# Chapter 10
# Creating and Defining Quality Metrics That Matter in Surgery

**Anai N. Kothari and Thomas A. Aloia**

## Creating and Defining Quality Metrics That Matter in Surgery

In March of 2013, Kirk Goldsberry and Eric Weiss introduced "The Dwight Effect" at the MIT Sports Analytics Conference. At that point, the NBA lagged behind other professional sports leagues in the adoption of advanced analytic techniques to evaluate in-game performance [1]. This was especially true for defensive performance, which was difficult to measure and effectively characterize. Since basketball has two key objectives – scoring points and preventing points –

A. N. Kothari
Department of Surgery, Division of Surgical Oncology, Medical College of Wisconsin, Milwaukee, WI, USA
e-mail: akothari@mcw.edu

T. A. Aloia (✉)
University of Texas MD Anderson Cancer Center, Houston, TX, USA
e-mail: thomas.aloia@ascension.org

177

not being able to assess the latter was a significant shortcoming. By combining optical tracking data with visual and spatial analytics, Goldsberry and Weiss were able to reframe how defense in the NBA could be measured [2, 3].

"The Dwight Effect" was the foundation for a new set of advanced defensive metrics that have since led to a transformation in the way basketball is played in the NBA [4]. However, it was not simply creating new metrics that led to this impact – the NBA was already awash in static measures of performance. Instead, insights were obtained by using a deeper understanding of the interaction of the players to identify novel independent variables that better correlated to performance outcomes. Creating and defining quality metrics "that matter" in surgery should have similar focus.

In healthcare, quality metrics are used in multiple ways. From benchmarking to quality improvement efforts to public reporting to reimbursement, quality measures are crucial to support assessment and improvement at the provider, hospital, system, and societal level. Particularly relevant to surgeons, the measurement movement has motivated hospitals and regulatory bodies to transparently report metrics that attempt to measure high-quality surgery. However, attempts to simply apply existing healthcare quality metrics to surgery are limited by inadequate adjustment of risk and incomplete consideration of the unique aspects of perioperative care. As a result, there is a strong incentive for surgeons to move from the sidelines to the playing field when it comes to quality measure design [5].

In this chapter, we introduce aspects of surgical care that can be measured and what data are available to create metrics. We then describe a framework for identifying quality metrics in surgery that matter to patients and providers and the key steps for creating/defining these metrics. Finally, we provide a design tool to create new metrics for surgical application.

# From Measuring Surgical Care to Designing Metrics

Approaching metric construction from a design-thinking perspective, the end goal of all measures is quantifying value to achieve the "quadruple aim" (improving patient experience, improving health of populations, reducing the per capita cost of healthcare, and improving the well-being of healthcare providers) [6, 7]. The traditional value equation accounts for outcomes and cost. We propose a modified value equation that further specifies the numerator by accounting for both quality (achieving a positive outcome) and safety (avoidance of harm) (Fig. 10.1) [5]. This can provide a foundation for identifying the inputs necessary to develop new surgical metrics.
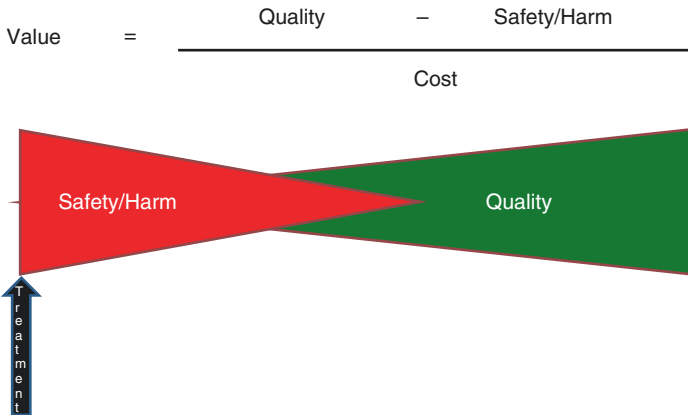


FIGURE 10.1 A framework for developing metrics for surgeons and surgical patients that emphasizes the pursuit of high value care by accounting for quality, safety/harm, and cost. (Aloia et al. [5])

## *Quality*

In 1966, Donabedian introduced a conceptual model for the assessment of quality. Donabedian proposed three major criteria of quality in medical care – structure, process, and outcomes. Each Donabedian component interacts with and influences the next, where structure is defined as the setting where care occurs, process refers to how care is delivered, and outcomes refer to the impact of care [8]. Of these, outcomes are seemingly the most important. However, while some outcomes like mortality are unmistakable, others can be less clear, making them challenging to specify. This has led to a reliance on using structural and process measures to define surgical quality metrics [9].

