KIDNEY CARE
PARTNERS

June 13, 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 NQF for undertaking this important work and offer comment on all six measures.

NQF 0260: Assessment of Health-Related Quality of Life (QoL) in Dialysis Patients (Witten and Associates, LLC)
KCP recognizes the importance of assessing the health-related quality of life for individuals with ESRD. Nevertheless we have an overarching concern about the measure, as well as specific concerns about the new specifications, evidence, performance gap, and validity.

- **OVERARCHING ISSUE.** Annual administration of the KDQOL is already required by Federal regulation, the Conditions for Coverage. KCP questions how endorsement of a measure for a process that is already mandated and surveyed will further improve patient care.

- **SPECIFICATIONS.** We support the changes to the exclusions that align them with the Conditions for Coverage, but KCP opposes eliminating the exclusion for patient refusal. First, the Conditions for Coverage permit patient refusal as long as it is documented. We believe approving a measure that directly conflicts with Federal regulation is problematic. Second, not accepting patient decisionmaking ignores patient autonomy; providers should not be forced to face intruding on patient decisionmaking vs. facing a penalty for poorer performance on this measure. We further note there is no performance gap when the specifications include patient refusal.

- **EVIDENCE.** As noted, KCP recognizes the importance of assessing health-related quality of life, but questions the lack of direct evidence for the measure. The developer cites KDOQI and the Institute of Medicine on the importance of functional assessment, however no peer-reviewed, empirical evidence is provided that the specifications (i.e., annual completion rate) are associated with higher quality.

- **PERFORMANCE GAP.** Based on the updated specifications, the performance range in 2015 was 16.7%-100%, with a median of 91.8% using "KDQOL-Complete" (K-C) data. Although the performance rate at the *patient-level* with the updated exclusion criteria (i.e., refusals = fail) is 84.8% (2015), 84.7% (2014), and 84.2% (2013), the performance rate with

1

refusals as an exclusion (old specifications) is 100% in 2013, 2014, 2015. KCP also further examined the data and notes the refusal exclusion appears stable over this period. We posit the change in specifications creates a gap where otherwise none exists, as well as puts the measure in conflict with the Conditions for Coverage.

- **VALIDITY.** KCP has two concerns about the measure's validity: the validity testing and the lack of risk adjustment.

  The developer performed validity testing on a sample that included all patients—i.e., those who refused, those who completed the survey, and those who met the exclusion criteria. It assessed association of completion with patients' KDQOL scores (linear fixed models with the score for each of the five scales as the dependent variable and facility completion rate as the main independent variable). *The models adjusted for patient-level characteristics of age, gender, race, and diabetes.* Based on this, it appears the measure was not tested as specified. First, all patients were used, even those who qualify for exclusions. Second, associations were examined, but the models were adjusted for patient-level characteristics even though the measure itself is not adjusted. Performance on the measure cannot be asserted as being associated with better quality (the five KDQOL scales) if the measure as specified is not used.

  The developer also notes, "This finding [association between completion and scores] is important because it is plausible that facilities with higher rates would be obtaining completed questionnaires from sicker patients, since it has been documented that individuals completing the QoL scores tend to be younger and healthier." Again, the developer draws this conclusion from analyzing a different data set and a risk-adjusted model. The measure is not whether an all-population, risk-adjusted measure of completion validates against the scale results: Testing and demonstration of validity must be of the measure as specified.

  Finally, KCP has expressed concern about NQF 0260 in other contexts (e.g., use in CMS' Comprehensive ESRD Care Initiative) because of the lack of risk adjustment for case mix. In fact, the developer's data demonstrate that case mix impacts a facility's score. Specifically, the developer presents data on the distribution of patient characteristics and the facility-level survey completion rate; the analysis uses refusals and completions, so comports with the proposed specifications. Facilities with more males will score, on average, 0.45% lower (per 10% difference) compared to facilities that have fewer males. Conversely, facilities with higher proportions of Asians—likely to exist in certain geographic areas—will score higher. We believe the lack of adjustment for the measure presents a significant threat to validity, particularly given a median performance of 91.8% with the updated specifications.

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 specifications, reliability, validity (risk model), and harmonization issues.

- **SPECIFICATIONS.** The specifications for the time period state "at least one year." KCP believes specifications should be unambiguous, so this construction is imprecise. We believe the time period should be an exact period, and we further believe the 1-year period is inappropriate based on the reliability testing data and, at minimum, should be a 4-year period.

  As we discuss further in the following section, KCP has significant concerns about the SMR's reliability for small- and medium-sized facilities. The SMR specifications do not

address a minimum sample size by excluding facilities of "x" or fewer patients, as we are aware other measures do.

