Benchmarks and Quality Measures

Posted on July 15, 2025

Introduction

Benchmarks and quality measures help healthcare organizations evaluate whether care is safe, effective, timely, patient-centered, efficient, and equitable. A quality measure defines a specific aspect of structure, process, outcome, patient experience, or population health. A benchmark provides a reference point against which performance can be compared. The original essay correctly emphasizes compatibility of definitions, statistical methods, collection periods, and populations. These conditions are essential because a comparison is meaningful only when the measured concept and calculation are sufficiently aligned.

Healthcare data come from electronic health records, claims, registries, surveys, laboratories, pharmacies, health information exchanges, public-health systems, and manual abstraction. The same clinical event may be represented differently across these sources. One database may use diagnosis codes, another laboratory values, and another clinician documentation. Data standardization is therefore not a technical detail. It determines whether an apparent difference reflects care quality or merely a difference in coding, population, timing, or missing information.

Quality Measures

The Centers for Medicare & Medicaid Services describes quality measures as tools used to quantify healthcare processes, outcomes, patient perceptions, and organizational structures associated with quality goals. Measures may support improvement, accountability, public reporting, and payment. A good measure has a clear purpose, evidence base, precise specification, reliable data source, and interpretation that stakeholders can understand. (Centers for Medicare & Medicaid Services, 2026a; Centers for Medicare & Medicaid Services, 2026b)

A measure is not the same as quality itself. Hospital mortality, readmission, vaccination, waiting time, patient experience, and medication safety each capture only part of care. Organizations should use a balanced set rather than treat one score as a complete judgment.

Benchmarks

A benchmark may be an organization’s previous performance, a national average, a high-performing peer group, an evidence-based target, or a threshold established by a program. Different benchmarks answer different questions. Comparing with last year shows internal improvement, while comparing

with similar organizations shows relative performance.

The benchmark must be appropriate to the measure and population. A small rural clinic should not automatically be compared with a tertiary referral hospital receiving more complex cases. Peer groups, case mix, service type, and data completeness affect fairness.

Structure Measures

Structure measures describe the environment in which care is delivered. Examples include nurse staffing, availability of specialist services, use of computerized prescribing, infection-control capability, or accreditation. They are based on the assumption that certain resources and systems support good care.

Structure does not guarantee outcome. A hospital may possess advanced technology but use it poorly. Structural measures are strongest when evidence connects the resource with improved process or outcome.

Process Measures

Process measures assess whether a recommended action occurred, such as vaccination, screening, medication reconciliation, timely antibiotic administration, or follow-up after hospitalization. They are often actionable because managers can redesign workflow when performance is low.

A process measure should reflect a clinically meaningful action rather than documentation burden alone. Exceptions and contraindications must be defined so that appropriate individualized care is not penalized.

Outcome Measures

Outcome measures evaluate results such as mortality, complications, functional improvement, infection, symptom control, or readmission. They matter directly to patients but are influenced by illness severity, social conditions, and events beyond one provider’s control.

Risk adjustment attempts to account for relevant differences in patient populations. It should not adjust away disparities that the measure is intended to reveal. Variables such as race or socioeconomic status require thoughtful use because adjustment can normalize unequal outcomes.

Patient-Reported Measures

Patient-reported outcome measures capture symptoms, function, or quality of life directly from patients. Patient-experience measures assess communication, respect, access, coordination, and other aspects of care. These measures add information not contained in claims or laboratory data.

Survey results depend on language, accessibility, response rate, timing, and sampling. Nonresponse can bias the result if people with different experiences are less likely to participate. Organizations should examine both overall scores and qualitative comments.

Data Compatibility

Compatible data use the same or mapped definitions, units, populations, time periods, and calculation logic. Compatibility does not always require identical systems. Different systems can exchange comparable information when standards and transformations are documented and validated.

Before combining data, analysts should create a specification that identifies the measure name, purpose, numerator, denominator, exclusions, risk adjustment, data source, period, and reporting unit. Without this specification, two teams may use the same label for different calculations.

Clinical Definitions

The original essay uses diabetes as an example. Diabetes can be identified through diagnosis codes, medication use, laboratory thresholds, or clinical registries. A measure of glycemic control may use a specific HbA1c period, eligible age range, exclusion criteria, and value threshold. If one source includes gestational diabetes and another excludes it, direct comparison is invalid.

Clinical definitions change as evidence and guidelines evolve. Measure versions must be recorded. Data calculated under different specifications should not be merged without assessing the change.

Numerator and Denominator

The numerator counts the events meeting the performance condition, while the denominator defines eligible opportunities or patients. A rate can change because the numerator improves or because the denominator is altered. Analysts should inspect both.

