KIDNEY CARE
PARTNERS

August 21, 2020

National Quality Forum
1099 14th Street NW
Suite 500
Washington DC 20005

RE: NQF Renal Project Spring 2020 Cycle

Kidney Care Partners (KCP) appreciates the opportunity to comment on the measures under consideration for endorsement in the National Quality Forum's (NQF) Spring 2020 Renal 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 NQF for undertaking this important work and offer comment on all three measures under review.

NQF 0369: Dialysis Facility Risk-Adjusted Standardized Mortality Ratio (SMR; CMS)
KCP believes mortality is an important outcome to measure, but has concerns about the SMR's reliability, validity (risk model), specifications, and has identified several harmonization issues with CMS's other standardized measures. In addition, the increasing numbers of Medicare Advantage (MA) patients in the ESRD quality programs—and the unavailability of outpatient claims data for these patients—threaten the validity of several CMS measures, including the SMR.

- **Reliability.** Based on the testing results, KCP has serious concerns about the SMR's reliability. In the most recent iteration of the measure, the overall IUR for the 4-year SMR was 0.5—a sizeable decline from the 2016 value of 0.59. That is, even with the 4-year SMR, one-half of a facility's score could be attributable to random noise and not signal. Though we recognize the characterization also depends on the analytic method, we note that statistical literature traditionally interprets a reliability statistic of 0.5-0.6 as "unacceptable."[1] Thus KCP thus believes that CMS should implement the measure adjusted to yield a reliable result (reliability statistic of 0.70 or greater), consistent with how the NQF bases its evaluation of measures and more generous than the literature.[2]

  Moreover, CMS did not provide testing data stratified by facility size for the measure iteration currently under review by NQF because it "is not required." Yet we note that even when using the 4-year period, prior SMR testing results indicated very poor reliability for small- and medium-sized facilities, with IURs of 0.30 and 0.45, respectively. Only large facilities had a reasonable IUR of 0.73 for 2010-2013 data. Given this history, we believe it's disingenuous, at best, not to provide reliability based on facility size merely because NQF

---

[1] Adams JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California:RAND Corporation. TR-653-NCQA, 2009.
[2] Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London: Routledge, p. 13; DeVellis, RF. (2012). *Scale development: Theory and applications*. Los Angeles: Sage. pp. 109–110; Adams, JL. (2009). The reliability of provider profiling. RAND Health.

"does not require" it. KCP believes penalizing facilities for performance due to random chance is not appropriate and that it is imperative that CMS provide the most recent reliability results stratified by facility size. Absent that information, we submit that the demonstrably unreliable SMR, as currently specified, is particularly unreliable and unsuitable for use in small facilities. KCP thus believes the measure must specifically require a minimum sample as identified through the developer's empirical testing to prevent small facilities from having scores that are highly subject to random variability. To further support this recommendation, Discern Health identified an approach using publicly available data to estimate the reliability based on facility size (Attachment A). Based on this analysis, the IUR for the smallest 20 percent of facilities is 0.449—meaning the SMR is unreliable for this population, wherein only 44.9% of their score is attributable to quality signal and more than half to random noise.

Regarding the performance period timeframe, we note that prior testing results for the 1-year SMR yielded IURs of 0.26-0.32 for each of 2010, 2011, 2012, and 2013—an "unacceptable" degree of reliability[3] where only about 30% of the variation in a score can be attributed to between-facility differences—yet the specifications at the time permitted the 1-year measure. The 4-year SMR yielded an IUR of 0.59 over the same time period, illustrating why it is critical that the performance period be precisely defined—and why that timeframe should at a minimum be a 4-year period.

