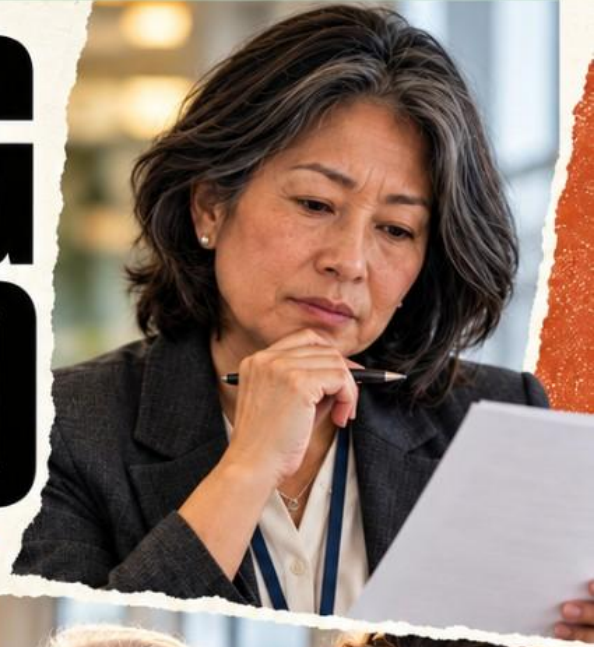


INVESTIGATIVE REPORT

OCTOBER 2026

THE WRONG SCORECARD



What are we calling success?

Hospital leadership, incentives, and accountability

Kelly Emrick

DHSc, PhD, MBA, BSRT(ARRT)R

Leadership: Are We Measuring the Wrong Performance Metrics?

What hospital leadership rewards reveal about success and accountability

Kelly Emrick, DHSc, PhD, MBA, BSRT(ARRT)R



The question behind the compensation debate

A hospital can meet its financial targets while leaving consequential questions about care unanswered. Leadership is tested by whether the board can link rewards to independently verified results for patients, the workforce, and the community.

A document-based investigation and proposed governance framework for United States nonprofit hospitals. Evidence reviewed through October 2026.

Abstract

Hospital leadership is evaluated through measures that make priorities visible. Yet the public record often lets readers see compensation and organizational performance without reconstructing the precise decision that connected them. This report examines that accountability gap using peer-reviewed compensation research, a 2026 policy brief, current compensation-survey reporting, and governance guidance. It asks a focused question: what evidence would justify describing a hospital leader as successful?

The evidence does not support a universal claim that boards ignore care or that higher executive compensation causes worse outcomes. Studies differ in their findings, use different samples, and frequently measure earlier periods. Associations with organizational scale recur; patient satisfaction also appears in the evidence. The more defensible concern is that readily measured growth and financial results can become persuasive substitutes for a complete account of mission performance.

The report proposes a six-domain leadership accountability scorecard that preserves financial sustainability while making patient outcomes, access, workforce capability, affordability, and governance visible. It rejects a single summary number as the sole basis for judgment. Material failures and unreliable data would trigger independent review under rules adopted before the performance period. The framework is an original, unvalidated governance proposal, not an established clinical instrument or a universal compensation formula.

Three findings for the board

First, compensation trends warrant scrutiny, but pay levels and pay ratios alone cannot establish motive, contractual incentives, clinical performance, or legal excess. Those are separate questions requiring separate evidence.

Second, including quality on a scorecard does not settle whether it matters. The board must inspect its weight, the difficulty of its targets, exceptions, measurement reliability, and the actual consequences of missing them.

Third, the relevant accountability unit is a documented decision. A credible board should explain what improved, what deteriorated, whose experience changed, what remains uncertain, and why those findings support the reward and remediation decisions it made.

Scope of the investigation

This is a targeted documentary investigation and critical synthesis, not a systematic review, original national dataset analysis, interview project, or audit of a named organization. We located primary studies and official guidance through targeted web searches, publisher pages, and PubMed or PubMed Central records. We checked sources for authorship, dates, methods, and fit to the claim. No hospital or executive is rated. Publication years are distinguished from observation years throughout.

Why this question matters now

The immediate issue for trustees is designing the reward system they are approving. In its September 2026 release, SullivanCotter reported a 4.1% median increase in executive base salary across its survey and a 5.4% increase in total cash compensation. A separate April 2026 pulse survey found that 45% of organizations changed incentive goal weights for fiscal 2026, often increasing financial emphasis or altering the balance between system and entity performance (SullivanCotter, 2026). These are vendor-reported survey findings,

not estimates for every American hospital; the release does not provide enough pulse survey detail to establish national representativeness.

A weighting change is a consequential governance decision. It defines which result can offset another, which leader benefits from a system average, and which local deterioration may disappear from the discussion. Trustees should be able to reconstruct that decision before debating whether an individual award looks large.