Denominator exclusions can be clinically necessary but can also be misused to improve scores. Exclusion reasons should be coded consistently and audited. Small denominators create unstable rates and may require suppression or multi-year aggregation.

Units and Coding

Laboratory units must be harmonized before comparison. Medication names may require standardized vocabularies, and diagnoses may be coded in different versions or systems. Dates, time zones, race and ethnicity categories, and provider identifiers also require consistent representation.

Mapping is not always one-to-one. A local code may combine concepts that a national standard separates. Analysts should document loss of detail rather than claim perfect equivalence.

Statistical Methods

Organizations should use the same formula, weighting, risk model, confidence method, and treatment of missing data. An unadjusted average cannot be compared fairly with a risk-standardized rate. A median waiting time differs from a mean, particularly when extreme delays exist.

Confidence intervals communicate uncertainty. A small difference in point estimates may not represent a meaningful difference. Control charts can help distinguish normal process variation from a sustained change.

Time Periods

Data should refer to comparable periods. Seasonal illness, policy changes, staffing, and public-health emergencies can affect performance. Comparing one winter quarter with a summer quarter may produce misleading conclusions.

Reporting lag also matters. Claims data may be complete months after care, while real-time clinical data are available sooner but may change as records are finalized. The measure should state when data are considered complete.

Population Compatibility

Age, diagnosis, severity, insurance, geography, and social conditions affect outcomes. Analysts should compare like populations or apply appropriate stratification and adjustment. Combining adults and children, inpatient and outpatient cases, or elective and emergency procedures may conceal important differences.

Population compatibility does not justify ignoring equity. Performance should also be stratified by race, ethnicity, language, sex, disability, rurality, and other relevant factors where data quality and privacy permit. Overall averages can hide groups experiencing poor care.

Data Completeness

A measure based on only a fraction of eligible cases may be biased. CMS programs commonly establish data-completeness and case-minimum requirements before a measure can be scored against a benchmark. The exact requirement varies by program and year.

Missingness should be analyzed, not simply removed. If follow-up outcomes are missing mainly for uninsured patients or people with unstable housing, the observed result may appear better than reality.

Data Validity

Validity asks whether the data represent what they claim to measure. A coded diagnosis may be entered for billing and may not indicate current disease severity. A checkbox may show that education was documented without proving understanding.

Validation can include chart review, source comparison, logic checks, and testing against known cases. Measure developers should evaluate whether performance differences reflect quality and whether the measure can be manipulated.

Reliability

Reliability concerns consistency. A measure should produce stable differences when true performance is stable and should not be dominated by random variation. Reliability depends on event frequency, sample size, measurement error, and variation among providers.

Low-reliability measures should not be used for high-stakes ranking or payment without caution. Combining periods or levels may improve stability but reduce timeliness.

Risk Adjustment

Risk adjustment estimates expected outcomes based on factors present before care. It can make comparisons fairer when providers serve different clinical populations. Variables should be clinically justified, measured consistently, and not influenced by the care being evaluated.

Models require periodic recalibration because populations and treatment change. Their performance should be examined across demographic groups. A model that systematically underpredicts risk for one group can create inequitable ratings.

Benchmark Selection

An internal benchmark is useful for continuous improvement but can normalize low performance. A national average provides context but represents ordinary rather than excellent care. Top-decile performance can inspire improvement but may reflect different resources or selection.

Evidence-based targets are preferable when a clear standard exists. Some measures have no reasonable goal of 100 percent because contraindications, patient preference, and clinical complexity matter. Targets should not encourage inappropriate treatment.

Health Information Exchange

A health information exchange enables secure electronic sharing of patient information among authorized organizations. HIE can reduce duplication, improve medication reconciliation, support transitions, and provide broader data for population health. It is not necessarily one regional database; models include centralized, federated, and network-based exchange.

Data quality affects usefulness. An incomplete medication list, duplicate patient, outdated allergy, or delayed laboratory result can create clinical risk. Receiving more data is not automatically better when clinicians cannot identify relevance or source.

Patient Matching

HIE depends on correctly linking records to the same person. Variation in names, addresses, dates, and identifiers can create duplicates or false matches. False negatives fragment the record, while false positives combine information from different people.

Matching methods require governance, validation, and correction procedures. Patients should have a way to report errors. Sensitive demographic data should be protected.

Interoperability

Interoperability means systems can exchange and use information. Technical transport is only one layer. Semantic interoperability requires shared meaning, and organizational interoperability requires policies, trust, workflow, and authority.

Standards such as FHIR and standardized clinical vocabularies can support exchange, but implementation differences remain. Conformance testing and version management are necessary.

