Performance Measurement in Chronic Kidney Disease

Kimberly A. Smith*† and Rodney A. Hayward*‡

Robert Wood Johnson Foundation Clinical Scholars Program, Department of Internal Medicine, Divisions of *General Internal Medicine and †Nephrology, University of Michigan, Ann Arbor, Michigan, and ‡Department of Veterans Affairs, VA Health Services Research & Development Center of Excellence, VA Ann Arbor Healthcare System, Ann Arbor, Michigan

There is a growing miasma over the real value of our health care dollar in the United States.1–3 Value is the ratio of quality to cost; in most economic sectors, lower quality products or services require a lower price to avoid extinction by market forces. The strong perception that suboptimal quality is widespread despite escalating health care costs has encouraged various national programs to measure the quality and expense of hospital care and, increasingly, outpatient services. Quality measures are well established for inpatients with myocardial infarction4–6 and outpatients with diabetes,7,8 and attention is now turning to chronic kidney disease (CKD).9 Blue Cross Blue Shield (BCBS) will begin a payment incentive program for CKD based on performance measures in the state of Michigan this year,10 the Medicare Quality Improvement Organizations are piloting three CKD measures,11 and the Joint Commission awarded its first CKD certification to the University of California San Diego Medical Center in 2010 based, in part, on performance measurement.12 Despite these efforts, performance measurement for CKD is still in its infancy. Here, we discuss the challenges of doing this well.

END-STAGE RENAL DISEASE INITIATIVES

Performance measurement and quality improvement are not new to nephrology. The Centers for Medicare & Medicaid Services (CMS) initiated their first quality improvement effort for a population with chronic illness in the 1990s: they charged the 18 regional end-stage renal disease (ESRD) networks to monitor and improve the quality of dialysis services. The publication of national guidelines provided a basis for the development of measures, comparative feedback on performance, and initiatives to improve quality.13,14 Moving forward, ESRD remains at the forefront of CMS’s quality efforts with the proposed rollout of the Quality Incentive Program in January 2012, which will be the first pay-for-performance program in a Medicare fee-for-service environment.15,16 In many respects, these early efforts successfully increased quality and accountability within the ESRD Program,17 but dialysis outcomes remain poor despite continued high costs, making a poor showing for value.18,19,13,20 As a result, payers are now shifting their gaze to CKD in the hopes of limiting progression to ESRD and bending the curve on long-term costs.21–24,14 In an effort to jump-start improvements in the treatment of patients with CKD, payers are proposing quality initiatives that are based on performance measures.10,11 Various stakeholders are now delineating details for the management of this population.

CURRENT CKD MEASURES

Relevant measures developed by national organizations rarely assess CKD directly. For example, the RAND Cor-
Table 1. Comparison of practice guidelines and performance measures

<table>
<thead>
<tr>
<th>Practice Guidelines</th>
<th>Performance Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Definition</strong></td>
<td>Tools to assess compliance with standards of clinical care</td>
</tr>
<tr>
<td>Recommendations for clinical care based on best available evidence and/or expert opinion</td>
<td></td>
</tr>
<tr>
<td><strong>Purpose</strong></td>
<td>Measure, report, and compare the quality of care</td>
</tr>
<tr>
<td>Objectively summarize complex and expanding medical literature for use by clinicians</td>
<td></td>
</tr>
<tr>
<td><strong>Focus</strong></td>
<td>Most essential recommendations backed by highest quality evidence and a commitment to improved performance</td>
</tr>
<tr>
<td>Comprehensive coverage of diverse aspects of care for a condition</td>
<td></td>
</tr>
<tr>
<td><strong>Specificity</strong></td>
<td>Detailed, as predetermined technical specifications are needed for objective and fair comparisons between groups</td>
</tr>
<tr>
<td>General, as physician is ultimately responsible for applying recommendations appropriately to specific patient circumstance</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Renal Physicians Association/Physician Consortium for Performance Improvement Chronic Kidney Disease Physician Performance Measurement Set (October 2007)

| Patients with hypertension and proteinuria who were prescribed ACE inhibitor or ARB therapy |
| Had the following laboratory testing ordered at least once: serum levels of calcium, phosphorus and intact PTH, and lipid profile |
| Received the influenza immunization during the flu season (September through February) |
| Referred for AV fistula at least once |
| Percentage of calendar months in which patients receiving ESA therapy have a hemoglobin <13 g/dl OR patients whose hemoglobin is ≥13 g/dl and have a documented plan of care |
| Blood pressure <130/80 mmHg OR blood pressure ≥130/80 mmHg with a documented plan of care |

