

Guiding Principles and Checklist for Population-Based Quality Metrics

Mahesh Krishnan,* Steven M. Brunelli,* Franklin W. Maddux,[†] Thomas F. Parker III,[‡] Douglas Johnson,[§] Allen R. Nissenson,* Allan Collins,^{||} and Eduardo Lacson Jr.[¶]

Abstract

The Centers for Medicare and Medicaid Services oversees the ESRD Quality Incentive Program to ensure that the highest quality of health care is provided by outpatient dialysis facilities that treat patients with ESRD. To that end, Centers for Medicare and Medicaid Services uses clinical performance measures to evaluate quality of care under a pay-for-performance or value-based purchasing model. Now more than ever, the ESRD therapeutic area serves as the vanguard of health care delivery. By translating medical evidence into clinical performance measures, the ESRD Prospective Payment System became the first disease-specific sector using the pay-for-performance model. A major challenge for the creation and implementation of clinical performance measures is the adjustments that are necessary to transition from taking care of individual patients to managing the care of patient populations. The National Quality Forum and others have developed effective and appropriate population-based clinical performance measures quality metrics that can be aggregated at the physician, hospital, dialysis facility, nursing home, or surgery center level. Clinical performance measures considered for endorsement by the National Quality Forum are evaluated using five key criteria: evidence, performance gap, and priority (impact); reliability; validity; feasibility; and usability and use. We have developed a checklist of special considerations for clinical performance measure development according to these National Quality Forum criteria. Although the checklist is focused on ESRD, it could also have broad application to chronic disease states, where health care delivery organizations seek to enhance quality, safety, and efficiency of their services. Clinical performance measures are likely to become the norm for tracking performance for health care insurers. Thus, it is critical that the methodologies used to develop such metrics serve the payer and the provider and most importantly, reflect what represents the best care to improve patient outcomes.

Clin J Am Soc Nephrol 9: 1124–1131, 2014. doi: 10.2215/CJN.11061013

Introduction

ESRD care in the United States is a data-dense therapeutic area because of, in part, the predominance of Medicare as a payer and the associated regulatory oversight, as well as the nature of the discipline of nephrology, which is a highly quantitative specialty. Because of the richness of available data, the ESRD program has been a vanguard for improvements in the delivery of health care for the complex chronically ill within Medicare. In 1983, ESRD became the first disease state to have a bundled payment system and was among the first to have a payment-for-performance system as mandated by the Medicare Improvements for Patients and Providers Act of 2008; this system was unique, because it was a withhold rather than a reward pay for performance. ESRD data are aggregated at both the dialysis facility and system levels and reported through a national registry in collaboration with the Centers for Medicare and Medicaid Services (CMS) and the National Institutes of Health. Using such data, the nephrology community has created and used population-based quality metrics to describe performance and drive improvements in care.

Maintaining a population-based view is essential when creating quality metrics, regardless of whether for driving improved outcomes, reporting, or reimbursement,

particularly at the unit of aggregation for which a measure is intended. The CMS has already signaled its intent to use such metrics and systems to improve outcomes throughout the Medicare program. The development and use of these metrics are shifting the care delivery paradigm from volume to value, emphasizing the three aims of better population outcomes, better individual patient experience of care, and constraint of health care costs. High-level guiding principles for identifying and creating population-based metrics have been described by senior Medicare leaders in the published literature (1,2). One entity assuming a prominent role for the translation of CMS principles into endorsement of actual facility-level metrics is the National Quality Forum (NQF), a nonprofit, nonpartisan, public service organization committed to the transformation of health care to be safe, equitable, and of the highest value. Measures considered for endorsement by the NQF may be proposed by the CMS after contractors have conducted literature reviews and Technical Expert Panels (TEPs) have opined. Measures may also be submitted to the NQF for endorsement by others in the community, including professional societies.

