



August 31, 2016

National Quality Forum  
 1030 Fifteenth Street, NW, Ste 800  
 Washington, DC 20005

RE: NQF Renal Project

Kidney Care Partners (KCP) appreciates the opportunity to comment on the measures under consideration for endorsement in the National Quality Forum’s (NQF) Renal Measures 2015-2017 Project. KCP is a coalition of members of the kidney care community that includes the full spectrum of stakeholders related to dialysis care – patient advocates, health care professionals, dialysis providers, researchers, and manufacturers and suppliers – organized to advance policies that improve the quality of care for individuals with both chronic kidney disease and end stage renal disease.

We commend the NQF Renal Standing Committee for its thoughtful deliberations. KCP supports and recognizes the importance and value of NQF's endorsement process to ensure the importance, reliability and validity of measures. Implementing parsimonious sets of measures that matter is becoming of increasing importance, making NQF's process more critical. We offer comment on all six measures.

**NQF 0260: Assessment of Health-Related Quality of Life (QoL) in Dialysis Patients (Witten and Associates, LLC)**

KCP supports the Committee’s recommendation against endorsement.

**NQF 0369: Dialysis Facility Risk-Adjusted Standardized Mortality Ratio (SMR; CMS)**

KCP supports the Committee’s recommendation against endorsement.

**NQF 1463: Standardized Hospitalization Ratio for Admissions (SHR; CMS)**

KCP believes hospitalization is an important outcome to measure, but has concerns the specifications, reliability, validity (risk model), and harmonization issues. We strongly encourage the Committee to reconsider the reliability testing data, which demonstrate significant reliability issues with the one-year SHR for small facilities, and comment specifically on the SHR’s reliability for such facilities.

**IUR for One-year SHR, Overall and by Facility Size, 2010-2013**

Facility Size (Number of patients)	2010		2011		2012		2013	
	IUR	N	IUR	N	IUR	N	IUR	N
All	0.72	5407	0.71	5583	0.70	5709	0.70	5864
Small (<=50)	0.54	1864	0.51	1921	0.48	1977	0.46	2028
Medium (51-87)	0.65	1702	0.63	1785	0.58	1825	0.57	1930
Large (>=88)	0.81	1841	0.81	1877	0.81	1907	0.82	1906

Although the overall reliability statistic for 2013 (and previous years) is 0.7, a level generally considered the minimum by NQF, the reliability statistics for medium and small facilities fall significantly short of the 0.7 threshold. CMS's own data indicate that for facilities  $\leq 50$  patients, more than half a facility's score (54%) is due to random noise and not a signal of quality. Even for medium facilities, the IUR is significantly below the 0.7 threshold, with 43% of a facility's score attributable to random noise and not signal. We note that the intended use for the SHR will be for public reporting and the penalty-based QIP; penalizing facilities for performance due to random chance is not appropriate. Given the poor reliability testing results, KCP did not support CMS's proposal to include it in the Quality Incentive Program (QIP) for Payment Year 2020.

#### **NQF 2977: Hemodialysis Vascular Access: Standardized Fistula Rate**

KCP recommends the developer consider modifications to improve the measure going forward.

- With respect to the specifications, the language in the previously endorsed AVF measure (#0257) specifically defines an autogenous AVF as using two needles has been replaced with an autogenous AVF "as the sole means of vascular access." KCP believes the specifications for #2977 are imprecise as to whether facilities would receive credit for patients using an AVF as the sole means of access, but who also have in place a graft or catheter that is no longer being used. We note patients with catheters remain at risk for infection and other adverse sequelae, so credit should not be given when a catheter is present, even if an AVF is being used. A numerator that specifies the patient must be on maintenance hemodialysis "using an AVF *with two needles and without a dialysis catheter present*" would remove ambiguity. In contrast, removal of an AV graft is complex and not without risk of complications, so KCP believes credit should be received for a patient who is using an AVF as the sole means of access, but who also may have a non-functioning AV graft present.
- KCP believes this measure improves on #0257, but we continue to believe two additional vasculature risk variables would strengthen the model: a history of multiple prior accesses and the presence of a cardiac device. We also note that the validity testing yielded an overall c-statistic of 0.71. We are concerned the model will not adequately discriminate performance – particularly that smaller units might look worse than reality. We believe a minimum c-statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

#### **NQF 2978: Hemodialysis Vascular Access: Long-Term Catheter Rate**

KCP supports the Committee's recommendation for endorsement.