Table 1. The evidence needed to reconstruct a leadership reward

Record	What it can reveal	What still needs verification
Compensation plan	Measures, weights, thresholds, caps, and discretion	The version approved before the period; later amendments and reasons
Performance packet	Reported results and comparison periods	Underlying records, exclusions, missingness, and subsidiary variation
Compensation decision	Award calculation and committee judgment	Treatment of missed targets, exceptions, conflicts, and remediation
Public filings and reports	Reported pay and selected organizational results	Comparable entities, periods, pay components, and outcome definitions

Original reporting framework. This table identifies records to request; it does not imply that private committee materials were obtained for this report.

A complete investigation would follow the same executive and organizational boundary across the plan, performance record, and final decision. A system chief executive may oversee many facilities, while the quality data describe one hospital. A payment reported in one year may reflect a longer performance period. Pairing those figures without reconciliation can produce an arresting but invalid comparison.

The most useful follow-up is therefore specific: request the measure dictionary, approved target schedule, actual calculation, independent validation, and explanation of discretion. Refusing to provide information limits what we can verify. It does not, by itself, establish wrongdoing.

The pay gap is a signal to investigate

A July (2026). A Baker Institute brief reported that the CEO-to-direct-patient-care compensation ratio in the largest bed-size decile rose from 35.9 in 2015 to 51.4 in 2022 (Jenkins, 2026). The underlying care compensation measure includes benefits and is standardized to a full-time equivalent. It is not an individual nurse salary. The brief also notes changing sample composition as hospitals consolidated.

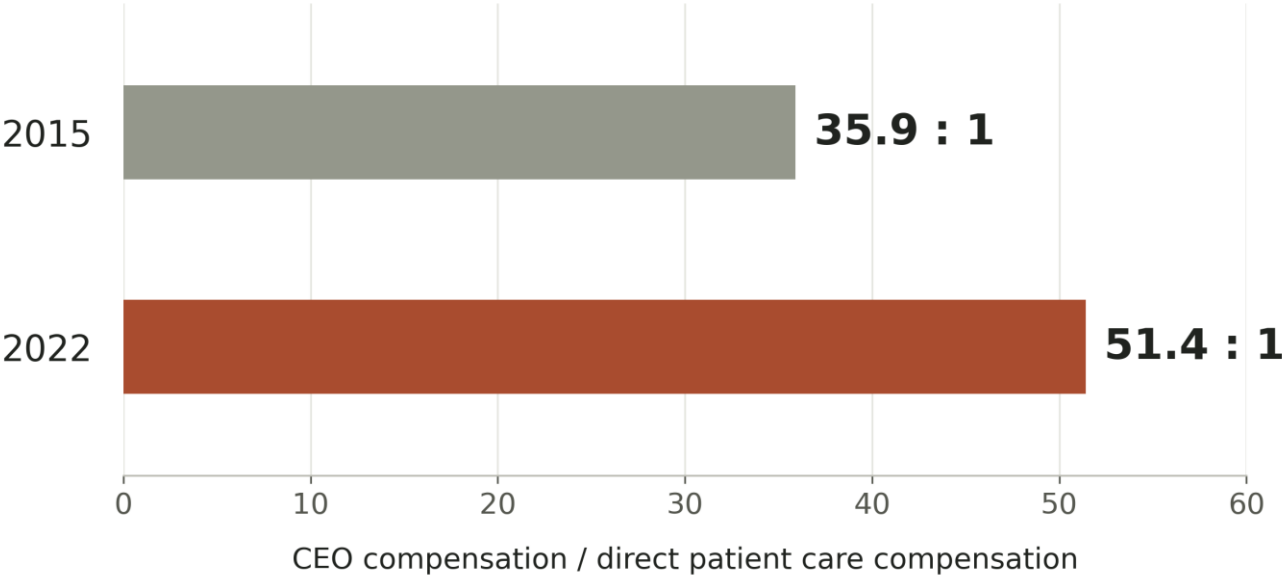


Figure 1. Published compensation ratios for the largest bed size decile, 2015 and 2022. Source: Jenkins (2026). Values are reproduced as reported, not recomputed from rounded dollar amounts. Groups are defined within each year; this is not a fixed cohort of the same organizations. No causal effect is estimated here.

What the ratio cannot answer

The ratio is a distributional observation. It does not identify how much of the difference reflects organizational complexity, executive labor markets, compensation design, changing workforce composition, or changes in the hospitals included. Nor does it show which patients benefited from the underlying decisions. Treating it as a verdict would replace one incomplete scorecard with another.

Its legitimate use is as a prompt for a board explanation. If leadership compensation outpaces direct care compensation, the committee should describe the responsibilities being rewarded, the relevant labor market, and the independently verified mission results. The workforce deserves an explanation that addresses the organization it actually experiences, including local staffing stability and access to necessary resources.

There is also a scale issue. A large percentage change in one executive package is not the same as a large share of total hospital spending. Don't assume either the symbolism of a pay decision or its budgetary impact from the headline alone. An investigation should report both when verified and avoid presenting executive pay reductions as a quantified solution to staffing or affordability problems without doing the calculation.

