# HRRP at 10 Years: Do Readmission Penalties Improve Hospital Performance?

Do penalties drive improvement? A 10-Year Analysis of Hospital Readmission Outcomes across 3,000 Institutions

Author: Jay Callery | Big Jacket Research

May 2026

Data: CMS Hospital Compare API, HRRP Penalty Files FY2016 to FY2025, VBP TPS
Pipeline: Python 3.12 | dbt-fusion 2.0 | BigQuery | Dagster 1.13.5
Repository: https://github.com/jcal-2/CMS-HRRP-Analysis

## Executive Summary

The Hospital Readmissions Reduction Program (HRRP) was designed to improve care quality by financially penalizing hospitals with excess readmissions. This analysis examines whether penalties achieved that objective, using 10 years of data (FY2016 to FY2025) across approximately 3,000 acute-care hospitals. The evidence is mixed: penalty severity has declined (suggesting some convergence toward expected rates), but 49% of tracked hospitals have been penalized every single year for a full decade.

**A cross-program analysis identifies 872 hospitals simultaneously penalized by both HRRP and Value-Based Purchasing (VBP), raising the question of whether penalty stacking reinforces disadvantage rather than driving improvement.**

### Key Findings at a Glance

| Metric | Value | Significance |
|---|---|---|
| Chronic penalty rate (all 10 years) | 49% (1,368 hospitals) | Largest single cohort in the data |
| Year-over-year persistence | 85.5% | FY2019 penalized, still penalized FY2025 |
| Chronic vs. never-penalized ERR | t=17.49, p=3.08e-62 | Cohen's d=1.38 (large effect) |
| Escape rate, high severity | 14.3% | 35% lower than low-severity hospitals |
| Compound-penalized (HRRP+VBP) | 872 hospitals | 69% chronic, avg 2.70 stars |
| Safety-net peer grouping benefit | 6pp vs. 3.4pp drop | Larger decline for safety-net hospitals |
| For-profit chronic rate | 23.4% | Higher than escaper rate (17.9%) |

## 1. Background

The Hospital Readmissions Reduction Program (HRRP), established under Section 3025 of the Affordable Care Act, began penalizing hospitals in FY2013 for excess 30-day readmissions. By FY2016, the program covered six condition and procedure groups: acute myocardial infarction, heart failure, pneumonia, chronic obstructive pulmonary disease, elective hip/knee replacement, and coronary artery bypass graft surgery.

The penalty mechanism is straightforward: hospitals with excess readmission ratios (ERR) above 1.0 for any measured condition face a reduction in Medicare base operating DRG payments, capped at 3%. The program runs alongside Value-Based Purchasing (VBP) and the Hospital-Acquired Conditions (HAC) Reduction Program, creating a triple-penalty environment for underperforming hospitals.

This study tracks the full FY2016 to FY2025 window (10 fiscal years) to evaluate the program's effectiveness from three angles. First, are penalties driving measurable improvement over time? Second, what separates hospitals that improve from those that do not? Third, what happens when hospitals face penalties from multiple CMS programs simultaneously?

Existing HRRP research has examined penalty rates and safety-net impact in individual fiscal years, but few studies have tracked the same hospitals longitudinally across a full decade or quantified the overlap between HRRP and VBP penalties at the hospital level. This analysis contributes both: a 10-year panel view of penalty persistence and a cross-program compound penalty identification that connects two programs CMS administers independently.

## 2. Methodology

### 2.1 Data Sources and Pipeline

Ten CMS API datasets were ingested programmatically alongside 10 years of HRRP penalty files and VBP Total Performance Scores. All data was loaded into Google BigQuery via a Python ingestion pipeline. A dbt pipeline (12 models: 8 staging, 2 fact, 2 dimension; 13 tests, all passing) transformed raw data into analytical tables. Dagster orchestrated the full end-to-end pipeline (4 assets, 23-second execution).

### 2.2 Cohort Definitions

Hospitals with complete 10-year tracking (n = 2,790) were classified into four behavioral cohorts based on penalty history:

| Cohort | Definition | n | % of Total |
|---|---|---|---|
| Chronic | Penalized all 10 years | 1,368 | 49.0% |
| Intermittent | Penalized 5 to 9 years | 672 | 24.1% |
| Escaper | Penalized early, escaped recently | ~200 | ~7% |
| Never Penalized | Penalized 0 years | 73 | 2.6% |

These cohort boundaries are analytical choices, not natural categories. A hospital penalized 9 of 10 years is classified as intermittent, not chronic, even though its trajectory is similar. The findings are robust to reasonable boundary adjustments; shifting the chronic threshold to 8+ years does not materially change the cohort performance profiles reported in section 3.4.

Hospitals that opened, closed, or merged during the study period are excluded from the 10-year tracking cohort (n = 2,790). If closures disproportionately affected the worst performers, which existing research suggests is plausible, the chronic cohort may understate the full scope of persistent penalty impact.

### 2.3 Statistical Methods

Group comparisons used Welch's t-test (unequal variance assumed) with Cohen's d for effect size. Significance threshold: $p < 0.05$. All reported p-values are two-tailed.

**Methodology Note:** CMS recalibrated the expected readmission calculation between FY2017 and FY2018, producing a visible jump in ERR values across all cohorts. This is a measurement artifact, not a change in care quality. All relevant charts annotate this discontinuity. Additionally, FY2016 and FY2017 HRRP files contain zero-coded ERR values for conditions not scored at a given hospital; these are excluded from cohort-level ERR averages to avoid understating readmission ratios. All ERR trajectory comparisons in this study use FY2018 as the baseline year.

## 3. Findings

### 3.1 Penalty Rate and Severity Over Time

The HRRP penalty rate has held between 75% and 83% for the entire decade (Figure 1). In every fiscal year since FY2016, at least three out of four eligible hospitals were penalized. However, penalty severity tells a different story: the average penalty among penalized hospitals peaked at 0.70% during FY2017 to FY2021, then dropped sharply to approximately 0.40% by FY2023.

The severity decline suggests partial progress: hospitals are clustering closer to the expected readmission threshold, producing smaller penalties. At the same time, the persistently high penalty rate raises a question about HRRP's design. Because the program uses a relative threshold (comparing hospitals to peers), even modest excess readmissions trigger a penalty. Whether the rate reflects sustained underperformance or a structural feature of the measurement model depends on which metric you prioritize.