Existing surgical quality metrics can be grouped based on the Donabedian framework. Postoperative mortality, complications, length of stay, and readmission are outcome indicators. Adherence to components of enhanced recovery after surgery programs and surgical care improvement project (SCIP) measures are examples of commonly used process indicators. Hospital and surgeon volume, nursing ratios, and external designations/accreditations are each structural indicators [10]. In the design of new metrics, using the Donabedian model can provide a template to organize these efforts. Balancing metric value with the work required to obtain data, it is recommended that measure sets contain a balanced portfolio of structure, process, and outcomes measures.

Data that can be used to define quality metrics are available through multiple existing internal and external sources. While impossible to detail all possible data sources, we highlight four major resources: clinical records, registries, billing data/claims, and federal agencies/programs.

Donabedian recognized the important role clinical records play in the assessment of quality. Specifically, patient records provide a narrative summary of how structure, process, and outcomes come together to impact individual patients. However, concerns surrounded their use due to incompleteness and inaccuracy. Many of these concerns have amplified

with the transition to the routine use of electronic clinical records [11]. While the electronic health record (EHR) comes with significant promise in the ability to obtain relevant quality improvement data given the availability of electronic documentation, prescription and test information, diagnostics, and many other elements, it is subject to inaccuracies during data entry [12]. Still, electronic health data are an important input to the design of quality metrics.

Clinical registries and databases provide important data for surgeons to use for designing quality metrics. These include institutional databases, local and regional collaborative data-sharing programs, and national datasets that aggregate outcome, process, and structural data [13]. Examples such as the American College of Surgeon's NSQIP program, the Michigan Bariatric Collaborative, and the ACS/NCI's National Care Database provide aggregated data that can be used to measure quality.

Several federal agencies and programs have been established to measure and report on the quality of care, including surgical quality [14]. The Center for Medicare and Medicaid Services (CMS) has several quality programs. These include CAHPS (Consumer Assessment of Healthcare Providers and Systems), Hospital Compare, and the Hospital Inpatient Quality Reporting Program (IQR). Another is the Agency of Health Quality and Research (AHRQ) that has developed several quality indicators, including prevention quality indicators (PQI), inpatient quality indicators (IQI), and patient safety indicators (PSI) [15]. While these are all established quality metrics in their own right, they can also provide a source of data to develop new metrics and adapt for local use.

## Safety

Quality is not the only aspect of care that contributes to the numerator of the value equation. Quality closely interacts with safety – together, determining an outcome. Unlike more well-defined models for measuring health quality, no univer-

sal approach exists for measuring patient safety. Instead, the focus is on "zero harm" or the avoidance of a negative outcome [16]. Interestingly, most measures that are labeled as surgical quality metrics are better defined as harm metrics. In fact, 95% of publicly reportable metrics in healthcare are harm metrics. For example, postoperative wound infections, deep vein thrombosis, pneumonia, and others are all harm events. Preventing these occurrences, when viewed in the context of the modified value equation, incompletely achieves the aim of high-value care, since they do directly incentivize surgeons to strive for higher-quality/positive outcomes.

There are other collateral damages that can arise from focusing solely on harm metrics including impaired patient access, arrested innovation, challenges in training, and surgeon burnout [17]. Harm metrics can lead to perceived high-risk patients not receiving the same care as their lower-risk counterparts. This is seen across multiple specialties where publicly reported metrics appear to influence decisions for offering surgical treatment [18]. Similarly, in an attempt to avoid harm and promote safety, the process for developing, testing, and implementing new techniques can be slowed [19]. For surgeons that are primarily being measured using harm metrics, there is a potential disincentive to educate and provide trainees the necessary autonomy to help them develop toward their own future independent practice. Finally, the emphasis on harm metrics can contribute to surgeon burnout – few enter surgical practice motivated to avoid harm but instead are intrinsically driven to achieve high quality. Therefore, creating metrics centered on achieving quality might have advantage over those focused on preventing harm [17]. Optimally, metrics that strive to improve quality while also mitigating harm should be prioritized. Ultimately, developing a balanced portfolio of harm and quality measures is more likely to achieve higher levels of value realization.