Historically, the process for measure development has been driven solely by the CMS, but recently, there

*DaVita Clinical Research, DaVita HealthCare Partners Inc., Denver, Colorado; [†]Clinical and Scientific Affairs, and [‡]Clinical Science, Epidemiology, and Research, Fresenius Medical Care North America, Waltham, Massachusetts; [§]Renal Ventures Management, Lakewood, Colorado; [¶]Dialysis Clinic, Inc., Nashville, Tennessee; and ^{||}Chronic Disease Research Group, Minneapolis, Minnesota

Correspondence: Dr. Mahesh Krishnan, DaVita Clinical Research, 825 South 8th Street, Suite 300, Minneapolis, MN 55404. E-mail: Mahesh.Krishnan@Davita.com

has been collaboration with other sister Health and Human Services agencies, such as the Agency for Healthcare Research and Quality and the Centers for Disease Control and Prevention (National Healthcare Safety Network Blood Stream Infection Measure). Quality measures for the ESRD program are derived from the six domains of quality measurement based on the national quality strategy (3). From these domains and legislatively mandated quality domains, the CMS selects a limited number of possible measures to be developed by an external contractor. The contractor then conducts a literature review and solicits nominations for experts from the community. The contractor subsequently appoints a committee chair, who then selects committee members to serve on a TEP. The TEP convenes in person and is supplied with the necessary background materials and sometimes, snapshots of the data. The TEP then recommends a select number of measures. These recommendations are converted to measure specifications, which are then sent to the NQF for endorsement. The newly developed measures are incorporated in each year’s proposed CMS ESRD rule and subject to public comment from the community. Given this process, this paper may serve as a guidance document for clinicians and professionals involved in quality measure development.

The NQF has provided rigorous criteria using a stepwise process to ensure that a given measure is reliable and valid for facility-level comparisons, and that adequate investigations have validated both the measure and the method used to apply the metric (4). However, to date, the most appropriate method of translating patient-level data available from the literature to population-based metrics that can be used as accountability measures has not been well defined. This article attempts to define that method for the United States dialysis community by reviewing the NQF criteria in a sequential fashion to create a parallel clinical performance measure (CPM) development checklist (Table 1).

NQF Criteria 1: Evidence, Performance Gap, and Priority (Impact) of a Measure—Importance to Measure and Report

A key descriptor for the development of a potential CPM is importance, which is often quantified with respect to

rationale, scope, and impact on care quality. The NQF defines this first importance criterion as the

extent to which the proposed measure focus is evidence-based, important to making significant gains in health-care quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance (4).

If a proposed measure does not present reasonable responses to address this importance, the criterion should be set aside and not used in performance payment models.

Checklist Item 1: Evidentiary Scope and Recognizing the Difference between Patient and Population Outcomes

Traditionally, prospective randomized clinical trials (RCTs) have been treated as the gold standard of evidence. Such studies result in an understanding of whether an individual patient may or may not benefit from a specific intervention. That finding differs from whether a population of patients (*e.g.*, a dialysis unit’s census) may, on average, benefit collectively from achieving an outcome. Given this finding, one must carefully evaluate the literature on both the strength of evidence and the applicability of the results to the population for which metrics are being created. Specifically, to ensure that the population studied reflects the population addressed by the proposed CPM, one must recognize that rigid inclusion and exclusion criteria limit the ability to extrapolate RCT findings in the real world (Table 2). Additionally, the unit of analysis or comparison, whether patient, clinic, or hospital level, may be entirely different from the analyses conducted or available for a CPM population based on patients’ country of origin or race/ethnicity. For CPM development, validation through prospective facility-based cluster analysis is ideal. Although generally considered weaker evidence than evidence from RCTs, observational studies of nephrology clinical databases provide insight into the scope of clinical practice variance as well as comparative clinical outcomes among them (5). The caveat comes in the form of maneuvers to detect and address biases potentially influencing the observed outcomes for particular interventions (Table 2).

Table 1. CPM development checklist

National Quality Forum Key Criteria	CPM Metric Checklist Items
Evidence, performance gap, and priority (impact)—importance of measurement and reporting (potential metric screening criterion)	(1) Evidentiary scope and recognizing the difference between patient and population outcomes; (2) performance gap; (3) impact and priority
Reliability and validity—scientific acceptability of measured properties	(1) Statistical reliability; (2) year-over-year reliability; (3) reliable technical precision; (4) facility-level CPM improvements correspond with patient improvements
Feasibility of a measure	(1) Infrastructure feasibility
Usability and use of a measure	(1) A plausible conceptual model must exist for quality improvement; (2) attributed patients; (3) measure transparency
Comparison with related or competing measures	Not covered in the current publication
CPM, clinical performance measure.	