Finally, to assess more directly the value of SMR in identifying facilities with extreme outcomes, CMS and UM-KECC crafted an additional metric of reliability termed the Profile-IUR (PIUR).[4] Per CMS, "The PIUR indicates the presence of outliers or heavier tails among the providers, which is not captured in the IUR itself. . . . [When] there are outlier providers, even measures with a low IUR can have a relatively high PIUR and can be very useful for identifying extreme providers."[5] The PIUR for the SMR was PIUR is 0.77, which CMS interprets as demonstrating that "the SMR is effective at detecting outlier facilities and statistically meaningful differences in performance scores across dialysis facilities."[6] We note that NQF's Scientific Methods Panel (SMP), none of whom were familiar with the PIUR, disagree that it is an appropriate measure of reliability for *any* measure used in the ESRD Quality Incentive Program (QIP), which are used to distinguish performance between providers falling in the *middle* of the curve to determine penalties. The SMP concluded that the IUR is and remains the appropriate measure of reliability for this purpose.

- **Validity and Handling of Medicare Advantage Patients.** In previous comments to CMS, KCP noted that many of the prevalent comorbidities in the final model had p-values significantly greater than 0.05. CMS responded that the large number of clinical factors in the model generates multicollinearity among covariates, likely resulting in some unexpected results. However, KCP remains concerned that this strategy results in a model that will not be generalizable. In the current model, for example, asthma is associated with a higher risk of death than critical illness myopathy, and 'obesity' is protective while 'mood disorders' are harmful. We posit these nonsensical findings are a function of collinearity and coding idiosyncrasy. Again, KCP supports prevalent comorbidity adjustment, but we are

[3] Adams JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California:RAND Corporation. TR-653-NCQA, 2009.
[4] He K, Dahlerus C, Xia L, Li Y, Kalbfleisch JD. The profile inter-unit reliability. *Biometrics.* 2019 Oct 23. doi: 10.1111/biom.13167. [Epub ahead of print.]
[5] Kalbfleisch JD, He K, Xia L, Li Y. Does the inter-unit reliability (IUR) measure reliability? *Health Services and Outcomes Research Methodology.* 2018;18(3):215-225. Doi: 10.1007/s10742-018-0185-4.
[6] CMS. SMR measures submission materials to NQF.

concerned that the proposed collection of adjusters will be less robust with each year that passes from initial model development.

KCP also notes that validity testing yielded a c-statistic for the SMR of 0.72. 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.

Finally, data provided by CMS indicate that at the end of 2017, 27 percent of dialysis patients had MA coverage (presumably higher now), and this varied widely across states—from about 2 percent in Wyoming to 34 percent in Rhode Island, and more than 44 percent in Puerto Rico (Table 2, Attachment B). Such geographic variation compromises the validity of the measures if MA patients are not accurately accounted for in the QIP and DFC metrics. Specifically, without changes to the current specifications, the evolving patient mix will introduce significant bias into measure calculations that could affect results for facilities with either very low or high MA patient populations. Recognizing this, KCP concurs with the need to change specifications for several CMS measures to accommodate the increase in MA patients and to avoid disparities in performance due to geography. KCP strongly believes, however, that greater transparency is required by CMS as it updates the relevant measures.

While the approach to handling MA patients varies considerably across CMS's metrics (Table 1, Attachment B), KCP recognizes the difficulty in construction and notes there appears to be a logical rationale for most of the decisions made because of the properties and intended purpose of each measure. Nevertheless, KCP strongly recommends that CMS perform a sensitivity analysis of performance with and without MA patients for each of the applicable QIP/DFC measures and *make the results publicly available*. Such data will provide an opportunity for KCP and others to offer potential, evidence-based mitigation strategies (e.g., a model that accounts for both populations, use of risk coefficients, as necessary).

CMS also should perform an analysis of risk model fit under the previous approach and the new in-patient-claims-only approach; currently we are unable to assess whether model fit improved or worsened with this approach. KCP is particularly concerned that limiting comorbidity data to inpatient claims might skew the models towards a sicker population, and that such a skew might reflect unfavorably on facilities that successfully keep hospitalization rates low. That is, because comorbidity adjustors developed exclusively from hospitalization data will necessarily underestimate the comorbidity profile of patients in facilities with low hospitalization rates, the "expected" hospitalization or mortality rates calculated for such facilities will be erroneously low, and the facilities' scores will be erroneously high. Only with transparency in these matters can the community assess the impact MA patient mix has on the QIP measures.