What the compensation model explains

Jenkins, Short, and Ho (2024) analyzed compensation changes between 2012 and 2019. Their decomposition attributed 44.5% of the measured increase to a shift in a modeled baseline, 28.5% to changes in relationships between observed characteristics and pay, and 27% to changes in those characteristics. Larger systems and profitability drove the latter components. Figure 2 translates that published decomposition into a visual reading guide.

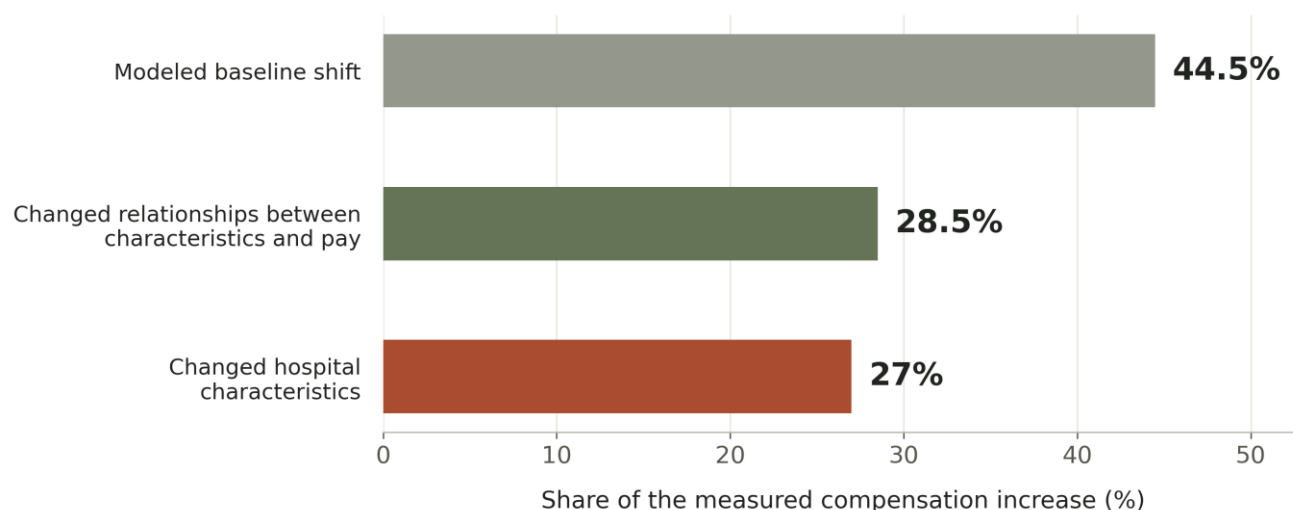


Figure 2. Statistical decomposition of the measured compensation increase, 2012–2019. Source: Jenkins et al. (2024). The baseline refers to the study’s modeled nonteaching organization with fewer than 100 beds, zero profit, and zero charity care. These percentages are explanatory components of a model, not executive bonus weights.

Association is not the compensation contract

An observed relationship between scale and pay does not show that a board awarded a merger bonus, that the executive pursued a transaction for personal gain, or that the resulting care improved or worsened. Establishing the incentive requires the contract and contemporaneous decision record; identifying its effects requires a stronger research design. The report therefore treats the model as evidence about patterns in compensation, not a finding about individual intent.

The governance implication is nonetheless practical. If scale is a major basis for benchmarking pay, the board should separately test whether added scale delivered the promised public value. Responsibility can increase after an acquisition even when its benefits remain unproven. A board may recognize that larger responsibility while still withholding claims of mission success until it verifies the relevant outcomes.

This distinction prevents a circular argument in which acquiring more facilities makes the job larger, the larger job justifies higher compensation, and the higher compensation is then presented as evidence that the acquisition was successful. None of those steps, on its own, establishes better care.

The evidence does not support a simple indictment

The strongest version of the argument must survive evidence that complicates it. In an earlier national study, Joynt et al. (2014) linked 2009 compensation to hospital characteristics and 2008 performance. Pay was associated with patient satisfaction and technology, but not the measured clinical outcomes, processes, or community benefit. Financial measures also lacked an association in that analysis. It would be inaccurate to summarize this finding as proof that every board rewards profit alone.

Mulligan et al. (2022), examining 2010 and 2015 data, likewise found a relationship with patient satisfaction and mixed findings for financial performance and uncompensated care. The authors interpreted their results as consistent with balancing financial health and mission. That interpretation belongs in the evidence record, alongside the limitations of observational measurement.