Variable	Potential Bias
Patient selection	Healthier patients perform better in clinical trials; the sickest patients are often excluded from randomized clinical trials; patients new to dialysis (incident versus prevalent patients)
Geographical disparities	Randomized clinical trials may enroll patients from a specific geographical location (<i>e.g.</i> , Japan) or single study site (lacking diversity in race/ethnicity), generating results that may not be extrapolated to the United States patient population; observational studies from clinical settings outside the United States (<i>e.g.</i> , Australia) may not translate to United States clinical settings
Clinical intervention	Improved patient outcomes can arise from uncontrolled for health care provider or facility-level practices rather than a treatment variable; study participants respond positively to personal attention and clinical care
Analytic confounders	Unit of analysis or comparison (<i>e.g.</i> , patient, clinic, or hospital level); person versus person-time calculations; complexity of statistical tools (<i>e.g.</i> , marginal structural models); role of competing events (<i>e.g.</i> , dead patients cannot be hospitalized)

The potential targets for additional measures must be directed at patient adverse events, such as mortality, hospitalizations, preventable readmissions, infections, anemia, and fractures, which are common high-impact events that create tremendous morbidity. Finally, the review of the evidence should end with an appropriate quality of evidence grade (6) (high, moderate, low, or very low) that is stated and transparent (Table 3). It is preferable to have determined the level of evidence required to proceed to the next step in CPM development *a priori*.

Medical education focuses on training physicians to take care of individual patients. Therefore, it is natural that, when clinicians are asked to create facility-level-based quality metrics, there may be a tendency to confuse what is important at a patient versus a facility level. For example, it may be appropriate for a patient to receive a bone biopsy for a given bone and mineral metabolism abnormality but that does not mean that all patients with renal metabolic bone disease should receive bone biopsies.

Additionally, it must be understood that measurement and outcome values may differ between different dialysis patient groups, such as incident and prevalent hemodialysis patients. Therefore, the metric should determine the eligible population and by understanding the distribution of the targeted CPM process or outcome, establish the level of variation or distance from the goal of getting these eligible patients to the

target. When it is practical, exceptional patients should be excluded from the calculation of the measure.

Checklist Item 2: Performance Gap

Ensuring that there is a true performance gap is a critical requirement for an appropriate CPM. For example, it is important to know when a given metric is no longer contributing to quality, which was suggested by VanLare and Conway (2) and Goodrich *et al.* (1). That is, when all dialysis facilities perform at such a high level of quality that the year-to-year variances are more reflective of the influx of a single new patient or some random event other than quality, then the ability to show significant addressable variances in quality is lost. For example, if facility-level mean patient hemoglobin of >12 g/dl for patients receiving erythropoietin is seen in only 1% of facilities, the particular quality metric can no longer be improved. Furthermore, if the delivered amount of dialysis as measured by Kt/V reaches 95%, it is hard to recognize additional improvements given the challenges with new patients and implementation of functional vascular access.

Other than these topped-out metrics, another problematic circumstance for showing a performance gap is when the CPM has not been well characterized and the optimal CPM value or range for patients with ESRD is unknown. This argument was recently made by the American Society

Characteristic	Study
Patient- and population-level evidence	Randomized clinical trials start with a high-quality GRADE but may be graded down after evaluation of study limitations, inconsistency of results with other randomized clinical trials, lack of direct evidence, lack of precision, or reporting bias
Population-level evidence	Observational studies start with a low-quality GRADE, but grading upward may be warranted with large treatment effects or evidence of dose response

GRADE quality of evidence definitions: high quality, additional research is very unlikely to change confidence in the estimate of effect; moderate quality, additional research is likely to have an important impact on confidence in the estimate of effect and may change the estimate; low quality, additional research is very likely to have an important impact on confidence in the estimate of effect and is likely to change the estimate; very low quality, any estimate of effect is very uncertain.

of Nephrology regarding the CMS Quality Improvement Program (QIP) hypercalcemia measure (7).

