NATIONAL QUALITY FORUM

MEASURE WORKSHEET

This document summarizes the evaluation of the measure as it progresses through NQF's Consensus Development Process (CDP). The information submitted by measure developers/stewards is included after the Brief Measure Information, Preliminary Analysis, and Pre-meeting Public and Member Comments sections.

Purple text represents the responses from measure developers. Red text denotes developer information has changed since the last measure evaluation review. Some content in the document is from Measure Developers.

To navigate the links in the worksheet: Ctrl + click link to go to the link; ALT + LEFT ARROW to return

Purple text represents the responses from measure developers.

Red text denotes developer information that has changed since the last measure evaluation review.

Brief Measure Information

NQF #: 3565

Corresponding Measures:

De.2. Measure Title: Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities

Co.1.1. Measure Steward: Centers for Medicare and Medicaid Services

De.3. Brief Description of Measure: The Standardized Emergency Department Encounter Ratio is defined to be the ratio of the observed number of emergency department (ED) encounters that occur for adult Medicare ESRD dialysis patients treated at a particular facility to the number of encounters that would be expected given the characteristics of the dialysis facility's patients and the national norm for dialysis facilities. Note that in this document an "emergency department encounter" always refers to an outpatient encounter that does not end in a hospital admission. This measure is calculated as a ratio but can also be expressed as a rate.

When used for public reporting, the measure calculation will be restricted to facilities with greater than 5 patient years at risk in the reporting year. This restriction is required to ensure patients cannot be identified due to small cell size.

1b.1. Developer Rationale: Emergency department (ED) encounters are an important indicator of care coordination and quality of life. In the general population studies have shown higher risk of an emergency department encounter subsequent to a discharge from an inpatient hospitalization or an outpatient emergency department encounter (e.g., see Hastings et al., 2008).

Rates of ED visits among end-stage renal disease (ESRD) dialysis patients have increased between 2007 and 2016. As reported by the USRDS, the unadjusted ED visit rate among HD patients increased from 2.6 to 3.0 per patient-year, and from 2.2 to 2.4 per patient-year for PD patients (USRDS ADR 2018), while the national percentage of ED visits among dialysis patients is 62% as of 2018 (FY2020 Dialysis Facility Report). More than half (55.0%) of all patients with ESRD visit the ED during their first year of dialysis, and patients with ESRD have a mean of 2.7 ED visits per patient-year (Lovasik et al., 2016). This rate is 6-fold higher than the national mean rates for US adults in the general population (Lovasik et al 2016). Furthermore, the Lovasik study notes that among Medicare beneficiaries with ESRD, 30% of hospital admissions that originate in the ED are for diagnoses that are often dialysis related such as complications of vascular access, congestive heart failure/fluid overload, septicemia, and hyperkalemia. A study by Zhang and colleagues (Zhang et al, 2019) reported that rates of ED visits among patients on thrice weekly in-center hemodialysis vary by dialysis schedule (Mon/Weds/Fri;

Tues/Thurs/Sat) and by day of week. For example the ED visit rate (without hospital admission) was highest on the day following the longer interdialytic interval over the weekend (Mondays), suggesting an association with facility structure and treatment schedule.

Cohen and colleagues (Cohen et al 2020) reported that missed dialysis treatments are associated with an over two-fold higher risk of an ED visit, suggesting an opportunity for dialysis facilities to establish or strengthen facility practices that can help to reduce skipped treatments through increased communication, care coordination, and patient education. This in turn has the potential to reduce avoidable ED visits.

Finally, the CMS Centers for Medicare and Medicaid Innovation's Comprehensive End Stage Renal Disease (ESRD) Care model emphasizes care coordination as a central feature of care delivery in order to reduce utilization and improve outcomes. During the second performance year, the original Wave 1 cohort of ESCOs (ESRD Seamless Care Organizations) experienced about a 3% reduction in ED use relative to the period before the CEC model was launched (Marrufo et al., CEC Annual Report Performance Year 2, 2019).

Measures of the frequency of ED use may help dialysis facility level efforts to prevent emergent unscheduled care and control escalating medical costs.

References:

Hastings NS., Oddone EZ., Fillenbaum G, Shane R J., Schmader KE. Frequency and predictors of adverse health outcomes in older Medicare beneficiaries discharged from the emergency department. Med Care. 2008 Aug; 46(8):771-7

Lovasik BP, Zhang R, Hockenberry JM, Schrager JD, Pastan SO, Mohan S, Patzer RE. Emergency Department Use and Hospital Admissions Among Patients With End-Stage Renal Disease in the United States. JAMA Intern Med. 2016 Oct 1; 176(10):1563-1565.

United States Renal Data System. 2018 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2018.

Dialysis Facility Reports –Sample Report FY2020. https://www.dialysisdata.org/sites/default/files/content/Methodology/DFR_SAMPLE_201907.pdf

Zhang S, Morgenstern H, Albertus P, Nallamothu B, He K, and Saran R. Emergency department visits and hospitalizations among hemodialysis patients by day of the week and dialysis schedule in the United States. PLOS ONE. https://doi.org/10.1371/journal.pone.0220966 August 15, 2019.

Marrufo G, Negrusa B, Ullman D, Hirth R, Messana J, Maughan B, Nelson J, Lindsey N, Gregory D, Svoboda R, Melin C, Chung A, Dahlerus C, Nahra T, Jiao A, McKeithen K, and Gilfix Z. Comprehensive End-Stage Renal Disease Care (CEC) Model. Performance Year 2 Annual Evaluation Report. Prepared for: Centers for Medicare & Medicaid Services. September 2019. https://innovation.cms.gov/Files/reports/cec-annrpt-py2.pdf

S.4. Numerator Statement: The observed number of outpatient Emergency Department encounters during the reporting period among eligible adult Medicare patients at a facility.

S.6. Denominator Statement: The expected number of Emergency Department encounters among eligible Medicare patients at the facility during the reporting period adjusted for the characteristics of the patients at the facility.

S.8. Denominator Exclusions: Exclusions that are implicit in the denominator definition include time at risk while a patient:

- Has Medicare Advantage coverage
- Has had ESRD for 90 days or less
- Is less than 18 years of age

The denominator also excludes patient time at risk for calendar months in which a patient is:

- Actively enrolled in hospice at any time during the calendar month

De.1. Measure Type: Outcome

S.17. Data Source: Claims, Registry Data

S.20. Level of Analysis: Facility

IF Endorsement Maintenance – Original Endorsement Date: Most Recent Endorsement Date:

IF this measure is included in a composite, NQF Composite#/title:

IF this measure is paired/grouped, NQF#/title:

De.4. IF PAIRED/GROUPED, what is the reason this measure must be reported with other measures to appropriately interpret results? The Standardized Emergency Department Encounter Ratio (SEDR) should be considered in conjunction with the Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) for Dialysis Facilities. These measures represent two different aspects of dialysis facilities' emergency department use that assesses complementary elements of care. The SEDR describes emergency department encounter rates with reference to the totality of patients being served by a given facility. The ED30 measure on the other hand estimates the number of index hospital discharges expected to be followed by an emergency department encounter within 4-30 days after the discharge given the observed number of hospital discharges for dialysis patients at the facility. A low SEDR, corresponding to low overall emergency department encounter rates, indicates that the facility has processes in place to avoid the need for unscheduled acute care. A low ED30 indicates that a facility is successful in managing the transition of care that occurs after a hospital discharge. This is analogous to how the NQF endorsed Standardized Hospitalization Ratio (#1463) and Standardized Readmission Ratio (#2496) might also be used together to evaluate facility processes of care.

Preliminary Analysis: New Measure

To maintain NQF endorsement endorsed measures are evaluated periodically to ensure that the measures still meets the NQF endorsement criteria ("maintenance"). The emphasis for maintaining endorsement is focused on how effective the measure is for promoting improvements in quality. Endorsed measures should have some experience from the field to inform the evaluation.

Criteria 1: Importance to Measure and Report

1a. Evidence

1a. Evidence. The evidence requirements for a health outcome measure include providing empirical data that demonstrate a relationship between the outcome and at least one healthcare structure, process, intervention, or service; if these data not available, data demonstrating wide variation in performance, assuming the data are from a robust number of providers and results are not subject to systematic bias. For measures derived from patient report, evidence also should demonstrate that the target population values the measured outcome, process, or structure and finds it meaningful.

Evidence Summary

- The developer states there are numerous dialysis care processes that can influence the likelihood of a patient requiring care in the ED that would be distinct from the need for hospitalization. These processes include:
    - Fluid manament/removal processes
    - Vascular access management
    - Management of electrolyte abnormalities

- The developer cites one study that showed a two-fold or higher increased risk in ED visits due to missed dialysis treatments. The developer also cites studies suggesting that improved health literacy, improved adherence to treatment schedules, and receipt of telehealth services can reduce ED utilization in dialysis patients, including high-risk dialysis patients (for telehealth).
- The developer also cites one study that reports ED visit rates were higher on the day following the longer interdialytic interval over the weekend.
- Lastly, the developer also cites the Centers for Medicare & Medicaid Services Comprehensive End Stage Renal Disease Care model, which has shown a 3% reduction in ED use relative to when the model was launched.

Question for the Committee:

- *Is there at least one intervention that the provider can undertake to achieve a change in the measure results?*

Guidance from Evidence Algorithm

Measure assesses performance on a health outcome (Box 1) → Empirical data suggests that a structure, process, intervention or service may improve the measured health outcome (Box 2) → (PASS)

Preliminary rating for evidence: ☒ **Pass** ☐ **No Pass**

1b. Gap in Care/Opportunity for Improvement and 1b. Disparities

Maintenance measures – increased emphasis on gap and variation

1b. Performance Gap. The performance gap requirements include demonstrating quality problems and opportunity for improvement.

- The developer provides the distribution of standardized emergency department encounter ratios (SEDRs) during a three year period from 2014 – 2017. The developer notes the range (0 – 4.30) of standardized ED visit rates variation of across facilities for clinician groups during the three year span.
  - The developer notes that the data include Medicare-certified dialysis facilities with eligible time at risk for the measure. Transplant-only facilities and Veteran Affairs (VA) facilities were excluded.
  - 2017: Number of facilities=6,691, Number of patients=382,039, Mean=1.00, Std Dev=0.34, Min=0.00, Max=4.30
- Developer assigned decile ranks to the 2017 data set of SEDRs
  - Decile 1: N=669, Min=0, Max=0.62
  - Decile 2: N=669, Min=0.62, Max=0.73
  - Decile 3: N=669, Min=0.73, Max=0.81
  - Decile 4: N=669, Min=0.81, Max=0.89
  - Decile 5: N=669, Min=0.89, Max=0.96
  - Decile 6: N=670, Min=0.96, Max=1.04
  - Decile 7: N=669, Min=1.04, Max=1.13
  - Decile 8: N=669, Min=1.13, Max=1.24
  - Decile 9: N=669, Min=1.24, Max=1.41
  - Decile 10: N=669, Min=1.41, Max=4.30

Disparities

- The developer provides data demonstrating the measure's ability to identify performance gaps based on the following factors: gender, age (>75 years), and race (Hispanic, Asian, and Native American.)

Questions for the Committee:

- *Is there a gap in care that warrants a national performance measure?*

Preliminary rating for opportunity for improvement: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 1: Importance to Measure and Report (including 1a, 1b, 1c)

1a. Evidence to Support Measure Focus: For all measures (structure, process, outcome, patient-reported structure/process), empirical data are required. How does the evidence relate to the specific structure, process, or outcome being measured? Does it apply directly or is it tangential? How does the structure, process, or outcome relate to desired outcomes? For maintenance measures –are you aware of any new studies/information that changes the evidence base for this measure that has not been cited in the submission? For measures derived from a patient report: Measures derived from a patient report must demonstrate that the target population values the measured outcome, process, or structure.

- **pass**
- **The evidence appears to be directly related to the outcome.**
- **No concerns**
- **meets**
- **Little causal evidence cited and some evidence related generally to ED use and not specifically for ESRD. A systematic study showed mixed results.only tev.**
- **yes. There is a at least one intervention that can be made**
- **evidence appears to be good for readmissions but tangental to ED visits**
- **Evidence support**
- **Evidence identified including study with increased risk in ED visits due to missed dialysis treatment**
- **Yes, evidence in favor of this measure is shown**

1b. Performance Gap: Was current performance data on the measure provided? How does it demonstrate a gap in care (variability or overall less than optimal performance) to warrant a national performance measure? Disparities: Was data on the measure by population subgroups provided? How does it demonstrate disparities in the care?

- **differences in gender, age and dual eligibile**
- **The perfromance data on the measure was provided and demonstrated disparities in care.**
- **No concerns**
- **There is significant gaps, and interventions that might help. Unclear which party provides the intervention and how to be coordinated based upon references.**
- **A large amount of data was extracted from the national ESRD data base. I did not see a gap analysis.**
- **Large gap between 1st and 10th deciles. Re: disparities, comment on gender/race but notably race was dropped re: risk adjustment**
- **Disparities were looked at and not included in the end measure**
- **Identifies gaps and disparities**

- **distribution of standardized emergency department encounter ratios (SEDRs) during a three year period from 2014 – 2017. The developer notes the range (0 – 4.30) of standardized ED visit rates variation of across facilities for clinician groups during the three year span. Identified measure's ability to identify performance gaps based on the following factors: gender, age (>75 years), and race (Hispanic, Asian, and Native American.)**
- **Moderate**

Criteria 2: Scientific Acceptability of Measure Properties

2a. Reliability: Specifications and Testing

2b. Validity: Testing; Exclusions; Risk-Adjustment; Meaningful Differences; Comparability; Missing Data

Reliability

2a1. Specifications requires the measure, as specified, to produce consistent (reliable) and credible (valid) results about the quality of care when implemented. For maintenance measures – no change in emphasis – specifications should be evaluated the same as with new measures.

2a2. Reliability testing demonstrates if the measure data elements are repeatable, producing the same results a high proportion of the time when assessed in the same population in the same time period and/or that the measure score is precise enough to distinguish differences in performance across providers. For maintenance measures – less emphasis if no new testing data provided.

Validity

2b2. Validity testing should demonstrate the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For maintenance measures – less emphasis if no new testing data provided.

2b2-2b6. Potential threats to validity should be assessed/addressed.

Measure evaluated by Technical Expert Panel (TEP)? ☒ Yes ☐ No

Evaluators:

- Franklin Maddux, MD, FACP,
- Andrew Narva, MD, FACP, FASN
- Michael Fischer, MD, MSPH
- Lori Hartwell

Renal TEP Review (Combined)

Renal TEP Summary:

This measure was reviewed by an NQF-convened Renal TEP. The summary is provided below. The developer also provided responses to the concerns raised by the Renal TEP, which can be found on the Standing Committee SharePoint site.

- **Measure Evidence**
  - Some TEP members stated there wasn't sufficient evidence supporting interventions that impact ED visits following hospitalization; however, there is good evidence for interventions

that reduce repeat hospitalizations. It may be reasonable that these interventions can impact ED re-visitations.
    - There was concern regarding attribution to dialysis facilities as not all ED visits are due to to dialysis care.
- **Measure Specifications**
    - Several TEP members commented that the population is clinically appropriate and congruent with the measure intent
    - There was a concern that excluding ED visits that ultimately result in a hospitalization could potentially lead to unintended consequences of unnecessary hospitalizations. However, given that there is a performance measure capturing hospitalizations, this concern was alleviated.
    - There was a question on whether non-ED urgent care is being captured and on the call the developer had indicated that non-ED urgent care was not captured.
    - There were also concerns regarding attribution. TEP members expressed that one visit to the outpatient dialysis facility does not allow for meaningfully impact on care to avoid a repeat ED visit.
- **Measure Exclusions**
    - There were comments that the exclusions are appropriate and relevant.
    - One TEP member shared that the measure should exclude ED visits that are not dialysis-related.
- **Validity Testing**
    - TEP members agreed that the correlations demonstrating validity are appropriate.
    - Some members felt that correlations are appropriate and consistent with dialysis care, but that the correlations are small.
- **Risk adjustment**
    - Generally the TEP was supportive of the risk adjustment model, however, several members expressed concern with the lack of SDS adjustment and the inclusion of all ED visits

Complex measure evaluated by Scientific Methods Panel? ☒ **Yes** ☐ **No**

Evaluators:

- David Nerenz, PhD, Co-chair
- Sean O'Brien, PhD
- Lacy Fabian, PhD
- Marybeth Farquhar, PhD, MSN, RN
- Joseph Kunisch, PhD, RN-BC, CPHQ
- Sam Simon, PhD
- Alex Sox-Harris, PhD, MS
- Eric Weinhandl, PhD, MS
- Paul Kurlansky, MD

Methods Panel Review (Combined)

Scientific Methods Panel Votes

- Reliability: H-2; M-6; L-1; I-0 (Pass)
- Validity: H-1; M-5; L-3; I-0 (Pass)

Methods Panel Evaluation Summary:

- **Specifications:**
    - No issues
- **Reliability Testing**

- **Method(s) of reliability testing:**
    - Conducted at facility-level using data from 2014-2017 among ~6,000 facilities
    - Measure Score reliability testing was conducted at the data source and level of analysis indicated
    - The developer used a measure of inter facility variation (IUR), which evaluates signal to noise ratio. Profile IUR (PIUR) was used to assess the measure's ability to capture outliers consistently. Assuming the measure is intended to flag outliers, then use of PIUR is appropriate.
    - The SMP raised interoperability concerns stating that because the PIUR is generally not interpretable as an IUR and does not appear to have another simple or direct interpretation, it raises concerns regarding how to determine what PIUR value corresponds to "acceptable reliability".
- **Reliability testing results:**
    - Overall IUR is low (0.62). IURs stratified by facility size were not provided.
    - The reliability of the measure to flag true outliers is generally acceptable (PIUR - 0.89).