Table 2. Reading the evidence without overstating it

Source and period	Contribution	Boundary
Joynt et al. (2014) 2008–2009	Patient satisfaction and technology were associated with pay	Older observational comparison; not a study of current contracts
Mulligan et al. (2022) 2010 and 2015	Mission-related and financial associations were not uniform	Measured associations do not identify each board's decision rule
Jenkins et al. (2024) 2012–2019	Separates modeled components of compensation growth	Decomposition is not causal attribution or a bonus formula
Jenkins (2026) 2015–2022	Describes pay differences by organization size	Policy brief; changing samples and a specific care labor definition

Professional guidance also undermines the claim that incentive plans lack quality. AHA Trustee Services guidance published in 2019 listed quality, patient engagement, safety, and financial results among commonly used measures, drawing on a 2018 survey (Cotter & Hastings, 2019). That historical guidance cannot describe every current plan. It shows why the next question must focus on how a measure operates, not whether its name appears on a slide.

When the proxy becomes the result

A proxy is useful when it gives the board a dependable view of the outcome it cares about. It becomes dangerous when people treat its improvement as sufficient evidence of success despite contrary information. Revenue, volume, bed count, market share, patient experience, and employee survey scores each answer a different question. None alone fully describes a hospital’s performance.

Growth requires an outcome test

Beaulieu et al. (2020) compared acquired hospitals with controls using data from 2007 through 2016. Acquisition was associated with modestly worse patient experience and no significant differential change in mortality or readmissions; process findings were inconclusive. This study does not evaluate executive compensation, and its historical average cannot establish what happened in a particular recent merger. It does challenge the assumption that organizational growth alone proves quality improvement.

A transaction scorecard should therefore distinguish completion from benefit. Closing a deal, integrating financial systems, or meeting a synergy budget is an implementation milestone. A board seeking evidence of public value should also follow service availability, patient outcomes, affordability, and workforce continuity at the affected facilities. It should document which promises have been tested and which remain projections.

A quality incentive also needs an outcome test

Figueroa et al. (2016) found no evidence of a mortality benefit associated with the early Hospital Value Based Purchasing program in their observational analysis of 2008–2013 admissions. Hospital reimbursement and executive compensation are different mechanisms. The study is not a verdict on every later incentive program. Its relevance here is narrower: naming a quality objective and attaching money to it do not establish that the mechanism delivered the intended outcome.

The design problem is translation. An annual bonus may depend on a short reporting window, while prevention, staffing stability, or access improvements develop over several years. A system average may improve while one community loses a service. A process completion percentage may rise without showing whether patients experienced a meaningful difference. Test those possibilities; don't presume them.

The board needs both results and explanation

The proposed alternative pairs outcome measures with operational evidence and a structured account of unexpected changes. When performance improves, leaders should identify the intervention and test plausible alternative explanations. When it deteriorates, the board should distinguish statistical variation, external shocks, avoidable operational failures, and deliberate misrepresentation. A scorecard should make those distinctions easier to investigate.

A more complete definition of success

Financial resilience is a condition for sustaining care. It should remain visible, with liquidity, operating performance, and investment capacity assessed in their local context. The error is allowing that necessary condition to stand in for the entire mission. Conversely, strong mission results do not remove the need to address an organization’s financial fragility.