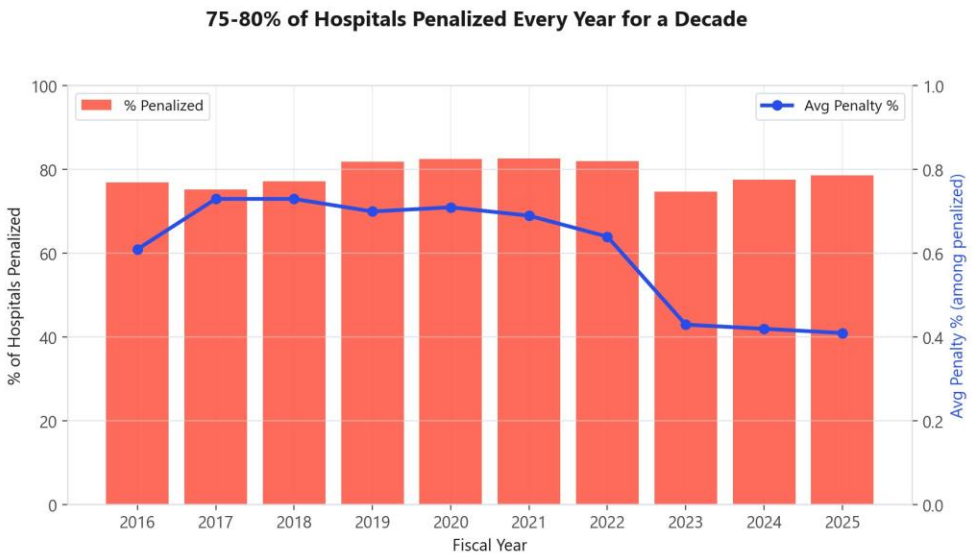



*Figure 1. Penalty rate remained at 75-80% while severity declined post-2022, suggesting hospital convergence toward expected readmission rates.*

## 3.2 Persistence: 49% Penalized All 10 Years

Among 2,790 hospitals with full 10-year tracking, 1,368 (49%) were penalized every single year (Figure 2). The distribution is heavily right-skewed: the single most common outcome is being penalized all 10 years, not escaping. An additional 21.7% were penalized 8 or 9 years. Only 73 hospitals (2.6%) were never penalized during the study period.

Year-over-year persistence reinforces this finding. Of 2,402 hospitals penalized in FY2019, 2,053 (85.5%) were still penalized in FY2025. The FY2024 to FY2025 transition matrix shows 2,140 hospitals remained penalized, while only 172 exited penalty status.

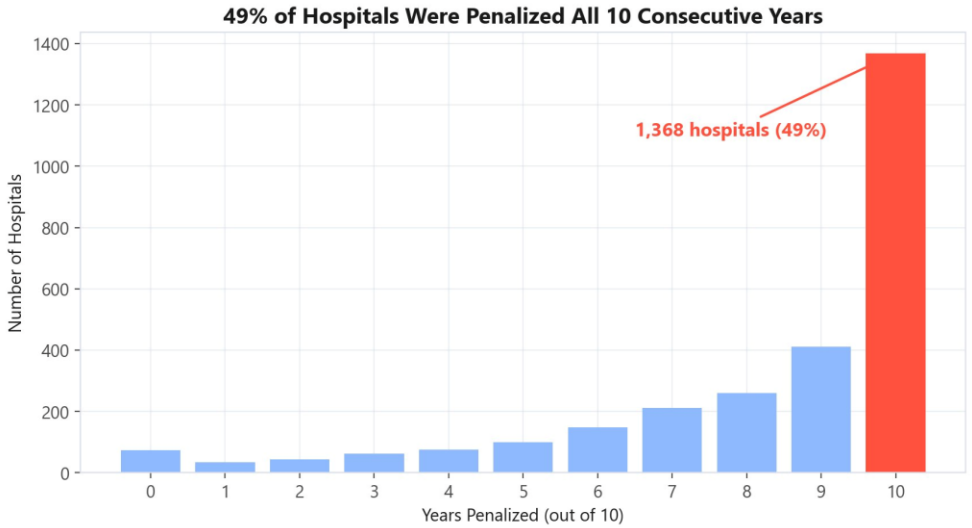



*Figure 2. Penalty persistence distribution across 2,790 hospitals with 10-year tracking. The tallest bar (1,368 hospitals) represents those penalized every single year.*

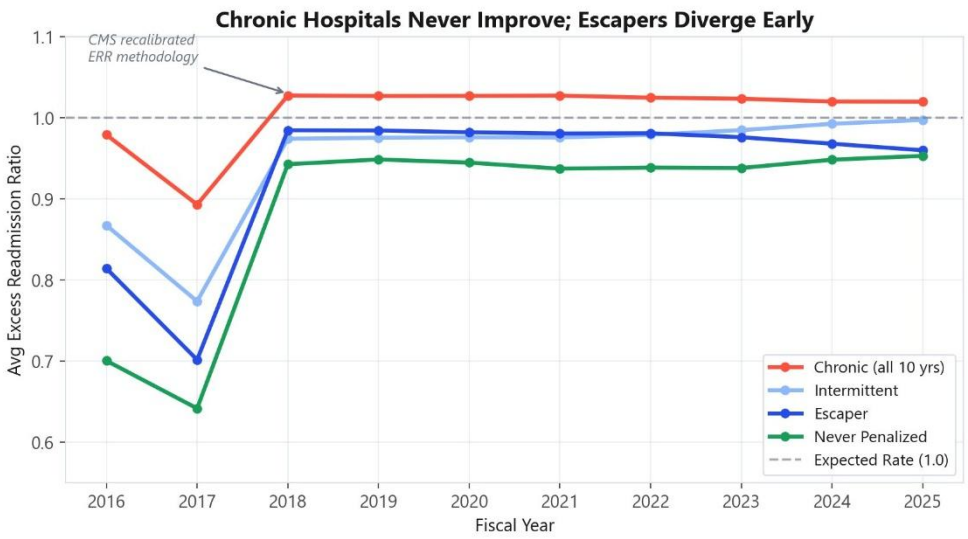
**Key Question:** If penalties are meant to incentivize improvement, what does it mean when 85.5% of penalized hospitals remain penalized six years later? The data alone cannot distinguish between hospitals that are not responding to incentives and hospitals facing structural barriers that penalties cannot address.

## 3.3 Cohort Trajectories: Where Improvement Shows Up (and Where It Does Not)

Tracking average ERR by cohort over time reveals four distinct behavioral patterns (Figure 3). Chronic hospitals maintained an ERR above 1.0 from FY2018 onward, with no visible improvement trajectory across seven subsequent years. Escapers, by contrast, started with similar ERR levels but diverged downward by FY2019, converging toward the never-penalized group.

This divergence suggests that penalties may work for some hospitals while having limited effect on others.

The intermittent cohort oscillates near the 1.0 threshold, consistent with hospitals near the penalty boundary who move in and out based on marginal shifts. The never-penalized group maintains a consistent ERR below 0.95.




The escaper trajectory is particularly instructive. Over the 10-year window, escapers reduced their average penalty severity from 0.40% to 0.00%. Their average ERR declined modestly from approximately 0.98 to 0.96 (post-FY2018 recalibration baseline). Chronic hospitals, by contrast, saw penalty severity drop only from 0.70% to 0.50%, with ERR stuck at approximately 1.02. They received slightly smaller penalties over time but never improved enough to escape. Notably, escapers maintained ERR values below 1.0 throughout the study period, raising the question of whether their exit reflects genuine improvement or structural characteristics that the pre-FY2019 methodology did not account for. Section 4.7 examines whether the escaper trajectory reflects operational improvement or structural differences in hospital size and business model.