The specifications do not exclude incident hospice patients. The NQF's Measure Applications Partnership (MAP) recently did not recommend the SMR, in part because the measure did not exclude patients who are already in hospice when they initiate dialysis. During the deliberations, it was noted that occasionally incident patients begin dialysis treatments while in hospice, but then choose to discontinue them after a period of time. KCP supports MAP's recommendation that patients who initiate dialysis while also in hospice be excluded from the SMR. As currently constructed, such patients are attributed to the facility providing the dialysis.

The SMR documentation indicates at least three expected deaths must occur for inclusion in the SMR calculations, but no justification or empirical analyses are offered to justify this threshold—e.g., how many clinics were excluded using this approach and what is the impact on scoring because of the exclusion?

Finally, the SMR specifications indicate the measures 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. We note that MAP also did not support the SMR because, in addition to the lack of a hospice exclusion, MAP felt "mortality rates would be more meaningful to consumers and actionable for facilities."

- **RELIABILITY.** Based on the testing results, KCP has serious concerns about the SMR's reliability. We note a reliability statistic of 0.70 is often considered as "good" reliability,[1] though we recognize the characterization also depends on the analytic method. Testing results for the 1-year SMR yielded IURs of 0.26-0.32 for each of 2010, 2011, 2012, and 2013—a low degree of reliability, where only about 30% of the variation in a score can be attributed to between-facility differences, yet the specifications permit this 1-year measure. The 4-year SMR yielded an IUR of 0.66 for 2009-2012 and only 0.59 for 2010-2013 data. Even with the 4-year SMR, less than 60% of a facility's score is attributable to between-facility differences for the overall sample. Moreover, 4-year SMR testing results specifically for small- and medium-sized facilities indicate very poor reliability, with IURs of 0.30 and 0.45, respectively. Only large facilities have a reasonable IUR of 0.73 for 2010-2013 data. As noted earlier, KCP thus believes the specifications must specifically require a minimum sample as identified through the developer's empirical testing.

- **VALIDITY.** KCP has strongly advocated for the use of prevalent co-morbidities in the SMR's risk model, and commends the developer for moving to incorporate prevalent co-morbidities in the specifications. We continue to be concerned about the validity of the Medical Evidence Form (CMS 2728) as a data source for incident co-morbidities, however, and urge that the Committee recommend that CMS assess this matter.

In previous comments to CMS, KCP noted that many of the prevalent co-morbidities in the final model had p-values significantly greater than 0.05—e.g., paralytic ileus (p=0.5007), episodic mood disorder NOS (p=0.8254). CMS responded that these were included because: "Most of the coefficient estimates for the prevalent co-morbidities are

---

[1] Adams, JL. The Reliability of Provider Profiling: A Tutorial. Santa Monica, California:RAND Corporation. TR-653-NCQA, 2009.

positive and statistically significant, but several do not obtain statistical significance. The very large number of clinical factors in the model expectedly generates multi-collinearity among co-variates, likely resulting in some unexpected results in direction of coefficient sign and levels of statistical significance. Inclusion of this set of prevalent co-morbidities reflects the consensus of the TEP that adjustment for all of these prevalent co-morbidities, in addition to incident co-morbidities, is important to reflect the initial and current health condition of the patient in risk adjustment."

We do not believe this approach is sufficient. Our conversations with TEP members for the SMR/SHR indicate they did not advocate for model building in a vacuum without accounting for the meaning of the coded co-morbid conditions, but rather for including as many co-morbid conditions as possible. This is a very different interpretation than is offered by the developer's explanation and far more appropriate when dealing with administrative coding habits that are not static over time. It may require, for example, grouping certain individual codes together to develop a more appropriate overarching description of true co-morbidity burden.

KCP is concerned that the strategy adopted for the SMR (and SHR) results in a model that will not be generalizable. Currently, for example, having thyroid cancer is protective to the same magnitude that atrial fibrillation is harmful. This makes no sense, and we posit is a function of collinearity and coding idiosyncrasy. Similarly, in the current model, osteomyelitis NOS-ankle is associated with a lower risk of death while ulcer of lower limb NOS is harmful. In actual medical practice, osteomyelitis is far worse than an ulcer of the lower limb. In the current model, lower extremity amputation is protective while 'status amput below knee' is harmful. Again, KCP supports prevalent co-morbidity adjustment, but we are concerned that the proposed collection of adjusters will be less robust with each year that passes from initial model development.

KCP also notes that while the SMR applies to all patients, the current list of co-morbidities does not account for those that may be unique to pediatrics. We recommend the Standing Committee suggest to the developer that such should be considered and included when indicated.

KCP also notes that the validity testing yielded a c-statistic for the SMR of 0.724. We are concerned the model will not adequately discriminate performance—particularly that smaller units, including pediatric 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.

