1 to 5) are the primary signals; higher is better.
HCRIS	Healthcare Cost Report Information System. CMS's repository of hospital cost reports submitted on Form 2552-10. Source for operating expenses, FTEs, discharges, revenue and margin.
IPPS	Inpatient Prospective Payment System. The CMS payment system for short-term acute care hospitals. The IPPS Final Rule Impact File is the source for wage index, beds, CMI, DSH percentage, and teaching intensity.
LOS	Length of Stay. Total inpatient days divided by total discharges. A cost-driver indicator.
MSPB	Medicare Spending Per Beneficiary. CMS Care Compare measure of risk-adjusted Medicare spending per episode (3 days pre-admission through 30 days post). A national-relative ratio.
POS	Provider of Services file. The CMS QIES quarterly release that lists every certified facility with its category, ownership, urban or rural designation, teaching switches, certified bed count, and certification dates.
PSI	Patient Safety Indicators. AHRQ-developed measures of in-hospital safety events. PSI_90 is the composite; PSI_03 to PSI_15 are individual indicators.
SIR	Standardized Infection Ratio. The observed-to-expected ratio for a Healthcare-Associated Infection measure. SIR > 1.0 = more infections than expected given case mix.

B. Key Terms

Term	Definition
Bed-size band	One of four buckets used as a peer-cohort axis: less than 100, 100 to 249, 250 to 499, 500 or more certified beds. The 500-plus band is merged into 250-plus for non-teaching hospitals where the 500-plus cell falls below the minimum cohort size.
Cohort size	Number of hospitals in a peer cohort. Enforced at 20 or more by the test assert_peer_cohort_minimum_size.
Cost composite	Per-CCN within-cohort percentile rank of cost signals (cost per CMI-adjusted discharge and operating CCR), simple-averaged. Range 0 to 1; higher = relatively more expensive within cohort.
Cost-driver indicator	A single percentile-ranked measure tied to one operational dimension (LOS, labor intensity, operating cost, margin pressure). Indicative of where a hospital's cost position diverges from peers; not diagnostic of why.
Doing More With Less list	The 10 hospitals in the top quartile of within-cohort quality AND the bottom quartile of within-cohort cost. The public-facing named list in this paper.
Indicative, not diagnostic	The methodological discipline of reporting percentile divergence (where a hospital sits relative to peers) without claiming causation, dollar attribution, or root cause. Applied to every numerical claim in this paper.
Margin pressure	1 minus operating margin. Higher value = thinner margin. Distinguishes hospitals with structural margin pressure (cost-cutting urgency) from profitable high-cost hospitals (waste-driven).
Median split	The cut used to define the four matrix quadrants: each axis is split at the cohort-relative 50th percentile.
Peer cohort	A group of hospitals defined by (teaching status x bed-size band x wage-index band). 28 cohorts in the 2024 vintage; smallest 27 hospitals. Every ranking in this paper is computed within peer cohort, not against the U.S. universe.
Percentile rank (within cohort)	A hospital's rank position on a signal, normalized to a 0 to 1 scale and computed using only members of the hospital's own peer cohort. 0 = lowest in cohort, 1 = highest.
Quadrant (Q1 to Q4)	The four buckets of the cost-quality matrix, defined by within-cohort median splits on both axes. Q1 = high quality, low cost; Q2 = high quality, high cost; Q3 = low quality, low cost; Q4 = low quality, high cost.
Quartile cut	A tighter version of the quadrant boundary, top quartile of one axis intersected with bottom quartile of the other axis, used to define the named Doing More With Less list.
Sign correction	The transformation applied to lower-is-better quality measures so they can be averaged with higher-is-better measures: corrected percentile = (1 minus percent_rank). Applied to mortality, readmission, complication, safety measures; not to HCAHPS.
Suppressed (composite)	A hospital for which the composite cannot be computed because of missing data on the relevant signals (or, for quality, fewer than two domains populated). Suppressed hospitals are excluded from quadrant placement and named lists but remain on the data layer.
Wage-index band	A peer-cohort axis based on quartiles of the FY 2026 IPPS wage index across the universe. Band 1 = lowest quartile (rural / low-cost-of-living areas); Band 4 = highest quartile (high-cost urban).

Winsorization	The data-cleaning operation of clipping outlier values to the 1st and 99th percentile bounds of the field's distribution. Applied at the cell level to HCRIS numeric data; each clipped value carries a flag.

C. Statistical Terms

Term	Definition	How It Appears in This Study
Percentile rank	Position of a value within a distribution, expressed on a 0 to 1 scale.	Every cost and quality signal in this paper is a within-cohort percentile rank.
Median split	A 50th-percentile cut on a distribution that produces two equal-size groups.	Used to define the four matrix quadrants on both axes.
Quartile	A 25th-percentile cut. Top quartile = top 25%; bottom quartile = bottom 25%.	Used for the tighter Doing More With Less cut and the cost-driver outlier flags.
Composite score	A summary measure produced by averaging multiple signal-level scores.	Cost composite averages 2 percentile-ranked cost signals; quality composite averages 5 domain scores.
Sign correction	Transforming a lower-is-better measure into a higher-is-better measure via (1 minus percent_rank).	Applied to mortality, readmission, complication, and safety measures so all quality signals point in the same direction.
Winsorization	Clipping outlier values to a specified percentile bound to limit their influence.	Applied at 1st and 99th percentile to HCRIS cell values; clipped rows carry an audit flag.
Suppression	Withholding a derived value when input coverage is insufficient.	Quality composite suppressed when fewer than 2 of 5 domains are populated; 39 hospitals affected.