Checklist Item 3: Impact and Priority

A CPM under consideration must either serve a large population or address a small population that has an unmet need. Thus, a key driver for prioritization of any proposed CPM is to have a large impact on a desirable outcome—either measures of effectiveness, such as mortality and hospitalization (Figure 1), or what matters most: health-related quality of life and experience of care. Ideally, the evidence should support not only patient-based improvements, but also facility-based improvements at a magnitude that justifies implementing a facility-level threshold-based CPM. Some aspects of assessing care may be related to the primary indication for treatment based on the major degree of morbidity or mortality. For example, kidney disease has several organ system complications that have been the target of therapeutic interventions, including bone and mineral disease, anemia, vascular access, fluid overload, and congestive heart failure. These areas may be areas for consideration if there is significant practice variability and variance in

outcomes, such that narrowing the gap would likely improve patients’ lives.

NQF Criteria 2: Scientific Acceptability of Measure Properties

For CPMs that meet the rigor of evidentiary screening, the details of the CPM become the next criteria for scrutiny. This stage of measure development and evaluation deals with the construct of the proposed CPM on whether the measure and related outcome link can be reliably reproduced. The NQF defines reliability as the “extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented” (4).

The fundamental basis of reliability is based on standardized data collection with two components. The measure is well defined and specified, such that its implementation across care delivery organizations is consistent, and it is reproducible within a given population. To evaluate CPM reliability, a deeper understanding of the statistics and nuances of the measure must be developed. This criterion goes to the direct impact of measuring a parameter and the