*Figure 3. ERR trajectories by cohort, FY2016 to FY2025. The 2017 to 2018 jump across all cohorts reflects a CMS methodology recalibration, not a care quality change. FY2016 and FY2017 values are further affected by zero-coded ERR entries; FY2018 is the reliable baseline for trend analysis.*

## 3.4 Multi-Dimensional Performance Gap

Chronic hospitals perform worse on every measured dimension, not just readmissions (Figure 4). The performance gap between chronic and never-penalized hospitals is statistically significant and large:

| Metric | Chronic | Never Penalized | Gap |
|---|---|---|---|
| Star Rating | 2.94 | 3.91 | -0.97 |
| Avg ERR | 1.018 | 0.895 | +0.123 |
| MSPB Ratio | 1.007 | 0.933 | +0.074 |
| HAC Score | 4,586 | 1,389 | +3,197 |
| Measures Above Expected | 2.67 | 0.30 | +2.37 |

Statistical test (ERR, chronic vs. never-penalized): **t = 17.49, p = 3.08e-62, Cohen's d = 1.38.** This is not a marginal difference; these are fundamentally different hospital populations.

HAC deterioration for chronic hospitals is independently significant (t = 9.72, p = 8.21e-22), confirming that chronic readmission penalties correlate with broader patient safety challenges.

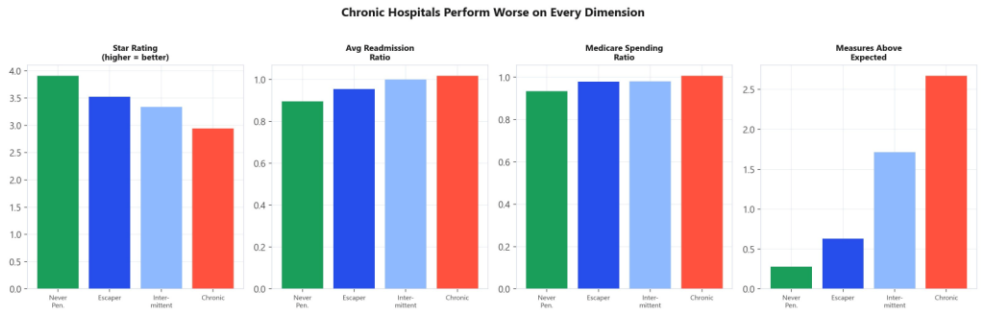



*Figure 4. Chronic hospitals underperform on all four dimensions: star rating, readmission ratio, Medicare spending, and measures above expected.*

## 3.5 Escape Rate Declines with Initial Severity

Hospitals entering the penalty system at higher severity levels are significantly less likely to escape (Figure 5). Low-severity hospitals (initial penalty 0 to 0.2%) escape at 22.0%, while high-severity hospitals (0.5%+) escape at only 14.3%, a 35% reduction in escape probability.

The severity-persistence relationship has a clear policy implication: early intervention, before penalties compound, offers the highest return. Once a hospital reaches high severity, the structural factors driving excess readmissions are deeply embedded.

Timing matters as much as severity. Hospitals that escape penalties tend to do so within the first 3 to 4 years. After that window, the probability of escape drops dramatically. This suggests that the behavioral response to penalties, if it occurs at all, happens early. Hospitals that have not improved by year 4 are unlikely to improve in years 5 through 10, regardless of continued financial pressure. These escape rates do not control for hospital size or eligibility changes; Section 4.7.4 examines whether smaller hospitals exit HRRP measurement thresholds rather than improving readmission performance.

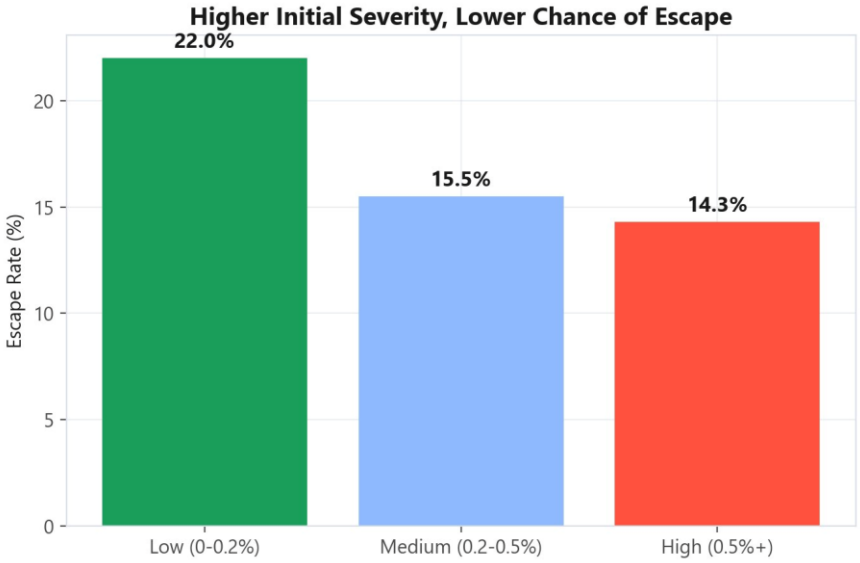



*Figure 5. Escape rate by initial penalty severity. Higher starting penalties predict lower escape probability.*

## 3.6 Regional and Ownership Disparities

Geography and ownership type both predict penalty outcomes. Regionally, the Northeast has the highest penalty rate at 87.2%, nearly 17 percentage points above the West at 70.3% (Figure 6). The Midwest and South cluster near the national average of 78.6%.

A volume versus rate distinction is worth noting: the South accounts for 44.5% of all chronic hospitals despite having a penalty rate near the national average. The region's large share of the hospital population means that even an average penalty rate produces the largest absolute count of chronically penalized institutions. Policy interventions targeting chronic hospitals by count, not rate, would disproportionately focus on the South.

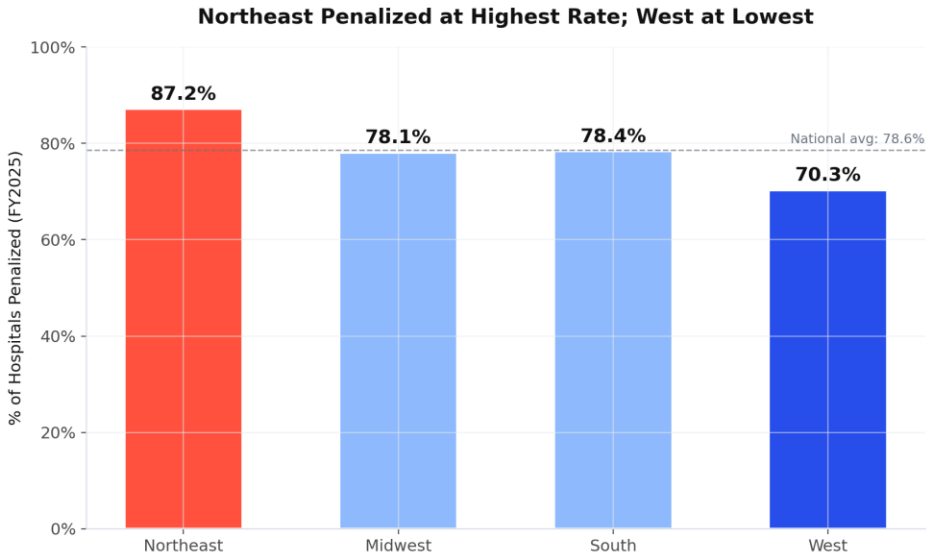



*Figure 6. Regional penalty rates in FY2025. The Northeast exceeds the national average by 8.6 percentage points; the West falls 8.3 points below.*

