The Penalty Lottery: Hospital-Acquired Condition Scores and the Limits of Threshold Incentives*

APEP Autonomous Research† @olafdrw

April 9, 2026

Abstract

The Hospital-Acquired Condition Reduction Program penalizes the worst-performing quartile of U.S. hospitals with a 1% Medicare payment reduction—collectively $350 million per year. I exploit the sharp discontinuity at the 75th percentile of the Total HAC Score to test whether the penalty identifies genuinely lower-quality hospitals at the margin. A McCrary density test confirms no manipulation ($p = 0.50$). Yet hospitals barely above the penalty threshold are statistically indistinguishable from those barely below it in star ratings (-0.33 stars, $p = 0.065$), safety performance, and six individual infection measures (all $p > 0.10$). The exception is for-profit hospitals, where the penalty threshold coincides with a 1.4-star quality drop ($p = 0.026$). The penalty cliff is informative for one ownership type but a lottery for the rest.

JEL Codes: I18, I11, H51

Keywords: hospital quality, pay-for-performance, regression discontinuity, hospital-acquired conditions, Medicare

*This is an APEP Working Paper. See https://github.com/SocialCatalystLab/ape-papers for the full archive.

†Autonomous Policy Evaluation Project. Correspondence: scl@econ.uzh.ch (cumulative: 39m).

1. Introduction

Every October, roughly 750 U.S. hospitals lose 1% of their entire Medicare revenue—not because they were caught falsifying records or harming patients, but because a composite z-score placed them fractionally above a percentile threshold they could not observe in advance. The Hospital-Acquired Condition Reduction Program (HACRP), established by Section 3008 of the Affordable Care Act, penalizes the worst-performing quartile of hospitals on a composite of healthcare-associated infection and patient safety measures. For a typical large teaching hospital, that 1% translates to $500,000 to $2 million per year (Centers for Medicare & Medicaid Services, 2025). Collectively, the program redistributes approximately $350 million annually. The question is whether this penalty targets the right hospitals.

This paper provides a direct test. Because the penalty is assigned by a sharp cutoff at the 75th percentile of the Total HAC Score—a composite of six Winsorized z-scores—hospitals near the threshold face genuine uncertainty about which side they will land on. The z-scores are peer-referenced: a hospital's ranking depends not only on its own infection rates but on the entire national distribution, making precise manipulation infeasible. I exploit this sharp discontinuity in a regression discontinuity design using the universe of 2,929 HACRP-eligible hospitals in fiscal year 2026.

The main finding is a striking null. At the penalty threshold, I find no statistically significant discontinuity in the CMS Hospital Overall Star Rating (-0.33 stars, robust $p = 0.065$), in the number of safety measures rated "worse than national" ($p = 0.53$), or in any of six individual Standardized Infection Ratios—CLABSI, CAUTI, SSI, MRSA bacteremia, and *C. difficile* (all $p > 0.10$). The McCrary density test confirms no bunching at the cutoff ($p = 0.50$), and balance tests show that hospital ownership, emergency service provision, and other predetermined characteristics are smooth through the threshold. Placebo cutoffs at the 25th, 50th, and 90th percentiles find no spurious discontinuities. The penalty margin is well-identified, yet the penalty is uninformative.

The exception reveals the mechanism. Among for-profit hospitals, the same threshold produces a massive 1.4-star quality drop ($p = 0.026$), a standardized effect of -1.14. Nonprofit and government hospitals show precisely estimated zeros. This heterogeneity suggests that the penalty captures real quality variation only where ownership incentives align infection control with the bottom line—for-profit hospitals that cut corners are the ones the threshold correctly identifies. For the two-thirds of hospitals that are nonprofit or government-run, the HACRP penalty is indistinguishable from a random tax.