All measures report the percentage of patients aged 18 years and older with a diagnosis of advanced CKD (stage 4 or 5, not receiving renal replacement therapy) who met the measure during the previous 12-month reporting period.
TYPES OF MEASURES

In the broadest sense, measurement can be implicit or explicit. Implicit measurement refers to expert practitioners reviewing the details of cases and using their clinical expertise and judgment to assess overall quality of care. Explicit measurement refers to the application of prespecified and clearly defined metrics to patient circumstance. Although implicit measurement recognizes the nuances of individual cases and can be a powerful method for distinguishing the quality of different providers, it is time-intensive and each review has a subjective component. For these reasons, most large-scale reporting systems use explicit measures, which are more limited in their ability to account for specific details but are more politically tenable because of the perception of objectivity, although it is frequently overlooked that criteria for explicit review are often derived subjectively. Explicit measures can assess aspects of quality ranging from processes of care to ultimate patient outcomes, each carrying advantages and disadvantages (Table 3).

In defining explicit measures, it is intuitively appealing to measure outcomes of ultimate clinical importance, such as death or progression to ESRD. Unfortunately here there are two major limitations that are difficult to overcome. The first is sample size; it requires a much larger sample size to detect differences in outcomes than in processes. Consider two clinics in which the prevalence of statin prescriptions is 40% versus 80%. One year of data on 25 patients per clinic will detect this large process difference with 80% power of discrimination. In contrast, given the same prescription practices, detecting a difference in cardiovascular events would require >30,000 patients per clinic; detecting a difference in mortality would require >1 million patients per clinic (Table 4). Process measures produce the clearest signals relative to the noise generated by patient variation, thereby requiring far fewer patients to detect meaningful differences.

The second major limitation is that risk for adverse outcomes often depends more on disease severity and health behaviors, which varies widely between patients, than on the quality of medical care. Whereas medical therapy may delay ESRD or death, the GFR at the time the clinician becomes involved is more influential. Also, patient behaviors like medication adherence, smoking, and substance use substantially impact outcomes. A measurement system that does not adequately account for these factors will, at best, systematically punish those who treat sicker patients and, at worst, lead physicians to avoid these patients altogether. Although case-mix adjustment attempts to account for disease severity—methods similar to All Patient Refined DRGs (APR-DRGs) used in hospital-based care—this relies on the identification and accurate measurement of important variables that often fall short. For CKD specifically, calibration of creatinine measurement and disputes about the prognostic value of the staging system further confound adjustment for disease severity. For all these reasons, none of the currently recommended CKD measures directly assesses ultimate clinical outcomes.

Instead, the focus of current measures is on intermediate outcomes and what are called processes of care. Two of the RPA/PCPI measures focus on intermediate outcomes: BP <130/80 and, among patients receiving erythropoietin-stimulating agents (ESA), hemoglobins <13 g/dl. The remaining measures assess processes of care such as ordering laboratory studies or prescribing inhibitors of the renin-angiotensin system. These types of measures are easier to associate with the quality of medical care provided by physicians or physician groups, facilitating assignment of accountability, interpretation of results, and identification of necessary actions for improvement.

CHALLENGES WITH EXISTING MEASURES

Despite theoretical advantages of intermediate outcome or process measures, pitfalls remain. These measures do not individually provide a global assessment of quality as a result of their narrow definitions and their typical focus on a specific aspect of management for a single disease. In addition, the targets for these measures, which are often expressed as summary statistics such as proportion of patients for whom a measure is met, can be controversial. Furthermore, process measures do not necessarily capture the distinction between performing a process and performing it well, such as the skill of a proceduralist. Thus, a simple process measure sometimes will not reflect quality adequately.