The regional gap becomes even more striking when examining cohort composition (Figure 7). The Northeast has the worst chronic-to-escaper ratio: 63.5% of its penalized hospitals are chronic, while only 8.8% are escapers. The West inverts this pattern (33.0% chronic, 22.5% escapers), suggesting that local market dynamics, payer mix, and regulatory environment shape outcomes beyond hospital-level management.

| Region | % Chronic | % Escapers | Ratio (Chronic/Escaper) |
|---|---|---|---|
| Northeast | 63.5% | 8.8% | 7.2x |
| South | 50.0% | 16.3% | 3.1x |
| Midwest | 41.6% | 16.0% | 2.6x |
| West | 33.0% | 22.5% | 1.5x |

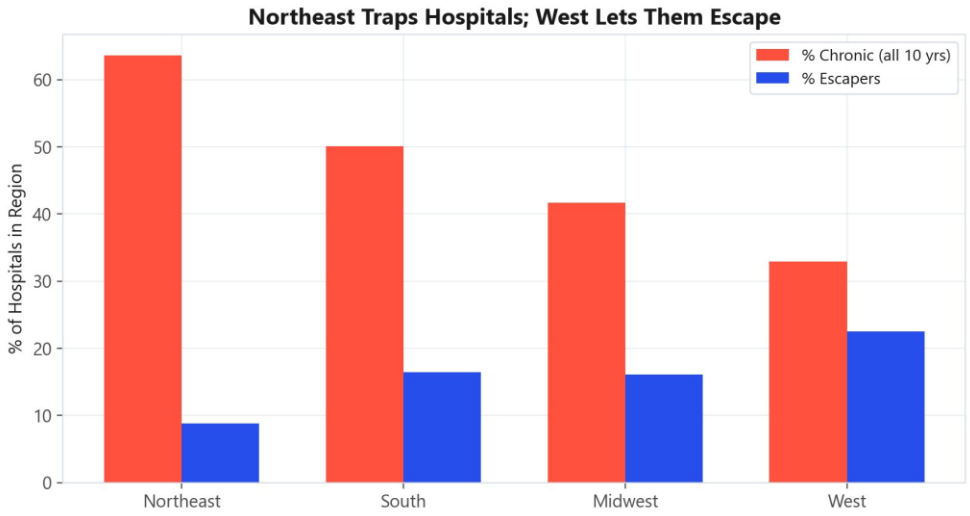



*Figure 7. The Northeast traps hospitals at a 7:1 chronic-to-escaper ratio; the West at 1.5:1.*

Ownership analysis produced a counterintuitive finding: For-Profit hospitals are more likely to be chronic (23.4% of chronic cohort) than escapers (17.9%), contrary to the expectation that profit-motivated institutions would invest

more aggressively to eliminate penalties. Government hospitals escape at the highest rate (21.0% of escapers vs. 11.7% of chronic).

Tracking penalized hospital counts by ownership over time (Figure 8) reveals a stable composition: Non-Profits consistently account for approximately 65% of all penalized hospitals, For-Profits 24%, and Government 11%. The ownership mix does not shift meaningfully across the decade, suggesting that the penalty system affects all ownership types proportionally. The FY2023 COVID dip is visible across all three groups.

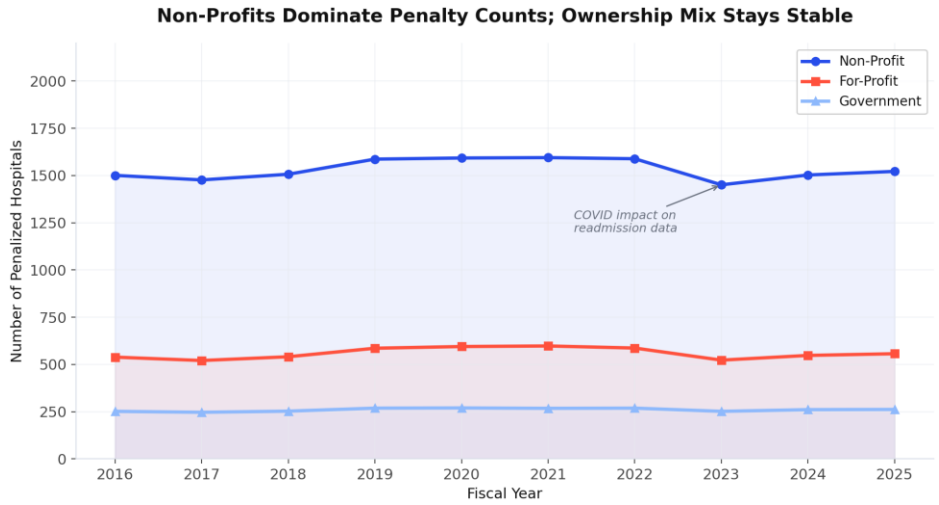



*Figure 8. Penalized hospital counts by ownership type, FY2016 to FY2025. The ownership composition remains stable, indicating penalties affect all types proportionally.*

## 3.7 Peer Grouping Helped Safety-Net Hospitals Modestly

CMS introduced peer grouping in FY2019 to compare safety-net hospitals (those with >=50% dual-eligible patients) against similar institutions rather than the full hospital population. The data shows a modest benefit (Figure 9):

- Safety-net penalty rate: 78.5% (FY2019) to 72.7% (FY2025), a 5.8 percentage point drop
- Non-safety-net penalty rate: 81.5% (FY2019) to 78.1% (FY2025), a 3.4 percentage point drop

The differential (approximately 6pp vs. 3.4pp) suggests peer grouping provided incremental help, but did not fundamentally change outcomes. The majority of safety-net hospitals remain penalized. The sharp dip visible in FY2023 likely reflects COVID-era disruptions to readmission measurement.