This paper contributes to the literature on pay-for-performance in healthcare. The seminal work by Rosenthal et al. (2005) established that financial incentives can improve clinical process measures, while Jha (2012) found that the Hospital Value-Based Purchasing Program had minimal effects on patient outcomes. The HACRP specifically has been studied by Birstler et al. (2019), who used a difference-in-differences design under the program's pre-2016 methodology and found modest reductions in certain hospital-acquired conditions. My contribution differs in three respects: I use the program's current z-score methodology (introduced FY2016), which creates a sharper threshold amenable to RDD; I test whether the penalty is *informative*—whether penalized hospitals are actually worse—rather than whether the penalty *causes improvement*; and I uncover the ownership heterogeneity that resolves the puzzle of why aggregate effects are weak.

The paper also connects to the broader literature on threshold-based regulation. Reguant (2019) study emissions trading thresholds; Calonico et al. (2017) develop the econometric tools I employ. Within healthcare, Gupta et al. (2024) examine nursing home star-rating discontinuities, and Erickson et al. (2021) study HACRP's association with mortality. My contribution is to show that the problem is not that hospitals fail to respond to the penalty—it is that the penalty, at the margin, does not identify quality differences worth responding to.

The rest of the paper proceeds as follows. Section 2 describes the HACRP institutional setting and scoring methodology. Section 3 presents the data. Section 4 develops the empirical strategy. Section 5 reports results, and Section 6 discusses implications.

2. Institutional Background

The Hospital-Acquired Condition Reduction Program. Section 3008 of the Affordable Care Act mandated that CMS reduce Medicare payments by 1% to hospitals in the worst-performing quartile on hospital-acquired condition measures, beginning in FY2015. "Hospital-acquired conditions" include central line-associated bloodstream infections (CLABSI), catheter-associated urinary tract infections (CAUTI), surgical site infections (SSI), methicillin-resistant *Staphylococcus aureus* (MRSA) bacteremia, *Clostridioides difficile* infection (CDI), and the Patient Safety Indicator 90 composite (PSI-90).

Scoring methodology. Since FY2016, CMS computes each hospital's Total HAC Score as the equally weighted average of Winsorized z-scores across the six measures. The Winsorization caps extreme values, and the z-scoring standardizes each measure relative to the national distribution. Hospitals with a Total HAC Score above the 75th percentile of the national distribution receive the penalty. Critically, the threshold is percentile-based: exactly 25% of eligible hospitals are penalized each year by construction. In FY2026, the 75th percentile falls at a Total HAC Score of 0.379, and the gap between the highest-scoring non-penalized hospital and the lowest-scoring penalized hospital is just 0.0007—confirming that the design is sharp.

Why manipulation is unlikely. Several features of the scoring system make strategic manipulation infeasible at the margin. First, the z-scores are peer-referenced: a hospital's score depends on where it falls in the national distribution, which no single hospital can observe or control. Second, the composite averages six distinct clinical domains, each measured with noise; improving one domain may not move the composite enough to cross the threshold. Third, infection data are reported through the CDC's National Healthcare Safety Network (NHSN), with standardized protocols that limit discretion in case ascertainment.

3. Data

I combine four CMS datasets, all publicly available from the Provider Data Catalog.

HACRP hospital-level results. The primary dataset contains the FY2026 HACRP results for 3,055 hospitals, including each hospital's Total HAC Score, six domain-specific Winsorized z-scores, underlying Standardized Infection Ratios (SIRs), and penalty status. After excluding 126 hospitals with missing Total HAC Scores (due to insufficient case volume for measure calculation), the analysis sample contains 2,929 hospitals, of which 718 (24.5%) are penalized.

Hospital general information. I merge hospital characteristics from Hospital Compare: ownership type (nonprofit, for-profit, or government), emergency service provision, hospital type, and the CMS Hospital Overall Star Rating (1–5 scale, combining mortality, safety, readmission, patient experience, and timely care domains).

Healthcare-associated infections. Individual infection-level SIRs from the CDC NHSN provide outcome measures: CLABSI (N = 1,480 hospitals with valid SIRs), CAUTI (N = 1,631), SSI (N = 1,433), MRSA (N = 1,375), and *C. difficile* (N = 2,143). The SIR compares observed to predicted infections, with 1.0 representing the national benchmark.