HIE and Quality Measurement

HIE data can broaden denominators, capture care across organizations, and reduce duplicate reporting. It may improve measurement of follow-up, medication, and utilization. However, participation may be incomplete, and data may be missing from organizations outside the network.

Analysts should state coverage and avoid presenting an HIE dataset as a complete population when it is not. Provenance should show where each item originated.

National Registries

A registry collects standardized information concerning a condition, procedure, device, or population. Registries can support benchmarking, safety surveillance, research, and improvement. They differ from HIE because their data are organized for a defined purpose rather than broad clinical exchange.

Registry participation and submission quality affect conclusions. Facilities need common specifications, validation, and feedback. Incorrect data can distort national estimates and unfairly compare institutions.

Measure Harmonization

Healthcare professionals often report similar but slightly different measures to multiple payers and programs. This creates burden and inconsistent incentives. The Core Quality Measures Collaborative promotes alignment of core sets across stakeholders. (Core Quality Measures Collaborative, 2026)

Harmonization should reduce duplication without forcing unlike purposes into one measure. A quality-improvement measure may need rapid local data, while a payment measure requires formal validation and audit.

Digital Quality Measures

Digital quality measures use electronically captured data and computable specifications. They can reduce manual abstraction and support timely feedback. CMS initiatives increasingly emphasize digital measurement and interoperable data.

Automation does not eliminate error. Electronic records may contain copied text, missing structured fields, or workflow differences. The algorithm should be tested against clinical review and monitored after system upgrades.

Benchmarking Process

The organization should define the improvement question, select a valid measure, verify data, identify an appropriate peer or target, analyze variation, and involve clinicians and patients in interpreting results. A low score should lead to investigation of both care and measurement.

Action plans need owners, timelines, and process indicators. Repeating the benchmark report without changing workflow does not improve quality.

Example: Diabetes Quality

A diabetes program might measure HbA1c control, blood-pressure management, kidney screening, eye examination, medication adherence, and patient-reported burden. The denominator must define diabetes type, age, enrollment, and exclusions. Laboratory values need consistent units and dates.

Results should be stratified to identify inequity. If one community has lower control because appointments, medication, or healthy food are inaccessible, education alone will not solve the gap.

Data Governance

Governance establishes ownership, definitions, access, stewardship, correction, retention, and accountability. A data dictionary should identify each field and measure. Changes should be approved and versioned.

Clinical, technical, statistical, legal, privacy, and patient perspectives are needed. Data teams cannot decide the meaning of care alone, while clinicians may not recognize technical limitations without collaboration.

Privacy and Security

Quality measurement uses sensitive health information. Access should be limited to purpose, and reporting should protect individuals from reidentification. Data sharing requires legal authority, security controls, vendor oversight, and breach response.

Privacy should not be used as a blanket excuse to avoid equity analysis. De-identified, aggregated, or controlled approaches can support legitimate improvement while reducing risk.

Unintended Consequences

Measures can encourage gaming, avoidance of high-risk patients, excessive documentation, or focus on measured conditions at the expense of unmeasured care. Public rankings may be misunderstood when differences are small or data are old.

Measure sets should be reviewed for burden and unintended behavior. CMS's Meaningful Measures approach emphasizes high-value measures, outcomes, alignment, and reduction of unnecessary burden.

Trust

Stakeholder trust grows when definitions are transparent, data can be corrected, limitations are acknowledged, and measures lead to useful action. Trust is damaged when organizations publish rankings without explaining uncertainty or change specifications without notice.

Patients and clinicians should be involved in selecting measures so that reporting reflects outcomes that matter. Technical validity is necessary but not sufficient for legitimacy.

Conclusion

Healthcare benchmarks and quality measures are valuable only when definitions, populations, methods, periods, and data sources are compatible. Standardization enables comparison, but it must be accompanied by validation, reliability assessment, risk adjustment, equity analysis, and transparent limitations.

Health information exchanges and registries can improve the completeness of clinical and quality data, yet incorrect matching, missing records, inconsistent coding, and delayed submission can create harmful conclusions. Interoperability requires shared meaning and governance, not merely data transfer.

Organizations should use benchmarks to guide improvement rather than to produce rankings alone. A balanced measure set, appropriate peer group, trustworthy data, and a clear action process can identify gaps and monitor progress. Incompatible or low-quality data can misdirect care, payment, and policy, making data quality itself a central component of healthcare quality.

References

Centers for Medicare & Medicaid Services. (2026a). *Quality measures and Measures Management System*.

Centers for Medicare & Medicaid Services. (2026b). *Meaningful Measures 2.0*.

Core Quality Measures Collaborative. (2026). *Core measure sets*.