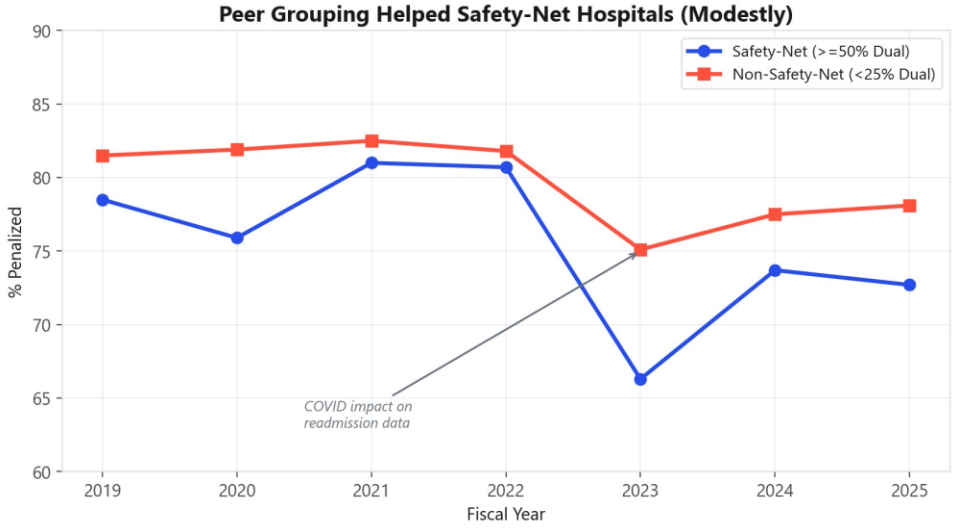



*Figure 9. Safety-net hospitals saw a larger penalty rate decline than non-safety-net after peer grouping was introduced in FY2019, but the gap remains modest.*

## 4. Compound Penalty Analysis: When HRRP Meets VBP

While HRRP and VBP are typically examined independently, this section quantifies the overlap between these two programs at the hospital level to understand what happens when penalties stack.

### 4.1 Identifying Compound-Penalized Hospitals

A hospital was classified as compound-penalized if it met both criteria simultaneously: (1) an active HRRP penalty (reduction percentage > 0%), and (2) a VBP Total Performance Score below the national median of 29.17. This identified 872 hospitals.

The national median was selected as the VBP threshold because it represents the point at which a hospital's total performance falls below the typical peer. Alternative thresholds (25th percentile, mean, or the actual VBP penalty formula) would change the count but not the directional finding: hospitals penalized by both programs perform materially worse than those penalized by one or neither.

### 4.2 Compound vs. Non-Compound Profile

Compound-penalized hospitals are systematically worse across every metric:

| Metric | Compound (n=872) | Single/None | Difference |
|---|---|---|---|
| Avg Star Rating | 2.70 | 3.63 | -0.93 |
| Avg ERR | 1.015 | 0.992 | +0.023 |
| MSPB Ratio | 1.023 | 0.964 | +0.059 |
| HAC Score | 4,915 | 3,442 | +1,473 |

**Risk:** 69% of compound-penalized hospitals are chronic (penalized all 10 HRRP years). This suggests compound penalty status is not a transient condition but a structural feature of these institutions.

### 4.3 Sales Tier Distribution

For commercial application, the 872 compound-penalized hospitals were segmented into priority tiers based on compound severity scores (combining HRRP penalty percentage, VBP TPS distance from median, and star rating):

| Tier | Count | Profile |
|---|---|---|
| Tier 1: Critical | 241 | Highest severity, greatest intervention need |
| Tier 2: High Priority | 360 | Significant multi-program penalty exposure |
| Tier 3: At Risk | 253 | Moderate compound penalty, potential for improvement |
| Tier 4: Emerging | 18 | Early-stage compound penalty indicators |

### 4.4 Enriched Dataset

The full compound-penalized hospital list (872 hospitals) is available as an enriched CSV with the following fields: sales tier, compound severity score, estimated HRRP loss, sales themes (e.g., "Readmission Reduction (urgent) | Quality Improvement Program"), and recommended vendor categories (e.g., "Discharge Planning / Care Coordination | Remote Patient Monitoring").

## 4.5 Top 15 Compound-Penalized Hospitals

The following hospitals rank highest on the compound severity score, which combines HRRP penalty percentage, VBP TPS distance from median, and star rating. Estimated HRRP losses are approximations based on penalty percentage and hospital size.

| Hospital | City | ST | Stars | HRRP % | VBP TPS | Est. Loss |
|---|---|---|---|---|---|---|
| Louis A Weiss Memorial Hospital | Chicago | IL | 1.0 | 1.73% | 12.9 | $865K |
| Sycamore Shoals Hospital | Elizabethton | TN | 2.0 | 2.00% | 15.8 | $1.0M |
| Anna Jaques Hospital | Newburyport | MA | 3.0 | 3.00% | 25.9 | $1.5M |
| Community Medical Center | Toms River | NJ | 1.0 | 2.29% | 20.0 | $1.1M |
| Johnson City Medical Center | Johnson City | TN | 1.0 | 1.11% | 9.4 | $555K |
| St Lucie Medical Center | Port St. Lucie | FL | 1.0 | 3.00% | 28.6 | $1.5M |
| ProMedica Monroe Regional Hospital | Monroe | MI | 1.0 | 0.85% | 7.7 | $425K |
| Sturdy Memorial Hospital | Attleboro | MA | 3.0 | 3.00% | 28.8 | $1.5M |
| Brown Univ Health Morton Hospital | Taunton | MA | 1.0 | 2.36% | 22.7 | $1.2M |
| Great Plains Regional Medical Ctr | Elk City | OK | 2.0 | 1.92% | 18.7 | $960K |
| UPMC St Margaret | Pittsburgh | PA | 3.0 | 2.10% | 20.8 | $1.1M |
| Hendrick Medical Center Brownwood | Brownwood | TX | 1.0 | 1.33% | 7.3 | $665K |
| Chan Soon-Shiong Med Ctr Windber | Windber | PA | 2.0 | 1.74% | 12.5 | $870K |
| Baystate Medical Center | Springfield | MA | 2.0 | 1.74% | 12.8 | $870K |
| Affiliate of Vitruvian Health | Cleveland | TN | 2.0 | 0.74% | 6.6 | $370K |

## 4.6 Worst-Performing Subgroup

For-Profit hospitals in the Northeast represent the most distressed subgroup: 2.20 average star rating, 93.3% chronic rate, and the highest average HRRP penalty at 0.77%. Top estimated annual HRRP financial losses reach approximately $1.5M per hospital (Anna Jaques Hospital MA, Sturdy Memorial Hospital MA, St Lucie Medical Center FL).

## 4.7 Financial Viability of Compound-Penalized Hospitals

The preceding sections establish which hospitals are compound-penalized and how they perform on quality metrics. This section asks a different question: can these hospitals afford to invest in interventions that would reduce their penalties? CMS Cost Report data (HCRIS Form 2552-10, FY2022 to FY2025) was matched to all 872 compound-penalized hospitals at a 100% match rate on facility CCN.

### Survivor Bias in the Chronic Cohort

Chronic-penalized hospitals are not the most financially distressed. Among the 872 compound-penalized hospitals, chronic institutions (penalized all 10 years) report a median operating margin of +0.34%, while intermittent hospitals report -1.62%.

| Cohort | Hospitals | With Margin | Avg Margin | Median Margin | Avg Beds |
|---|---|---|---|---|---|
| Chronic | 601 | 595 (99%) | +0.66% | +0.34% | 299 |
| Intermittent | 271 | 264 (97%) | -1.67% | -1.62% | 257 |

This pattern is consistent with survivorship bias. Hospitals that lack the financial capacity to absorb a decade of penalties have already exited through closure, acquisition, or service line reduction. The chronic cohort is not a group that failed to respond to incentives. It is the group that survived despite them.