Intermediate outcome measures often require much smaller sample size than ultimate outcomes but do share the other major limitation: if disease severity is not adequately adjusted, the measurement may not reflect true differences in quality. Consider the problems that accompany the apparently straightforward intermediate outcome of achieving a BP <130/80. First, the majority of CKD in the United States results from poorly controlled hypertension and/or diabetes. Although undoubtedly some of these patients suffer from poor-quality health care, many simply have more severe disease or sought medical attention late in its course. With the development of CKD, BP control becomes only more challenging. Many patients will never achieve tight control because of severe disease, contraindications to medications, access to care, or nonadherence. With current measures, the easiest path to improved performance would be to drop patients with difficult-to-control hypertension.

As suggested by Kerr et al., tightly linked measures may overcome these limitations by tying a specific process, such as intensifying treatment to an indication for an intermediate outcome that is not yet at target. With these measures, providers get credit for responding appropriately to clinical situations, which should decrease the incentive to drop patients who are less likely to reach targets. As an example, the RPA/PCPI BP measure grants credit for either achieving the goal BP or appropriately responding to an elevated BP (BP <130/80 mmHg or BP ≥130/80 mmHg with a documented plan of care).
Table 3. Comparison of performance measures types

<table>
<thead>
<tr>
<th>Type</th>
<th>Definition</th>
<th>Examples</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>Specific action of provider</td>
<td>Assessing proteinuria</td>
<td>Directly assesses provider action</td>
<td>Value dependent on link to important outcomes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Prescribing ACE therapy</td>
<td>Occurs in short time period</td>
<td>Focus is narrow</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Sensitive to quality differences even with small sample sizes</td>
<td>“Right” performance level may be unknown</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Difficult to assess how well process is performed</td>
</tr>
<tr>
<td>Intermediate</td>
<td>Outcome not of direct importance but linked to</td>
<td>Blood pressure $&lt; 130/80$</td>
<td>Assesses effectiveness of interventions</td>
<td>Value dependent on link to important outcomes</td>
</tr>
<tr>
<td>outcome</td>
<td>ultimate clinical outcomes</td>
<td>Hemoglobin $&lt; 13$ g/dl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&quot;Tightly linked&quot;</td>
<td>Achievement of an intermediate outcome</td>
<td>Blood pressure $&lt; 130/80$ OR</td>
<td>Combines advantages of process and intermediate outcome measures</td>
<td>Can reflect factors outside of a provider’s control</td>
</tr>
<tr>
<td></td>
<td>Or appropriate response to abnormal value</td>
<td>appropriate response to a higher value</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcome</td>
<td>Health status of patients</td>
<td>Progression to ESRD</td>
<td>Reflects the ultimate goal of the health care system</td>
<td>Long time line for many conditions</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Death</td>
<td>Provides global assessment of quality</td>
<td>Influenced by many providers, types of care, and factors other than health</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Meaningful to consumers</td>
<td>care</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Requires large samples sizes to detect meaningful differences</td>
</tr>
</tbody>
</table>
course, collecting data to identify a plan of care is complex and, without these data, the measure reduces to the less informative, dichotomous intermediate outcome. The RPA/PCPI measure of anemia management makes a similar attempt with the same limitations: percentage of calendar months in which patients receiving ESA therapy have a hemoglobin <13 g/dl or patients whose hemoglobin is ≥13 g/dl have a documented plan of care.

A lack of evidence linking a process or intermediate outcome measure to clinically important outcomes also undermines its validity. Simply ordering a test or writing a prescription does not necessarily improve health. Health outcomes will not improve if no one reviews or appropriately responds to a test result or if the prescribed therapy is not taken or is ineffective. To ensure the focus remains on maximizing health benefits for patients, there must be strong evidence linking measures to clinically meaningful outcomes. For example, one RPA/PCPI measure assesses the prescription of renin-angiotensin system inhibitors among patients with hypertension and proteinuria; randomized controlled trials in this population show delayed progression to ESRD.49 Subsequent evidence60–62 led to a 2007 update that highlighted the need to individualize treatment goals but continued to recommend specific hemoglobin targets.73 An early performance measure encouraged these targets,69 fortunately, during the RPA/PCPI review process, this measure was eliminated and replaced with one that supported a less controversial upper hemoglobin limit (Table 2).74 More recent evidence further validates early concerns about the potential hazards of ESA therapy in CKD.75–77 This history warns against the adoption of performance measures without high-quality evidence that achieving the metric will improve clinical outcomes.