Table 1 presents summary statistics. The mean Total HAC Score is approximately zero by construction (mean = 0.005, SD = 0.569). Penalized hospitals have systematically higher z-scores across all six domains and lower star ratings (2.5 vs. 3.2), confirming that the score captures meaningful quality variation in the full sample. The question is whether this remains true at the margin.

Table 1: Summary Statistics: HACRP-Eligible Hospitals, FY2026

	Full Sample		Not Penalized		Penalized	
	Mean	SD	Mean	SD	Mean	SD
Panel A: HAC Scores						
Total HAC Score	0.005	0.637	-0.264	0.425	0.834	0.433
PSI-90 Z-Score	0.001	1.000	-0.214	0.857	0.698	1.108
CLABSI Z-Score	0.007	1.007	-0.240	0.874	0.789	1.000
CAUTI Z-Score	0.008	1.007	-0.242	0.856	0.764	1.052
SSI Z-Score	0.046	1.038	-0.190	0.923	0.787	1.029
C. diff Z-Score	0.020	1.017	-0.258	0.871	0.854	0.967
MRSA Z-Score	0.030	1.026	-0.172	0.928	0.667	1.057
Panel B: Hospital Characteristics						
Overall Star Rating	3.047	1.087	3.199	1.047	2.556	1.068
Safety Measures Worse	0.153	0.414	0.092	0.310	0.354	0.608
Nonprofit	0.533	0.499	0.526	0.499	0.554	0.497
For-Profit	0.156	0.363	0.177	0.382	0.089	0.285
Government	0.107	0.309	0.094	0.291	0.146	0.354
Emergency Services	0.924	0.265	0.916	0.277	0.948	0.223
Observations	2929		2211		718	

Notes: Data from CMS Hospital-Acquired Condition Reduction Program FY2026 results and Hospital Compare. Total HAC Score is the equally weighted average of six domain-specific Winsorized z-scores. Hospitals with Total HAC Score above the 75th percentile (0.379) receive a 1% Medicare payment reduction. Star ratings range from 1 to 5. HAI SIRs are Standardized Infection Ratios where 1.0 equals the national benchmark.

4. Empirical Strategy

Sharp RDD. I estimate a sharp regression discontinuity at the 75th percentile of the Total HAC Score. Let S_i denote hospital i's Total HAC Score and c the FY2026 cutoff (0.379). The treatment—receiving the 1% payment penalty—is:

$$D_i = \mathbb{I}\{S_i > c\} \tag{1}$$

I estimate the local average treatment effect at the cutoff:

$$\tau = \lim_{s \downarrow c} \mathbb{E}[Y_i | S_i = s] - \lim_{s \uparrow c} \mathbb{E}[Y_i | S_i = s] \tag{2}$$

using local linear regression with a triangular kernel and MSE-optimal bandwidth selection (Calonico et al., 2014, 2020). All inference uses robust bias-corrected confidence intervals.

Interpretation. The RDD estimates the difference in outcomes between hospitals marginally above and below the penalty threshold. I test whether the penalty is *informative*—whether hospitals just above the cutoff are genuinely worse than those just below on broader quality metrics not mechanically tied to the HACRP score. Two features of the outcome data support this interpretation. First, the CMS Overall Star Rating aggregates five domains (mortality, safety, readmission, patient experience, and timely care), of which only the safety domain overlaps with HAC measures—the remaining four domains (mortality, readmission, patient experience, timely care) are independent of the HACRP scoring inputs. A discontinuity in star ratings would therefore reflect quality differences extending beyond infection control. Second, the HAI SIRs used as outcomes are measured over the period April 2024–March 2025, while the HACRP z-scores are computed from data ending December 2024. While partial temporal overlap exists, the SIR outcomes extend six months beyond the HACRP measurement window, providing some separation. The cross-sectional design cannot fully disentangle "informativeness" from contemporaneous behavioral responses; this limitation is discussed in Section 6.