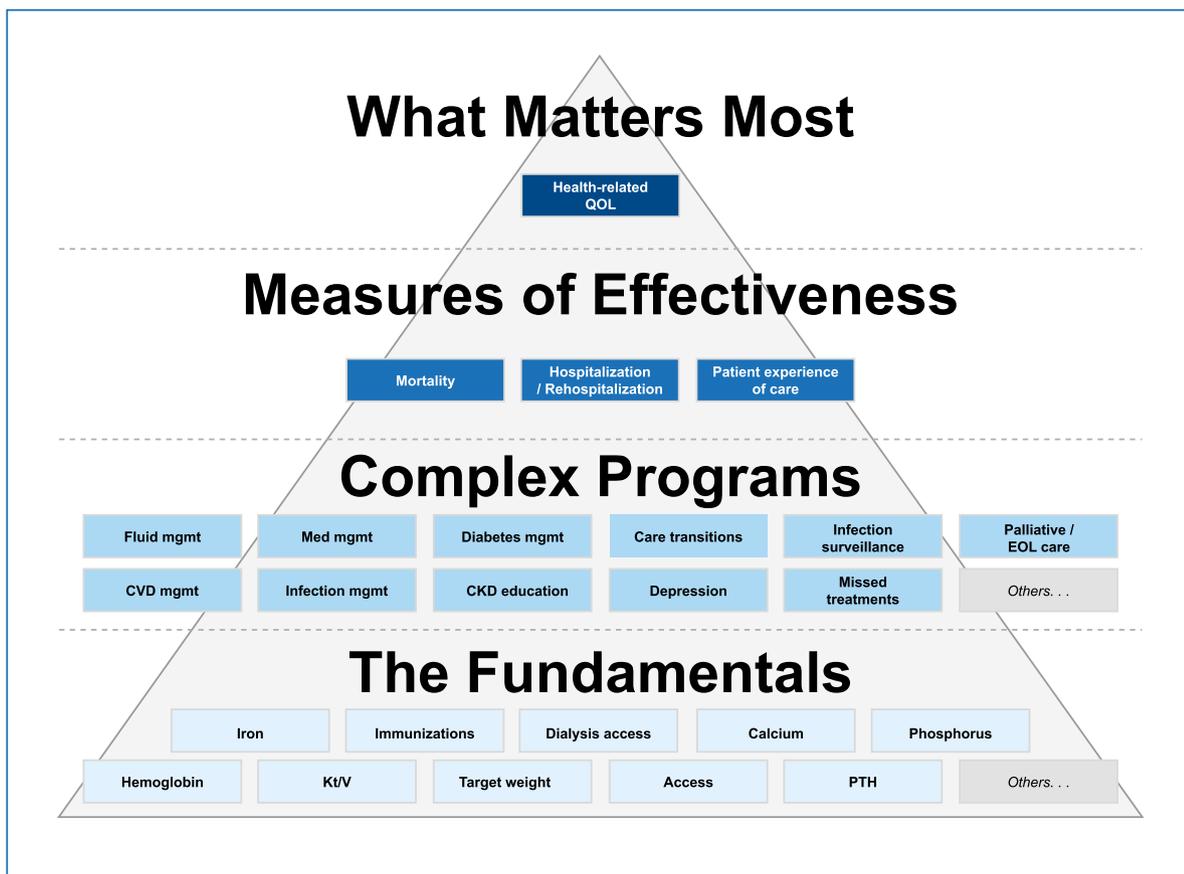


Figure 1. | The fundamentals are the lowest tier of performance measures and might evolve from quality improvement measures to rudimentary quality control measures. The next level consists of more complex programs and requires integrated care paradigms between dialysis facility staff (nephrologists, nutritionists, and social workers) and health care providers external to the clinic. Measures of effectiveness require persistent consideration of the fundamentals and complex programs and leads to what matters most: health-related quality of life. CVD, cardiovascular disease; EOL, end of life; Med, medical; Mgmt, management; PTH, parathyroid hormone; QOL, quality of life. Modified from reference 10, with permission.

variability on a population or unit care level. The challenge is whether a measure is comparable across smaller versus larger providers with good reliability.

Checklist Item 1: Statistical Reliability

It is essential that CPM developers understand that population-based metrics should be tested at the facility level. Furthermore, within a given unit of aggregation, an understanding should be established regarding the frequency distribution curve of outcomes. If the data are normally distributed, the use of means is appropriate across a facility population. If the data are asymmetrically distributed, medians may be a more appropriate metric for representing the central tendency of a population. The distribution and range of values within this distribution will impact determination of quality thresholds and, more importantly, payment impact thresholds, where truly underperforming outliers should be targeted for intervention.

Apart from different clinical processes used during the transition to dialysis, the distributions of potential CPM candidates vary between patients new to dialysis and established dialysis patients, such as differences between these patient groups in rates of mortality, hospitalization, or quality of life.

Checklist Item 2: Year-over-Year Reliability

An important aspect of continuous quality improvement is recognition of the right timeframe for the universe-level baseline to which a given CPM will be compared. The baseline should be as contemporary to the start of the measurement period as possible, and it should have a definition of stability year over year before the measurement period. Intraclass correlation can be used to measure the consistency of a facility's performance over time; intraclass correlation will also provide a good representative hierarchy of facilities' performances based on the metric. The year-over-year SD will provide information about distribution stability based on the range of available values.

Without a standard baseline, it is difficult to determine relative improvement. Facilities should receive credit for not just achieving the predefined threshold but also, significant improvement to achieving the threshold, particularly for facilities that start out at the bottom of the distribution curve for that metric.

Furthermore, the metric of interest at the facility level must have a relatively small natural variation. Related to the issue of the minimum size of a dialysis facility, imprecise measurements made in small numbers of patients are subject to random temporal variation, which may obscure the ability to gauge actual variations in quality. For example, standardized mortality ratios for smaller facilities may be impacted by one or two deaths (*i.e.*, even from a motor vehicle accident), such that 1 year could seem extremely good and the next year could seem extremely bad. Therefore, some measures may not be statistically feasible based on year-to-year variation.

Statistical reliability also comes into play when considering the minimum number of patients treated at a facility for the CPM to be a valid quality measure. There is a minimum patient number required to allow for confidence that facility-level outcomes truly reflect quality and not just simply random variation associated with a very small

sample size. As constructed, the ESRD QIP has a minimum unit size required to be scored: 10 patients. Most active CPMs indicate that there is a 1-patient exemption, where the unit with 10 or fewer eligible patients meets the goal as long as only 1 patient fails to meet the threshold. Unfortunately, for example, when a CPM threshold is set at $\geq 95\%$, a facility must have at least 20 eligible patients to have 1 patient fail and still meet the goal. Thus, facilities with 11–19 eligible patients must be perfect. It may be better to generalize that at least one exception is allowable for each performance measure before a facility is penalized for that measure. After the initial exception, the minimum number of eligible patients required for a facility to be reliably graded for a particular measure should be statistically determined. Candidate CPMs, where the number of eligible patients required is too large to reliably identify poor performers, are unlikely to be clinically meaningful.