### Validated Tiers: Penalty Severity Meets Financial Capacity

To identify hospitals that are both heavily penalized and financially capable of investing in improvement, the 241 Tier 1 Critical hospitals were split based on financial viability (positive operating margin and 100+ beds):

| Subgroup | Count | Median Beds | Median Revenue | Median Margin | Median FTE |
|---|---|---|---|---|---|
| Tier 1 Qualified | 123 | 259 | $394M | +8.55% | 1,237 |
| Tier 1 Marginal | 118 | 156 | $153M | -4.20% | 648 |

The 123 Tier 1 Qualified hospitals are large, well-resourced institutions operating at healthy margins despite carrying the highest compound penalty severity in the dataset. They have the financial capacity to fund readmission reduction programs. The penalty has not driven them to invest.

### Scale Disadvantage Across All Tiers

Sub-100 bed hospitals are disproportionately financially distressed regardless of penalty tier. Across all tiers combined, 67% of sub-100 bed hospitals operate at negative margins, compared to 49% of hospitals with 300+ beds.

| Tier | <100 Beds: Pos / Neg | 100-299: Pos / Neg | 300+: Pos / Neg |
|---|---|---|---|
| Tier 1, Critical | 11 / 17 | 80 / 55 | 43 / 35 |
| Tier 2, High Priority | 16 / 24 | 93 / 84 | 63 / 78 |
| Tier 3, At Risk | 30 / 38 | 57 / 50 | 36 / 40 |
| Tier 4, Emerging | 0 / 6 | 1 / 3 | 3 / 5 |

Tier 4 Emerging is the most financially distressed subgroup (22% positive margin). These are recently penalty-prone institutions that may represent the next wave of closures rather than future chronic survivors.

### The Escaper Paradox: Smaller, Not Healthier

Section 3.3 identified escapers as hospitals that reduced their ERR and exited penalty status, suggesting penalties drove operational improvement. Financial profiling complicates this narrative.

| Metric | Chronic (n=1,381) | Escaper (n=486) | Escaper as % of Chronic | p-value (MWU) |
|---|---|---|---|---|
| Median Bed Count | 196 | 58 | 30% | 4.5e-70 |
| Median Revenue | $320M | $132M | 41% | 1.7e-36 |
| Median FTE | 1,111 | 432 | 39% | 1.0e-46 |
| Median Margin | -0.80% | -0.44% | indistinguishable | 0.74 (ns) |

Escapers are not financially healthier than chronic hospitals. They are dramatically smaller across three independent size measures (beds, revenue, FTE), all at extreme significance levels ($p < 1e\text{-}30$). Operating margins are statistically indistinguishable.

This size gap suggests an alternative explanation for the escape trajectory. Small, low-volume hospitals may fall out of HRRP scoring eligibility rather than improving their readmission performance. Mechanisms include insufficient case volume to compute reliable ERR, Critical Access Hospital designation (which exempts from HRRP), and service line closures that remove readmission denominators.

Chronic hospitals, with median bed counts of 196 and revenues of $320M, generate enough case volume across measured conditions to remain in the HRRP denominator permanently. They cannot shrink their way out of penalty eligibility.

Controlling for bed count via OLS regression sharpens this finding further. The cohort effect on operating margin remains non-significant ($p = 0.085$). However, revenue and FTE effects reverse: at equivalent bed counts, escaper hospitals generate $147M more in net patient revenue ($p < 1e\text{-}12$) and employ 420 more FTE ($p < 1e\text{-}9$) than chronic hospitals. This is a Simpson's Paradox result: escapers appeared to have lower revenue and fewer staff in the raw comparison only because they are smaller. Bed-for-bed, they are more revenue-dense and labor-dense than chronic hospitals.

The combination of smaller size and higher per-bed intensity suggests escapers concentrate in a different business model, not just a smaller version of chronic hospitals. Two mechanisms fit this pattern: Critical Access Hospital (CAH) designation, which exempts facilities from HRRP entirely and reimburses on a cost-plus basis (explaining higher FTE-per-bed), and service mix shifts toward outpatient or specialty lines that fall outside HRRP's inpatient readmission denominators (explaining higher revenue-per-bed).

A five-mechanism probe using scale-normalized metrics confirms this interpretation across 1,866 hospitals with matched HCRIS financials:

| Mechanism | Chronic | Escaper | Direction | MWU p-value |
|---|---|---|---|---|
| Discharges per bed | 49.4 | 38.6 | chronic > escaper | < 1e-27 |
| Revenue per discharge | $33,813 | $52,931 | escaper > chronic | < 1e-55 |
| Expense per discharge | $34,856 | $53,990 | escaper > chronic | < 1e-57 |
| FTE per discharge | 0.125 | 0.188 | escaper > chronic | < 1e-46 |
| HRRP measures scored | 5.39 of 6 | 4.41 of 6 | chronic > escaper | direct evidence |

The measure-coverage gap is direct evidence of volume-threshold dropout. Escapers are scored on an average of 4.41 of 6 HRRP conditions, compared to 5.39 for chronic hospitals. Eighteen percent of escapers are scored on 3 or fewer conditions, compared to 0.6% of chronic hospitals. Fewer scored conditions means fewer opportunities to trigger the excess readmission penalty, regardless of clinical performance on the conditions that remain.

The per-discharge metrics reveal a distinct institutional profile. Escapers earn 57% more revenue per discharge ($52,931 vs $33,813 median), consistent with higher-acuity case mix or higher-payer-mix facilities. They employ 50% more FTE per discharge (0.188 vs 0.125), a pattern consistent with teaching hospitals. They run lower occupancy per bed (38.6 vs 49.4 discharges per bed), consistent with service lines that generate fewer but higher-value inpatient cases.

The ERR trajectory reinforces this interpretation. After correcting for a FY2016/2017 data artifact (see Section 2.3), escapers maintained an average ERR below 1.0 across the entire study period (0.982 in FY2018, declining to 0.960 in FY2025). They did not arrive at lower ERR through a visible improvement event. The 0.04-point gap between escapers and chronic hospitals was present from the start of the panel and widened only modestly. Escapers spent a median of 6 years penalized (IQR 4 to 8) before exiting, indicating they were genuinely entrenched in the penalty system, not borderline cases. Their exit correlates with the FY2019 introduction of peer grouping, which reclassified hospitals into more favorable comparison groups based on dual-eligible patient proportions.

**Research Implication:** Before attributing escaper trajectories to operational improvement driven by penalty incentives, future analyses should control for hospital size and flag CAH designation. The financial data suggests that the escape signal reflects structural business model differences and measurement eligibility changes rather than care quality improvement within the same hospital model. A research-grade evaluation of HRRP effectiveness should restrict comparisons to PPS hospitals with matched bed counts.