- **Validity Testing**
    - **Method(s) of validity testing:**
        - Face validity and score-level empirical validity testing were conducted
        - The validity testing focused on comparing the worse than expected group to all others (as or better than expected).
    - **Validity testing results:**
        - The SMP stated that the face validity results were acceptable but had concerns about the empirical validity testing results generally being quite weak despite performing as expected.
        - The results below show mean quality measure performance scores for facility mortality rates (SMR), transfusion events (STrR), AV Fistula rates (SFR), Percentage of Prevalent Patients Waitlisted (PPPW), Standardized Hospitalization Mortality Ratio (SHR), and Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge for Dialysis Facilities (ED30) by classification of facilities as 'better than/as expected' versus 'worse than expected' for SEDR (Table 4).

Table 4. Classification of SEDR and mean facility performance scores for Related Measures, 2017

		SEDR Classification		
Measure	Facilities Missing	Better than /As Expected	Worse than Expected	As Hypothesized?
SMR	310	1.00	1.08	Yes
STrR	619	0.98	1.14	Yes
SFR	395	63.49	62.12	Yes
PPPW	161	19.59	14.07	Yes
SHR	163	0.99	1.01	Yes

ED30	92	1.00	1.46	Yes

- **Exclusions:**
    - No exclusions indicated
- **Risk adjustment Summary:** Method – Statistical Modeling
    - Conceptual rationale for the SDS factors was included
        1. Forward stepwise regression, bootstrapped, Bayesian false discovery rate (FDR)
        2. The c-stat of the adjustment model is modest (0.62)
    - SDS factors were included in the model
        1. Employment status, ethnicity, duals
        2. Gender is only variable maintained in the final model
    - Rationale for not including many SES variable is "Adjusting for these patient factors could have the unintended consequence of creating or reinforcing disparities and limiting access to care. The primary goal should be to implement quality measures that result in the highest quality of patient care and equitable access for all patients to that care."
- **Meaningful Differences:**
    - Overall, most are flagged as expected (about 94%), while <1% are better than expected, and approximately 5% are flagged as worse than expected.

Questions for the Committee regarding reliability:

- *Do you have any concerns that the measure can be consistently implemented (i.e., are measure specifications adequate)?*
- *Are the IUR values acceptable?*
- *Is the PIUR method appropriate for determining acceptable reliability?*
- *The Scientific Methods Panel is satisfied with the reliability testing for the measure. Does the Committee think there is a need to discuss and/or vote on reliability?*

Questions for the Committee regarding validity:

- *Do you have any concerns regarding the validity of the measure (e.g., exclusions, risk-adjustment approach, etc.)?*
- *Do you agree with the developer's decision, based on their analysis, to not include SES factors in their risk-adjustment model?*
- *Does this measure identify meaningful differences in quality?*
- *Can stakeholders make judgements about quality of care when 94% of facilities are "as expected"?*
- *The Scientific Methods Panel is satisfied with the validity analyses for the measure. Does the Committee think there is a need to discuss and/or vote on validity?*

Preliminary rating for opportunity for Reliability: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Specifications precise unambiguous and complete (Box 1)→ Empirical reliability testing conducted (Box 2)→ Testing conducted at computed measure score level (Box 4)→ Method described and appropriate (Box 5) → Level of certainty or confidence that measure scores are reliable (Box 6) →MODERATE

Preliminary rating for opportunity for Validity: ☐ **High** ☒ **Moderate** ☐ **Low** ☐ **Insufficient**

Specifications consistent with evidence (Box 1)→Potential threats to validity assessed (Box 2) →Empirical validity testing of measure as specified (Box 3) →Testing performed with measure score (Box 6) → Method described and appropriate (Box 7) →Level of certainty or confidence that measure score is a valid indicator of quality (Box 8) →Moderate

Committee Pre-evaluation Comments:
Criteria 2: Scientific Acceptability of Measure Properties (including all 2a, 2b, and 2c)

2a1. Reliability-Specifications: Which data elements, if any, are not clearly defined? Which codes with descriptors, if any, are not provided? Which steps, if any, in the logic or calculation algorithm or other specifications (e.g., risk/case-mix adjustment, survey/sampling instructions) are not clear? What concerns do you have about the likelihood that this measure can be consistently implemented?

- **None**
- **The algorithm logic seemed clear.**
- **No concerns**
- **Are all ED equal in staff, facility and equipment to manage a ED ESRD patient? Would hospitalization occur as a result, and then that patient excluded from the measure?**
- **Two methods to show reliability are examined with one esp. sensitive to outliers (PIUR). Too much focus on outliers? s about consistent implementation.**
- **should be ok for implementation**
- **Moderate**
- **no concerns**
- **Overall IUR is low (0.62).**
- **Moderate**

2a2. Reliability - Testing: Do you have any concerns about the reliability of the measure?

- **none**
- **None.**
- **No concerns**
- **no**
- **Appropriate.**
- **I am honestly a little mystified by the PIUR but per some comments from the SMP, there may be concerns about interpretation of PIUR.**
- **no SES risk adjustment**
- **no concerns**
- **No**
- **Nothing specific**

2b1. Validity -Testing: Do you have any concerns with the testing results?

- **none**

- **None.**
- **No concerns**
- **If 94% are as expected, what improvement can be made?**
- **Too many interrelationships with other measures?**
- **face validity ok**
- **weak**
- **no concerns**
- **Suggest only including ED visits related to dialysis care as care for vascular access, hyperkalemai, heart failure exacerbations, etc. The reason would be to enhance the improvement opportunities that dialysis provides can implement to impact this measure**
- **Moderate**

2b4-7. Threats to Validity (Statistically Significant Differences, Multiple Data Sources, Missing Data)2b4. Meaningful Differences: How do analyses indicate this measure identifies meaningful differences about quality? 2b5. Comparability of performance scores: If multiple sets of specifications: Do analyses indicate they produce comparable results? 2b6. Missing data/no response: Does missing data constitute a threat to the validity of this measure?

- **no**
- **Not in my opinion.**
- **No concerns**
- **no**
- **No Concerns.**
- **re: SES factors and risk adjustment--same struggle as with other measures. While we believe there may be differences related to SDS/SES factors, including them in risk adjustment models typically creates concern for increasing disparities in access rather than demonstrating differences in quality of care provided. Additionally, would be interested to hear from other committee members re: the 94% "as expected" result.**
- **unsure if SES adjustments would create threat - possibly so**
- **94% of the differences are expected**
- **No further comments in this area**
- **2b5**

2b2-3. Other Threats to Validity (Exclusions, Risk Adjustment)2b2. Exclusions: Are the exclusions consistent with the evidence? Are any patients or patient groups inappropriately excluded from the measure?2b3. Risk Adjustment: If outcome (intermediate, health, or PRO-based) or resource use performance measure: Is there a conceptual relationship between potential social risk factor variables and the measure focus? How well do social risk factor variables that were available and analyzed align with the conceptual description provided? Are all of the risk-adjustment variables present at the start of care (if not, do you agree with the rationale provided)? Was the risk adjustment (case-mix adjustment) appropriately developed and tested? Do analyses indicate acceptable results? Is an appropriate risk-adjustment strategy included in the measure?

- **yes**
- **I believe that the risk-adjustment strategy is appropriate.**
- **No concerns**
- **Are ED equal and consistent for this specialized high need group in such a way that the data is accurate and equal across sites?**
- **Risk adjustment approach is logical though the SES analysis suspect that the data available may not be correct.may hey are not**
- **Social risk factor variables were mostly excluded**
- **Risk adjustment for SES was discarded for inclusion**
- **Exclusions appropriate and support the risk adjustment model**
- **Appears adequate**
- **2b2**

Criterion 3. Feasibility

Maintenance measures – no change in emphasis – implementation issues may be more prominent

3. Feasibility is the extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

- The developer states that all applicable data are readily available and sourced electronically.
- The developer notes that the data are collected by and used by healthcare personnel during care processes. Subsequently, another personnel is responsible for coding the data.

Questions for the Committee:

- *Are the required data elements routinely generated and used during care delivery?*

Preliminary rating for feasibility: ☒ **High** ☐ **Moderate** ☐ **Low** ☐ **Insufficient**

Committee Pre-evaluation Comments:
Criteria 3: Feasibility

3. Feasibility: Which of the required data elements are not routinely generated and used during care delivery? Which of the required data elements are not available in electronic form (e.g., EHR or other electronic sources)? What are your concerns about how the data collection strategy can be put into operational use?

- **none**
- **The data elements seem reasonably feasible.**
- **No concerns**
- **none**
- **All of the data can be assessed electronically but appears to involve different sources.**
- **Electronic data. No concerns**
- **none**
- **Feasibile**
- **all applicable data are readily available and sourced electronically**
- **High feasibility**

Criterion 4: Usability and Use

Maintenance measures – increased emphasis – much greater focus on measure use and usefulness, including both impact/improvement and unintended consequences

4a. Use (4a1. Accountability and Transparency; 4a2. Feedback on measure)

4a. Use evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4a.1. Accountability and Transparency. Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

Current uses of the measure

Publicly reported? ☐ **Yes** ☒ **No**

Current use in an accountability program? ☐ **Yes** ☒ **No** ☐ **UNCLEAR**

OR

Planned use in an accountability program? ☒ **Yes** ☐ **No**

Accountability program details

- The developer states that CMS will consider implementing the measure as part of CMS' Dialysis Facility public reporting program.

- The developer states that the purpose of this program is to help dialysis patients and their caregivers understand the quality of care provided by dialysis facilities and to be able to compare selected aspects of care between dialysis facilities.
- The developer does not provide a timeframe of use.

4a.2. Feedback on the measure by those being measured or others. Three criteria demonstrate feedback: 1) those being measured have been given performance results or data, as well as assistance with interpreting the measure results and data; 2) those being measured and other users have been given an opportunity to provide feedback on the measure performance or implementation; 3) this feedback has been considered when changes are incorporated into the measure

- The developer does not provide any information regarding feedback received on the measure during its development.

***Questions for the Committee**:*

- *Does the SC have any concerns related to the plan for use?*

Preliminary rating for Use: ☒ **Pass** ☐ **No Pass**

4b. Usability (4a1. Improvement; 4a2. Benefits of measure)

4b. Usability evaluate the extent to which audiences (e.g., consumers, purchasers, providers, policymakers) use or could use performance results for both accountability and performance improvement activities.

4b.1 Improvement. Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated.

- The developer does not provide any information regarding the rationale for improvement

4b2. Benefits vs. harms. Benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

- The developer does not provide any information regarding the benefits or harms of the measure.

***Questions for the Committee**:*

- *None*

Preliminary rating for Usability and use: ☐ **High** ☐ **Moderate** ☐ **Low** ☒ **Insufficient**

RATIONLE: The developer does not provide any information regarding the rationale for improvement or the potential benefits or harms of the measure.

Committee Pre-evaluation Comments:
Criteria 4: Usability and Use

4a1. Use - Accountability and Transparency: How is the measure being publicly reported? Are the performance results disclosed and available outside of the organizations or practices whose performance is measured? For maintenance measures - which accountability applications is the measure being used for? For new measures - if not in use at the time of initial endorsement, is a credible plan for implementation provided?4a2. Use - Feedback on the measure: Have those being measured been given performance results or data, as well as assistance with interpreting the measure results and data? Have those being measured or other users been given an opportunity to provide feedback on the measure performance or implementation? Has this feedback has been considered when changes are incorporated into the measure?

- **yes**
- **Providers can always use more assistance with interpreting the measure results and data.**
- **No concerns**
- **meets**
- **No use data available.**
- **Not yet in use.**
- **Since this might be used for CMS penalties in the future, not sure the attribution fits perfectly**
- **Not being used today**
- **The developer needs to provide informtaion regarding the rationale for improvement on the potential benefits or harms of the measure**
- **4a2 yes**

4b1. Usability – Improvement: How can the performance results be used to further the goal of high-quality, efficient healthcare? If not in use for performance improvement at the time of initial endorsement, is a credible rationale provided that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations?4b2. Usability – Benefits vs. harms: Describe any actual unintended consequences and note how you think the benefits of the measure outweigh them.

- **none**
- **Yes, the performance results can be used to further the goal of high-quality, efficient healthcare.**
- **No concerns**
- **not much there.**
- **There is no data to support usability.**
- **If vast majority of facilities are acceptable per the measure currently, may not be that useful to drive quality**
- **Possibly create potential for high risk patients to be rejected from centers**
- **Not being used to day but CMS will consider**
- **Appears usable**
- **4b2- Benefits overweight the harms**

Criterion 5: Related and Competing Measures

Related or competing measures

The following measures are all related, though not necessarily competing:

- 1463 : Standardized Hospitalization Ratio for Dialysis Facilities (SHR)
- 2505 : Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health
- 3566: Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) (currently undergoing endorsement review with SEDR)

Harmonization:

- The developer states that these measures are not completely harmonized, as each measure assesses different outcomes and/or target populations as reflected in the measure specifications.
- The developer states that the SEDR measures the overall rate of ED use while the ED30 focuses on ED utilization closely following a hospitalization. Both SEDR and ED30 apply to the same target population - adult Medicare-covered dialysis patients who have had ESRD for more than 90 days.
- The developer states that the SEDR measures overall outpatient acute care services while SHR measure inpatient acute care services. The SHR measures includes pediatric patients.
- The developer indicates differences in exclusions and for risk adjustment as well.

Committee Pre-evaluation Comments: Criterion 5: Related and Competing Measures

5. Related and Competing: Are there any related and competing measures? If so, are any specifications that are not harmonized? Are there any additional steps needed for the measures to be harmonized?

- **No concerns**
- **related, but not competing.**
- **There is a related measure undergoing evaluation.**
- **multiple related measures**
- **no**
- **Not competing and slightly different from current measures**
- **No further comments**
- **Yes**

Public and Member Comments

Comments and Member Support/Non-Support Submitted as of: June 12, 2020

Kidney Care Partners (KCP) appreciates the opportunity to comment on the measures under consideration for endorsement in the National Quality Forum's (NQF) All-Cause Admissions and Readmissions Project, Spring 2020 Cycle. KCP is a coalition of more than 30 organizations, comprised of patient advocates, dialysis professionals, care providers, researchers, and manufacturers, dedicated to working together to improve

quality of care for individuals with Chronic Kidney Disease (CKD) and End-Stage Renal Disease (ESRD). This letter addresses the two new measures submitted for review within the project, the Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities (NQF 3565) and the Standardized Ratio for Emergency Department Encounters Occurring within 30 Days of Hospital Discharge (ED30) for Dialysis Facilities (NQF 3566).