Evidence for the screening and treatment of mineral bone disease in CKD is limited similarly. The authors of the 2009 KDIGO CKD-mineral bone disease guideline highlight that the guideline encompasses many aspects of care for which there is little or no evidence to inform recommendations.33 In fact, out of 49 recommendations, only 2 are supported by evidence from well-designed trials with patient-centered outcomes as end points. The authors contend there are few opportunities for developing clinical performance measures from the mineral bone disease guidelines.

Consider the performance measure that assesses whether serum levels of calcium, phosphorus, and intact PTH are measured at least annually (Table 2), a measure derived from a 2003 K/DOQI guideline. The parallel 2009 KDIGO guideline rates the evidence as 1C, designating a strong recommendation but low-quality evidence, cautioning that the true effect may be substantially different from the estimate of the effect.79,33 Despite evidence in ESRD that abnormalities in these labs reflect abnormal bone turnover and associate with cardiovascular events and mortality,80 there is no evidence that changing therapy based on the results improves outcomes.

Data are even more limited for CKD.81 Even if it is assumed these mineral tests should be ordered routinely, should they be checked annually or at some other interval? Notably, both primary care physicians and nephrologists perform poorly on these performance measures.82–84 Does this represent ignorance of the guidelines, disagreement with the guidelines, or individualization of treatment goals? Even more important, high-stakes performance measures drive performance toward achieving what is measured, and this may result in adverse consequences. In the above example, consider the burden to patients and the costs of screening for and treating mineral and bone disorders without strong evidence for improved outcomes.

Beyond this, what if a measure inadvertently encourages potentially harmful therapy? For example, again consider the measure of whether a CKD patient achieves a BP <130/80 mmHg. As Lewis summarized recently, this target derives primarily from the Modification of Diet in Renal Disease study, which randomly assigned patients with CKD to a mean arterial pressure (MAP) of ≤92 or ≤107.85,86 Randomization to a MAP allows for a wide range of SBP and DBP combinations; although the mean BP in the low MAP group was 126/77, the SBP ranged from 98 to 154. These data, along with observational studies, define the goal BP of <130/80, ignoring the actual range of SBP achieved in the study. Attempting to achieve SBP <130 runs the risk of increasing the diastolic hypotension, which associates with increased risk for cardiovascular events in the general population—the exact opposite of the intended goal.87,88 Although whether this association is causal remains controversial,85 it is appropriate to drive all patients toward this SBP target without

---

Table 4. Sample size required for 80% power to detect a quality difference between two clinics related to statin prescription (true prescription prevalence: 40 versus 80%)

<table>
<thead>
<tr>
<th>Type of Measure</th>
<th>Measure</th>
<th>Sample Size Required (No. Patients per Clinic per Year)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>Prescription of statins</td>
<td>25</td>
</tr>
<tr>
<td>Intermediate outcome</td>
<td>Cardiovascular events</td>
<td>30,000</td>
</tr>
<tr>
<td>Outcome</td>
<td>Mortality</td>
<td>1,000,000</td>
</tr>
</tbody>
</table>
Although many aspects of CKD management require further study before performance measurement is appropriate, measurement is more tenable in the areas with both high-quality evidence and general consensus about the value of treatment. For example, there is broad support for the use of renin-angiotensin system inhibitors to delay progression of CKD. The RPA/PCPI measure of prescription of these medications for patients with hypertension and proteinuria reflects this critical aspect of CKD care. As data become available, this measure should expand to include other important CKD populations. There is also broad agreement that arteriovenous fistulas in late stage CKD are preferable for hemodialysis access in most situations. One RPA/PCPI measure assesses referral for arteriovenous fistula placement for appropriate patients with CKD. Future measures could address other elements of the complex system that culminates in a functional vascular access for hemodialysis. So, although performance measurement for CKD remains in an early stage of development, the past decade has witnessed substantial progress (Figure 1).