#### **NQF 2979: Standardized Transfusion Ratio for Dialysis Facilities (STrR; CMS)**

During the last project, this Standing Committee reviewed the STrR as #2699 and did not recommend it. As we discuss further in the section on Validity, we do not believe the new measure addressed the Committee's concerns about hospital- and physician-related factors. Overall, we remain concerned about the reliability, as well as the specifications and validity. We strongly encourage the Committee to reconsider the reliability testing data, which document reliability issues with the STrR for small facilities, and comment specifically on the STrR's reliability for such facilities.

- **RELIABILITY.** KCP has significant concerns about the results from the reliability testing for the STrR. KCP notes a reliability statistic of 0.70 is often considered as "good"

reliability, though the characterization also depends on the analytic method. Reliability testing, overall, for the STRR yielded IURs of 0.60-0.66 across all facilities for each of 2011, 2012, 2013, and 2014. Such values indicate about 65% of the variation in a score can be attributed to between-facility differences (signal) and about 35% to within-facility differences (noise) – a moderate degree of reliability. However, when looking exclusively at small (defined as <=46) and medium (47-78) facilities, the IURs are substantially lower. Specifically, the IURs ranged from 0.30-0.41 and 0.50-0.56 for small and medium facilities, respectively, over the same time period. In other words, approximately 60-70% of a small facility’s score is due to random noise. KCP believes the specifications must specifically require a minimum sample as identified through the developer’s empirical testing.

- **SPECIFICATIONS.** CMS has revised the measure specifications to more “conservatively” define transfusion events, such that all inpatient transfusion events must include, at a minimum, an appropriate ICD-9 Procedure Code or Value Code to be captured in the measure – inpatient transfusion events for claims that include only 038 or 039 revenue codes without an accompanying procedure or value code are not captured in the numerator. The specifications also specify a maximum of one event per day and that an event not be defined by the number of units of blood transfused.

KCP supports and appreciates the need to refine and tighten how transfusion events are counted and applauds CMS’s intent in undertaking these revisions, but we do not believe the proposed solution is a valid representation of transfusion events.

Importantly, there is no existing coding requirement that procedure or value codes be used, which means valid transfusion claims that include only revenue codes will be missed. KCP believes the proposed specification changes result in a measure with significant threats to validity.

Current transfusion coding practices clearly vary by hospital,<sup>1</sup> and hospital coding practices are beyond dialysis facilities’ sphere of control. For example, we are aware of hospitals that exclusively use revenue codes and do not use the procedure or value codes. In-patients at this type of hospital will appear to have no transfusion events assigned to the dialysis facility, whereas those at a hospital that uses the procedure and/or value codes will have recorded events. Simply put, facilities within given catchment areas will be differentially affected by hospital coding variations, which clearly impact measure scoring. We are particularly concerned that the revisions, if implemented, will result in increased variability in performance across dialysis facilities wholly due to external factors and not performance. Facilities will appear to have “poor” performance because of higher than expected numbers of transfusions – and will expend time and resources to improve – when in fact the score is merely a reflection of coding practices.

Again, KCP strongly supports the need to refine how transfusion events are defined, and we urge the Standing Committee to recommend the developer continue considering alternative models to define transfusion events. Alternatively, the Committee could suggest that CMS consider revising hospital transfusion coding rules to require that the ICD-9/ICD-10 procedure and value codes necessary for the validity of the proposed methodology be universally included in claims.

---

<sup>1</sup> Weinhandl ED, Gilbertson DT, Collins AJ. Dialysis facility-level transfusion rates can be unreliable due to variability in hospital-level billing patterns for blood. Chronic Disease Research Group poster, ASN. 2014.

Additionally, the testing documentation notes that facilities with 10 or fewer patients were excluded, but we note the specifications do not state this. Again, KCP believes that a minimum size exclusion should be indicated and, as the developer's results document, and we discuss in the following section, reliability is poor even when the facility size is significantly greater than 10 patients.