Information on the risk model states that determination of a prevalent co-morbidity requires at least two outpatient claims or one inpatient claim, but no justification or empirical analyses are offered to support this algorithm over other approaches. We are aware this approach has been validated for diabetes,[2] but we are not that it has been validated for the large number of other co-morbidities or is broadly generalizable.

Finally, the risk model includes ambiguous language. The submission indicates patient characteristics included in the stage 1 model include "nursing home status in *previous* year." It is unclear if this means patients moving into a nursing home for the first time *during* the measurement year would not be adjusted for "nursing home status."

---

[2] Hebert PL, Geiss LS, et al. Identifying persons with diabetes using Medicare claims data. *Am J Med Qual.* 1999;14(4):270-277.

Specifically, it is unclear as to whether the look-back is *one year prior to the given event* (inclusive of the data year) or if this verbiage means the look-back is *in the previous calendar year* (not inclusive of the data year). KCP believes such ambiguity should be addressed and that the current reporting year be included, not just the previous one.

- **HARMONIZATION ISSUES.** 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 <=45, 46-85, >=86 for the <u>1-year</u> reliability analyses, but were <=135, 136-305, and >=306 for the <u>4-year</u> analyses. For the SHR, <=50, 51-87, and >=88 were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as <=46, 47-78, and >=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.

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

KCP believes hospitalization is an important outcome to measure, but has concerns about the specifications, reliability, validity (risk model), and harmonization issues. Many of our comments have been articulated in the context of those we make on the SMR, but owing to the NQF's electronic portal for measure-by-measure comments, we repeat them for the SHR.

- **SPECIFICATIONS.** KCP has strongly advocated for the use of prevalent co-morbidities in the SHR's risk model, and commends the developer for moving to incorporate prevalent co-morbidities in the specifications. We continue to be concerned about the validity of the Medical Evidence Form (CMS 2728) as a data source for incident co-morbidities, however, and urge that the Committee recommend that CMS assess this matter.

    The SHR specifications for the time period also state "at least one year." Again, as a principle, KCP believes specifications should be unambiguous. We believe the time period should be an exact period.

    As we discuss in the reliability section, KCP has significant concerns about the reliability of the 1-year SHR for small and medium facilities. The SHR specifications do not address a minimum sample size by excluding facilities of "x" or fewer patients, as we are aware other measures do.

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

    Finally, the SHR 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.

- **RELIABILITY.** We again note a reliability statistic of 0.70 is often considered as "good" reliability, though we recognize the characterization also depends on the analytic method. Again, based on the results from the reliability testing, we have significant concerns about the reliability of the 1-year SHR for small and medium facilities (IUR range of 0.46-0.65, depending on the year). The SHR specifications do not address a minimum sample size by excluding facilities of "x" or fewer patients, as we are aware other measures do. As noted earlier, KCP thus believes the specifications must specifically require a minimum sample as identified through the developer's empirical testing.

- **VALIDITY.** KCP has strongly advocated for the use of prevalent co-morbidities in the SHR's risk model, and commends the developer for moving to incorporate prevalent co-morbidities in the specifications. We continue to be concerned about the validity of the 2728 as a data source for incident co-morbidities, however, and urge that the Committee recommend that CMS assess this matter.

  In previous comments to CMS, KCP noted that many of the prevalent co-morbidities in the final model had p-values significantly greater than 0.05—e.g., paralytic ileus (p=0.5007), episodic mood disorder NOS (p=0.8254). CMS responded that these were included because: "Most of the coefficient estimates for the prevalent co-morbidities are positive and statistically significant, but several do not obtain statistical significance. The very large number of clinical factors in the model expectedly generates multi-collinearity among co-variates, likely resulting in some unexpected results in direction of coefficient sign and levels of statistical significance. Inclusion of this set of prevalent co-morbidities reflects the consensus of the TEP [Technical Expert Panel] that adjustment for all of these prevalent co-morbidities, in addition to incident co-morbidities, is important to reflect the initial and current health condition of the patient in risk adjustment."

  We do not believe this approach is sufficient. Our conversations with TEP members indicate they did not advocate for model building in a vacuum without accounting for the meaning of the coded co-morbid conditions, but rather for including as many co-morbid conditions as possible. This is a very different interpretation than is offered by the developer's explanation and more appropriate when dealing with administrative coding habits that are not static over time. It may require, for example, grouping certain individual codes together to develop an appropriate overarching description of true co-morbidity burden.