I. Overarching Concerns

KCP recognizes the importance of assessing emergency department (ED) utilization by individuals with ESRD. Nevertheless, we have numerous concerns about the proposed Standardized Ratio for ED Encounters Occurring within 30 Days of Hospital Discharge (ED30) and Standardized ED Encounter Ratio for Dialysis Facilities (SEDR) metrics. We believe the measures as currently specified will not improve the quality of care or outcomes for dialysis patients—and may in fact exacerbate existing sociodemographic status (SDS) and geographic disparities. Below we detail several overarching concerns and make several recommendations applicable to both metrics; concerns specific to the individual measures are then addressed.

i. Medicare Advantage (MA) Patients. Unlike CMS's other standardized measures for dialysis facilities, the SEDR and ED30 (and Standardized Transfusion Ratio) exclude MA patients because their numerator case identification relies on outpatient claims, which are largely unavailable for these patients. We appreciate the difficulty CMS faces adapting its measures to the changing Medicare environment, but have substantial concerns with this approach. Specifically, we believe the exclusion of MA patients will create an untenable scenario in which these ED measures will effectively address a population that diverges considerably from that of the other QIP measures. This may be of particular importance with the ED30 measure, as CMS promotes it as the complement to the Standardized Readmission Ratio for Dialysis Facilities (NQF 2496), wherein the two measures together provide a full picture of patients who require emergent care following hospital discharge. But as the SRR includes MA patients and the ED30 does not, the denominator populations are inherently different, and the picture provided by these complementary measures would be misleading. Additionally, CMS notes in its measure submission materials 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. We believe that such variability in coverage patterns compromises the validity of the measures, putting states, regions, and individual facilities with a low proportion of MA patients at a substantial disadvantage with the ED measures.

ii. All-Cause Construct. As proposed, ED30 and SEDR capture all ED visits by ESRD patients, regardless of cause. KCP strongly objects to this construction, believing that it is too expansive in scope and will unfairly penalize dialysis facilities for random ED visits that are beyond their control and sphere of influence. Our analysis of ED encounters during 2015 (prior to implementation of ICD-10 diagnosis coding), showed that approximately 30 percent of encounters among dialysis patients were accompanied by principal discharge diagnoses in the range from 780.x to 799.x (Symptoms, Signs, And Ill-Defined Conditions). This lack of specificity about the nature of morbidity in the ED demonstrates that ED encounters cannot be readily attributed to any one health care provider, let alone an outpatient dialysis provider.

iii. Ratio Construct. As we have done with CMS's other standardized ratio measures (the SMR, SHR, SRR, and STrR), KCP again strongly recommends that ratio measures be avoided and that risk-adjusted rates or year-over-year normalized rates be used. For the ED30 and SEDR measures in particular, we note that there is precedent for this approach; specifically, CMS has developed and actively maintains stewardship of two NQF-endorsed home health ED utilization measures (NQF 0173 and 2505) that use the type of risk-adjusted rate to which we're referring.

iv. Exclusions. KCP recommends incorporating two additional exclusions into the ED30 and SEDR measure specifications: 1) ESRD patients who seek care in an ED for any reason (including those related to ESRD and dialysis care) after missing their most recent scheduled dialysis session; and 2) ESRD patients who reside in/are discharged to a Long-Term Care or Skilled Nursing Facility. We make the former recommendation on

the basis that it is unreasonable to penalize a facility for medical issues for which it has not had the opportunity to intervene or arising from lack of adherence to prescribed care, and the latter because a dialysis facility should not be held accountable for medical decisions made by another provider (i.e., the LTC or SNF) and are beyond its realm of control.

v. Urgent Care Centers. KCP recommends that urgent care center revenue codes be included in the ED30 and SEDR numerators. The ED measures are inextricably tied to geographic locale, including but not limited to availability of EDs vs. urgent care centers. Because urgent care is not encompassed by the two measures (with the exception of centers located within an existing emergency room), facilities where an ED option is more readily available geographically than urgent care will be inordinately penalized by these measures as compared to facilities with the same patient mix where urgent care is available. We believe this will exacerbate existing SDS and geographic disparities of the type documented by the December 2016 report issued by the Office of the Assistant Secretary for Planning and Evaluation.1

vi. Risk Models. We note that risk model testing yielded an overall C-statistic of 0.665 for the ED30 and 0.61 for the SEDR, raising concerns that the models will not adequately discriminate performance. Smaller units, in particular, might look worse than their actual performance. We reiterate our long-held position that a minimum C-statistic of 0.8 is a more appropriate indicator of a model's goodness of fit, predictive ability, and validity to represent meaningful differences among facilities.

II. Standardized Ratio for ED Encounters Occurring within 30 Days of Hospital Discharge (ED30)

KCP has identified a number of concerns and makes recommendations specific to the ED30, as follows:

i. Reliability. KCP posits the ED30 is not reliable as specified. Reliability testing for measure yielded an overall IUR of 0.451 across all facilities, indicating that only 45 percent of the variation in a score can be attributed to between-facility differences (signal) and 55 percent to within-facility differences (noise)—by statistical convention, a "poor" degree of measure reliability.2,3 KCP believes it is incumbent on CMS to address the measure's empirically demonstrated lack of reliability and use an adjuster or otherwise account for the poor reliability before the measure receives further consideration.

Moreover, we fear the reliability for small facilities in particular might be substantially lower than the overall IURs, as has been the case with other CMS standardized ratio measures. To illustrate our concern, the Standardized Hospitalization Ratio for Dialysis Facilities (NQF 1463) was reported in 2013 (the most recent stratified data provided by CMS) to have an overall IUR of 0.70. However, the IUR was only 0.46 ("poor" reliability) for the nearly 35 percent of facilities (n = 2,028) meeting CMS's definition of "small" (<=50 patients, for the SHR). Without evidence to the contrary, KCP is concerned that the ED30 reliability is similarly lower for small facilities, effectively rendering the metric meaningless for use in performance measurement in this sizeable group of providers. Consistent with our previous stance on this matter, we believe it is incumbent on CMS to demonstrate reliability for all facilities by providing data by facility size and use its testing data to assess the impact of a "small numbers"

(1 U.S. Department of Health and Human Services, Office of Assistant Secretary for Planning and Evaluation Report to Congress: Social Risk Factors and Performance Under Medicare's Value-Based Purchasing Programs, December 2016. https://aspe.hhs.gov/system/files/pdf/253971/ASPESESRTCfull.pdf. Last accessed May 19, 2020.

2 A reliability statistic of 0.70 is generally considered as "acceptable" reliability.

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

effect on reliability and to empirically determine appropriate facility-level exclusion parameters and adjust the specifications accordingly.

Finally, we note that CMS has incorporated a new reliability statistic into its testing protocol, the "Profile IUR", or "PIUR". The PIUR, which itself is quite low for this measure at 0.570, was developed by CMS's measure developer contractor UM-KECC to address the unacceptably low measure reliability "that can

result when many facilities have outcomes similar to the national norm, even though the measure is still very useful to identify facilities with extreme outcomes." However, NQF's Scientific Methods Panel (SMP) noted in its April 1, 2020 conference call that the QIP measures are not intended to identify facility outliers, but rather to distinguish performance between providers. The Panel disagreed with the developer's assertion that the PIUR is an appropriate measure of reliability for the QIP measures, maintaining that the applicable statistic is the IUR. We concur with this assessment and further propose that a measure incapable of discerning performance between providers approximating the norm is not a meaningful or valid measure.

ii. Stratification of Reliability Results by Facility Size. KCP notes that unlike testing results provided for its other standardized ratio measures, CMS has provided no stratification of ED30 reliability scores by facility size; we are thus unable to discern how widely reliability varies across the spectrum of facility sizes. In particular, we are concerned that the reliability for small facilities is substantially lower than the overall IUR of 0.45 (already poor), as has been the case with other standardized ratio measures. For instance, the Standardized Transfusion Ratio for Dialysis Facilities (STrR) measure (NQF 2979) was found to have an overall IUR of 0.60—a "moderate" degree of reliability—however, the IUR for the STrR was only 0.3 for small facilities ("poor" reliability), which were defined by CMS for this measure as <=46 patients. KCP is thus concerned that the already-unacceptably low overall ED30 reliability (IUR = 0.45) is likely even lower for small facilities, effectively rendering the metric meaningless for use in performance measurement in this group of providers. We believe it highly likely that small facilities with as few as one or two patients who utilize ED services will be unfairly characterized as poor performers. KCP believes it is incumbent on CMS to demonstrate reliability for all facilities by providing data by facility size.

iii. Meaningful Differences in Performance. KCP posits that validity of the ED30 is low. An essential component of NQF's evaluation of validity is a demonstration of meaningful differences in performance. Testing results indicate that the ED30 can only distinguish differences in performance in less than 6 percent of facilities—specifically, 2.85 percent of facilities were classified as "better than expected" and 3.05 percent as "worse than expected." Simply put, the measure is unable to assess meaningful variations in performance in the overwhelming majority (94.10 percent) of facilities. This inability to discriminate between facilities illustrates the futility of using this measure, as specified, in a public reporting or value-based purchasing program—end-users will ultimately be unable to effectively compare or make informed decisions about the quality of care provided in various facilities. Again, KCP recognizes the importance of assessing ED utilization by individuals with ESRD; however, testing results do not support the premise that the proposed ED30 metric will provide a valid (or reliable, as just noted) representation of quality.

III. Standardized ED Encounter Ratio (SEDR) for Dialysis Facilities

KCP had identified a number of concerns and makes recommendations specific to the SEDR, as below.

i. Reliability. Reliability testing for the SEDR yielded an overall IUR ranging from 0.62 to 0.63—a decrease from a previous version of the measure we reviewed 2017, then 0.64 to 0.72. We have significant concerns with a measure for which reliability has demonstrably decreased. And as with the ED30, reliability statistics were not stratified by facility size, again raising concerns about inadequate measure performance in small facilities, as has been the case with other CMS standardized ratio measures. With no evidence to the contrary, we cannot simply assume that the SEDR will provide reliable, meaningful information in this group of providers and urge CMS to supply reliability data by facility size.

Finally, as with the ED30, KCP concurs with the SMP's conclusion that the developer's proposal to use the PIUR in lieu of a poor or declining IUR is wholly inappropriate. We again posit that a measure incapable of discerning performance between providers approximating the norm is not a meaningful or valid measure.

ii. Stratification of Reliability Results by Facility Size. As with the ED30, CMS has not provided stratification of SEDR reliability scores by facility size, making it impossible to discern how widely reliability varies across the spectrum of facility sizes. Again, we are concerned that the reliability for small facilities may be substantially lower than the overall IUR, as has been the case with other standardized ratio measures and that small

facilities with even one or two patients who utilize ED services might be unfairly characterized as poor performers. KCP believes it is incumbent on CMS to demonstrate reliability for all facilities by providing data by facility size.

iii. Meaningful Differences in Performance. KCP posits that the validity of the SEDR is low. Again, an essential component of the NQF's evaluation of validity is a demonstration of meaningful differences in performance. Empirical testing indicates that the SEDR can only distinguish differences in performance in approximately 5.65 percent of facilities (0.60 percent were characterized as "better than expected" and 5.05 percent as "worse than expected"); the measure was unable to assess meaningful variations in performance in the overwhelming majority (94.35 percent) of facilities. This inability to discriminate between facilities illustrates the futility of using this measure, as specified, in a public reporting or value-based purchasing program—end-users will ultimately be unable to effectively compare or make informed decisions about the quality of care provided in various facilities. We also note that the SEDR discrimination is substantially more skewed towards poor performers than the ED30, providing additional evidence that the model is not performing well. We reiterate our recognition of the importance of assessing ED utilization by individuals with ESRD. Testing results, however, do not support the validity (or reliability, as noted above) of the SEDR; it will not provide an accurate and meaningful representation of quality as currently specified.

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

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.

Board of Nephrology Examiners Nursing Technology

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.

US Renal Care

Vertex

Vifor Pharma

Combined Renal Technical Expert Panel Evaluation

Measure Number: 3565

Measure Title: Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities

1. Measure Evidence (Sections 1a. in submission form – see Evidence attachment)

1a. To what extent does the evidence provided in the submission form support the relationship of the readmission outcome to clinical processes or structures of care in dialysis facilities?

- **TEP Member #2:** Not all causes of emergency department encounters are due to dialysis care.
- **TEP Member #3:** I assume that the question is not the correct one for this measure as the measure is for ED and not readmission?

  As such, the evidence that there is opportunity redesign a delivery system to avoid excessive use of emergency departments for non-emergent issues has been clearly shown in the ESCO data where alternative health access and delivery has been encouraged and incented through the program. I would disagree that this is strictly a dialysis facility issue but an series of delivery system interventions that are frequently not within the dialysis facility. The disintermediation of a dialysis facility decisions regarding how and where the next site of care is for patients with ESRD requiring dialysis obtain their care. Some of the key emergency department visits are related to the effects of a flawed basic delivery system that allows a long interdialytic interval. As such the payment model and the current systems of care may account for some of the potential improvement in this measure. The ability of the dialysis facility to avoid harmful vascular access injury, missed treatments, residual fluid or salt loading during the treatment are areas under the direct control of the facility and treatment staff, but many of the other issues are either social or outside the dialysis facility and in the hands of the local health delivery models that occur.
- **TEP Member #1:** There is good evidence that some ER visits may be related to quality of care in the dialysis facility. However, there are also many reasons why a dialysis patient goes to the ER that is unrelated to the dialysis facility. Based on the analyses presented, there appears to be a substantial facility-level performance gap.

  My big concern is attribution. How are these ER events tied to care provided by the dialysis facility? Restrict to dialysis-related complications?
- **TEP Member #4:** The evidence is strong that ER visit sare an indicator of the dialysis care process. It is not unusual for dialysis providers to suggest that, because they are not responsible for all aspects of a

patient's care, they should not be held accountable by measures such as this. However, dialysis patients receive care that is often fragmented and the failure to collect information such as this would not be in their interests.

2. Measure Specifications (Sections S.4 – S.7 in submission form)

2a. To what extent is the measure population clinically appropriate?

- **TEP Member #3:** The target population is appropriate.
- **TEP Member #4:** The measure excludes patients who have been recently hospitalized and excludes visits that result in an admission. This is appropriate.

2b. To what extent are the definitions and codes used to identify the measure population clinically consistent with the intent of the measure?

- **TEP Member #1:** I have several concerns about specifications. First, HD and PD patients need to be separated as their patterns of ER use is likely very different and a given facility may have a different proportion of PD vs. HD patients. Second, why not include those ER visits that lead to hospitalization? Especially since those that lead to observation stays are included? One could game the system by having all of your HD patients who arrive in the ER admitted to the hospital. Both for completeness and the lost opportunity to directly admit a patient and avoid costs of ER. Third, why was the $1200 figure chosen? Fourth, I disagree with these two points about facility attribution b/c it continues to attribute patient ER visits to a dialysis facility even though the facility is no longer responsible for care of patient: When a patient transfers from one facility to another, the patient continues to be attributed to the original facility for 60 days and then is attributed to the destination facility. Patients who withdrew from dialysis or recovered renal function remain assigned to their treatment facility for 60 days after withdrawal or recovery Fifth, how is non-ER urgent care being captured? Six, the revenue codes for ER visits should be confirmed to ensure that all ER visits are being captured.
- **TEP Member #2:** It's appropriate to know the ER visits. The measure has some issues of not distinguishing what is dialysis related.
- **TEP Member #3:** The measure is a ratio, so there is room to improve the nature of the measure to avoid the problems an expected rate of ED visits. This is a very local issue of how EDs are utilized and is frequently different based on access to care issues within a community. Not all communities have the same methods within the delivery system to divert care to other settings and if the goal is to change the nature of the delivery system the dialysis facility is a component of that but not the heart of the local system which would require different subspecialties to put in place alternative sites of care for certain conditions that occur with some frequency.
- **TEP Member #4:** They are consistent.

3. Measure Exclusions (Sections S.8 – S.9 in submission form and 2b2.1 – 2b2.3 of Testing attachment)

3a. To what extent are exclusions identified and clinically relevant for the measure intent?

- **TEP Member #3:** The exclusions are all defensible. The question is whether they provide adequate focus on the patients or activities that make for a stable measure with meaning throughout the facility. I don't consider the exclusions that are related to the cusp of the time periods as being addressed in my review of the measure. I think the attempt to exclude people recently hospitalized is a good thought, but whether the exclusion is enough to address the question of these patients is not clear.
- **TEP Member #4:** Exclusions are appropriate, notably hospice care.
- 3b. To what extent are the exclusions, if any, consistent with the evidence?
- **TEP Member #3:** The existing exclusions are consistent with evidence that they may influence the SEDR in an unintended way.
- **TEP Member #4:** Consistent.

3c. To what extent do the exclusions, if any, represent a large proportion of patients that could bias the measured population?

- **TEP Member #1:** The exclusions are reasonable.
- **TEP Member #2:** The measure needs to exclude ER visits that are not dialysis related.
- **TEP Member #3:** The population of Medicare Advantage is growing will continue to grow following implementation of the 21st Century CURES Act. This will instill a biased population that is included in SEDR and should change the type of patient in the calculation and the services like transportation that the person has. This will influence what care settings are accessed beyond the ED.
- **TEP Member #4:** Developers have addressed issue of Medicare Advantage patients exclusion due to lack of data.