The submission also indicates the minimum data requirement for the STrR is 10 patient-years at risk, which differs from the SHR, which uses 5 patient-years at risk. No justification or empirical analyses are offered to justify the selected threshold or the difference.

Finally, the STrR specifications indicate the measure can be expressed as a rate, but is calculated as a ratio. KCP prefers normalized rates or year-over-year improvement in rates instead of a standardized ratio. We believe comprehension, transparency, and utility to all stakeholders is superior with a scientifically valid *rate* methodology.

- **VALIDITY.** In addition to KCP's concerns about the specifications and the threat to validity of variable capture of transfusion events depending on hospital coding practices, KCP has several concerns about the co-variates (or lack thereof) and risk model.

NQF did not endorse the STrR in 2015, in part because this Standing Committee raised concern that the measure did not adjust for hospital- and physician-related transfusion practices. Physicians independently, or following hospital protocols, make decisions about whether or not to transfuse a specific patient, so it is important to account for the variability these factors create. The revised measure does not incorporate these factors into the risk model, so KCP's concurrence with the Committee's original concern remains.

KCP notes that while the SMR and SHR have been revised to incorporate prevalent co-morbidities into their risk models, the STrR has not been so revised; only incident co-morbidities, derived from the Medical Evidence Form (CMS 2728), are considered. This approach means the STrR risk model only reflects those conditions present upon when the patient initiates dialysis; failure to appropriately account for prevalent co-morbidities is a threat to validity. In the harmonization section, we also note that CMS adjusts for 2728-derived co-morbidities for SHR and SMR differently than it does for the STrR. Finally, as we have noted before, we continue to be concerned about the validity of the 2728 as a data source and urge that the Committee recommend that CMS assess this matter.

KCP notes that the validity testing yielded an overall c-statistic of 0.65. We are concerned the model will not adequately discriminate performance – particularly that smaller units might look worse than reality. We believe a minimum c-statistic of 0.8 is a more appropriate indicator of the model's goodness of fit and validity to represent meaningful differences among facilities and encourage continuous improvement of the model.

- **HARMONIZATION ISSUES.** The new SMR and SHR risk models adjust for each incident co-morbidity (from the 2728) separately, instead of using a "co-morbidity index." The model also approaches diabetes as a single co-morbidity rather than four separate indicators (currently on insulin, on oral medications, without medications, diabetic retinopathy). The STrR has not been similarly revised. KCP believes the Standing Committee should recommend that the developer harmonize the STrR with the other measures so that each incident co-morbidity is examined separately (i.e., unbundled, as compared to the current measure) and diabetes is approached as a single co-morbidity

(i.e., bundled, as compared to the current risk model).

The risk models for the groupings used for patient age and duration of ESRD differ among the SMR, SHR, and STrR. For example, the age groups for the SMR is n=3, but for the SHR and STrR the age groupings are the same, but n=6. Similarly, the number of groups for ESRD duration for the SMR (n=4) differs from that for the SHR (n=6). No justification or empirical analyses are offered to justify these differences.

There also are significant inconsistencies in how facility size is defined when assessing reliability for the SMR, SHR, and STrR. Specifically, for the SMR, the definitions were  $\leq 45$ , 46-85,  $\geq 86$  for the 1-year reliability analyses, but were  $\leq 135$ , 136-305, and  $\geq 306$  for the 4-year analyses. For the SHR,  $\leq 50$ , 51-87, and  $\geq 88$  were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as  $\leq 46$ , 47-78, and  $\geq 79$ , respectively. We understand reliability for a given measure depends on sample size, but find the varying demarcations analytically troubling. We posit a more appropriate analytic approach would be to analyze reliability using consistent "bins" of size (i.e., small, medium, and large are consistently defined) and identify the facility size at which reliability for that particular measure can be confidently inferred – and then reflect the minimum size in the actual specifications.

KCP again thanks you for the opportunity to comment on this important work. If you have any questions, please do not hesitate to contact Lisa McGonigal, MD, MPH (lmcgon@msn.com or 203.298.0567).

Sincerely,

/s/  
Sara Love Rawlings, JD  
Executive Director