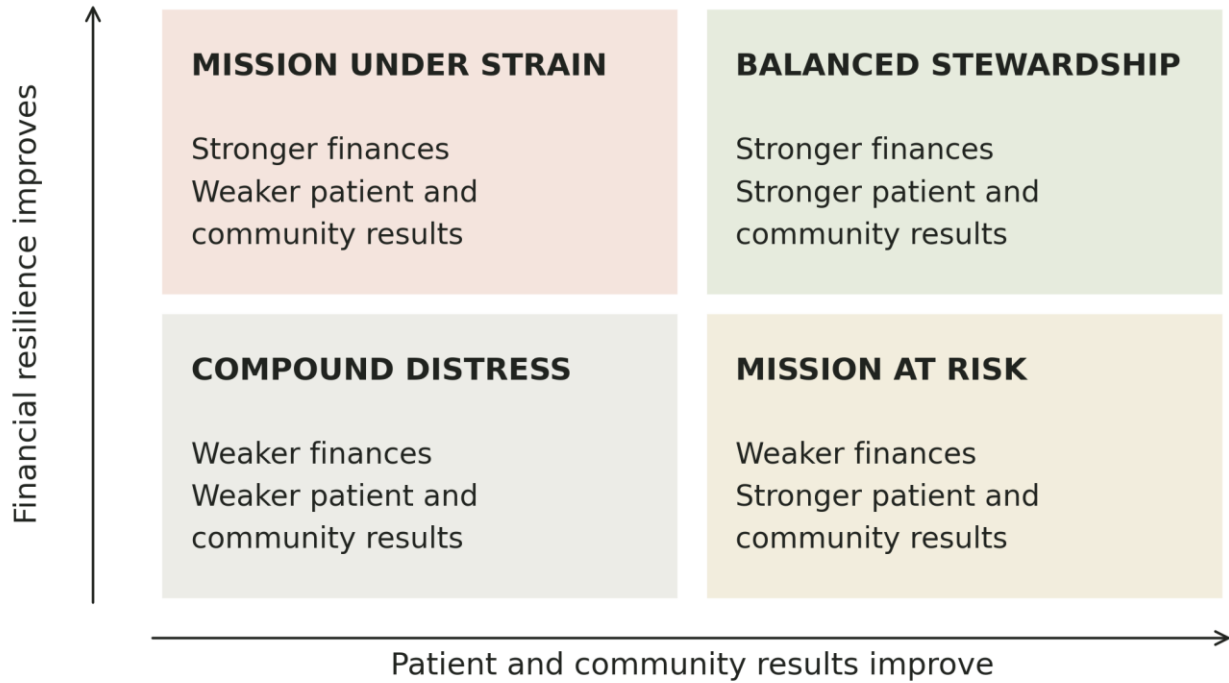


Figure 3. Original conceptual matrix by Kelly Emrick. The quadrants organize board discussion; they are not validated classifications, measured thresholds, or ratings of actual hospitals. The supporting scorecard must keep patient and community results disaggregated.

The matrix separates two questions that are too easily collapsed. Is the organization financially capable of sustaining its commitments? Are those commitments producing credible results for the people it serves? The same reported margin could coexist with very different answers to the second question. The same clinical achievement could rest on very different financial foundations.

The Institute for Healthcare Improvement’s governance framework supports oversight of all dimensions of quality, with attention to trustee knowledge, improvement systems, and board culture (Daley Ullem et al.,

2018). The model proposed here extends that governance logic into a transparent leadership evaluation process. Its contribution is a practical arrangement of evidence and decision rights, not a claim that a new set of domains has already been shown to improve outcomes.

A scorecard grounded in care delivery

The proposed scorecard has six domains. Tables 3 and 4 supply a starter measurement dictionary, not a universal specification. After consulting clinical leaders, patients, staff, and community representatives, each organization should select a small, defensible set of measures. Define measures before assessing performance, with a named owner, reporting period, minimum sample rule, and source of independent verification.

Table 3. Proposed care delivery domains

Domain and accountable owner	Candidate measure and definition	Balancing evidence and review
Patient outcomes and safety Chief medical and nursing officers	Risk-adjusted outcomes using the selected measure’s published specification; selected harm rates with explicit cases and exposure denominators	Show uncertainty and service level variation. Reconcile case review with coded records. Do not reward fewer reported incidents without checking reporting behavior.
Timely access and equity Operations and access leaders	Median and 90th percentile days from accepted referral to completed appointment, by service; unresolved referrals reported separately	Track cancellations, rejected referrals, transfers, and completion by relevant patient groups. Shorter waits among completed visits can conceal excluded patients.
Workforce capability and safety Nursing and human resources leaders	Voluntary direct care departures divided by average direct care headcount; monthly trend with role and facility detail	Pair retention with vacancies, overtime, competence, worker safety, and patient outcomes. Audit involuntary or transferred staff classifications.

Original proposed dictionary. Candidate measures require local validation, governance approval, and appropriate privacy protections. No universal numerical targets are asserted.

Make denominators part of the conversation

A denominator is not a technical footnote. It determines which people count. An access measure based only on completed appointments cannot describe the experience of people who never obtained one. A workforce measure can change when positions are outsourced or reclassified. A harm rate can change when surveillance becomes more thorough. Trustees need a concise explanation of those choices whenever a measure affects a leadership evaluation.

For equity, report meaningful differences where data quality and sample size allow, while retaining an explicit category for missing information. Risk adjustment can improve comparison across clinical populations; stratification can expose unequal experiences. Neither should make a disadvantaged group disappear from view. Small samples require uncertainty, aggregation across appropriate periods, or qualitative review rather than unsupported rankings.

A scorecard grounded in stewardship

Table 4. Proposed stewardship domains

Domain and accountable owner	Candidate measure and definition	Balancing evidence and review
Affordability and community benefit Finance and community health leaders	Financial assistance approvals divided by completed eligible applications; separately disclose dollars at a defined accounting basis and progress on community commitments	Review application abandonment, denial reasons, patient billing complaints, and service availability. Distinguish spending from demonstrated community outcomes.
Financial sustainability Chief financial officer	Operating income divided by operating revenue, with consistent treatment of unusual items; liquidity and capital maintenance shown separately	Assess payer mix, deferred maintenance, essential services, and savings sources. Explain whether improvement shifted burdens to patients, staff, or future periods.
Governance and trust Board quality and audit committees	Overdue material corrective actions divided by all material actions due; validation status and unresolved findings shown explicitly	Independent confirmation of closure, protected staff escalation, and review of data corrections. An administratively closed case does not prove the problem was resolved.

Original proposed dictionary. Owners are accountable for the measure; the verification function should be sufficiently independent of the person whose award depends on it.