- **Specifications and Harmonization Issues.** Prior SMR specifications indicated that "The time window can be specified from one to four years. Currently, the measure is publicly reported using four years of data." This language has now been removed from the specifications. While we have previously noted that this prior construction was imprecise and KCP believes specifications should be unambiguous, we are concerned that there is no longer any mention at all of the performance period timeframe. We believe the time period

should be exact; we further believe the 1-year period is inappropriate based on the reliability testing data (see above) and, at minimum, should be a 4-year period.

Additionally, the risk model groupings and parameters used for patient age and duration of ESRD differ among CMS's standardized measures falling under the NQF Renal Project domain. For example, the age groups for the SMR is n=3, but for Standardized Transfusion Ratio (STrR, NQF 2979) and Standardized Fistula Rate (SFR, NQF 2977) there are 5 and 4 age groupings, respectively. Similarly, the number of groups for ESRD duration for the SMR and SFR are both n=4 but the groupings differ dramatically, and both differ from the STrR (n=6). No justification or empirical analyses are offered to justify these differences.

NQF 2977: Hemodialysis Vascular Access—Standardized Fistula Rate (SFR; CMS)

The NQF Renal Standing Committee did not reach consensus for the *Evidence* criterion for the SFR secondary to concerns from some members that the measure may effectively be "topped out" at 64% for all patients for whom an arteriovenous fistula (AVF) is clinically appropriate, and because NKF's Kidney Disease Outcomes Quality Initiative downgraded the evidence supporting the relevant guideline to "expert opinion." Nevertheless, because vascular access may be the most important measure for patients making decisions about dialysis facilities, KCP believes it remains important to include a fistula measure in the ESRD QIP and we continue our general support of this measure. If the Standing Committee does not feel the empiric evidence is sufficient for this measure, we encourage them to consider NQF's "Insufficient Evidence with Exception" pathway.

We also offer the following technical considerations for the developer:

- KCP notes that the validity testing for the SFR yielded an overall c-statistic of 0.705; as previously noted, 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.

- As with the SMR, KCP strongly recommends that CMS perform a sensitivity analysis of performance with and without Medicare Advantage patients for the SFR and *make the results publicly available* to provide stakeholders the opportunity to offer potential mitigation strategies (e.g., a model that accounts for both populations, use of risk coefficients, as necessary).

- As described above for the SMR, the numbers and parameters for the risk model "patient age" and "duration of ESRD" groupings differ among CMS's standardized measures in the Renal Project; no justification or empirical analyses are offered to justify these differences.

NQF 2978: Hemodialysis Vascular Access—Long-Term Catheter Rate (LTCR; CMS)

KCP generally supports this measure.

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 203.530.9524).

Sincerely,

Kidney Care Partners:

Akebia Therapeutics, Inc.
American Kidney Fund, Inc.
American Nephrology Nurses Association
American Renal Associates
American Society of Nephrology
American Society of Pediatric Nephrology
Amgen, Inc.
Ardelyx
AstraZeneca
Atlantic Dialysis Management Services, LLC
Baxter International, Inc.
B. Braun Medical, Inc.
Cara Therapeutics, Inc.
Centers for Dialysis Care
DaVita, Inc.
Dialysis Patient Citizens, Inc.
DialyzeDirect
Fresenius Medical Care North America
Fresenius Medical Care Renal Therapies Group
Greenfield Health Systems
Kidney Care Council
National Kidney Foundation, Inc.
National Renal Administrators Association
Nephrology Nursing Certification Commission
Renal Physicians Association
Renal Support Network
Rockwell Medical
Rogosin Institute
Satellite Healthcare, Inc.
U.S. Renal Care, Inc.
Vertex Pharmaceuticals
Vifor Pharma Ltd.