## *Cost*

The third element and the denominator of the value equation is cost. Measurement of cost quantifies the financial burden associated with rendering a healthcare service, usually in a single episode of care. The challenge in measurement of cost is that it is an opaque term and can describe several items including patient out-of-pocket payments, charges, prices, provision of care costs, indirect costs, and acquisition costs. We propose a framework that focuses on three "real-dollar" domains: [1] patient-borne cost, [2] third-party payors, and [3] institutional cost [20].

Patient-borne cost summarizes the direct and indirect expenses taken on by patients for the care they receive. For example, out-of-pocket costs can be estimated by using copays and deductibles. These direct expenses can be assessed using patient-level billing data. Another type of cost incurred by the patient is indirect and more difficult to assess. Examples of indirect patient costs include lost wages and travel costs. Third-party payor costs focus on reimbursement contributions from insurance companies or governmental health plans. Characterizing and measuring these can be complex, particularly before the initiation of treatment due to lack of transparency into, plan maximums, charge to reimbursement ratios, stop loss provisions, and other differences in third-party contracts. However, post-therapy accounting has become more transparent as penetration of electronic billing platforms embedded in electronic health records has increased access to precise payor funds flow data. Institutional cost includes all of the procurement and production expenses to provision care, such as equipment, pharmacy, staff, services, time, infrastructure, information technology, and many other inputs, both direct and indirect.

# Checklist for Creating Surgical Metrics

As described, the volume of data available to surgeons continues to grow at an exponential pace and comes from multiple sources. Data alone, however, do not lead to actual insight, change, and improvement. Instead, data must be translated into usable metrics. This process can be facilitated using a consistent framework. This level of standardization has the advantage of avoiding the temptation to create unneeded and redundant metrics – which is a common practice when faced with increasing available data streams [6].

Several national groups including the National Quality Forum (NQF), CMS, and Physician Consortium for Performance Enhancement (PCPI) have developed guiding principles for developing new measures. Leveraging the expertise of these regulatory organizations can help ensure the creation of high-quality surgical metrics. CMS has formulated a standardized approach that is used across all of the agency's quality programs and initiatives [21]. This blueprint can be broken down into five stages: conceptualization, specification, testing, implementation, and evaluation (Fig. 10.2).

## *Conceptualization*

The initial step to developing a new surgical metric is considering how it will enhance the healthcare system [22]. High-quality metrics should be meaningful to multiple stakeholders including providers, administrators, and patients. To accomplish this, focusing on high-impact areas with real opportunity for improvement is essential. Other considerations include minimizing the burden on providers to both use and collect the measure, prioritizing electronic data to specify the metric, reducing care delivery disparities, and aligning the metric with other quality improvement programs (both local and national) [23].
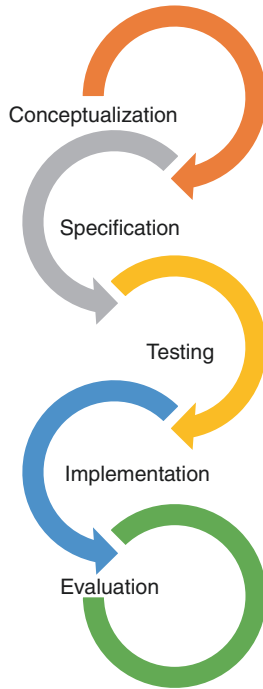
Figure 10.2 A checklist for developing surgical quality metrics that matter

The conceptualization phase includes information gathering, engaging subject and content experts, and a public comment process. Information gathering is arguably the most important step of the entire process and focuses on obtaining data that will eventually be used to justify the metric's implementation. This requires a comprehensive literature search and understanding what existing clinical guidelines are already in place. The focus should be on creating a metric that leads to better population health, better care, and/or more affordable care [24].