Do not mistake the dictionary for the decision

A measure dictionary establishes a common language. The board still needs to decide how much improvement is meaningful, which comparisons are fair, and which findings require escalation. Targets should reflect baseline performance, achievable improvement, relevant external standards, and statistical reliability. This report assigns no national cutoff for an acceptable margin, turnover rate, wait time, or compensation ratio.

CMS describes hospital value-based purchasing through domains of clinical outcomes, person and community engagement, safety, efficiency, and cost reduction (Centers for Medicare & Medicaid Services, n.d.). Those measures can contribute useful evidence. Their availability does not make a Medicare payment methodology a ready-made executive compensation formula, nor does it cover every local obligation.

An organization should show absolute performance and change over time. Rewarding improvement alone can favor a leader who starts from a weak baseline; rewarding level alone can favor inherited advantages. The board should state its reasoning for considering both and should identify the resources and authority available to the leader during the period. Accountability is stronger when responsibility and opportunity are visible together.

The average can conceal the failure

Consider two entirely hypothetical organizations scored across the six domains. Both have an equal weighted average of 80. Organization A has stronger financial and stewardship scores but weaker safety, access, and workforce scores. Organization B scores 80 in every domain. A single total declares a tie while concealing a decision that trustees would need to examine closely.

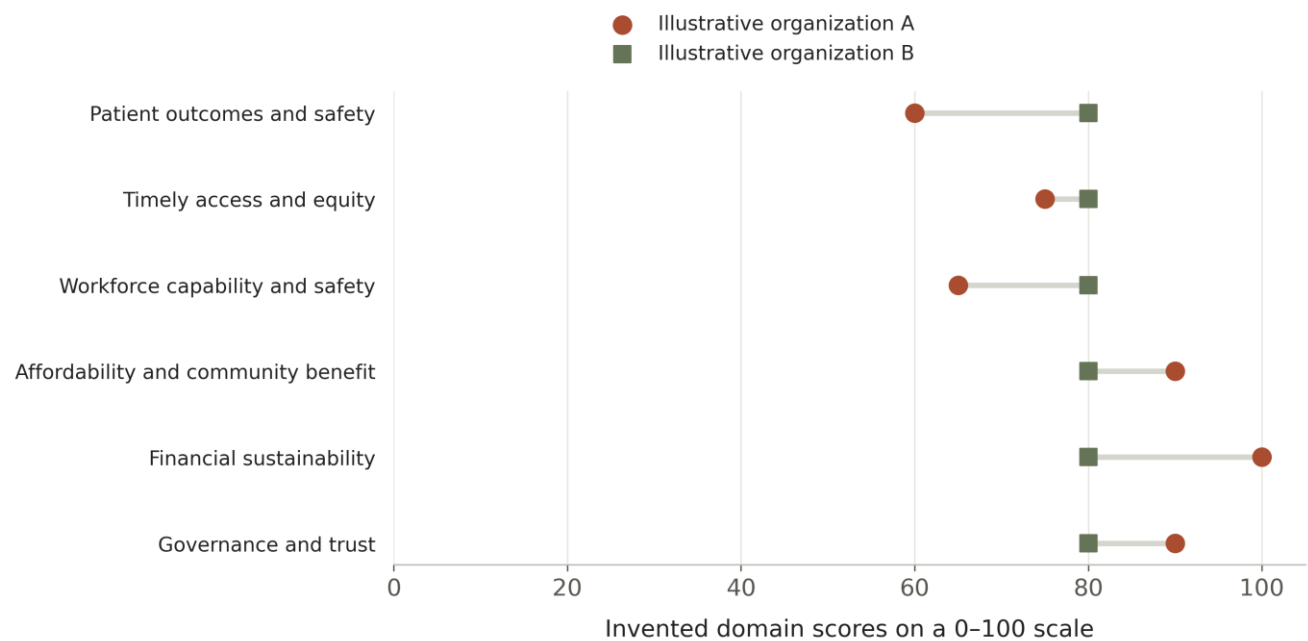


Figure 4. Invented teaching example, not observed hospital data. A = 60, 75, 65, 90, 100, 90; B = 80 in each domain. Each sum is 480, and each arithmetic mean is 80. Equal weights and the 0–100 scale are used only to illustrate information loss, not recommended for adoption.

Preserve the profile before calculating a total

The proposed scorecard retains the domain profile, trend, uncertainty, and unresolved exceptions on the first page of the board packet. A summary score may be useful for a defined purpose, but it should not erase a material safety or integrity concern. A financial surplus should not automatically cancel such a finding through arithmetic.

Equally, a single adverse event should not mechanically determine an executive’s pay. The board needs independently reviewed facts, the problem’s severity and recurrence, the leader’s authority and response, and the data’s reliability. The proposal is a review requirement under prospective rules, not automatic punishment or a substitute for professional judgment.

Before linking this framework to compensation, test whether the conclusion changes when weights, time windows, or measure definitions change within plausible ranges. If small technical changes reverse the leadership rating, the score’s apparent precision is misleading. Report that instability rather than selecting the version that supports a preferred result.