As a part of the development process, it may be advisable to monitor a candidate CPM over time to evaluate year-to-year stability of facility rates during a historical period before the measure is developed into an adequately modeled standardized measure. It would be ideal to show a particular baseline cross-sectional distribution that identifies underperforming facilities and indicates that these same facilities have consistently underperformed. This longitudinal monitoring should also be performed with the desired facility-specific outcome of a CPM. The observed year-to-year deviation will provide an estimate of facility-specific common cause variability and hint at the magnitude of change required to indicate when there is a true effect from an intervention.

Checklist Item 3: Reliable Technical Precision

When evaluating CPM reliability, current technology limitations should be considered. Although the best example in the current QIP is based on laboratory measures, with additional study, it may expand to other forms of technology in the future, such as online clearance-based adequacy measurements or novel assessments of fluid status (*e.g.*, bioimpedance analysis or blood volume monitoring) (8). There are two key issues that require evaluation. Is a particular method/technology preferred? (If multiple methods/brands are available, are they comparable?) How precise are the measurements contributed by each specific technology/method? For example, the current QIP specifies the use of serum calcium measurements, but specifications for testing method may not be homogeneous. It is essential for the CMS to consider whether the data are standardized for electrolyte testing in serum versus plasma. Part of the consideration includes assay manufacturer specifications and established current use in hospitals and dialysis facilities. Will establishing a rule that is very specific to serum potentially exclude a viable and comparable testing in plasma? Other considerations include the following questions. What calibrations are being used for a given laboratory? What constitutes the upper and lower limits of normal? However, in the case of the hypercalcemia measure for payment year 2016, it uses a 3-month rolling average as a metric (9), thus potentially accounting for the clinical response time and the laboratory variation characteristic of the assay.

Checklist Item 4: Improvements in Facility-Level CPMs Should Be Relevant to Patients

We suggest that outcome measures, defined as quality of life, experience of care, mortality, and hospitalization, could be based on the high-level patient-focused facility quality hierarchy (10) (Figure 1). This hierarchy provides an example for a process that could be applied by a provider to improve primary outcomes. This same theme of patient-centered outcomes is inherent in the Kidney Disease Improving Global Outcomes guidelines. Specifically, such a construct is used in determining the quality of data to consider and the data's evidence grade (11).

Validation of Metrics Needs to Occur at the Facility Level

Often, patient-level outcomes are extrapolated to facilities. We believe that extrapolation is not valid. Instead, we suggest that one should first examine the association of the facility-level outcomes of the surrogate measure of a primary outcome (usually, mortality; however, it could be other outcomes). If there is directional correlation (the better the metric, the better the primary outcome), then the measure passes the first validity test.

A second test of validity involves showing a $\Delta\Delta$ (12). Unit-level changes in the quality metric should correlate with a hard metric of interest (mortality, hospitalization, or quality of life) at the facility level over time. Ideally, this analysis would be performed prospectively, but, at a minimum, it should be performed retrospectively.

NQF Criteria 3: Feasibility of a Measure

The NQF defines feasibility of a measure as the

extent to which the required data are readily available or could be captured without undue burden and can be implemented for performance measurement. For clinical measures, the required data elements are routinely generated and used during care delivery (*e.g.*, blood pressure, lab test, diagnosis, medication order). The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified. Demonstration that the data collection strategy (*e.g.*, source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (*e.g.*, already in operational use, or testing) demonstrates that it is ready to put into operational use (4).

Of note, availability of data should be tempered with accuracy of data and comparability of data from different sources, such that feasibility does not trump reliability.

Checklist Item 1: Infrastructure Feasibility

The lack of a functional information infrastructure and data exchange methodology with the CMS through the Consolidated Renal Operations in a Web-Enabled Network is a major barrier to harmonizing the quality measure process. Optimally, the complete set of data should be available and the rejection rate should be <1%. The data

should be collected under stable business rules for a period of 1–2 years to determine an appropriate baseline and allow for evaluation of year-over-year stability as discussed.

This standardization of data collection would also facilitate validation of the measure after deployment. Specifically, such standardized data collection allows for cluster-based observational studies or randomized trials that would help to determine the impact of measures that are being pragmatically and prospectively tested. This method is akin to a plan–do–study–act cycle on a larger scale.

NQF Criteria 4: Usability and Use

The NQF defines usability and use as the

extent to which potential audiences (*e.g.*, consumers, purchasers, providers, policymakers) are using or could use performance results for both accountability and performance improvement to achieve the goal of high-quality, efficient healthcare for individuals or populations (4),

and it defines use for improvement as

progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations. The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists) (4).