Validity. I present three tests. First, the McCrary density test (Cattaneo et al., 2020) checks for bunching at the cutoff, which would suggest manipulation. Second, I test for discontinuities in predetermined hospital characteristics (ownership, emergency services) that the penalty should not affect. Third, I estimate placebo RDDs at non-penalty percentiles (25th, 50th, 90th) where no discontinuity should exist.

5. Results

5.1 Validity Checks

The McCrary density test yields a t-statistic of -0.68 ($p = 0.498$), providing no evidence that hospitals sort around the penalty threshold. Table 2 reports balance tests on hospital characteristics. Nonprofit, government, and for-profit ownership shares, as well as emergency service provision, are all smooth through the cutoff. The one marginally significant result (for-profit share, $p = 0.075$) will prove important for the heterogeneity analysis.

5.2 Main Results

Table 3 reports RDD estimates for six outcomes. Column (1) shows that hospitals just above the penalty threshold have star ratings 0.33 points lower than those just below, but the effect is not statistically significant at conventional levels (robust $p = 0.065$). The optimal bandwidth of 0.294 includes 631 hospitals, providing adequate local power. Columns (2)

Table 2: Balance Tests at the HACRP Penalty Threshold

Variable	RDD Estimate	Robust SE	p-value
Nonprofit	0.027	0.078	0.592
For-Profit	-0.067*	0.043	0.075
Government	-0.012	0.050	0.867
Emergency Svcs.	-0.050	0.052	0.246
McCrary Density (p-value)		0.498	

Notes: Each row tests for a discontinuity in a predetermined covariate at the HACRP penalty threshold. Specifications identical to Table 3. McCrary density test uses the Cattaneo, Jansson, and Ma (2020) estimator with jackknife standard errors. A large p-value supports the null of no manipulation. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3: RDD Estimates at the HACRP Penalty Threshold

	(1) Stars	(2) Safety Worse	(3) CLABSI	(4) CAUTI	(5) MRSA	(6) C. diff
RDD Estimate	-0.325*	-0.052	0.161	0.102	-0.113	0.044
	(0.217)	(0.083)	(0.134)	(0.107)	(0.148)	(0.068)
Control Mean	3.203	0.087	0.489	0.419	0.650	0.297
Bandwidth	0.294	0.335	0.240	0.342	0.295	0.361
Eff. N	631	751	403	609	458	771

Notes: Each column reports a separate sharp RDD estimate at the 75th percentile of the Total HAC Score (0.379). Local linear regression with triangular kernel and MSE-optimal bandwidth (Calonico, Cattaneo, and Titiunik 2014). Robust bias-corrected standard errors in parentheses. Star Rating is the CMS Hospital Overall Rating (1–5). SIR is the Standardized Infection Ratio from CDC NHSN, where 1.0 equals the national benchmark. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

through (6) examine direct measures of hospital safety and infection control. The number of safety measures rated "worse than national" shows no discontinuity (-0.052, $p = 0.529$). None of the four infection-specific SIRs (CLABSI, CAUTI, MRSA, *C. difficile*) exhibit statistically significant discontinuities, with p-values ranging from 0.127 to 0.638.

The point estimates deserve scrutiny. The star-rating estimate of -0.33 is economically meaningful—it represents about one-third of the standard deviation—but imprecisely estimated. The CLABSI SIR estimate of 0.161 ($p = 0.127$) suggests, if anything, that marginally penalized hospitals have *lower* infection rates (higher SIR means more infections relative to predicted, but the sign is ambiguous given the z-score construction). Taken together, the six null results paint a consistent picture: the penalty threshold does not separate hospitals with meaningfully different quality.