  KCP is concerned the strategy adopted for the SHR (and SMR) results in a model that will not be generalizable. Currently, for example, having thyroid cancer is protective to the same magnitude that atrial fibrillation is harmful. This makes no sense, and we posit is a function of collinearity and coding idiosyncracy. Similarly, in the current model osteomyelitis NOS-ankle is associated with a lower risk of death, while ulcer of lower limb NOS is harmful. In actual medical practice, osteomyelitis is far worse than an ulcer of the lower limb. In the current model, lower extremity amputation is protective while 'status amput below knee' is harmful. Again, KCP supports prevalent co-morbidity adjustment, but we are concerned that the proposed collection of adjusters will be less robust with each year that passes from initial model development.

  KCP also notes that the validity testing yielded an overall c-statistic for the SHR 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.

Information on the risk model states that determination of a prevalent co-morbidity requires at least two outpatient claims or one inpatient claim, but no justification or empirical analyses are offered to support this algorithm over other approaches. As noted for the SMR, we are aware this approach has been validated for diabetes, but we are not that it has been validated for the large number of other co-morbidities or is broadly generalizable.

Finally, the risk model includes ambiguous language. The submission indicates patient characteristics included in the stage 1 model include "nursing home status in *previous* year." It is unclear if this means patients moving into a nursing home for the first time *during* the measurement year would not be adjusted for "nursing home status." Specifically, it is unclear as to whether the look-back is *one year prior to the given event* (inclusive of the data year) or if this verbiage means the look-back is *in the previous calendar year* (not inclusive of the data year). KCP believes such ambiguity should be addressed and that the current reporting year be included, not just the previous one.

- **HARMONIZATION ISSUES.** 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 <=45, 46-85, >=86 for the <u>1-year</u> reliability analyses, but were <=135, 136-305, and >=306 for the <u>4-year</u> analyses. For the SHR, <=50, 51-87, and >=88 were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as <=46, 47-78, and >=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.

NQF 2977: Hemodialysis Vascular Access: Standardized Fistula Rate
As with the catheter measure, KCP used the existing arteriovenous fistula (AVF) measure, NQF 0257, for context in our review.

- **SPECIFICATIONS**. The language in #0257 that 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 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 sequellae, so credit should not be not 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* <u>and</u> *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 notes the 90-day ESRD requirement has been removed from the denominator statement as compared to #0257, which means the "clock" for the measure starts on the first day of dialysis in a non-hospital setting—but that the permitted timeframe for catheter use in the numerator is still 90 days; we support this change. Additionally, we commend the developer for adding an exclusion for patients with limited life expectancy and for now unambiguously identifying the four subcategories, both approaches that KCP had recommended.

- **VALIDITY.** KCP believes this measure improves on #0257 and commends the developer for accepting KCP's recommendation in previous comments to remove the co-variate alcohol dependence from the model's risk variables. 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.

    KCP notes 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
As with the AVF measure, KCP used the existing catheter measure, NQF 0256, for context in our review.

- **SPECIFICATIONS**. As with the AVF measure, KCP notes the 90-day ESRD requirement has been removed from the denominator statement as compared to #0256, which means the "clock" for the measure starts on the first day of dialysis in a non-hospital setting—but that the permitted timeframe for catheter use in the numerator is still 90 days; we support this change. Additionally, we commend the developer for adding an exclusion for patients with limited life expectancy and for now unambiguously identifying the four subcategories, both approaches that KCP had recommended.

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 addresses the Committee's concerns about hospital- and physician-related factors. We comment on the specifications, reliability, validity (risk model), and harmonization issues.

- **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,[3] 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.

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.

- **RELIABILITY**. In addition to our concerns that the specifications pose a threat to the validity of the updated STrR, KCP also has concerns about the reliability testing for these revised specifications.

  KCP again 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

---

[3] 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.

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. As noted earlier, KCP thus believes the specifications must specifically require a minimum sample as identified through the developer's empirical testing.

- **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 <=45, 46-85, >=86 for the <u>1-year</u> reliability analyses, but were <=135, 136-305, and >=306 for the <u>4-year</u> analyses. For the SHR, <=50, 51-87, and >=88 were used. Finally, for STrR reliability analyses, small, medium, and large facilities were defined as <=46, 47-78, and >=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,

AbbVie
Akebia
American Kidney Fund
American Nephrology Nurses Association
American Renal Associates
American Society of Nephrology
American Society of Pediatric Nephrology
Amgen
Astra Zeneca
Baxter
Board of Nephrology Examiners Nursing Technology
Centers for Dialysis Care
DaVita
Dialysis Clinic, Inc.
Dialysis Patient Citizens
Fresenius Medical Care
Fresenius Medicare Care Renal Therapies
Greenfield Health Systems
Keryx
Kidney Care Council
National Renal Administrators Association
Nephrology Nursing Certification Commission
Northwest Kidney Centers
NxStage Medical
Renal Physicians Association
Rogosin Institute
Sanofi
Satellite Healthcare
U.S. Renal Care