These criteria can serve to determine how prior evidence indicates that use of the measure will impact the landscape and gauge how changes in the measure affect quality, safety, and efficiency.

Checklist Item 1: A Plausible Conceptual Model Must Exist for Quality Improvement

A conceptual model should exist, where the facility or health care provider is able to impact the metric with quality improvement efforts and continuous quality improvement methods: the sphere of influence should mirror the sphere of responsibility of the health care provider. For example, take a hospital readmission metric. For a hospital that is responsible for every discharge, it could be potentially liable for every readmission. The hospital's influence contributes 100% of the denominator of such a metric. In contrast, a readmission metric for a nursing home or dialysis unit is more complex. Only a subset of hospital discharges attributable to a facility may be directly under the control of the facility. In this situation, a conceptual model should align the subset of hospital discharges that are cause-specific and under the influence of a facility.

For a quality metric to be of value, data must be available in at least quarterly intervals for health care providers to understand trends and implement changes in a continuous

quality improvement process. This finding is true for simple metrics based on electronic medical record data, such as laboratory values. However, for more advanced metrics, the method becomes complicated. Consider transfusions, which are administered almost exclusively in hospitals. Without real-time access to data and with the lack of control over transfusion administration, it is challenging for dialysis clinics to implement processes that can effectively minimize patients' needs for them.

Finally, it is important to track improvement in facility-based proportions of patients meeting the CPM target and ensure that these improvements enhance facility-level outcomes. One example is the two-phase study by Lacson *et al.* (13,14), where relevant potentially actionable patient-based factors were initially identified (13) and then, the facility-level impact of these factors as a CPM was evaluated (14). However, taken together, these reports (13,14) provide a conceptual framework as the basis for future evaluation of candidate CPMs. The discussion must consider dynamics within the facility and how the stakeholders—not just the facility staff but the physicians, patients, and families—may interact to bring about change within the facility using a plausible model of care.

Checklist Item 2: Attributed Patients

For a health care provider (whether a facility or a physician) to be accountable for any given outcome, the business rules that define attribution of a patient population to that provider must be clearly defined. For example, patients who have some minimum number of health care delivery episodes with a given provider can be attributed. For the ESRD QIP, at least 12 months of dialysis claims data from a center are required to place the mantle of accountability for a quality metric on a given provider.

The operational responsibility—the entity accountable for the clinical process and the patient outcome—must be clearly delineated. Under the current structure (where the facility is responsible for its CPMs), it is important to consider the measures, such as administering the prescribed dose of medication, that are truly under the purview of the facility staff.

Checklist Item 3: Measure Transparency

Prospective CPMs should be developed in a transparent manner. The details of a metric, including its validation, should be clearly available to all stakeholders. Transparency would include detailed specification of the measure, expected lifecycle, benchmarking, and directional aspirations for the care represented by the measure.

NQF Criteria 5: Comparison with Related or Competing Measures

The final area of the NQF checklist is comparison with related or existing measures.

If a measure meets the above criteria and there are endorsed or new related measures (either the same measure focus or the same target population) or competing measures (both the same measure focus and the same target population), the measures are compared to address harmonization and/or selection of the best measure (4).

This process is important to avoid duplication or conflict. Unlike the other components of the checklist mentioned above, this last criterion requires a measure-by-measure review, which is best conducted in the context of the TEP convened to discuss a specific measure domain. Thus, it is beyond the scope of our manuscript.

Conclusion

The current process for quality measurement development is slowly making the change from expert opinion (extrapolating patient-level outcomes) to true facility-based management. Given the impact on physician behavior, it is not wise to compromise (when there is no available metric) simply for the sake of measuring something. The risks, benefits, and unintended consequences should be analyzed and acknowledged (with measures developed to a state of maturity, where they are most useful) before embedding them in the payment system or delivery model for care.

The guiding principles outlined here may serve as a checklist to help guide clinicians and CPM developers to make the transition from taking care of individual patients to implementing truly effective facility-based metrics. This checklist is applicable to ESRD but has broader implications for other chronic disease states, where an organized health delivery system would create better quality, ensure safer care, and improve efficiencies. ESRD is in the forefront of health care delivery by translating raw data into, first, useful and actionable metrics and then, improved results as an example for all patients in a value-based health care delivery system.