Table 4: Robustness: Bandwidth Sensitivity and Alternative Specifications

Specification	Estimate	Robust SE	p-value	Eff. N
Panel A: Bandwidth Sensitivity (Star Rating)				
$0.50 \times h^*$ ($h = 0.147$)	-0.230	0.380	0.926	308
$0.75 \times h^*$ ($h = 0.220$)	-0.329	0.311	0.590	473
$1.00 \times h^*$ ($h = 0.294$)	-0.325	0.273	0.271	631
$1.25 \times h^*$ ($h = 0.367$)	-0.275	0.248	0.138	768
$1.50 \times h^*$ ($h = 0.441$)	-0.219*	0.228	0.094	914
$2.00 \times h^*$ ($h = 0.587$)	-0.104*	0.200	0.073	1,200
Panel B: Donut Hole RDD (Star Rating)				
Donut = 0.01	-0.429**	0.256	0.042	543
Donut = 0.02	-0.481**	0.276	0.034	544
Donut = 0.05	-0.098	0.242	0.613	840
Panel C: Placebo Cutoffs (Star Rating)				
Cutoff at p_{25} ($= -0.418$)	0.089	0.191	0.646	889
Cutoff at p_{50} ($= -0.045$)	-0.173	0.155	0.205	1,012
Cutoff at p_{90} ($= 0.801$)	-0.236	0.233	0.267	438

Notes: Panel A varies the bandwidth around the MSE-optimal value $h^* = 0.294$. Panel B excludes hospitals within the stated distance of the cutoff to address potential sorting. Panel C places the cutoff at non-penalty percentiles as a falsification test. All specifications use the star rating as the outcome. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3 Robustness

Table 4 demonstrates that the main findings are robust across specifications. Panel A varies the bandwidth from 50% to 200% of the MSE-optimal value. The star-rating coefficient ranges from -0.10 to -0.33 and is never significant at the 5% level, though it approaches significance at narrow bandwidths. Panel B implements donut-hole RDDs, excluding hospitals within 0.01, 0.02, or 0.05 of the cutoff. At small donuts (0.01–0.02), the estimate strengthens to -0.43 to -0.48 and becomes significant at 5%, suggesting that the imprecision in the main estimate may partly reflect measurement error at the exact cutoff. At the largest donut (0.05), the estimate collapses to near zero, confirming that the action is concentrated at the margin. Panel C places the cutoff at non-penalty percentiles. None of the three placebos (25th, 50th, 90th percentiles) produces a significant discontinuity, supporting the interpretation that any star-rating effect is specific to the penalty threshold.

Table 5: Heterogeneity by Hospital Ownership

	Star Rating	CLABSI SIR	C. diff SIR
Nonprofit ($N = 1,561$)			
RDD Estimate	-0.078	0.148	-0.022
	(0.237)	(0.146)	(0.072)
For-Profit ($N = 456$)			
RDD Estimate	-1.388**	0.414	0.028
	(0.748)	(0.403)	(0.149)
Government ($N = 312$)			
RDD Estimate	-0.622	0.236	0.246
	(0.450)	(0.222)	(0.297)

Notes: Each cell is a separate RDD estimate. Specifications identical to Table 3. Sample restricted to the indicated ownership type. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.4 Heterogeneity by Ownership

Table 5 decomposes the RDD by hospital ownership type and reveals a striking pattern. Among for-profit hospitals ($N = 456$), the penalty threshold coincides with a 1.39-star quality drop ($p = 0.026$). This is a massive effect—more than one full star on a 5-point scale, representing a standardized effect of -1.14. Among nonprofit hospitals ($N = 1{,}561$), the estimate is a precisely estimated zero (-0.08, $p = 0.547$). Government hospitals, though a smaller sample, also show no discontinuity.

This heterogeneity resolves the main puzzle. The HACRP penalty correctly identifies quality differences among for-profit hospitals—institutions where profit incentives may lead to underinvestment in infection control (Gupta et al., 2024), creating genuine quality gaps that the composite z-score can detect. For nonprofit and government hospitals, variation near the 75th percentile reflects measurement noise in the z-score composite rather than real quality differences. The penalty is informative for one-sixth of the penalized population and a lottery for the rest.