4. Validity Testing (Sections 2b.1.2 – 2b.1.4 of Testing attachment)

4a. To what extent are the magnitudes and directions of the correlations with other measures what you would expect?

- **TEP Member #1:** Seems reasonable and expected. The Methods Panel raised concerns about reliability - namely that this measure can only be used to detect outliers.
- **TEP Member #3:** The implications of the relationships to other measures is quite interesting to look at but the interpretations look highly speculative to me. There are numerous plausible explanations for direct or inverse relationships to certain measures. Although directionally most of these made some sense, the degree to which they are truly correlated or the impact of major health delivery system disruption like the current COVID-19 pandemic may have huge influence on whether the baseline expected rates have anything to do with reality.
- **TEP Member #4:** Expect meaningful correlations with other care indicators including SMR, StrR, SFR. The correlations between SEDR and ED30 is analogous to the correlation between SHR and SRR.

5. Risk Adjustment (Sections 2b.3 of Testing attachment)

5a. To what extent are the covariates (factors) included in the risk-adjustment model clinically relevant and consistent with the measure's intent?

- **TEP Member #1:** I don't understand why sex is included as risk adjustment but the sociodemographic factors below are not. The rationale seems inconsistent. The Methods Panel also raised this concern.
- "Race, ethnicity, dual eligible status and area deprivation are not included in the final risk adjusted model. Other studies have reported associations between patient-level race, ethnicity and dual eligible status and acute care utilization, however it is unclear whether these differences are due to underlying biological or other patient factors or represent disparities in care."
- **TEP Member #2:** The risk model of validity testing yields a result that is suboptimal. In addition, the validity testing could only tell differences in about 6 percent of dialysis facilities.
- **TEP Member #3:** The lack of geographic or racial risk adjustment is a gap that is true in all of these measures if you are trying to understand the real world application of these measures creating influence in the behaviors of an dialysis facility and its staff with respect to the patients that they are treating.
- **TEP Member #4:** They are relevant and consistent.

Combined Methods Panel Scientific Acceptability Evaluation

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3565

Measure Title: Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities

Type of measure:

☐ **Process** ☐ **Process: Appropriate Use** ☐ **Structure** ☐ **Efficiency** ☐ **Cost/Resource Use**

☒ **Outcome** ☐ **Outcome: PRO-PM** ☐ **Outcome: Intermediate Clinical Outcome** ☐ **Composite**

Data Source:

☒ **Claims** ☐ **Electronic Health Data** ☐ **Electronic Health Records** ☒ **Management Data**
☐ **Assessment Data** ☐ **Paper Medical Records** ☐ **Instrument-Based Data** ☒ **Registry Data**
☒ **Enrollment Data** ☐ **Other**

Level of Analysis:

☐ **Clinician: Group/Practice** ☐ **Clinician: Individual** ☒ **Facility** ☐ **Health Plan**
☐ **Population: Community, County or City** ☐ **Population: Regional and State**
☐ **Integrated Delivery System** ☐ **Other**

Measure is:

☒ **New** ☐ **Previously endorsed (**NOTE: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.)

RELIABILITY: SPECIFICATIONS

1. **Are submitted specifications precise, unambiguous, and complete so that they can be consistently implemented?** ☒☐ **Yes** ☒☐ **No**

   Submission document: "MIF_3565" document, items S.1-S.22

   NOTE*: NQF staff will conduct a separate, more technical, check of eCQM specifications, value sets, logic, and feasibility, so no need to consider these in your evaluation.*

2. **Briefly summarize any concerns about the measure specifications.**
   - **Panel Member #1** None
   - **Panel Member #3** This is a classical measure specification pertaining to an outcome in the population of dialysis patients with Medicare Parts A and B. Outpatient acute care in the emergency department and observation status has been increasing in recent years, although the significance of this increase is unclear to me.
   - Many patients with acute complications are advised by their dialysis facilities to visit the emergency department, rather than to wait until the next outpatient dialysis treatment. If the patient visits the emergency room, he or she may be admitted to the inpatient hospital, admitted into outpatient observation status, or evaluated and discharged. The latter two classes are the focus of the metric, whereas the first class is not.
   - Therefore, strictly speaking, the measure mixes the incidence of acute complications, the idiosyncratic advice of dialysis facility staff, the likelihood of a local hospital to admit dialysis patients in the emergency department to the inpatient hospital, and implicitly, the relative supply of urgent care versus emergency medicine in the local market. For this reason, the measure is stealthily difficult to interpret. All of this is very important for the NQF SMP to consider, especially as more acute care moves from inpatient to outpatient settings.
   - I do have one specific concern, aside from the general concerns expressed above. The flow chart appears to indicate that a patient-month within 2 months after a month with an inaptient claim could be retained in the measure denominator. It seems that this definition permits inclusion of a dialysis patient with both Medicare as a secondary payer and a recent hospitalization. This is very problematic.

- **Panel Member #6** May want to clarify what an “emergeny department encounter” is as in one place on the form, it refers to an “outpatient encounter that does not end in a hospital admission.” Elsewhere on the form the developer indicates that “observational stays” are included in the measure numerator details. Not sure why they included it.
- **Panel Member #8** Specifications are not clearly documented, coefficients are not provided.

RELIABILITY: TESTING

Submission document: “MIF_xxxx” document for specifications, testing attachment questions 1.1-1.4 and section 2a2

3. **Reliability testing level** ☒ **Measure score** ☐ **Data element** ☐ **Neither**
4. **Reliability testing was conducted with the data source and level of analysis indicated for this measure** ☒ **Yes** ☒ **No**
5. If score-level and/or data element reliability testing was NOT conducted or if the methods used were NOT appropriate, was **empirical <u>VALIDITY</u> testing** of <u>patient-level data</u> conducted?

   ☐ **Yes** ☒ **No**
6. **Assess the method(s) used for reliability testing**

   Submission document: Testing attachment, section 2a2.2

   - **Panel Member #6** Used IUR and profile IUR. Appropriate method of reliability testing for detecting outliers.
   - **Panel Member #8** Developer used a measure of inter facility variation (IUR), which evaluates signal to noise ratio. Profile IUR (PIUR) was used to assess the measure’s ability to capture outliers consistently. Assuming the measure is intented to flag outliers, then use of PIUR is appropriate.
   - **Panel Member #9** The developers report: (1) inter unit reliability (IUR) which is the conventional proportion of signal variation definition of reliability and (2) profile inter unit reliability (PIUR) which is a relatively recent method. The PIUR addresses how well the measure can identify providers in the tails of the performance distribution but the interpretation is not straightforward. Conceptually, it involves identifying providers who have scores above a threshold (i.e. low performance) and then calculating the proportion of these providers who would have scores above this threshold again if performance was re-estimated in a different random sample of patients from the same provider-specific patient population while holding each provider's underlying true performance fixed. After determining this "reflagging probability" quantity, the PIUR is calculated as the value of IUR that would yield this reflagging probability in a hypothetical measurement scenario in which true and estimated performance values are distributed according to a random effects model with normally distributed true performance values. If this type of hierarchical model is a good approximation of truth, then IUR and PIUR would be estimating the same quantity and so whatever threshold numerical value corresponds to "acceptable reliability" for IUR results ould also be applied when evaluating PIUR results. However, the motivation for using PIUR is the assumption that true performance is not normally distributed e.g. the number of providers with extremely high or low true performance may be higher than what would be expected under a normal distribution. When the PIUR is applied to datasets in which true performance is non-normal, my impression is that it cannot be interpreted as estimating the same quantity as the IUR (i.e it is not estimating the squared correlation between true and estimated values or the proportion of signal variation). In fact the true PIUR may be much higher than the true IUR. Because the PIUR is not in general interpretable as an IUR and because it does not appear to have another simple or direct interpretation, this raises the question of how to determine what PIUR value corresponds to "acceptable reliability".
7. **Assess the results of reliability testing**

Submission document: Testing attachment, section 2a2.3

- **Panel Member #1** Overall IUR is low (0.62). IURs stratified by facility size were not provided. This suggests that the meausure is too unreliable be used to distinguished true between facility differences (signal vs noise). However, the reliability of the measure to flag true outliers is good (.89). Thus, the use of the measure for this specific purpose appears to be supported.
- **Panel Member #2** The value obtained for the IUR was moderate. The PIUR value was larger and demonstrates that the SEDR has high reliability for detecting outlier facilities.
- Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities?
- **Panel Member #3** The estimated IUR is equal to 0.62-0.63 in 2014-2017, whereas the estimated PIUR is equal to 0.89-0.91 in the same interval.
- **Panel Member #6** Appropriate interpretation of testing results.
- **Panel Member #8** The underlying signal to noise ratio (IUR) among approximately 6000 facilities across multiple years was consistently marginal (r=0.62 – 0.63), however, PIUR was much higher (0.89 – 0.91), indicating more acceptable discrimination of outliers.

8. Was the method described and appropriate for assessing the proportion of variability due to real differences among measured entities? NOTE: If multiple methods used, at least one must be appropriate.

   Submission document: Testing attachment, section 2a2.2

   ☒ **Yes**

   ☐ **No**

   ☐ **Not applicable** (score-level testing was not performed)

9. Was the method described and appropriate for assessing the reliability of ALL critical data elements?

   Submission document: Testing attachment, section 2a2.2

   ☐ **Yes**

   ☒ **No**

   ☐ **Not applicable** (data element testing was not performed)

10. **OVERALL RATING OF RELIABILITY** (taking into account precision of specifications and all testing results):

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has not been conducted)

    ☒ **Low** (NOTE: Should rate LOW if you believe specifications are NOT precise, unambiguous, and complete or if testing methods/results are not adequate)

    ☐ **Insufficient** (NOTE: Should rate INSUFFICIENT if you believe you do not have the information you need to make a rating decision)

11. **Briefly explain rationale for the rating of OVERALL RATING OF RELIABILITY and any concerns you may have with the approach to demonstrating reliability.**

    - **Panel Member #1** As stated above, the overall IUR is low. This suggests that the meausure is too unreliable be used to distinguished true between facility differences (signal vs noise) across the distribution of performance. However, the reliability of the measure to flag true outliers is good. Thus, the use of the measure for this specific purpose appears to be supported. If the measure title or description were clear about the intended use of the measure (flagging outliers), then I would rate reliability higher.

- **Panel Member #2** The IUR values are moderate; the PIUR values are high. Since no data element reliability testing was done, the values for measure score reliability become the overall ratings for reliability.
- **Panel Member #3** I must disclose that I do not fully understand the advantages, disadvantages, and vulnerabilities of the PIUR statistic. The IUR statistic is relatively low. However, the steward suggests that the PIUR statistic offers a window into the ability of the measure to reliably identify outlying facilities. The PIUR statistic is in the neighborhood of 0.9, thus suggesting to me that the measure can reliably identify facilities with high excess mortality.
- **Panel Member #5** The value obtained for the IUR is moderate in size. The PIUR is larger and demonstrates that the SEDR is effective at detecting outlier facilities and statistically meaningful differences in performance scores across dialysis facilities.
- **Panel Member #6** No concerns.
- **Panel Member #8** Overall reliability (not PIUR) was substandard, but if the use case for the measure is identifying outliers, PIUR was good.
- **Panel Member #9** Moderate rating is based on an IUR estimate of 0.62. The estimated PIUR was 0.89 which sounds quite high but I am not sure 100% how to judge this.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

12. **Please describe any concerns you have with measure exclusions.**

    Submission document: Testing attachment, section 2b2.

    - **Panel Member #1** None
    - **Panel Member #2** None
    - **Panel Member #3** The exclusion of Medicare Part C enrollees is appropriate, given lack of availability of outpatient facility claims. The challenge posed by
    - **Panel Member #6** No concerns.

13. **Please describe any concerns you have regarding the ability to identify meaningful differences in performance.**

    Submission document: Testing attachment, section 2b4.

    - **Panel Member #1 - Approximately 5% are flagged as worse than expected.**
    - **Panel Member #2** As noted above, the measure can only reliably identify extreme high or low outliers. It cannot identify meaningful differences in performance within the large main body of the distribution of scores.
    - **Panel Member #3** I have no specific concerns.
    - **Panel Member #8** None – measure appears to have good discriminiation between deciles of performance.
    - **Panel Member #6** No concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**

    Submission document: Testing attachment, section 2b5.

    - **Panel Member #1** N/A
    - **Panel Member #6** No concerns.
    - **Panel Member #2** N/A
    - **Panel Member #3** This item is not applicable.

15. **Please describe any concerns you have regarding missing data.**

Submission document: Testing attachment, section 2b6.

- **Panel Member #1** None
- **Panel Member #2** No significant concerns
- **Panel Member #3** I do not fully understand the genesis of such a high percentage of patients with less than 6 months of Medicare coverage during the preceding year. This lacks face validity to me.
- **Panel Member #6** No concerns.

16. **Risk Adjustment**

16a. **Risk-adjustment method** ☐ **None** ☒ **Statistical model** ☐ **Stratification**

16b. **If not risk-adjusted, is this supported by either a conceptual rationale or empirical analyses?**

☐ Yes ☐ No ☒ Not applicable

16c. **Social risk adjustment:**

16c.1 Are social risk factors included in risk model? ☒ Yes ☒ No ☐ Not applicable

16c.2 Conceptual rationale for social risk factors included? ☒ Yes ☒ No

16c.3 Is there a conceptual relationship between potential social risk factor variables and the measure focus? ☐ Yes ☒ No

16d. **Risk adjustment summary:**

16d.1 All of the risk-adjustment variables present at the start of care? ☒ Yes ☐ No

16d.2 If factors not present at the start of care, do you agree with the rationale provided for inclusion? ☒ Yes ☐ No

16d.3 Is the risk adjustment approach appropriately developed and assessed? ☒ Yes ☐ No

16d.4 Do analyses indicate acceptable results (e.g., acceptable discrimination and calibration) ☒ Yes ☒ No

16d.5.Appropriate risk-adjustment strategy included in the measure? ☒ Yes ☒ No

16e. **Assess the risk-adjustment approach**

- **Panel Member #1** Generally adequate methods. Rationale for not including many SES variable is "Adjusting for these patient factors could have the unintended consequence of creating or reinforcing disparities and limiting access to care. The primary goal should be to implement quality measures that result in the highest quality of patient care and equitable access for all patients to that care." As discussed, the wisdom of this decision may depend on the use of the measure (QI vs accountability). The C-stat of the adjustment model is 0.62
- **Panel Member #2** The developer presents a detailed and strong case for the influence of both patient-level and area-level social and economic factors on ED use, and therefore on ED use in the context of this measure, independent of the quality of care provided by dialysis facilities. Therefore, all the essential conditions for using these factors in adjustment laid out in the NQF 2014 SES Expert Panel report have been met. In spite of this, the developer has chosen to not include patient-level and area-level SDS and SES factors in the risk adjustment model. This is not acceptable.
- **Panel Member #3** The risk adjustment approach is logical. The challenge of the approach relates to the simple fact that physician/supplier claims apparently do not inform comorbidity identification.
- **Panel Member #6** Approach is appropriate**.** C-statistic for the measure is 0.61
- **Panel Member #8** Modest c-statistic indicates despite the high # of risk factors, case mix may not be adequately controlled

VALIDITY: TESTING

17. **Validity testing level:** ☒ **Measure score** ☐ **Data element** ☐ **Both**

18. **Method of establishing validity of the measure score:**

    ☒ **Face validity**

    ☒ **Empirical validity testing of the measure score**

    ☐ **N/A (score-level testing not conducted)**

19. **Assess the method(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.2

    - **Panel Member #1** The validity testing focused on comparing the worse than expected group to all others (as or better than expected). What I liked about this method is it mapped on the most reliable use of the measure – flagging outliers.
    - **Panel Member #2** The face validity results are acceptable; the empirical validity testing results are generally quite weak, but in the predicted directions for the most part.
    - **Panel Member #3** The steward has assessed the correlation of the measure with other dialysis facility-level measures.
    - **Panel Member #6** Method used to assess validity is appropriate.
    - **Panel Member #8** Hypthesis testing; very robust approach.

20. **Assess the results(s) for establishing validity**

    Submission document: Testing attachment, section 2b2.3

    - **Panel Member #1** The hypothesized relationships were supported.
    - **Panel Member #2** The face validity results are acceptable; the empirical validity testing results are generally quite weak, but in the predicted directions for the most part.
    - **Panel Member #3** The associations in Table 4 of Measure Testing are difficult to interpret and ostensibly unimpressive. The measure appears to have weak associations with other measures (excluding "ED30").
    - **Panel Member #6** Appropriate interpretation of results.
    - **Panel Member #8** Hypothesis testing indicates measure is valid.