Acknowledgments

We thank Donna E. Jensen of DaVita Clinical Research for editorial support.

Disclosures

M.K. and S.M.B. are employees of DaVita Clinical Research, a wholly owned subsidiary of DaVita HealthCare Partners Inc. F.W.M. and E.L. are employees of Fresenius Medical Care North America. T.F.P. is an employee of Renal Ventures Management. D.J. is an employee of Dialysis Clinic Inc. A.R.N. is employed by DaVita HealthCare Partners, Inc. A.C. has participated on advisory boards for Amgen, NxStage, Keryx, and ZS Pharma and consulted for Amgen, Keryx, and ZS Pharma.

References

1. Goodrich K, Garcia E, Conway PH: A history of and a vision for CMS quality measurement programs. *Jt Comm J Qual Patient Saf* 38: 465–470, 2012
2. VanLare JM, Conway PH: Value-based purchasing—national programs to move from volume to value. *N Engl J Med* 367: 292–295, 2012
3. Centers for Medicare and Medicaid Services: Centers for Medicare and Medicaid Services End-Stage Renal Disease Quality Incentive Program 2016 Payment System Proposed Rule, 2013. Available at: <http://www.cms.gov/Outreach-and-Education/Outreach/NPC/Downloads/2013-08-14-ProposedRule-NPC.pdf>. Accessed December 9, 2013
4. National Quality Forum: 2013 National Quality Forum Measure Evaluation Criteria, 2013. Available at: http://www.qualityforum.org/docs/measure_evaluation_criteria.aspx. Accessed December 13, 2013
5. Krishnan M, Wilfehrt HM, Lacson E Jr.: In data we trust: The role and utility of dialysis provider databases in the policy process. *Clin J Am Soc Nephrol* 7: 1891–1896, 2012

6. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; GRADE Working Group: GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 336: 924–926, 2008
7. RE: CMS-1352-P: Medicare Program; End-Stage Renal Disease Prospective Payment System, Quality Incentive Program, and Bad Debt Reductions for All Medicare Providers; Proposed Rule, 2013. Available at: http://www.asn-online.org/policy/webdocs/12.8.31_submittedqipcomments.docx.pdf. Accessed September 16, 2013
8. Parker TF 3rd, Hakim R, Nissenson AR, Krishnan M, Bond TC, Chan K, Maddux FW, Glascock R: A quality initiative. Reducing rates of hospitalizations by objectively monitoring volume removal. *Nephrol News Issues* 27: 30–36, 2013
9. Centers for Medicare & Medicaid Services (CMS); HHS: Medicare program; end-stage renal disease prospective payment system, quality incentive program, and bad debt reductions for all Medicare providers. Final rule. *Fed Regist* 77: 67450–67531, 2012
10. Nissenson AR: Improving outcomes for ESRD patients: Shifting the quality paradigm [published online ahead of print November 7, 2013]. *Clin J Am Soc Nephrol*
11. Uhlig K, Macleod A, Craig J, Lau J, Levey AS, Levin A, Moist L, Steinberg E, Walker R, Wanner C, Lameire N, Eknoyan G: Grading evidence and recommendations for clinical practice guidelines in nephrology. A position statement from Kidney Disease: Improving Global Outcomes (KDIGO). *Kidney Int* 70: 2058–2065, 2006
12. Wolfe RA, Hulbert-Shearon TE, Ashby VB, Mahadevan S, Port FK: Improvements in dialysis patient mortality are associated with improvements in urea reduction ratio and hematocrit, 1999 to 2002. *Am J Kidney Dis* 45: 127–135, 2005
13. Lacson E Jr., Wang W, Hakim RM, Teng M, Lazarus JM: Associates of mortality and hospitalization in hemodialysis: Potentially actionable laboratory variables and vascular access. *Am J Kidney Dis* 53: 79–90, 2009
14. Lacson E Jr., Wang W, Lazarus JM, Hakim RM: Hemodialysis facility-based quality-of-care indicators and facility-specific patient outcomes. *Am J Kidney Dis* 54: 490–497, 2009

Published online ahead of print. Publication date available at www.cjasn.org.