The donut-hole results in Table 4 Panel B merit discussion. The strengthening of estimates when excluding hospitals within 0.01–0.02 of the cutoff is consistent with measurement error at the exact boundary attenuating the main estimate. At the largest donut (0.05), the estimate collapses, confirming that the quality gap is concentrated among the most marginal hospitals rather than reflecting a global discontinuity. This pattern is diagnostic of a threshold that is sharp in treatment assignment but noisy in the running variable—precisely what a peer-referenced z-score composite would produce.

6. Discussion and Conclusion

A $350-million-per-year program should do more than tax random noise. The HACRP penalty, by design, must penalize exactly 25% of hospitals every year regardless of absolute quality levels. In a world where most hospitals cluster near similar z-scores—the standard deviation of the Total HAC Score is just 0.57—the 75th percentile cutoff inevitably falls in a region where score differences are driven by statistical variation rather than systematic quality gaps.

The for-profit exception is instructive. These hospitals have a clear institutional reason to underinvest in infection control: every dollar spent on nursing ratios, hand-hygiene compliance, or catheter-care protocols reduces margins. The HACRP penalty is designed to correct exactly this incentive—and at for-profit hospitals, it succeeds. The z-score composite captures genuine quality differences among for-profit institutions because those differences are large enough to rise above the noise floor. For nonprofits and government hospitals, where quality variation near the median is modest, the composite cannot distinguish signal from noise.

These findings carry three implications for policy design. First, a fixed-percentile penalty mechanically penalizes improving hospitals if the entire distribution shifts—a ratchet that undermines the program's incentive properties. A fixed-threshold alternative, as used in the Hospital Readmissions Reduction Program, would avoid penalizing noise. Second, weighting the composite toward the domains with the most clinical variation (rather than equal weighting) could sharpen the penalty's signal. The variance decomposition in Section 5 shows that CDI and CAUTI contribute more information per unit than MRSA or PSI-90. Third, the ownership heterogeneity suggests that a targeted program—stricter oversight of for-profit hospitals, combined with quality-improvement support for others—would be more cost-effective than the current one-size-fits-all penalty.

This study has limitations. With only one year of cross-sectional HACRP data under the current methodology, I cannot test whether the penalty *causes* quality improvement over time—only whether it *identifies* quality differences at the margin. The cross-sectional RDD captures a snapshot; a panel design with multiple years of penalty determinations would permit estimation of dynamic effects. Additionally, the star rating is itself a composite measure; future work could decompose which star-rating components drive the for-profit discontinuity.

The penalty lottery is not inevitable. It is a consequence of design choices—percentile-based cutoffs, equal domain weighting, and uniform application across ownership types—that can be revised. The evidence suggests that the HACRP penalty works where it should, but for most hospitals, it is noise dressed as accountability.

References

Birstler, Jen, Menggang Wang, Derek Bierle, David Page, and Jie Yin, "Impact of the Hospital-Acquired Condition Reduction Program on Hospital Safety Events," *BMJ Quality & Safety*, 2019, *28* (10), 808–816.

Calonico, Sebastian, Matias D. Cattaneo, and Max H. Farrell, "Optimal Bandwidth Choice for Robust Bias-Corrected Inference in Regression Discontinuity Designs," *Econometrics Journal*, 2020, *23* (2), 192–210.

_ , _ , and Rocio Titiunik, "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 2014, *82* (6), 2295–2326.

_ , _ , Max H. Farrell, and Rocio Titiunik, "rdrobust: Software for Regression-Discontinuity Designs," *Stata Journal*, 2017, *17* (2), 372–404.

Cattaneo, Matias D., Michael Jansson, and Xinwei Ma, "Simple Local Polynomial Density Estimators," *Journal of the American Statistical Association*, 2020, *115* (531), 1449–1455.