21. **Was the method described and appropriate for assessing conceptually and theoretically sound hypothesized relationships?**

    Submission document: Testing attachment, section 2b1.

    ☒ **Yes**

    ☐ **No**

    ☐ **Not applicable** (score-level testing was not performed)

22. **Was the method described and appropriate for assessing the accuracy of ALL critical data elements?**
    NOTE that data element validation from the literature is acceptable.

    Submission document: *Testing attachment, section 2b1.*

    ☒ **Yes**

    ☐ **No**

    ☒ **Not applicable** (data element testing was not performed)

23. **OVERALL RATING OF VALIDITY taking into account the results and scope of all testing and analysis of potential threats.**

    ☐ **High** (NOTE: Can be HIGH only if score-level testing has been conducted)

    ☒ **Moderate** (NOTE: Moderate is the highest eligible rating if score-level testing has NOT been conducted)

☒ **Low** (NOTE: Should rate LOW if you believe that there are threats to validity and/or relevant threats to validity were not assessed OR if testing methods/results are not adequate)

☐ **Insufficient** (NOTE: For instrument-based measures and some composite measures, testing at both the score level and the data element level is required; if not conducted, should rate as INSUFFICIENT.)

24. **Briefly explain rationale for rating of OVERALL RATING OF VALIDITY and any concerns you may have with the developers' approach to demonstrating validity.**

- **Panel Member #1 –** The methods were reasonable and results were
- **Panel Member #2** As in the case of reliability, only score-level validity testing was done, so the moderate rating for score-level validity is the same as the moderate rating for overall validity.
- **Panel Member #3** The data that are presented offer little to the support the characteriziation of the measure as an outcome measure that reflects health status.
- **Panel Member #5** The results below show mean quality measure performance scores for SMR, STrR, SFR, PPPW, SHR, and ED30 by classification of facilities as 'better than/as expected' versus 'worse than expected' for SEDR (Table 4). Taken together these results provide validation support for SEDR. Performance on key quality measures that were expected to be related to ED use was also related to facility flagging in the respective 'better than/as expected' or 'worse than expected' categories.
- **Panel Member #6** No concerns regarding developers' approach to demonstrating validity.
- **Panel Member #8** Hypothesis testing indicates the measure has good validity, although the c-statistic was modest.
- **Panel Member #9** Reporting rates of ER encounters at the level of dialysis facilities makes sense to me if there is evidence to suggest that variation in risk-adjusted ER rates across dialysis facilities is largely related to care provided by the dialysis facility. In absence of such evidence or compelling arguments, my default assumption is that variation in ER rates would be largely determined by care provided by providers other than the dialysis facility. Thus, an excellent dialysis facility may have poor outcomes because their patients tend to receive below-average care from their other care providers (not the dialysis facility). The developers mention that this topic ("the degree to which performance on a measure is under control of the dialysis facility") was discussed by a TEP in 2016 but the fact that it was a discussion topic is not in itself sufficiently convincing.
- If I am understanding correctly, the statistical method used to classify hospitals as better or worse than expected based on the "empirical null distribution" appears to account for the idea that variation in risk-adjusted ER rates may not be fully explained by care provided by dialysis facilities and this may partially mitigate my concern. If we were able to observe true facility-specific risk-adjusted ER rates, then what range of risk-adjusted ER rates rates would be regarded as normal/acceptable and value would be large enough to raise a concern about the dialysis facility?

ADDITIONAL RECOMMENDATIONS

25. **If you have listed any concerns in this form, do you believe these concerns warrant further discussion by the multi-stakeholder Standing Committee? If so, please list those concerns below.**

Panel Member #2 The measure has been shown to be reliable for the purpose of identifying extreme outliers (e.g., top or bottom 5% of the score distribution). NQF endorsement should reflect that limitation. The measure should not be used for other purposes based on an "NQF-endorsed" status.

The decision to exclude significant individual-level and area-level SDS and SES factors from the adjustment model is not acceptable. The risk adjustment model should include these factors.

Panel Member #8 Appropriateness of measure exclusions should be reviewed by standing committee.

1. Evidence and Performance Gap – Importance to Measure and Report

Extent to which the specific measure focus is evidence-based, important to making significant gains in healthcare quality, and improving health outcomes for a specific high-priority (high-impact) aspect of healthcare where there is variation in or overall less-than-optimal performance. ***Measures must be judged to meet all sub criteria to pass this criterion and be evaluated against the remaining criteria.***

1a. Evidence to Support the Measure Focus – See attached Evidence Submission Form

xxxxxxxxxx.docx

1a.1 For Maintenance of Endorsement: Is there new evidence about the measure since the last update/submission?

Do not remove any existing information. If there have been any changes to evidence, the Committee will consider the new evidence. Please use the most current version of the evidence attachment (v7.1). Please use red font to indicate updated evidence.

1a. Evidence (subcriterion 1a)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities

Date of Submission: 4/9/2020

1a.1.This is a measure of: (*should be consistent with type of measure entered in De.1*)

Outcome

☒ Health outcome: Emergency department utilization that does not result in hospitalization

☐ Patient-reported outcome (PRO): Click here to name the PRO

PROs include HRQoL/functional status, symptom/symptom burden, experience with care, health-related behaviors. (*A PRO-based performance measure is not a survey instrument. Data may be collected using a survey instrument to construct a PRO measure.)*

☐ Intermediate clinical outcome (*e.g., lab value*): Click here to name the intermediate outcome

☐ Process: Click here to name what is being measured

☐ Appropriate use measure: Click here to name what is being measured

☐ Structure: Click here to name the structure

☐ Composite: Click here to name what is being measured

1a.12 LOGIC MODEL Diagram or briefly describe the steps between the healthcare structures and processes (e.g., interventions, or services) and the patient's health outcome(s). The relationships in the diagram should be easily understood by general, non-technical audiences. Indicate the structure, process or outcome being measured.

Emergency Department (ED) utilization is an important indicator of patient morbidity and quality of life. As reported by the USRDS, from 2007 – 2016 the unadjusted ED visit among HD patients increased from 2.6 to 3.0 per patient-year, and from 2.2 to 2.4 per patient-year for PD patients (USRDS ADR 2018). More than half (55.0%) of all patients with end-stage renal disease (ESRD) visit the ED during their first year of dialysis, and patients with ESRD have a mean of 2.7 visits per patient-year [1]. This rate is 6-fold higher than the national mean rates for US adults in the general population [2], while the national percentage of ED visits among

dialysis patients is 62% as of 2018 (FY2020 Dialysis Facility Report). Measures of the frequency of ED use at the dialysis facility level may help efforts to prevent emergent unscheduled care and control escalating medical costs. There are numerous dialysis care processes that can influence the likelihood of a patient requiring care in the ED that would be distinct from the need for hospitalization (i.e. the ED is not merely a gateway to hospital admission). These processes include:

(1) Inadequate processes related to fluid management/removal: Inadequate control of total body fluid balance and fluid removal can result in fluid overload and congestive heart failure, increasing the possibility of the need for ED use and emergent dialysis. Conversely, overly aggressive fluid removal can lead to hypotension and in extreme situations, the patient may become unresponsive (i.e. syncope). When this happens, patients are often sent to the ED for additional evaluation, but are rarely admitted.
(2) Inadequate management of vascular access: vascular access thrombosis or bleeding, or malfunction of a central venous catheter may require urgent intervention. If facilities do not have established processes of care to manage these access related complications, patients may be referred to the ED for intervention, but would not necessarily require hospital admission. Furthermore, inadequate infection prevention processes can lead to bacteremia or septicemia, increasing the possibility of the need for ED use.
(3) Inadequate management of electrolyte abnormalities: Failure to maintain processes to ensure adequate dialysis and nutritional counseling can lead to hyperkalemia, increasing the possibility of the need for ED use and emergent dialysis. Once potassium is controlled, patients can often be discharged from the ED without requiring hospitalization.

Dialysis Facility Reports –Sample Report FY2020.
https://www.dialysisdata.org/sites/default/files/content/Methodology/DFR_SAMPLE_201907.pdf

United States Renal Data System. 2018 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2018.

****RESPOND TO ONLY ONE SECTION BELOW -EITHER 1a.2, 1a.3 or 1a.4) ****

1a.2 FOR OUTCOME MEASURES including PATIENT REPORTED OUTCOMES- State the rationale supporting the relationship between the health outcome (or PRO) to at least one healthcare structure, process (e.g., intervention, or service).

Among Medicare beneficiaries, 30% of hospital admissions that originate in the ED are for diagnoses that are often dialysis related such as complications of vascular access, congestive heart failure/fluid overload, septicemia, and hyperkalemia [1]. Recent research points to many additional opportunities to further reduce unnecessary ED use in this population. Programs developed to impact dialysis provider practices have been shown to improve intermediate outcomes (reduced catheter vascular access [3], small solute adequacy, anemia management), hospitalization, and mortality.

Cohen and colleagues [9] reported that missed dialysis treatments are associated with an over two-fold higher risk of an ED visit,

suggesting an opportunity for dialysis facilities to establish or strengthen facility practices that can help to reduce skipped treatments through increased communication, care coordination, and patient education. This in turn has the potential to reduce avoidable ED visits. Given the association between missed dialysis treatments and increased risk of an ED visit [4], dialysis facility interventions that improve adhearance to the

treatment schedule would be expected to decrease ED utilization. Other interventions, such as telehealth, have been demonstrated to reduce ED utilization in high-risk dialysis patients [5].

Zhang and colleagues [10] reported that rates of ED visits among patients on thrice weekly in-center hemodialysis vary by dialysis schedule (Mon/Weds/Fri; Tues/Thurs/Sat) and by day of week. For example the ED visit rate (without hospital admission) was highest on the day following the longer interdialytic interval over the weekend (Mondays), suggesting an association with facility structure and treatment schedule.

In the general population, outpatient ED visits were reported to have increased more slowly for Medicare patients being treated by patient-centered medical home practices when compared to non-patient-centered medical homes[6]. A comparable example that may hold promise of reducing ED use among ESRD dialysis patients is the current CMS Centers for Medicare and Medicaid Innovation's Comprehensive End Stage Renal Disease (ESRD) Care model that emphasizes care coordination as a central feature of care delivery in order to reduce utilization and improve outcomes. During the second performance year, the original Wave 1 cohort of ESCOs (ESRD Seamless Care Organizations) experienced about a 3% reduction in ED use relative to the period before the CEC model was launched [11].

Finally, low health literacy has been associated with increased use of ED services [7] and some studies have indicated that patient education interventions can reduce ED utilization [8].

References:

1. Lovasik, B.P., et al., Emergency Department Use and Hospital Admissions Among Patients With End-Stage Renal Disease in the United States. *JAMA Intern Med*, 2016. 176(10): p. 1563-1565.
   Patients with end-stage renal disease (ESRD) have the highest risk for hospitalization among those with chronic medical conditions, including heart failure, pulmonary disease, or cancer.1 However, to our knowledge, no study has examined use of the emergency department (ED) among the national Medicare population with ESRD. We sought to describe ED visits and hospitalizations through the ED and to determine the sociodemographic and clinical characteristics of patients with ESRD who use ED services in the United States.

2. Centers for Disease Control and Prevention. National hospital ambulatory medical care survey: 2011 emergency department summary tables. http://www.cdc.gov/nchs/fastats/injury.htm 2011 [cited 2017 January 9].

3. Ng LJ, Chen F, Pisoni RL, Krishnan M, Mapes D, Keen M, Bradbury BD. Hospitalization risks related to vascular access type among incident US hemodialysis patients. Nephrol Dial Transplant. 26(11):3659-66, 2011

   BACKGROUND: The excess morbidity and mortality related to catheter utilization at and immediately following dialysis initiation may simply be a proxy for poor prognosis. We examined hospitalization burden related to vascular access (VA) type among incident patients who received some predialysis care.

   METHODS: We identified a random sample of incident US Dialysis Outcomes and Practice Patterns Study hemodialysis patients (1996-2004) who reported predialysis nephrologist care. VA utilization was assessed at baseline and throughout the first 6 months on dialysis. Poisson regression was used to estimate the risk of all-cause and cause-specific hospitalizations during the first 6 months.

RESULTS: Among 2635 incident patients, 60% were dialyzing with a catheter, 22% with a graft and 18% with a fistula at baseline. Compared to fistulae, baseline catheter use was associated with an increased risk of all-cause hospitalization [adjusted relative risk (RR) = 1.30, 95% confidence interval (CI): 1.09-1.54] and graft use was not (RR = 1.07, 95% CI: 0.89-1.28). Allowing for VA changes over time, the risk of catheter versus fistula use was more pronounced (RR = 1.72, 95% CI: 1.42-2.08) and increased slightly for graft use (RR = 1.15, 95% CI: 0.94-1.41). Baseline catheter use was most strongly related to infection-related (RR = 1.47, 95% CI: 0.92-2.36) and VA-related hospitalizations (RR = 1.49, 95% CI: 1.06-2.11). These effects were further strengthened when VA use was allowed to vary over time (RR = 2.31, 95% CI: 1.48-3.61 and RR = 3.10, 95% CI: 1.95-4.91, respectively). A similar pattern was noted for VA-related hospitalizations with graft use. Discussion. Among potentially healthier incident patients, hospitalization risk, particularly infection and VA-related, was highest for patients dialyzing with a catheter at initiation and throughout follow-up, providing further support to clinical practice recommendations to minimize catheter placement.

4. Chan, K. E.;Thadhani, R. I.;Maddux, F. W. Adherence barriers to chronic dialysis in the United States**.** *J Am Soc Nephrol.* 2014 25(11):2642-8 doi:10.1681/asn.2013111160

   Hemodialysis patients often do not attend their scheduled treatment session. We investigated factors associated with missed appointments and whether such nonadherence poses significant harm to patients and increases overall health care utilization in an observational analysis of 44 million hemodialysis treatments for 182,536 patients with ESRD in the United States. We assessed the risk of hospitalization, emergency room visit, or intensive-coronary care unit (ICU-CCU) admission in the 2 days after a missed treatment relative to the risk for patients who received hemodialysis. Over the 5-year study period, the average missed treatment rate was 7.1 days per patient-year. In covariate adjusted logistic regression, the risk of hospitalization (odds ratio [OR], 3.98; 95% confidence interval [95% CI], 3.93 to 4.04), emergency room visit (OR, 2.00; 95% CI, 1.87 to 2.14), or ICU-CCU admission (OR, 3.89; 95% CI, 3.81 to 3.96) increased significantly after a missed treatment. Overall, 0.9 missed treatment days per year associated with suboptimal transportation to dialysis, inclement weather, holidays, psychiatric illness, pain, and gastrointestinal upset. These barriers also associated with excess hospitalization (5.6 more events per patient-year), emergency room visits (1.1 more visits), and ICU-CCU admissions (0.8 more admissions). In conclusion, poor adherence to hemodialysis treatments may be a substantial roadblock to achieving better patient outcomes. Addressing systemic and patient barriers that impede access to hemodialysis care may decrease missed appointments and reduce patient morbidity.

5. Minatodani, D. E.;Berman, S. J. Home telehealth in high-risk dialysis patients: a 3-year study. *Telemed J E Health.* 2013 19(7):520-2 doi:10.1089/tmj.2012.0196

   OBJECTIVE: This study is a continuation of a previous pilot project that demonstrated improved health outcomes and significant cost savings using home telehealth with nurse oversight in patients with end-stage renal disease undergoing chronic dialysis. We are reporting the results of a larger sample size over a 3-year study period to test the validity of our original observations.

   SUBJECTS AND METHODS: Ninety-nine patients were included in this study; 43 (18 females, 25 males) with a mean age of 58.6 years were enrolled in the remote technology (RT) group, and 56 (26 females, 30 males) with a mean age of 63.1 years were enrolled in the usual-care (UC) group. Health resource outcome measures included hospitalizations, emergency room (ER) visits, and number of days hospitalized. Economic analysis was conducted on hospital and ER charges.

RESULTS: Hospitalizations (RT, 1.8; UC, 3.0), hospital days (RT, 11.6; UC, 25.0), and hospital and ER charges (RT, $66,000; UC, $157,000) were significantly lower in the RT group, as were hospital and ER charges per study day (RT, $159; UC, $317).

CONCLUSIONS: The results support our previous findings, that is, home telehealth can contribute to improved health outcomes and cost of care in high-risk dialysis patients.