From a scorecard to an accountable decision

The proposed governance cycle separates four functions: defining the rules, verifying evidence, adjudicating results, and explaining the decision. A committee can delegate analytical work while retaining responsibility for the judgment. It should not ask the executive whose award is at issue to be the final authority on whether the evidence is reliable.



Figure 5. Original proposed board review cycle. The return arrow represents revision for a future performance period, not retrospective rewriting of targets to produce a desired award.

IRS guidance describes a rebuttable presumption of reasonable compensation when an authorized body without conflicts approves the arrangement in advance, relies on appropriate comparability information, and documents its basis adequately and promptly. If those conditions are not met, the IRS applies a facts-and-circumstances approach (Internal Revenue Service, n.d.). This legal process should not be confused with proof that the organization has fulfilled its clinical or community mission.

For mission accountability, the committee should retain the original targets, dated amendments, validation findings, and a written explanation of discretion. Any proposed deferral, reduction, recovery provision, or contract change requires prospective design and applicable legal review. The framework supplies governance questions; it does not prescribe legal outcomes for an existing agreement.

The public explanation should make the reasoning understandable without disclosing protected patient information or confidential employment details. It should identify the performance domains, the basis for the judgment, significant unresolved limitations, and the corrective commitments. Publishing a large set of indicators without explaining the decision is disclosure without accountability.

The reporting trail a board should demand

A defensible investigation follows the evidence from source record to board decision. The following audit trail is a proposed working protocol for trustees, internal audit, and subsequent reporting. It makes missing evidence visible and gives management a fair opportunity to explain discrepancies before drawing conclusions.

Table 5. Proposed verification and challenge protocol

Test	Evidence to inspect	Decision consequence
Match the unit and time	Executive role, legal entities, facilities, performance dates, award dates, and reporting years	Do not compare mismatched boundaries or describe an old outcome as current.

Test	Evidence to inspect	Decision consequence
Reconcile the numbers	Source extracts, measure code or specifications, exclusions, corrections, and denominator counts	Flag unreconciled measures as unverified; investigate before using them to justify an award.
Examine the exceptions	Overrides, goal revisions, unusual items, missing facilities, and low volume groups	Explain how each exception affects the result; apply the same rules consistently.
Challenge the interpretation	Alternative explanations, comparison groups, clinical review, and staff or patient evidence	Separate verified observation from causal inference and management attribution.
Document the response	Management explanation, independent findings, committee rationale, owners, and deadlines	Record disagreement and remediation; verify completion rather than accepting a status label.

Interview questions for the next reporting stage

Ask the compensation chair to identify a recent instance in which a mission-related result changed an award, and request the supporting calculation. Ask the quality chair which adverse trend could prevent the committee from treating the year as successful. Ask the chief financial officer to reconcile reported compensation and performance periods. Ask clinical and workforce leaders where the system average differs most from local experience.

Ask patient and community representatives which commitments matter most to them and whether those commitments appear in the board’s evaluation. Interviews would supply context and leads, not substitute for records. If a subsequent investigation names an organization, it should provide the relevant findings and a meaningful opportunity to respond before publication. No such interviews or outreach are represented as completed in this report.

The first 90 days of implementation

The safest starting point is a shadow scorecard: use the proposed framework alongside the existing evaluation process without immediately changing compensation. This creates time to test definitions, surface missing data, and learn whether the proposed review rules produce consistent decisions. Ninety days can establish governance and a baseline; it cannot demonstrate long-term clinical effectiveness.

Table 6. Proposed implementation sequence

Period and lead	Work to complete	Reviewable output
Days 1–30 Board quality and compensation chairs	Inventory current measures and contracts. Map responsibilities. Hear from clinical, workforce, patient, and community representatives.	Approved scope, measure dictionary, data owners, conflict rules, and list of evidence gaps.
Days 31–60 Quality analytics and internal audit	Reconcile a baseline. Test denominators, local variation, and missingness. Simulate review of material failures and changes in weights.	Shadow scorecard with uncertainty, documented exceptions, and sensitivity analysis.
Days 61–90 Joint board review	Compare independent reviewers’ conclusions. Resolve definitions. Agree on prospective rules and a schedule for longer follow-up.	Record the pilot decision, remediation plan, disclosure outline, and conditions for future use.

Validation before a compensation consequence

The pilot should test whether two qualified reviewers reach similar conclusions from the same evidence, whether the measures remain stable when definitions are reapplied, and whether the work required is proportionate to its decision value. It should also identify measures that can be improved administratively without improving the underlying service. A scorecard that is easy to manipulate is not ready for high-stakes use.

Longer follow-up should examine whether the framework changes decisions, resource allocation, and actual care. A comparison with other periods or organizations needs a credible design that accounts for baseline differences and concurrent changes. Until that work is done, claims should remain modest: the framework may make governance more transparent and testable, but its effect on patient outcomes is unknown.