## 5. Policy Implications

1. **Persistence raises questions about incentive design.** When 49% of hospitals are penalized all 10 years and 85.5% remain penalized over a 6-year window, the program reliably identifies struggling hospitals. Whether persistent penalties continue to motivate improvement, or whether they become an accepted cost of doing business, is a question the data raises but cannot fully answer.
2. **Peer grouping is necessary but insufficient.** Safety-net hospitals saw approximately 6 percentage points more improvement than non-safety-net hospitals after peer grouping was introduced in FY2019, but the majority remain penalized. Addressing structural disadvantage (payer mix, community health burden, resource constraints) requires more than peer comparison.
3. **Early severity predicts long-term persistence.** Hospitals entering with high initial penalties have a 35% lower escape rate than those entering at low severity (14.3% vs. 22.0%). Policy interventions and vendor solutions should target hospitals at first penalty, before institutional patterns calcify.
4. **Regional disparities need investigation.** The Northeast's 7:1 chronic-to-escaper ratio vs. the West's 1.5:1 ratio suggests that local market dynamics, state Medicaid policies, and regional health infrastructure play significant roles beyond individual hospital management decisions.
5. **Compound penalties warrant closer examination.** The 872 hospitals penalized by both HRRP and VBP simultaneously are the lowest performers in the Medicare system (2.70 avg stars, 69% chronic). Whether

penalty stacking intensifies the incentive to improve or compounds existing disadvantage is a critical design question for CMS.

### 5.1 The data suggests four concrete policy responses worth examining

These responses are informed by the data patterns observed in this study but remain untested hypotheses. Each would require further analysis, pilot implementation, and evaluation before adoption.

- **Aggregate penalty cap across programs.** HRRP and VBP currently operate independently, meaning a hospital could lose 3% from HRRP plus another 2 to 3% from VBP. Capping the combined reduction (for example, at 3% total) would prevent compound hemorrhaging of the resources needed to fund improvement: care coordinators, discharge planning staff, and follow-up programs.
- **Extend peer grouping to VBP.** CMS introduced peer grouping for HRRP in FY2019 and it measurably helped safety-net hospitals (section 3.7). VBP still compares all hospitals against national benchmarks, which disadvantages hospitals serving complex, high-acuity populations. Applying the same peer grouping logic to VBP would bring consistency across programs.
- **Reinvestment mandates.** Instead of penalty revenue flowing back to CMS's general fund, redirect a portion into technical assistance grants for compound-penalized hospitals. The Partnership for Patients program demonstrated that targeted support can drive measurable improvement where penalties alone have not.
- **Improvement trajectory scoring.** Rather than penalizing absolute performance, reward improvement velocity. A hospital that moves from a TPS of 15 to 25 is making meaningful progress even if it remains below the national median. Trajectory-based scoring would shift the incentive from "be above the line" to "move in the right direction," which better matches the reality that improvement from a low baseline takes multiple years.

## 6. Limitations of study

- Observational study: all findings are correlational. Penalty persistence does not prove penalties caused worse outcomes; confounders (community health, payer mix, staffing) likely contribute.
- ERR methodology change between FY2017 and FY2018 limits direct year-over-year comparisons across that boundary. Post-2018 trends within cohorts remain valid.
- Dual-eligible proportion data is only available from FY2019 onward, preventing a full pre/post analysis of peer grouping impact.
- COVID-19 disrupted readmission measurement in FY2022 to FY2024. CMS excluded some performance periods, which may affect penalty rates in those years.
- The compound penalty analysis uses VBP TPS median as a threshold. Different threshold choices would produce different hospital counts.
- Estimated HRRP financial losses are approximations based on publicly available penalty percentages and hospital size proxies; actual losses depend on Medicare case mix and volume.
- The penalty severity decline observed post-2022 may partially reflect COVID-era measurement disruptions rather than genuine hospital improvement. This trend should be re-evaluated as post-pandemic performance data stabilizes.
- The for-profit ownership finding (higher chronic rate than expected) does not control for hospital size, case mix, geography, or payer mix. The raw ownership percentages may reflect confounders rather than ownership-driven behavior differences.
- Hospitals that closed during the study period are excluded from longitudinal tracking. If financial pressure from persistent penalties contributed to closures, the surviving cohort understates the program's full impact on the most vulnerable institutions.

## 7. Conclusion

### Do CMS penalties improve hospital performance?

The answer after 10 years is: for some hospitals, yes. Severity has declined, escapers demonstrate real improvement trajectories, and peer grouping has provided incremental relief for safety-net institutions.

But for nearly half of all tracked hospitals, penalties have become a permanent condition rather than a catalyst for change. The 1,368 hospitals penalized every year, and the 872 penalized by multiple CMS programs simultaneously, suggest the incentive model works best for hospitals closest to the threshold and least well for those furthest from it.

The data points toward supplementing penalties with targeted support: early intervention at first penalty, expanded peer grouping, and cross-program coordination. The goal is not to soften accountability, but to match the intensity of the intervention to the depth of the challenge.

Here are some concrete actions to take:

1. **Early intervention at first penalty.** The data shows escape rates drop dramatically after year 3 to 4. Hospitals that haven't improved by then almost never do. The action: deploy support (care coordination programs, discharge planning resources, remote monitoring) the moment a hospital receives its first penalty, not after a decade of compounding.
2. **Expanded peer grouping.** Peer grouping works for HRRP (6pp improvement for safety-net hospitals) but doesn't exist for VBP. The action: extend the same peer comparison methodology to Value-Based Purchasing so hospitals aren't being benchmarked against facilities serving fundamentally different patient populations.
3. **Cross-program coordination (aggregate penalty cap).** HRRP and VBP operate independently, stacking penalties on the same 872 hospitals without either program accounting for the other. The action: cap combined penalties across programs (e.g., 3% total) so compound-penalized hospitals retain enough operating margin to fund the improvements both programs are trying to incentivize.
4. **Match intervention intensity to challenge depth.** The current design applies the same mechanism (financial penalty) regardless of whether a hospital is 0.1% above the threshold or has been at the 3% cap for a decade. The action: trajectory-based scoring that rewards improvement velocity, not just absolute position. A hospital moving from 1.10 ERR to 1.03 ERR is making real progress even if it's still above 1.0.

# Appendix:

# Glossary

## A. Acronyms

| Acronym | Full Term |
|---|---|
| ACA | Affordable Care Act |
| AMI | Acute Myocardial Infarction |
| CABG | Coronary Artery Bypass Graft |
| CMS | Centers for Medicare & Medicaid Services |
| COPD | Chronic Obstructive Pulmonary Disease |
| DRG | Diagnosis-Related Group |
| ERR | Excess Readmission Ratio |
| FY | Fiscal Year |
| HAC | Hospital-Acquired Conditions Reduction Program |
| HF | Heart Failure |
| HRRP | Hospital Readmissions Reduction Program |
| MSPB | Medicare Spending Per Beneficiary |
| PN | Pneumonia |
| TPS | Total Performance Score |
| VBP | Value-Based Purchasing |

## B. Key Terms Used in This Analysis

### Penalty Program Terms

| Term | Definition |
|---|---|
| Penalty (HRRP) | A reduction in Medicare base operating DRG payments applied to hospitals with excess readmission ratios above 1.0 in any measured condition. Maximum reduction is capped at 3% of total Medicare reimbursement. |
| Penalty Rate | The percentage of eligible hospitals receiving any HRRP penalty in a given fiscal year. In this study, the rate ranged from 74.7% to 82.6%. |
| Penalty Severity | The average penalty percentage among hospitals that were penalized. Distinct from penalty rate: a hospital can be penalized (rate) at a small or large amount (severity). |
| Base Operating DRG Payment | Medicare's per-case reimbursement to hospitals, determined by the patient's diagnosis-related group. HRRP penalties are applied as a percentage reduction to this payment across all Medicare discharges, not just readmission-related cases. |
| 30-Day Readmission | An unplanned return to any acute-care hospital within 30 days of discharge. CMS counts readmissions to any hospital, not just the original discharging facility. |
| Peer Grouping | A CMS methodology introduced in FY2019 that compares a hospital's readmission performance against hospitals with similar proportions of dual-eligible patients, rather than against the full national population. |