6. Pines, J. M.;Keyes, V.;van Hasselt, M.;McCall, N. Emergency department and inpatient hospital use by Medicare beneficiaries in patient-centered medical homes. *Ann Emerg Med.* 2015 65(6):652-60 doi:10.1016/j.annemergmed.2015.01.002

STUDY OBJECTIVE: Patient-centered medical homes are primary care practices that focus on coordinating acute and preventive care. Such practices can obtain patient-centered medical home recognition from the National Committee for Quality Assurance. We compare growth rates for emergency department (ED) use and costs of ED visits and hospitalizations (all-cause and ambulatory-care-sensitive conditions) between patient-centered medical homes recognized in 2009 or 2010 and practices without recognition.

METHODS: We studied a sample of US primary care practices and federally qualified health centers: 308 with and 1,906 without patient-centered medical home recognition, using fiscal year 2008 to 2010 Medicare fee-for-service data. We assessed average annual practice-level payments per beneficiary for ED visits and hospitalizations and rates of ED visits and hospitalizations (overall and ambulatory-care-sensitive condition) per 100 beneficiaries before and after patient-centered medical home recognition, using a difference-in-differences regression model comparing patient-centered medical homes and propensity-matched non-patient-centered medical homes.

RESULTS: Comparing patient-centered medical home with non-patient-centered medical home practices, the rate of growth in ED payments per beneficiary was $54 less for 2009 patient-centered medical homes and $48 less for 2010 patient-centered medical homes relative to non-patient-centered medical home practices. The rate of growth in all-cause and ambulatory-care-sensitive condition ED visits per 100 beneficiaries was 13 and 8 visits fewer for 2009 patient-centered medical homes and 12 and 7 visits fewer for 2010 patient-centered medical homes, respectively. There was no hospitalization effect.

CONCLUSION: From 2008 to 2010, outpatient ED visits increased more slowly for Medicare patients being treated by patient-centered medical home practices than comparison non-patient-centered medical homes. The reduction was in visits for both ambulatory-care-sensitive and non-ambulatory-care-sensitive conditions, suggesting that steps taken by practices to attain patient-centered medical home recognition such as improving care access may decrease some of the demand for outpatient ED care.

7. Green, J. A.;Mor, M. K.;Shields, A. M.;Sevick, M. A.;Arnold, R. M.;Palevsky, P. M.;Fine, M. J.;Weisbord, S. D. Associations of health literacy with dialysis adherence and health resource utilization in patients receiving maintenance hemodialysis. *Am J Kidney Dis.* 2013 62(1):73-80 doi:10.1053/j.ajkd.2012.12.014

BACKGROUND: Although limited health literacy is common in hemodialysis patients, its effects on clinical outcomes are not well understood.

STUDY DESIGN: Observational study.

SETTING & PARTICIPANTS: 260 maintenance hemodialysis patients enrolled in a randomized clinical trial of symptom management strategies from January 2009 through April 2011.

PREDICTOR: Limited health literacy.

OUTCOMES: Dialysis adherence (missed and abbreviated treatments) and health resource utilization (emergency department visits and end-stage renal disease [ESRD]-related hospitalizations).

MEASUREMENTS: We assessed health literacy using the Rapid Estimate of Adult Literacy in Medicine (REALM) and used negative binomial regression to analyze the independent associations of limited health literacy with dialysis adherence and health resource utilization over 12-24 months.

RESULTS: 41 of 260 (16%) patients showed limited health literacy (REALM score, </=60). There were 1,152 missed treatments, 5,127 abbreviated treatments, 552 emergency department visits, and 463 ESRD-related hospitalizations. Limited health literacy was associated independently with an increased incidence of missed dialysis treatments (missed, 0.6% vs 0.3%; adjusted incidence rate ratio [IRR], 2.14; 95% CI, 1.10-4.17), emergency department visits (annual visits, 1.7 vs 1.0; adjusted IRR, 1.37; 95% CI, 1.01-1.86), and hospitalizations related to ESRD (annual hospitalizations, 0.9 vs 0.5; adjusted IRR, 1.55; 95% CI, 1.03-2.34).

LIMITATIONS: Generalizability and potential for residual confounding.

CONCLUSIONS: Patients receiving maintenance hemodialysis who have limited health literacy are more likely to miss dialysis treatments, use emergency care, and be hospitalized related to their kidney disease. These findings have important clinical practice and cost implications.

8. Morgan, S. R.;Chang, A. M.;Alqatari, M.;Pines, J. M. Non-emergency department interventions to reduce ED utilization: a systematic review. *Acad Emerg Med.* 2013 20(10):969-85 doi:10.1111/acem.12219

OBJECTIVES: Recent health policy changes have focused efforts on reducing emergency department (ED) visits as a way to reduce costs and improve quality of care. This was a systematic review of interventions based outside the ED aimed at reducing ED use.

METHODS: This study was designed as a systematic review. We reviewed the literature on interventions in five categories: patient education, creation of additional non-ED capacity, managed care, prehospital diversion, and patient financial incentives. Studies written in English, with interventions administered outside of the ED, and a comparison group where ED use was an outcome, were included. Two independent reviewers screened search results using MEDLINE, Cochrane, OAIster, or Scopus. The following data were abstracted from included studies: type of intervention, study design, population, details of intervention, effect on ED use, effect on non-ED health care use, and other health and financial outcomes. Quality of individual articles was assessed using Grading of Recommendations Assessment, Development, and Evaluation (GRADE) guidelines.

RESULTS: Of 39 included studies, 34 were observational and five were randomized controlled trials. Two of five studies on patient education found reductions in ED use ranging from 21% to 80%. Out of 10 studies of additional non-ED capacity, four showed decreases of 9% to 54%, and one a 21% increase. Both studies on prehospital diversion found reductions of 3% to 7%. Of 12 studies on managed care, 10 had decreases ranging from 1% to 46%. Nine out of 10 studies on patient financial

incentives found decreases of 3% to 50%, and one a 34% increase. Nineteen studies reported effect on non-ED use with mixed results. Seventeen studies included data on health outcomes, but 13 of these only included data on hospitalizations rather than morbidity and mortality. Seven studies included data on cost outcomes. According to the GRADE guidelines, all studies had at least some risk of bias, with four moderate quality, one low quality, and 34 very low quality studies.

CONCLUSIONS: Many studies have explored interventions based outside the ED to reduce ED use in various populations, with mixed evidence. Approximately two-thirds identified here showed reductions in ED use. The interventions with the greatest number of studies showing reductions in ED use include patient financial incentives and managed care, while the greatest magnitude of reductions were found in patient education. These findings have implications for insurers and policymakers seeking to reduce ED use.

9. Cohen D, Gray K, Colson C, Van Wyck D, Tentori F, and Brunell S. Impact of Rescheduling a Missed Hemodialysis Treatment on Clinical Outcomes. Kidney Med. 2(1):12-19.Published online December 11, 2019

Rationale & Objective: Among patients treated with in-center hemodialysis (HD), missed treatments are associated with higher subsequent rates of hospitalization and other adverse outcomes compared with attending treatment. The objective of this study was to determine whether and to what degree attending a rescheduled treatment on the day following a missed treatment ameliorates these risks.

Study Design: Retrospective, observational.

Setting & Participants: Included patients were those who were, as of any of 12 index dates during 2014, adult Medicare beneficiaries treated with in-center HD (vintage ≥ 90 days) on a Monday/Wednesday/Friday schedule.

Exposure: Treatment attendance on the index date and the subsequent day.

Outcomes: Hospital admissions, emergency department visits, mortality, blood pressure, and anemia measures, considered during the 7- and 30-day periods following exposure.

Analytical Approach: In parallel analyses, patients who missed or rescheduled treatment were each matched (1:5) to patients who attended treatment on the index date on the basis of index day of week and propensity score. Within the matched cohorts, outcomes were compared across exposures using repeated-measures generalized linear models.

Results: Compared with attending treatment (N = 19,260), a missed treatment (N = 3,852) was associated with a 2.09-fold higher rate of hospitalization in the subsequent 7 days; a rescheduled treatment (N = 2,128) was associated with a 1.68-fold higher rate of hospitalization than attending (N = 10,640). Compared with attending treatment, hospitalization rates were 1.39- and 1.28-fold higher among patients who missed and rescheduled treatment, respectively, during the 30-day outcome period. Emergency department visits followed a similar pattern of associations as hospitalization. No statistically significant associations were observed with respect to mortality for either missed or rescheduled treatments compared with attending treatment.

Limitations: Possible influence of unmeasured confounding; unknown generalizability to patients with non-Medicare insurance.

Conclusions: Attending a rescheduled in-center HD treatment attenuates but does not fully mitigate the adverse effects of a missed treatment.

10. Zhang S, Morgenstern H, Albertus P, Nallamothu B, He K, and Saran R. Emergency department visits and hospitalizations among hemodialysis patients by day of the week and dialysis schedule in the United States. PLOS ONE. https://doi.org/10.1371/journal.pone.0220966 August 15, 2019.

Background and objective: Previous reports indicated that patients on thrice-weekly hemodialysis (HD) had higher mortality rates after the 3-day interdialytic interval. However, day-of-the-week patterns of emergency department (ED) visits and hospitalizations remain under-investigated.

Methods: We conducted a retrospective cohort study of HD patients on thrice-weekly dialysis, using 2013 data from the United States Renal Data System (USRDS). We estimated crude incidence rates of ED visits and hospitalizations by day of the week and dialysis schedule (Monday, Wednesday, Friday or Tuesday, Thursday, Saturday). Using Poisson regression, we estimated case-mix adjusted rate ratios of all-cause ED visits and hospitalizations, and adjusted rates of cause-specific ED visits and hospitalizations.

Results: We identified 241,093 eligible HD patients in 2013, who had 514,773 ED visits and 301,674 hospitalizations that year. Three distinct but related patterns of outcome events were observed. Crude and adjusted incidence rates of all-cause, cardiovascular, and infection-related ED visits and hospitalizations, but not vascular-access-related events, were higher on all three HD treatment days ("dialysis-day effect"). Rates for ED visits and hospitalizations were lower on weekends than weekdays, rising appreciably from Sunday to Monday for both dialysis schedules ("post-weekend effect"); and rates were highest after the long 3-day interval between dialysis sessions for both dialysis schedules ("interdialytic-gap effect"). In contrast, rates of hospitalizations not preceded by an ED visit were nearly the same Monday through Friday and lower on weekends for both dialysis schedules.

Conclusions: Higher rates of ED visits and hospitalizations on the days of HD sessions and early in the week are a public-health concern that should stimulate research to explain these patterns and reduce the excessive morbidity and associated costs among patients on thrice-weekly HD, while improving quality of care and patient experience with dialysis.

11. Marrufo G, Negrusa B, Ullman D, Hirth R, Messana J, Maughan B, Nelson J, Lindsey N, Gregory D, Svoboda R, Melin C, Chung A, Dahlerus C, Nahra T, Jiao A, McKeithen K, and Gilfix Z. Comprehensive End-Stage Renal Disease Care (CEC) Model. Performance Year 2 Annual Evaluation Report. Prepared for: Centers for Medicare & Medicaid Services. September 2019. https://innovation.cms.gov/Files/reports/cec-annrpt-py2.pdf

1a.3. SYSTEMATIC REVIEW(SR) OF THE EVIDENCE (for INTERMEDIATE OUTCOME, PROCESS, OR STRUCTURE PERFORMANCE MEASURES) If the evidence is not based on a systematic review go to section 1a.4) If you wish to include more than one systematic review, add additional tables.

What is the source of the systematic review of the body of evidence that supports the performance measure? A systematic review is a scientific investigation that focuses on a specific question and uses explicit, prespecified scientific methods to identify, select, assess, and summarize the findings of similar but separate studies. It may include a quantitative synthesis (meta-analysis), depending on the available data. (IOM)

☐ Clinical Practice Guideline recommendation (with evidence review)

☐ US Preventive Services Task Force Recommendation

☐ Other systematic review and grading of the body of evidence (*e.g., Cochrane Collaboration, AHRQ Evidence Practice Center*)

☐ Other

Source of Systematic Review: • **Title** • **Author** • **Date** • **Citation, including page number** • **URL**	**N/A**
Quote the guideline or recommendation verbatim about the process, structure or intermediate outcome being measured. If not a guideline, summarize the conclusions from the SR.	N/A
Grade assigned to the **evidence** associated with the recommendation with the definition of the grade	N/A
Provide all other grades and definitions from the evidence grading system	N/A
Grade assigned to the **recommendation** with definition of the grade	N/A
Provide all other grades and definitions from the recommendation grading system	N/A
Body of evidence: • Quantity – how many studies? • Quality – what type of studies?	N/A
Estimates of benefit and consistency across studies	N/A
What harms were identified?	N/A
Identify any new studies conducted since the SR. Do the new studies change the conclusions from the SR?	N/A

1a.4 OTHER SOURCE OF EVIDENCE

If source of evidence is NOT from a clinical practice guideline, USPSTF, or systematic review, please describe the evidence on which you are basing the performance measure.

1a.4.1 Briefly SYNTHESIZE the evidence that supports the measure. A list of references without a summary is not acceptable.

N/A

1a.4.2 What process was used to identify the evidence?

N/A

1a.4.3. Provide the citation(s) for the evidence.

N/A

1b. Performance Gap

Demonstration of quality problems and opportunity for improvement, i.e., data demonstrating:

- considerable variation, or overall less-than-optimal performance, in the quality of care across providers; and/or
- Disparities in care across population groups.

1b.1. Briefly explain the rationale for this measure *(e.g., how the measure will improve the quality of care, the benefits or improvements in quality envisioned by use of this measure)*

If a COMPOSITE (e.g., combination of component measure scores, all-or-none, any-or-none), SKIP this question and answer the composite questions.

Emergency department (ED) encounters are an important indicator of care coordination and quality of life. In the general population studies have shown higher risk of an emergency department encounter subsequent to a discharge from an inpatient hospitalization or an outpatient emergency department encounter (e.g., see Hastings et al., 2008).

Rates of ED visits among end-stage renal disease (ESRD) dialysis patients have increased between 2007 and 2016. As reported by the USRDS, the unadjusted ED visit rate among HD patients increased from 2.6 to 3.0 per patient-year, and from 2.2 to 2.4 per patient-year for PD patients (USRDS ADR 2018), while the national percentage of ED visits among dialysis patients is 62% as of 2018 (FY2020 Dialysis Facility Report). More than half (55.0%) of all patients with ESRD visit the ED during their first year of dialysis, and patients with ESRD have a mean of 2.7 ED visits per patient-year (Lovasik et al., 2016). This rate is 6-fold higher than the national mean rates for US adults in the general population (Lovasik et al 2016). Furthermore, the Lovasik study notes that among Medicare beneficiaries with ESRD, 30% of hospital admissions that originate in the ED are for diagnoses that are often dialysis related such as complications of vascular access, congestive heart failure/fluid overload, septicemia, and hyperkalemia. A study by Zhang and colleagues (Zhang et al, 2019) reported that rates of ED visits among patients on thrice weekly in-center hemodialysis vary by dialysis schedule (Mon/Weds/Fri; Tues/Thurs/Sat) and by day of week. For example the ED visit rate (without hospital admission) was highest on the day following the longer interdialytic interval over the weekend (Mondays), suggesting an association with facility structure and treatment schedule.

Cohen and colleagues (Cohen et al 2020) reported that missed dialysis treatments are associated with an over two-fold higher risk of an ED visit, suggesting an opportunity for dialysis facilities to establish or strengthen facility practices that can help to reduce skipped treatments through increased communication, care coordination, and patient education. This in turn has the potential to reduce avoidable ED visits.

Finally, the CMS Centers for Medicare and Medicaid Innovation's Comprehensive End Stage Renal Disease (ESRD) Care model emphasizes care coordination as a central feature of care delivery in order to reduce utilization and improve outcomes. During the second performance year, the original Wave 1 cohort of ESCOs (ESRD Seamless Care Organizations) experienced about a 3% reduction in ED use relative to the period before the CEC model was launched (Marrufo et al., CEC Annual Report Performance Year 2, 2019).

Measures of the frequency of ED use may help dialysis facility level efforts to prevent emergent unscheduled care and control escalating medical costs.

References:

Hastings NS., Oddone EZ., Fillenbaum G, Shane R J., Schmader KE. Frequency and predictors of adverse health outcomes in older Medicare beneficiaries discharged from the emergency department. Med Care. 2008 Aug; 46(8):771-7

Lovasik BP, Zhang R, Hockenberry JM, Schrager JD, Pastan SO, Mohan S, Patzer RE. Emergency Department Use and Hospital Admissions Among Patients With End-Stage Renal Disease in the United States. JAMA Intern Med. 2016 Oct 1; 176(10):1563-1565.