At the end of the pilot, the board should be prepared to remove a weak measure. More indicators are not automatically more accountability. The best set is small enough to understand, broad enough to expose serious tradeoffs, and sufficiently reliable to support a decision that can withstand challenge.

Limitations and the judgment that follows

This investigation relies on published studies, official guidance, and a vendor's public survey release. It does not have access to individual compensation contracts, private committee deliberations, or the underlying national datasets. Consequently, it cannot establish how a particular award was calculated, whether an executive acted in response to a specific incentive, or whether compensation drove a clinical or workforce outcome.

The sources span different periods and methods. Older observational studies remain useful for identifying recurrent questions but cannot be treated as a census of 2026 practice. The 2026 policy brief describes earlier data. The current survey release offers timely context but does not resolve the representativeness of its separate pulse findings. No pooled effect estimate or national prevalence estimate is calculated in this report.

Hospital performance is also multidimensional and context-dependent. Rural, teaching, pediatric, safety net, and integrated systems face different operating conditions. A performance measure can be valid for one comparison and misleading for another. Data lag, small samples, coding changes, and shifting organizational boundaries further limit inference. The proposed model should be adapted and evaluated rather than imposed as a uniform ranking system.

What the investigation establishes

The record supports sustained scrutiny of the relationship between leadership rewards and mission performance. It also requires restraint: compensation patterns do not prove misconduct, and the presence of quality measures does not prove that a board's incentives are effective. The evidence leaves a consequential gap between observing pay and explaining the decision that produced it.

The practical response is to make that decision reviewable. Boards should preserve the full performance profile, specify what can and cannot offset a material failure, independently validate the evidence, and explain how they used discretion. Financial resilience belongs in this account because sustainable care requires it. So do patient outcomes, access, workforce capability, affordability, and trustworthy governance.

The wrong scorecard lets leaders confidently declare success while leaving those questions unanswered. A better scorecard doesn't eliminate judgment. It gives judgment a visible evidentiary basis, exposes uncertainty, and makes decision-makers answerable for the tradeoffs they accept.

References

- Beaulieu, N. D., Dafny, L. S., Landon, B. E., Dalton, J. B., Kuye, I., & McWilliams, J. M. (2020). Changes in quality of care after hospital mergers and acquisitions. *New England Journal of Medicine*, 382(1), 51–59. <https://doi.org/10.1056/NEJMsa1901383>
- Centers for Medicare & Medicaid Services. (n.d.). *Linking quality to payment*. Provider Data Catalog. Retrieved October 25, 2026, from <https://data.cms.gov/provider-data/topics/hospitals/linking-quality-to-payment>
- Cotter, T., & Hastings, K. (2019, March). *Sample performance standards for evaluating annual incentive plans* [Excerpt from Key action steps for the compensation committee in 2019]. AHA Trustee Services. <https://trustees.aha.org/sample-performance-standards-evaluating-annual-incentive-plans>
- Daley Ullem, E., Gandhi, T. K., Mate, K., Whittington, J., Renton, M., & Huebner, J. (2018). *Framework for effective board governance of health system quality* [White paper]. Institute for Healthcare Improvement. <https://www.ihl.org/library/white-papers/framework-effective-board-governance-health-system-quality>
- Figueroa, J. F., Tsugawa, Y., Zheng, J., Orav, E. J., & Jha, A. K. (2016). Association between the Value-Based Purchasing pay-for-performance program and patient mortality in US hospitals: Observational study. *BMJ*, 353, i2214. <https://doi.org/10.1136/bmj.i2214>
- Internal Revenue Service. (n.d.). *Rebuttable presumption – Intermediate sanctions*. Retrieved October 25, 2026, from <https://www.irs.gov/charities-non-profits/charitable-organizations/rebuttable-presumption-intermediate-sanctions>
- Jenkins, D. (2026, July 16). *The growing gap between hospital CEO and direct care pay*. Rice University’s Baker Institute for Public Policy. <https://doi.org/10.25613/WSJY-BZ17>
- Jenkins, D., Short, M. N., & Ho, V. (2024). The determinants of nonprofit hospital CEO compensation. *PLOS ONE*, 19(7), e0306571. <https://doi.org/10.1371/journal.pone.0306571>
- Joynt, K. E., Le, S. T., Orav, E. J., & Jha, A. K. (2014). Compensation of chief executive officers at nonprofit US hospitals. *JAMA Internal Medicine*, 174(1), 61–67. <https://doi.org/10.1001/jamainternmed.2013.11537>
- Mulligan, K., Choksy, S., Ishitani, C., & Romley, J. A. (2022). What do nonprofit hospitals reward? An examination of CEO compensation in nonprofit hospitals. *PLOS ONE*, 17(3), e0264712. <https://doi.org/10.1371/journal.pone.0264712>
- SullivanCotter. (2026, September 24). *Health care executive compensation growth stabilizes as organizations sharpen focus on strategic leadership investment* [Press release].