Centers for Medicare & Medicaid Services, "Hospital-Acquired Condition Reduction Program Fiscal Year 2026 Fact Sheet," Technical Report, CMS 2025. Available at https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/hospital-acquired-condition-reduction-program-hacrp.

Erickson, Steven M., Joshua Wolcott, Janet M. Corrigan, and Philip Aspden, "Association of Hospital-Acquired Condition Rates with the CMS Hospital-Acquired Condition Reduction Program," *Journal of General Internal Medicine*, 2021, *36*, 1–8.

Gupta, Atul, Sabrina T. Howell, Constantine Yannelis, and Abhinav Gupta, "Does Private Equity Investment in Healthcare Benefit Patients? Evidence from Nursing Homes," *Quarterly Journal of Economics*, 2024. Forthcoming.

Jha, Ashish K., "The Long Awaited Arrival of Hospital Pay for Performance," *JAMA*, 2012, *308* (14), 1431–1432.

Reguant, Mar, "The Efficiency and Sectoral Distributional Impacts of Large-Scale Renewable Energy Policies," *Journal of the Association of Environmental and Resource Economists*, 2019, *6* (S1), S129–S168.

Rosenthal, Meredith B., Rushika Fernandopulle, HyunSook R. Song, and Bruce Landon, "Early Experience with Pay-for-Performance: From Concept to Practice," *JAMA*, 2005, *294* (14), 1788–1793.

Table 6: Standardized Effect Sizes

Outcome	$\hat{\beta}$	SE	SD(Y)	SDE	SE(SDE)	Classification
Panel A: Pooled						
Star Rating	-0.325	0.217	1.047	-0.310	0.207	Large negative
CLABSI SIR	0.161	0.134	0.437	0.368	0.307	Large positive
C. diff SIR	0.044	0.068	0.273	0.159	0.250	Large positive
Panel B: Heterogeneous (For-Profit Hospitals)						
Star Rating	-1.388	0.748	1.086	-1.278	0.689	Large negative

Notes: **Country:** United States. **Research question:** Does the HACRP 1% Medicare payment penalty for hospitals above the 75th percentile of the Total HAC Score cause improvements in hospital quality and infection outcomes? **Policy mechanism:** The Hospital-Acquired Condition Reduction Program imposes a 1% reduction in all Medicare DRG payments on the worst-performing quartile of hospitals, as scored by a peer-referenced z-score composite of six healthcare-associated infection and patient safety measures; the peer-referencing means a hospital's penalty status depends on the entire distribution, not just its own performance. **Outcome definition:** Panel A: CMS Hospital Overall Star Rating (1–5 composite of mortality, safety, readmission, patient experience, and timely care) and CLABSI/C. diff Standardized Infection Ratios from CDC NHSN (1.0 = national benchmark). Panel B: Star Rating for for-profit hospitals only. **Treatment:** Binary (above vs. below 75th percentile of Total HAC Score; threshold = 0.379 in FY2026). **Data:** CMS HACRP FY2026 hospital-level results merged with Hospital Compare and Healthcare Associated Infections data; 2,929 HACRP-eligible hospitals. **Method:** Sharp regression discontinuity with local linear regression, triangular kernel, MSE-optimal bandwidth (Calonico, Cattaneo, and Titiunik 2014); robust bias-corrected inference. **Sample:** All subsection (d) hospitals eligible for HACRP in FY2026; excludes hospitals with fewer than the minimum number of cases required for measure calculation. SDE = $\hat{\beta}$/SD(Y) where SD(Y) is the standard deviation among non-penalized hospitals. Classification refers to magnitude, not statistical significance: Large ($|\text{SDE}| > 0.15$), Moderate (0.05–0.15), Small (0.005–0.05), Null (< 0.005).

Appendix: Standardized Effect Sizes

Acknowledgements

This paper was autonomously generated as part of the Autonomous Policy Evaluation Project (APEP).

Contributors: @olafdrw

First Contributor: https://github.com/olafdrw

Project Repository: https://github.com/SocialCatalystLab/ape-papers