United States Renal Data System. 2018 USRDS annual data report: Epidemiology of kidney disease in the United States. National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases, Bethesda, MD, 2018.

Dialysis Facility Reports –Sample Report FY2020. https://www.dialysisdata.org/sites/default/files/content/Methodology/DFR_SAMPLE_201907.pdf

Zhang S, Morgenstern H, Albertus P, Nallamothu B, He K, and Saran R. Emergency department visits and hospitalizations among hemodialysis patients by day of the week and dialysis schedule in the United States. PLOS ONE. https://doi.org/10.1371/journal.pone.0220966 August 15, 2019.

Marrufo G, Negrusa B, Ullman D, Hirth R, Messana J, Maughan B, Nelson J, Lindsey N, Gregory D, Svoboda R, Melin C, Chung A, Dahlerus C, Nahra T, Jiao A, McKeithen K, and Gilfix Z. Comprehensive End-Stage Renal Disease Care (CEC) Model. Performance Year 2 Annual Evaluation Report. Prepared for: Centers for Medicare & Medicaid Services. September 2019. https://innovation.cms.gov/Files/reports/cec-annrpt-py2.pdf

1b.2. Provide performance scores on the measure as specified (current and over time) at the specified level of analysis. *(This is required for maintenance of endorsement. Include mean, std dev, min, max, interquartile range, scores by decile. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities include.) This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

We calculated SEDR for each year 2014-2017 (Table 1). We included all Medicare-certified dialysis facilities with eligible time at risk for the measure. We excluded transplant-only facilities and Veteran Affairs (VA) facilities. The distribution of the SEDR for each year is shown in the table below (restricted to facilities with at least 5 patient years at risk). Standardized ED Visit rates vary widely across facilities. For example, for the 6,691 facilities included in 2017, the SEDR varied from 0.00 to 4.30 (Table 1). For each year the mean value was 1.00 and the SD was 0.34 to 0.35. The second table (Table 2) shows the deciles of the SEDR for 2017.

Table 1.SEDR Performance Score Descriptives, 2014-2017

2014: Number of facilities=6,056, Number of patients=371,677, Mean=1.00, Std Dev=0.35, Min=0.00, Max=3.64

2015: Number of facilities=6,251, Number of patients=374,473, Mean=1.00, Std Dev=0.35, Min=0.00, Max=6.15

2016: Number of facilities=6,435, Number of patients=379,138, Mean=1.00, Std Dev=0.34, Min=0.00, Max=3.77

2017: Number of facilities=6,691, Number of patients=382,039, Mean=1.00, Std Dev=0.34, Min=0.00, Max=4.30

Table 2. Deciles of Standardized ED Visit Ratio, 2017

Decile 1: N=669, Min=0, Max=0.62

Decile 2: N=669, Min=0.62, Max=0.73

Decile 3: N=669, Min=0.73, Max=0.81

Decile 4: N=669, Min=0.81, Max=0.89

Decile 5: N=669, Min=0.89, Max=0.96

Decile 6: N=670, Min=0.96, Max=1.04

Decile 7: N=669, Min=1.04, Max=1.13

Decile 8: N=669, Min=1.13, Max=1.24

Decile 9: N=669, Min=1.24, Max=1.41

Decile 10: N=669, Min=1.41, Max=4.30

1b.3. If no or limited performance data on the measure as specified is reported in 1b2, then provide a summary of data from the literature that indicates opportunity for improvement or overall less than optimal performance on the specific focus of measurement.

N/A

1b.4. Provide disparities data from the measure as specified (current and over time) by population group, e.g., by race/ethnicity, gender, age, insurance status, socioeconomic status, and/or disability. *(This is required for maintenance of endorsement. Describe the data source including number of measured entities; number of patients; dates of data; if a sample, characteristics of the entities included.) For measures that show high levels of performance, i.e., "topped out", disparities data may demonstrate an opportunity for improvement/gap in care for certain sub-populations. This information also will be used to address the sub-criterion on improvement (4b1) under Usability and Use.*

Our results indicate potential differences in emergency department utilization. Differences are observed by age, sex (females), race (Blacks, Native Americans, Asian/PI), dual Medicare-Medicaid status, and employment status (unemployed). For example, the risk of an emergency room visit decreases with age until about 75 years of age and then increases. Females had a higher risk of an emergency department encounter compared to males (38% higher). Black patients also had a higher risk (16% higher) of an emergency department visit compared to whites, as do Native Americans (4% higher). However, Asian/Pacific Islander patients had a lower risk (15% lower). Hispanic patients had a higher risk (3%) of an emergency department encounter compared to non-Hispanic patients. Patients who were unemployed (at ESRD incidence) had a 14% higher risk of an emergency department encounter, compared to employed patients (employed at ESRD incidence). Finally, patients dually covered by Medicare and Medicaid had a 23% higher risk of an emergency department encounter compared to patient with Medicare only. While there are notable differences by younger age, race, Hispanic ethnicity, sex and insurance status, it is unclear if these disparities in emergency department encounters are based on different clinical risk factors for these subgroups or differences in care quality. Refer to Risk Adjustment section (2b4) for further analyses on race, ethnicity, sex and socioeconomic status.

Age:

Age, Hazard Ratio = 0.97, p<0.0001.

Age Squared, Hazard Ratio = 1.0002, p<0.0001.

Sex:

For Female: Hazard Ratio = 1.38, p<0.0001.

Male was used as the reference group.

Race:

White was used as the reference group.

For Black: Hazard Ratio =1.16, p<0.0001.

For Native Americans: Hazard Ratio =1.04, p<0.0001.

For Asian/PI: Hazard Ratio =0.85, p<0.0001.

For Other race: Hazard Ratio = 1.03, p-value =0.06

Ethnicity:

Non-Hispanic/Unknown Hispanic was used as the reference group.

For Hispanic: Hazard Ratio = 1.03, p-value =<0.0001.

Employment Status:

Employed was used as the reference group.

For Unemployed: Hazard Ratio =1.14, p<0.0001.

For Other/Unknown*: Hazard Ratio =1.10, and the p<0.0001.

* Other/Unknown group includes Homemaker, Retired due to age/preference, retired due to disability, Medical leave of absence, or missing employment status.

Medicare Coverage:

Non-dual eligible was used as the reference group.

Dual eligible: Hazard Ratio = 1.23, and the p-value <0.0001.

Area-Level SES/SDS:

Area Deprivation Index Value, Hazard Ratio = 1.002 p<0.0001.

1b.5. If no or limited data on disparities from the measure as specified is reported in 1b.4, then provide a summary of data from the literature that addresses disparities in care on the specific focus of measurement. Include citations. Not necessary if performance data provided in 1b.4

N/A

2.3 For maintenance of endorsement

Risk adjustment: For outcome, resource use, cost, and some process measures, risk-adjustment that includes social risk factors is not prohibited at present. Please update sections 1.8, 2a2, 2b1,2b4.3 and 2b5 in the Testing attachment and S.140 and S.11 in the online submission form. NOTE: These sections must be updated even if social risk factors are not included in the risk-adjustment strategy. You MUST use the most current version of the Testing Attachment (v7.1) -- older versions of the form will not have all required questions.

Measure Testing (subcriteria 2a2, 2b1-2b6)

Measure Number (*if previously endorsed*)**:** Click here to enter NQF number

Measure Title: Standardized Emergency Department Encounter Ratio for Dialysis Facilities

Date of Submission: 1/5/2020

Type of Measure:

1. DATA/SAMPLE USED FOR ALL TESTING OF THIS MEASURE

Often the same data are used for all aspects of measure testing. In an effort to eliminate duplication, the first five questions apply to all measure testing. If there are differences by aspect of testing,(e.g., reliability vs. validity) be sure to indicate the specific differences in question 1.7.

1.1. What type of data was used for testing? (*Check all the sources of data identified in the measure specifications and data used for testing the measure. Testing must be provided for all the sources of data specified and intended for measure implementation.* ***If different data sources are used for the numerator and denominator, indicate N [numerator] or D [denominator] after the checkbox.***)

Measure Specified to Use Data From: (*must be consistent with data sources entered in S.17*)	Measure Tested with Data From:
☐ abstracted from paper record	☐ abstracted from paper record
☒ claims	☒ claims
☒ registry	☒ registry
☐ abstracted from electronic health record	☐ abstracted from electronic health record
☐ eMeasure (HQMF) implemented in EHRs	☐ eMeasure (HQMF) implemented in EHRs

☐ other: Click here to describe	☐ other: Click here to describe

1.2. If an existing dataset was used, identify the specific dataset (*the dataset used for testing must be consistent with the measure specifications for target population and healthcare entities being measured; e.g., Medicare Part A claims, Medicaid claims, other commercial insurance, nursing home MDS, home health OASIS, clinical registry*).

Data are derived from an extensive national ESRD patient database, which is primarily based on CROWNWeb facility-reported clinical and administrative data (including CMS-2728 Medical Evidence Form, CMS-2746 Death Notification Form, and CMS-2744 Annual Facility Survey Form and patient tracking data), the Renal Management Information System (REMIS), the Medicare Enrollment Database (EDB), and Medicare claims data. In addition the database includes transplant data from the Scientific Registry of Transplant Recipients (SRTR), and data from the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Business Intelligence Center (QBIC) (which includes Provider and Survey and Certification data from Automated Survey Processing Environment (ASPEN)), and the Dialysis Facility Compare (DFC).

The database is comprehensive for Medicare patients not enrolled in Medicare Advantage. Medicare Advantage patients are included in all sources but their Medicare payment records are limited to inpatient claims. Non-Medicare patients are included in all sources except for the Medicare payment records. Tracking by dialysis provider and treatment modality is available for all patients including those with only partial or no Medicare coverage.

Information on hospitalizations is obtained from Part A Medicare Inpatient Claims Standard Analysis Files (SAFs), and past-year comorbidity data are obtained from multiple Part A types (inpatient, home health, hospice, skilled nursing facility claims) only.

1.3. What are the dates of the data used in testing? January 2014-December 2017

1.4. What levels of analysis were tested? (*testing must be provided for <u>all</u> the levels specified and intended for measure implementation, e.g., individual clinician, hospital, health plan*)

Measure Specified to Measure Performance of: (*must be consistent with levels entered in item S.20*)	Measure Tested at Level of:
☐ individual clinician	☐ individual clinician
☐ group/practice	☐ group/practice
☒ hospital/facility/agency	☒ hospital/facility/agency
☐ health plan	☐ health plan
☐ other: Click here to describe	☐ other: Click here to describe

1.5. How many and which <u>measured entities</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of measured entities included in the analysis (e.g., size, location, type); if a sample was used, describe how entities were selected for inclusion in the sample*)

Table 1. Number of facilities and median facility size by year

Year	Total Facilities	Total Patients	Median Patients Per Facility

2014	6,816	372,826	54
2015	6,992	375,586	53
2016	7,259	380,423	52
2017	7,550	383,414	51

1.6. How many and which <u>patients</u> were included in the testing and analysis (by level of analysis and data source)? (*identify the number and descriptive characteristics of patients included in the analysis (e.g., age, sex, race, diagnosis); if a sample was used, describe how patients were selected for inclusion in the sample*)

Medicare dialysis patients were included in the testing and analysis for each of the four years from 2014-2017 of which there were 372,826, 375,586, 380,423, and 383,414 patients respectively.

Table 2. Descriptives of Patient Characteristics Included in the Measure

Patient Demographics	Percent
Age	
Patient Age: 18-24	0.7
Patient Age: 25-44	11.2
Patient Age: 45-59	26.8
Patient Age: 60-74	39.4
Patient Age: 75+	22.0
Sex (% female)	43.8
ESRD due to Diabetes (%)	46.5
Medicare coverage(%)	
Medicare primary + Medicaid	39.8
Medicare primary + no Medicaid	48.5
Medicare secondary/Other	11.7
Time since Start of ESRD	
91 days-6 months	11.1
6 months-1 year	13.3
1-2 years	16.6
2-3 years	14.8
3-5 years	18.4
5+ years	25.7
Employment status 6 months prior to ESRD (%)	
Unemployed	22.9
Employed	19.7
Other/Unknown *	57.4
Race (%)	
White	59.2

Patient Demographics	Percent
Black	33.1
Asian/Pacific Islander	4.8
Native American/Alaskan Native	1.3
Other/Unknown	1.6
Ethnicity (%)	
Hispanic	15.6
Non-Hispanic/Unknown	84.4

* Other/Unknown groups includes Homemaker, Retired due to age/preference, retired due to disability, Medical leave of absence, or missing employment status. Note: Some categories do not sum to 100% due to rounding.

1.7. If there are differences in the data or sample used for different aspects of testing (e.g., reliability, validity, exclusions, risk adjustment), identify how the data or sample are different for each aspect of testing reported below.

N/A

1.8 What were the social risk factors that were available and analyzed? For example, patient-reported data (e.g., income, education, language), proxy variables when social risk data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate) which do not have to be a proxy for patient-level data.

Patient level:

- Employment status 6 months prior to ESRD
- Sex
- Race
- Ethnicity
- Medicare dual eligible
- ZIP code level – Area Deprivation Index (ADI) from Census data (2009-2013). Based on patient zip-code.

Data on patient level SDS/SES factors obtained from Medicare claims and administrative data.

2a2. RELIABILITY TESTING

***Note**: If accuracy/correctness (validity) of data elements was empirically tested, separate reliability testing of data elements is not required – in 2a2.1 check critical data elements; in 2a2.2 enter "see section 2b2 for validity testing of data elements"; and skip 2a2.3 and 2a2.4.*

2a2.1. What level of reliability testing was conducted? (*may be one or both levels*)

☐ **Critical data elements used in the measure** (*e.g., inter-abstractor reliability; data element reliability must address ALL critical data elements*)

☒ **Performance measure score** (e.g., *signal-to-noise analysis*)

2a2.2. For each level checked above, describe the method of reliability testing and what it tests (*describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used*)

The reliability of the Standardized Emergency Department Encounter Ratio (SEDR) was assessed using data among eligible Medicare ESRD dialysis patients during 2014-2017. If the measure were a simple average across individuals in the facility, the usual approach for determining measure reliability would be a one-way analysis of variance (ANOVA), in which the between and within facility variation in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of a measure that is attributable to the between-facility variation. The SEDR, however, is not a simple average and we instead estimate the IUR using a bootstrap approach, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by ANOVA.

Here we describe our approach to calculating IUR. Let T1,…,TN be the SEDR for these facilities. Within each facility, select at random and with replacement B bootstrap samples. Our numerical experiments reveal that B=100 is sufficient. That is, if the ith facility has ni subjects, randomly draw with replacement ni subjects from those in the same facility, find their corresponding SEDRi and repeat the process B (say, 100) times. Thus, for the ith facility, we have bootstrapped SEDRs of T_i1^*,…, T_i200^*. Let S_i^* be the sample variance of this bootstrap sample. From this it can be seen that

$$s_{t,w}^2 = \frac{\sum_{i=1}^{N}[(n_i - 1)S_i^{*2}]}{\sum_{i=1}^{N}(n_i - 1)}$$

is a bootstrap estimate of the within-facility variance in the SEDR, namely, $\sigma_{t,w}^2$. Calling on formulas from the one way analysis of variance, an estimate of the overall variance of T_i is

$$s_t^2 = \frac{1}{n'(N-1)} \sum_{i=1}^{N} n_i (T_i - \bar{T})^2$$

where

$$\bar{T} = \sum n_i T_i / \sum n_i$$

is the weighted mean of the observed SEDR and

$$n' = \frac{1}{N-1} \left(\sum n_i - \sum n_i^2 \Big/ \sum n_i \right)$$

is approximately the average facility size (number of patients per facility). Note that s_t^2 is the total variation of SEDR and is an estimate of $\sigma_b^2 + \sigma_{t,w}^2$, where σ_b^2 is the between-facility variance, the true signal reflecting the differences across facilities. Thus, the estimated IUR, which is defined by

$$IUR = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_{t,w}^2}$$

can be estimated with $(s_t^2 - s_{t,w}^2)/s_t^2$.

The measure calculation is only reported for facilities with at least 5 patient years at risk.