## *Specification*

Following conceptualization, the next step is specifying the measure. This includes detailing the elements of the metric, defining the type of metric, and determining necessary data sources. Both CMS and NQF outline that quality measures should include a title/description, numerator, denominator, exclusions, and rationale [21].

As an example, the NQF-endorsed quality measure "perioperative temperature management" illustrates how to apply these principles to quality metric specification.

The title/description of the perioperative temperature management measure is "Percentage of patients, regardless of age, who undergo surgical or therapeutic procedures under general or neuraxial anesthesia of 60 minutes duration or longer for whom at least one body temperature greater than or equal to 35.5 degrees Celsius (or 95.9 degrees Fahrenheit) was achieved within the 30 minutes immediately before or the 15 minutes immediately after anesthesia end time." Title/descriptions should clearly describe the population of interest and specify the objective [25].

The numerator for perioperative temperature management is "Patients for whom at least one body temperature greater than or equal to 35.5 degrees Celsius (or 95.9 degrees Fahrenheit) was achieved within the 30 minutes immediately before or the 15 minutes immediately after anesthesia end time." Numerators specify what is necessary to achieve compliance with the measure.

The denominator for perioperative temperature management is "All patients, regardless of age, who undergo surgical or therapeutic procedures under general or neuraxial anesthesia of 60 minutes duration or longer." The denominator exclusions are "monitored anesthesia care and peripheral nerve block." Quality measure denominators describe the total population a metric will be applicable to and highlight those excluded from the measure.

Specifying the metric also involves identifying the necessary data sources to calculate the measure. Measures can be

based on a single source or multiple sources of data inputs including administrative data, electronic clinical data, standardized patient assessments, medical records, surveys, and registries.

## Testing

Prior to launch, a rigorous assessment of the technical and scientific merit of the measure should be conducted [21]. This is based on four general criteria: importance, scientific acceptability, feasibility, and usability. During this stage, exploration of data collection in a real-world setting is reviewed. Can the data actually be collected? Does collecting data for the metric create substantial hardship (financial cost, number of people needed to maintain the data, etc.)? And, most importantly, does the data being collected measure what was intended? The testing phase is often iterative and requires multiple cycles prior to moving on to the next stage.

## Implementation

The implementation stage includes endorsement and complete rollout of the new metric. There are multiple consensus groups that can endorse a new metric including national organizations (ACS, NQF), specialty societies, and local/regional groups. The endorsement process can be long and happen in parallel to the actual rollout of a new measure.

Rollout planning includes preparing for audit and validation, provider education, and pilot programs. In fact, the implementation stage is primarily an education phase – where developers ensure end users understand the purpose of the measure and how to use it. Pilot programs offer a gradual rollout and can help provide feedback from stakeholders using the measure to further improve usability/compliance.

## *Evaluation*

The final stage in metric creation includes the actual use of the measure, ensuring a process for evaluation, and continued maintenance. While a tremendous amount of energy and effort are required to place a new surgical metric into practice, that is just the beginning of the process. The most impactful and consistently used metrics are subject to constant scrutiny. This ensures they remain receptive to changes in literature, public feedback, and maintain scientific validity. Above all, the evaluation phase challenges all metrics to remain relevant and promote quality improvement.

Evaluation of new metrics includes active, ongoing information surveillance. This is similar to the information gathering conducted during development. Many surgical metrics are example of this. For instance, the Surgical Care Improvement Project (SCIP) was developed as a national program to help improve surgical care. SCIP developed several performance measures to help reduce surgical site infections, cardiovascular complications, venous thromboembolism, and respiratory complications [26]. Following broad implementation, many of the SCIP measures were studied to understand if they were actually achieving their intended aim [27]. While they may have contributed to improve surgical care, over time the adherence to the metrics was close to 100% making the impact of the measures difficult to interpret. Ultimately, this ongoing evaluation translated to change, and the SCIP measures were retired in 2015.

Other important considerations during the continued evaluation phase include reassessing the data collected (are there better ways or improved data inputs?), comparisons to other similar measures (are there places of overlap?), and maintenance reviews (should the metric be retained, revised, retired, suspended, removed?).