### Cohort Definitions

| Term | Definition |
|---|---|

| | |
|---|---|
| Chronic | A hospital penalized in all 10 consecutive fiscal years studied (FY2016 to FY2025). Represents 49% of hospitals with full tracking (n = 1,368). |
| Intermittent | A hospital penalized in 5 to 9 of the 10 fiscal years. These hospitals oscillate near the penalty threshold. |
| Escaper | A hospital that was penalized in early years of the study period but exited penalty status in later years, demonstrating measurable improvement. |
| Never Penalized | A hospital that received no HRRP penalty in any of the 10 fiscal years (n = 73). Used as the control group for performance comparisons. |
| Escape Rate | The percentage of initially penalized hospitals that exit penalty status. Calculated within severity subgroups (low: 22.0%, medium: 15.5%, high: 14.3%). |
| Persistence Rate | The percentage of hospitals penalized in one year that remain penalized in a subsequent year. The 85.5% figure tracks FY2019 to FY2025. |

## Hospital Classification Terms

| Term | Definition |
|---|---|
| Safety-Net Hospital | A hospital where 50% or more of patients are dual-eligible (qualifying for both Medicare and Medicaid). These hospitals typically serve lower-income, higher-acuity populations. |
| Dual-Eligible | A patient who qualifies for both Medicare (age/disability) and Medicaid (income). High dual-eligible proportions indicate a hospital serves economically disadvantaged populations. |
| For-Profit | An investor-owned hospital operated to generate returns for shareholders. In this study, for-profit hospitals were more likely to be chronic (23.4%) than escapers (17.9%). |
| Non-Profit | A tax-exempt hospital, typically community-based or affiliated with a health system or religious organization. Comprises the majority of chronic hospitals (64.5%). |
| Government | A publicly owned hospital (federal, state, or county). Government hospitals had the highest escape rate among ownership types (21.0% of escapers vs. 11.7% of chronic). |
| Star Rating | The CMS Overall Hospital Quality Star Rating, a 1 to 5 composite score incorporating safety, readmission, patient experience, timely care, and mortality measures. Higher is better. |

## Performance Metrics

| Term | Definition |
|---|---|
| Excess Readmission Ratio (ERR) | The ratio of a hospital's predicted readmissions to its expected readmissions for a given condition. An ERR of 1.0 means performance matches expectations. Above 1.0 triggers a penalty; below 1.0 means better than expected. |
| Medicare Spending Per Beneficiary (MSPB) | A ratio comparing a hospital's Medicare spending to the national median. Values above 1.0 indicate higher-than-average spending per patient episode. |
| Hospital-Acquired Condition (HAC) Score | A composite measure of patient safety events occurring during hospitalization (infections, falls, complications). Higher scores indicate more adverse events. Used by the HAC Reduction Program to penalize the worst-performing quartile. |
| Total Performance Score (TPS) | The VBP composite score combining clinical outcomes, patient experience, safety, and efficiency measures. The national median used in this study was 29.17; hospitals below this threshold were classified as VBP-penalized. |
| Compound Severity Score | A composite metric created for this analysis combining HRRP penalty percentage, VBP TPS distance from median, and star rating into a single score for ranking compound-penalized hospitals. |
| Measures Above Expected | The count of readmission measures for which a hospital's ERR exceeds 1.0. Chronic hospitals averaged 2.67 measures above expected vs. 0.30 for never-penalized. |

## HRRP Measured Conditions and Procedures

HRRP tracks 30-day readmission rates for the following six condition and procedure groups. A hospital can be penalized for excess readmissions in any combination of these:

| Condition/Procedure | Abbreviation | What It Covers |
|---|---|---|
| Acute Myocardial Infarction | AMI | Heart attacks. Readmissions within 30 days of discharge for heart attack treatment. |
| Heart Failure | HF | Chronic condition where the heart cannot pump blood efficiently. One of the highest-volume readmission categories. |
| Pneumonia | PN | Lung infections requiring hospitalization. Includes both community-acquired and hospital-acquired cases. |
| Chronic Obstructive Pulmonary Disease | COPD | Progressive lung disease including emphysema and chronic bronchitis. Added to HRRP in FY2015. |
| Elective Hip/Knee Arthroplasty | THA/TKA | Planned total hip or knee replacement surgery. Added to HRRP in FY2015. Represents a surgical (vs. medical) readmission category. |
| Coronary Artery Bypass Graft | CABG | Open-heart surgery to bypass blocked coronary arteries. Added to HRRP in FY2017. The most recent addition to the program. |

## C. Statistical Terms

This analysis uses standard inferential statistics to compare hospital cohorts. The following terms appear in the findings:

| Term | Definition | How It Appears in This Study |
|---|---|---|
| Welch's t-test | A statistical test comparing the means of two groups when the groups may have different variances and different sample sizes. More robust than the standard t-test for unequal groups. | Used to compare chronic vs. never-penalized hospital ERR values. |
| t-statistic | The test output measuring how many standard errors separate the two group means. Larger values indicate greater separation between groups. | t = 17.49 for chronic vs. never-penalized ERR (very large separation). |
| p-value | The probability of observing a result this extreme if there were no real difference between groups. Smaller values indicate stronger evidence of a genuine difference. | p = 3.08e-62 (effectively zero, meaning the difference is not due to chance). |
| Statistical Significance | A result is statistically significant when the p-value falls below a pre-set threshold (typically 0.05, or 5%). It means the observed difference is unlikely to be random. | All reported comparisons use $p < 0.05$ as the threshold. |
| Cohen's d | A measure of effect size quantifying the magnitude of difference between two groups in standard deviation units. 0.2 = small, 0.5 = medium, 0.8+ = large. | d = 1.38 for chronic vs. never-penalized ERR (large effect). |
| Effect Size | A standardized measure of how large a difference or relationship is, independent of sample size. Unlike p-values, effect size tells you whether the difference matters practically, not just statistically. | Cohen's d is the effect size measure used throughout. |
| Two-Tailed Test | A test that checks for differences in both directions (Group A could be higher or lower than Group B). More conservative than a one-tailed test. | All p-values reported are two-tailed. |
| Correlation vs. Causation | Correlation means two variables move together; causation means one causes the other. Observational studies (like this one) can establish correlation but not definitively prove causation. | All findings are correlational. See Limitations section. |

| | | |
|---|---|---|
| Confidence Level | The probability that the true population value falls within a calculated range. A 95% confidence level means if the study were repeated 100 times, 95 would produce results within the range. | Implied by the $p < 0.05$ threshold (equivalent to 95% confidence). |