**PATIENTS, PROVIDERS, AND DATA**

Once appropriate measures are identified, the problem remains that specific subgroups of patients may receive most if not all of the benefit from an intervention. For example, some patients are at very high risk of ESRD or cardiovascular death, whereas for others, CKD may be an incidental diagnosis of little consequence long term. Ideally, performance measures and quality improvement initiatives would target the highest risk patients for whom aggressive management is worth the risk and cost. Unfortunately, there is not yet a reliable tool to stratify risk of CKD progression. In a blunt effort to tailor performance measurement, HEDIS excludes patients older than 75 years. Because a substantial number of CKD patients are over this age, with a growing number progressing to ESRD, this may be inappropriate for CKD performance measurement, especially when an elderly patient is high functioning and otherwise healthy. In contrast, RPA/PCPI measures target patients with an eGFR <30 ml/min per 1.73 m²; although this certainly identifies a high-risk population, it eliminates many relevant patients such as those with substantial proteinuria, rapid decline in GFR, or a GFR between 30 and 45 ml/min per 1.73 m². Until validated risk-stratification models are available for CKD, the identification of high-risk patients may need to rely on a combination of eGFR, rate of progression, and/or magnitude of proteinuria.

Increasingly, primary care providers and nephrologists are working together to manage CKD patients, but the roles and responsibilities of each also remain unclear. Primary care providers will almost certainly become more involved in the management of CKD in the future, which demands consideration of whether the current guidelines are appropriate for the primary care setting. Primary care providers are often unaware of CKD guidelines, written by nephrology stakeholders, and do not follow recommended practices. This is not surprising given the guideline length, complexity, and focus on conditions that primary care providers are unlikely to manage such as acidosis, bone disease, or anemia of CKD. Whereas nephrologists often dedicate their attention to CKD alone, primary care providers must balance several concurrent illnesses and preventive interventions. Yarnall et al. estimated that it would require approximately 8 h/d to provide recommended preventive care to a typical patient panel, leaving no time for acute or chronic disease management. The current CKD guidelines do not adequately summarize or prioritize CKD interventions but rather dilute the most critical messages in a sea of recommendations mainly relevant to nephrologists.

Even after defining appropriate measures, defining the patient population, and assigning accountability, the challenge of data collection remains for the majority of clinical practices that still lack an adequate informatics infrastructure. Ultimately, the hope is that measures reflect real differences in quality, but oversimplification of measures to ease data collection and interpretation may discard the details necessary to identify true distinctions. All too often, logistical challenges—rather than clinical importance or validity—dictate the choice.

![Figure 1: Time line of important events in development of performance measures for CKD.](image-url)
of measures. As described previously, it is far easier to simplify the BP or anemia measures by eliminating credit for plans of care when intermediate outcome targets are unmet. The national push toward universal electronic medical records may facilitate integrated measures that better represent true quality. Given the priority for quality monitoring and improvement, developers should proactively design systems that can capture necessary data elements as they are needed.

**CONCLUSION**

The value of health care in the United States is increasingly the focus of modern interest. Attention must shift from whether to measure value toward how to measure it in an efficient and meaningful way. Although some providers view performance measurement as an unwarranted intrusion into the sanctity of the physician-patient relationship, these efforts are here to stay and continue to gain momentum. As a group, nephrologists must actively engage to ensure that measures optimize benefit to patients with CKD.

Performance measurement is an evolving discipline. To advance the field and to improve the care of patients, measures must encourage quality improvement rather than incentivize providers to drop challenging patients or risk harm by having to enact unproven treatments. Unfortunately, high-quality evidence for many interventions in nephrology is scarce, providing a less than adequate foundation on which to build performance metrics. Given the potential for unintended consequences, it is important to start with a small set of well-designed measures based on good evidence rather than a large number of measures that are not—measures that may distract or impede our focus on making the quality movement work for nephrology.

Last, it is necessary to regularly assess whether this process is accomplishing its goal—to improve value for patients with CKD. Other national initiatives have produced mixed results, although measures, methods, and quality improvement strategies continue to evolve. As experts in the care of patients with CKD, nephrologists have an obligation to proactively shape the future of performance measurement and quality improvement.

**ACKNOWLEDGMENTS**

The authors thank James P. Smith for helpful comments on versions of this work.

Both authors are supported by the Robert Wood Johnson Foundation Clinical Scholars Program. R.H. is also supported by the Department of Veteran Affairs Cooperative Studies Program (CSP #465 FS) and the Measurement Core of the Michigan Diabetes Research & Training Center (NIDDK of the National Institutes of Health [P60 DK-20572]).

**DISCLOSURES**

None.

**REFERENCES**


20. Rocco MV, Frankenfield DL, Hopson SD,