# Model for Patient-Centric Surgical Outcome Measure Development

A possible way to facilitate the conceptualization of new metrics is through surgical societies. These groups are preassembled expert panels and include stakeholders with significant domain expertise. To aid in this process, a structured template can be helpful to allow individuals that may not have formal training in measure development an opportunity to actively participate. We developed a novel tool that leverages the components of the modified value equation to inform the discussion (Table 10.1) [5]. This tool separates patient-centered outcomes into the following domains: safety/harm, quality, short-term utility/disutility, long-term utility/disutility. Use of this tool has been shown to rapidly produce focused procedure-specific metric sets that can be refined through fit testing with patients [5].

Table 10.1 A tool for the development of new surgical quality metrics

| |
|---|
| Relevance: Is it a meaningful measure that identifies potential for improvements? |
| Scientific soundness: Is it a scientifically valid, accurate, and reproducible measure? Is there clinical evidence to support its use? Can it provide a process–outcome link? |
| Feasibility: Is it fiscally and logistically workable? Can it be precisely specified and conducted within confidentiality parameters? Is it auditable? |
| Comprehensiveness: How extensive is the information yielded through the measure? |
| Quality metric: Survival, resolution of symptoms, and/or degree of recovery |
| 1. Select a procedure |
| 2. List the most common diagnoses that indicate that procedure |

Table 10.1  (continued)

3. Symptom burden: Describe the most common symptoms patients with those diagnoses present with

4. Life interference: Of the symptoms in #3, list 2–3 that are the most disabling to the patient (ability to eat, walk, work, care for self, care for others, enjoy life)

5. Circle the symptoms in #4 that improve/resolve with a "technically successful" procedure in more than 50% of cases

6. Write a metric that addresses one of the circled items

7. Define metric failure

Safety metric: Short-term complication (including readmission)

| 8. List the 3 most common 30-day surgical complications for the procedure | 9. List the median complication grade that occurs with each complication: 5 – death; 4 – organ failure, ICU; 3 – rescue procedure (IR or OR); 2 – medical management at bedside; 1 – no specific intervention) | 10. Multiply across, then circle the row with the highest number |
|---|---|---|

First most common (5points)

Second most common (3 points)

Third most common (1 point)

11. Write a metric that addresses the circled item

12. Define metric failure

Safety metric: Long-term disutility

13. List the most common surgically induced disability present at 6–12 months postoperatively

14. Write a metric that addresses the item

15. Define metric failure

Aloia et al. [5]

## Conclusion

The perfect surgical metric is likely unattainable. However, creating and defining metrics that matter is a worthwhile effort for surgeons to engage in. With ongoing external pressure for transparent reporting, ensuring surgical metrics are meaningful is of paramount importance. Surgical leadership in developing, specifying, and implementing new measures is crucial. As seen in both healthcare and non-healthcare applications, performance metrics have important consequences – they can reshape the game. A systematic and rigorous approach to metric development can provide assurance that any resultant changes are for the better.

## References

1. Goldsberry K, Weiss E. The Dwight effect: a new ensemble of interior defense analytics for the NBA. Sports Aptitude, LLC Web 2013;1–11.
2. Franks A, Miller A, Bornn L, Goldsberry K. Counterpoints: advanced defensive metrics for nba basketball [Internet]. In: 9th annual MIT Sloan sports analytics conference, Boston, MA. lukebornn.com; 2015. Available from: http://www.lukebornn.com/papers/franks_ssac_2015.pdf.
3. Sampaio J, McGarry T, Calleja-González J, Jiménez Sáiz S, Schelling I Del Alcázar X, Balciunas M. Exploring game performance in the national basketball association using player tracking data. PLoS One. 2015;10(7):e0132894.
4. Verrier J. You can't stop NBA offenses—and now, you can't even Hope to contain them [Internet]. The Ringer. 2018 [cited 2020 Nov 29]. Available from: https://www.theringer.com/nba/2018/10/30/18038802/nba-defense-offensive-boom.
5. Aloia TA, Jackson T, Ghaferi A, Dort J, Schwarz E, Romanelli J. Developing minimally invasive procedure quality metrics: one step at a time. Surg Endosc. 2019;33(3):679–83.
6. Henry LR, von Holzen UW, Minarich MJ, et al. Quality measurement affecting surgical practice: utility versus utopia. Am J Surg. 2018;215(3):357–66.

7. Bodenheimer T, Sinsky C. From triple to quadruple aim: care of the patient requires care of the provider. Ann Fam Med. 2014;12(6):573–6.

8. Berwick D, Fox DM. "Evaluating the quality of medical care": Donabedian's classic article 50 years later. Milbank Q. 2016;94(2):237.

9. Birkmeyer JD, Dimick JB, Birkmeyer NJO. Measuring the quality of surgical care: structure, process, or outcomes? J Am Coll Surg. 2004;198(4):626–32.

10. Merkow RP, Bilimoria KY, Ko CY. Surgical quality measurement: an evolving science. JAMA Surg. 2013;148(7):586–7.

11. Tang PC, Ralston M, Arrigotti MF, Qureshi L, Graham J. Comparison of methodologies for calculating quality measures based on administrative data versus clinical data from an electronic health record system: implications for performance measures. J Am Med Inform Assoc. 2007;14(1):10–5.

12. Chan KS, Fowles JB, Weiner JP. Electronic health records and the reliability and validity of quality measures: a review of the literature. Med Care Res Rev. 2010;67(5):503–27.

13. Stey AM, Russell MM, Ko CY, Sacks GD, Dawes AJ, Gibbons MM. Clinical registries and quality measurement in surgery: a systematic review. Surgery. 2015;157(2):381–95.

14. Burstin H, Leatherman S, Goldmann D. The evolution of healthcare quality measurement in the United States. J Intern Med. 2016;279(2):154–9.

15. Elixhauser A, Pancholi M, Clancy CM. Using the AHRQ quality indicators to improve health care quality. Jt Comm J Qual Patient Saf. 2005;31(9):533–8.

16. Thomas EJ. The harms of promoting "Zero Harm". BMJ Qual Saf. 2020;29(1):4–6.

17. Aloia TA. Should zero harm be our goal? Ann Surg. 2020;271(1):33.

18. Shahian DM, Jacobs JP, Badhwar V, D'Agostino RS, Bavaria JE, Prager RL. Risk aversion and public reporting. Part 1: observations from cardiac surgery and interventional cardiology. Ann Thorac Surg. 2017;104(6):2093–101.

19. Marcus RK, Lillemoe HA, Caudle AS, et al. Facilitation of surgical innovation: is it possible to speed the introduction of new technology while simultaneously improving patient safety? Ann Surg. 2019;270(6):937–41.

20. Porter ME, Kaplan RS, Frigo ML. Managing healthcare costs and value. Strategic Finance [Internet]; 2017. Available from:

http://search.proquest.com/openview/0147acbebd6fb625f241bcf9b3a60f77/1?pq-origsite=gscholar&cbl=48426.

21. Centers For Medicare & Medicaid Services. Blueprint for the CMS measures management system. 2017.

22. Conceptualization [Internet]. [cited 2020 Nov 27]. Available from: https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/MMS/MSP-Conceptualization.

23. Krishnan M, Brunelli SM, Maddux FW, et al. Guiding principles and checklist for population-based quality metrics. Clin J Am Soc Nephrol. 2014;9(6):1124–31.

24. McGlynn EA, Adams JL. What makes a good quality measure? JAMA. 2014;312(15):1517–8.

25. National Quality Forum Measure. Perioperative temperature management – national quality strategy domain: patient safety – meaningful measure area: preventable healthcare harm. Quality ID #424 (NQF 2681) [Internet] 2018. Available from: https://qpp.cms.gov/docs/QPP_quality_measure_specifications/CQM-Measures/2019_Measure_424_MIPSCQM.pdf.

26. Stulberg JJ, Delaney CP, Neuhauser DV, Aron DC, Fu P, Koroukian SM. Adherence to surgical care improvement project measures and the association with postoperative infections. JAMA. 2010;303(24):2479–85.

27. Fry DE. Surgical site infections and the surgical care improvement project (SCIP): evolution of national quality measures. Surg Infect. 2008;9(6):579–84.