To assess more directly the value of SEDR in identifying facilities with extreme outcomes, we also computed an additional metric of reliability, termed the profile IUR (PIUR) [1]. The PIUR was developed since the IUR can be quite small if there are many facilities which have outcomes similar to the national norm, even though the measure is still very useful to identify facilities with extreme outcomes [2]. The PIUR is based on the measure's ability to consistently flag the same facilities. We proceed in two steps: first, we evaluate the ability of a measure to consistently profile facilities with extreme outcomes; second, we use the IUR to calibrate PIUR. Specifically, we consider a sample-splitting approach: within each facility randomly split patients into two equal-sized subgroups. For a given threshold (e.g. p-value or z-score in a hypothesis testing procedure), determine whether each facility is identified as extreme based on the first and the second subgroups. Repeat this process 100 times to estimate the probability that, given a facility is classified as extreme based on the first subgroup, it is also classified as extreme based on the second subgroup. This empirical reflagging rate is calibrated to give the PIUR by determining the IUR value that would yield this reflagging rate in the absence of outliers. The PIUR measures reliability in terms of the probability of reflagging rates but is on the same scale as IUR. The PIUR is substantially larger than the IUR when the data include many outliers or extreme values that are not captured in the IUR itself.

1. 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]

2. Kalbfleisch JD, He K, Xia L, Li Y. Does the inter-unit reliability (IUR) measure reliability?, Health Services and Outcomes Research Methodology, 2018 Sept. 18(3), 215-225. Doi: 10.1007/s10742-018-0185-4.

3. He K, Kalbfleisch JD, Yang Y, Fei Z. Inter-unit reliability for nonlinear models. Stat Med. 2019 Feb 28;38(5):844-854. doi: 10.1002/sim.8005. Epub 2018 Oct 18.

2a2.3. For each level of testing checked above, what were the al results from reliability testing? (e.*g., percent agreement and kappa for the critical data elements; distribution of reliability statistics from a signal-to-noise analysis*)

Table 3: IUR for one-year SEDR, 2014-2017

Year	IUR	PIUR	N
2014	0.63	0.89	6,056
2015	0.63	0.91	6,251
2016	0.62	0.90	6,433
2017	0.62	0.89	6,691

As noted above, the PIUR measures reliability in terms of reflagging rates but is placed on the same scale as IUR. The higher PIUR compared to the IUR indicates the presence of outliers or heavier tails among the providers, which is not captured in the IUR itself. If there are no outliers, one should expect the PIUR to be similar to the IUR; but in cases where there are outlier providers, even measures with a low IUR can have relatively high PIUR and can be very useful for identifying extreme providers.

2a2.4 What is your interpretation of the results in terms of demonstrating reliability? (i.*e., what do the results mean and what are the norms for the test conducted?*)

The value obtained for the IUR is moderate in size. The PIUR is larger and demonstrates that the SEDR is effective at detecting outlier facilities and statistically meaningful differences in performance scores across dialysis facilities.

2b1. VALIDITY TESTING

2b1.1. What level of validity testing was conducted? (*may be one or both levels*)

☐ **Critical data elements** (*data element validity must address ALL critical data elements*)

☒ **Performance measure score**

☒ **Empirical validity testing**

☒ **Systematic assessment of face validity of <u>performance measure score</u> as an indicator** of quality or resource use (*i.e., is an accurate reflection of performance on quality or resource use and can distinguish good from poor performance*) **NOTE**: Empirical validity testing is expected at time of maintenance review; if not possible, justification is required.

2b1.2. For each level of testing checked above, describe the method of validity testing and what it tests (*describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)*

Face Validity: In May 2016, we presented a preliminary version of the SEDR measure to a CMS Technical Expert Panel (TEP) for clinical validity. The nine member TEP was composed of clinical nephrologists, ED physicians, a renal nurse, and ESRD patients. The TEP discussions were informed by a review of relevant literature and related ED and hospital measures as part of the environmental scan we prepared for the TEP. Potential measures were evaluated using the criteria for clinical performance measures adopted by the National Quality Forum (NQF) and CMS (importance, scientific acceptability, feasibility, and usability). During the discussion, the TEP considered:

- Relevant measures endorsed by the National Quality Forum (NQF), or reported in the Dialysis Facility Reports (DFRs)
- Components of a potential ED measure, such as the location of the patient prior to the ED encounter, the method by which the patient was directed to the ED, presenting complaint, severity of illness, and outcome of the ED encounter
- The degree to which performance on a measure is under control of the dialysis facility
- The potential need for exclusion criteria and/or risk adjustment
- Data availability and additional analyses

The TEP discussed different ED outcomes and recommended limiting an ED encounter measure to visits that do not result in an inpatient admission because ED visits resulting in hospitalization are already captured through the respective NQF endorsed Standardized Hospitalization Ratio (SHR) for Admissions and the Standardized Readmission Ratio (SRR) for dialysis facilities measures. In addition, the TEP agreed that

observation stays should be included in an ED measure. Ultimately, the TEP indicated that ED encounters that do not result in admission are not well monitored as a quality indicator and panelists believed this measure would provide facilities with a more complete picture of their performance on key clinical outcomes of mortality, hospitalization, readmission, and ED usage. The TEP consensus supported the clinical validity of the measure. Finally, in June 2017 a final model that included extensive risk adjustment for prevalent comorbidities was presented to the TEP for review. The TEP voted unanimously in support of the final fully risk adjusted SEDR measure. See the section on risk adjustment for further detail on prevalent comorbidity risk adjustment.

Empirical validity testing - validation of performance measure scores:

To validate SEDR we first stratified facilities into the 'better than/as expected' and 'worse than expected' categories of SEDR. Next we calculated mean performance scores for several quality measures: Standardized Mortality Ratio (SMR), Standardized Transfusion Ratio (STrR), Standardized Fistula Rate (SFR), Percentage of Prevalent Patients Waitlisted (PPPW), Standardized Hospitalization Ratio (SHR), and Emergency Department Visit within 30 days of discharge (ED30). We then compared mean performance scores across the two strata of 'better than/as expected' 'and 'worse than expected' categories for SEDR. Statistically significant outliers (i.e., better and worse than expected) were determined using the method described in section 2b4.1 to flag facilities as better than expected and worse than expected based on the national average, at the $p<0.05$ level.

We expect better mean performance on the above quality measures for facilities classified as 'better than/as expected' for SEDR compared to facilities classified as 'worse than expected.' Compared to facilities that perform 'worse than expected', facilities that perform 'better than/as expected' on SEDR are likely to have more successful care coordination and other processes of care in place that may help patients avoid an ED visit:

- SMR: We expect to observe a lower mean standardized mortality ratio for facilities in the 'better than/as expected' category for SEDR compared to facilities classified as 'worse than expected.' Facilities with a higher rate of ED utilization may not have care processes in place to support management of acute care.
- STrR: We expect to observe a lower mean standardized transfusion event ratio for facilities in the 'better than/as expected' category for SEDR compared to facilities classified as 'worse than expected.' Facilities that have a lower STrR likely have processes of care in place to support robust anemia management and other care processes compared to facilities with a higher STrR.
- Standardized Fistula Rate (SFR): We expect to observe a higher mean standardized fistula rate for facilities in the 'better than/as expected' category for SEDR compared to facilities classified as 'worse than expected.' AVFs are typically considered to be the preferred vascular access due to lower risk of infection and potential need for hospitalization or other acute care. Higher standardized fistula rates suggests facilities are successful at creating AVFs due to more robust processes to coordinate care outside of the dialysis facility. Facilities that do a better job at care coordination reduce the likelihood that patients will experience a preventable and unscheduled acute event resulting in an ED visit.
- PPPW: We expect to observe a higher mean standardized percentage of prevalent patients on the waitlist for facilities in the 'better than/as expected' category for SEDR compared to facilities classified as 'worse than expected.' Facilities that have a higher standardized percentage of patients on the transplant waitlist suggest they may have more robust processes to coordinate care outside of the dialysis facility with other providers and the transplant center, compared to facilities with lower percentages. This includes the facility taking steps to ensure patients maintain sufficient health status in order to be placed on the waitlist. Therefore, facilities that have higher standardized waitlist percentages are likely deploying effective care coordination and other care processes that may reduce the likelihood of patients getting preventable and unscheduled acute care from the ED.

- SHR: We expect that facilities classified as 'worse than expected' for SEDR will have a standardized hospitalization ratio that is close to the national norm. SEDR only captures outpatient ED visits that do not result in an admission which, by definition, is a different patient subpopulation than SHR. Patients that require acute care from the ED without an admission likely have lower acuity medical needs that can be handled in an outpatient setting without admission. Therefore we do not expect SEDR flagging to be related to how facilities perform on SHR.
- ED30: We expect to observe a lower mean ED30 ratio for facilities classified as 'better than/as expected' for SEDR compared to facilities classified as 'worse than expected' since both measures are a reflection of outpatient ED use. However the measures represent two different aspects of dialysis patients' emergency department use that assess complementary elements of facility care. A low SEDR, corresponding to low overall emergency department encounter rates, indicates that the facility has processes (e.g. patient/staff education, assistance with primary care, frequent evaluation of target weight) in place to avoid the need for unscheduled acute care. A low ED30 indicates that a facility is successful in managing the transition of care (e.g. medication reconciliation, evaluation of target weight, assistance with follow up appointments) that occurs after a hospital discharge. [Note, ED30 is a companion measure to SEDR and is also being submitted to NQF]

2b1.3. What were the statistical results from validity testing? (*e.g., correlation; t-test*)

The results below show mean quality measure performance scores for SMR, STrR, SFR, PPPW, SHR, and ED30 by classification of facilities as 'better than/as expected' versus 'worse than expected' for SEDR (Table 4).

Table 4. Classification of SEDR and mean facility performance scores for Related Measures, 2017

		SEDR Classification		
Measure	Facilities Missing	Better than /As Expected	Worse than Expected	As Hypothesized?
SMR	310	1.00	1.08	Yes
STrR	619	0.98	1.14	Yes
SFR	395	63.49	62.12	Yes
PPPW	161	19.59	14.07	Yes
SHR	163	0.99	1.01	Yes
ED30	92	1.00	1.46	Yes

2b1.4. What is your interpretation of the results in terms of demonstrating validity? (i.*e., what do the results mean and what are the norms for the test conducted?*)

On average the standardized mortality ratio was 8% higher than the national average for facilities that were 'worse than expected,' and no different from the national average (SMR = 1.00) for facilities that were 'better than/as expected' for SEDR.

On average the standardized transfusion event ratio was 14% higher than the national average for facilities classified as 'worse than expected'. This suggests that facilities which have lower numbers of transfusion events likely have better processes of care in place to support robust anemia management and other care processes, thus reducing patient utilization of the ED for some acute care needs.

Overall the average the SFR was 63.49% in facilities classified as 'better than/as expected' and 62.12% in facilities classified as 'worse than expected.' The results reinforce the observation that patients with AVFs have lower risk of infection and potential need for acute care or hospitalization compared to patients with other access types, such as long-term catheter. Higher facility standardized fistula rates suggests facilities may be doing a better job at care coordination, reducing patient utilization of the ED for many acute care needs. While the difference in fistula rates was small between facilities this may reflect that national trends in AVF rates have generally plateaued across many US dialysis facilities.

The mean facility standardized percentage of patients waitlisted (PPPW) in facilities classified as 'better than/as expected' was 19.59% compared to facilities classified as 'worse than expected' (14.07%), suggesting that facilities that have higher rates of patients on the transplant waitlist may have more robust processes to coordinate care outside of the dialysis facility with other providers. These facilities are likely deploying more effective care coordination and other care processes that may reduce the likelihood of patients utilizing the ED for many acute care needs.

Facilities classified as 'better than/as expected' and those classified as 'worse than expected' for SEDR performed similarly on SHR. The mean SHR was 0.99 for 'better than/as expected' and 1.01 for facilities classified as 'worse than expected' for SEDR indicating flagging for SEDR is not related to how facilities perform on SHR. This suggests that both measures are capturing different patient subpopulations and different facets of facility care quality.

The ED30 ratio on average for facilities classified as 'better than/as expected' for SEDR was the same as national average (1.00), while facilities classified as 'worse than expected' had an ED30 ratio 46% higher than the national average. These results reinforce that both ED30 and SEDR assess complementary elements of care that are likely reflected by internal processes that support greater access to care and other clinical triaging of patients that may be experiencing onset of an acute event, which may help reduce patient utilization of the ED for preventable acute care needs.

Taken together these results provide validation support for SEDR. Performance on key quality measures that were expected to be related to ED use was also related to facility flagging in the respective 'better than/as expected' or 'worse than expected' categories.

2b2. EXCLUSIONS ANALYSIS

NA ☐ no exclusions — *skip to section 2b3*

2b2.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

We calculated a Pearson correlation to assess the association between the SEDR measure with and without the hospice exclusion. Additionally, we calculated the number and percentage of patient years at risk, and ED visits excluded for patients actively enrolled in Hospice.

Exclusions that are implicit in the denominator definition include patient time at risk in which the patient:

- Has Medicare Advantage coverage
- Has had ESRD for 90 days or less
- Is less than 18 years of age

We also exclude patient time at risk where the patient was:

- Actively enrolled in hospice during the calendar month of the ED encounter

Based on input from the May 2016 TEP, we additionally excluded pediatric patients, hospice patients, and patients in their first 90 days of ESRD treatment. A majority of TEP members proposed excluding pediatric patients due to substantial differences in both the pediatric population comorbidities as well as reasons for seeking care in the ED when compared to the adult population. Hospice patients were excluded to allow for ED care that may be palliative in nature and directed by other providers outside of the dialysis facility. These concerns are relevant in the context of the measure's potential applications, which are to identify poor-performing facilities for quality improvement purposes.

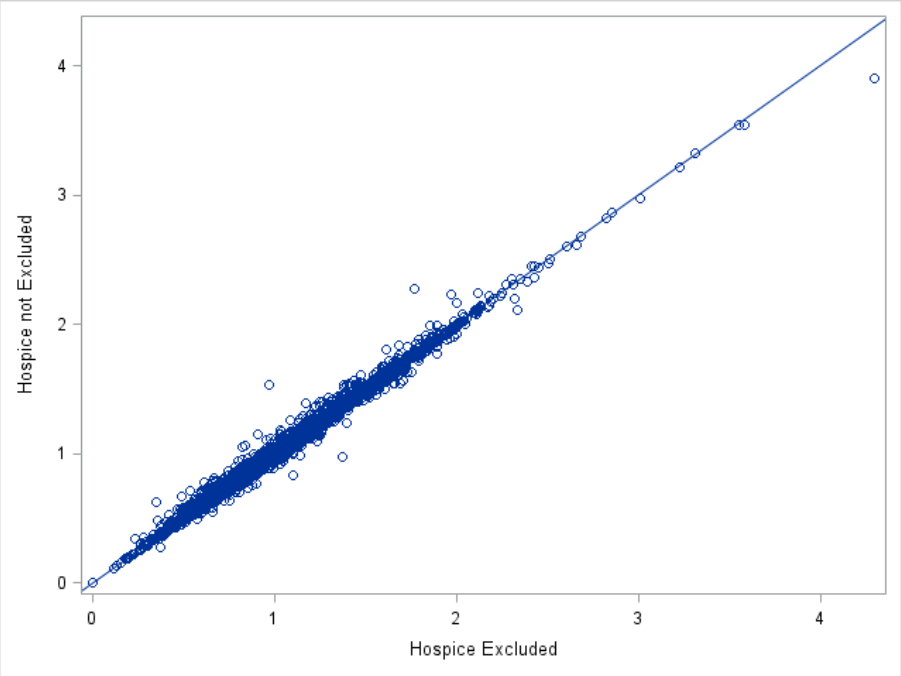
Medicare Advantage patients are excluded from SEDR as outpatient ED claims are not available, therefore we do not have information on ED utilization for this subpopulation of Medicare patients. See section 2b.6 for further discussion and assessment of missing data due to absence of outpatient ED and other outpatient claims information for MA patients.

2b2.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 1,294 patient years at risk excluded in 2017 due to active enrollment in hospice, which represents 0.43% of total years before the exclusion. This excludes 4,119 (0.82%) of ED visits during this time period (2017).

As shown in Figure 1, we compared each facility's SEDR with and without the hospice exclusion and found the two measures to be highly correlated (overall Pearson correlation coefficient [r] =0.9962, p<0.0001).

Figure 1. Correlation between SEDR with and without the hospice exclusion (2017)



2b2.3. What is your interpretation of the results in terms of demonstrating that exclusions are needed to prevent unfair distortion of performance results? (*i.e., the value outweighs the burden of increased data collection and analysis. Note: **If patient preference is an exclusion**, the measure must be specified so that the effect on the performance score is transparent, e.g., scores with and without exclusion*)

The measure with and without the hospice exclusion criteria is highly correlated suggesting the overall impact on the measure's validity in not substantial. However, this exclusion is necessary to account for any differences in the proportion of hospice patients between facilities.

2b3. RISK ADJUSTMENT/STRATIFICATION FOR OUTCOME OR RESOURCE USE MEASURES

If not an intermediate or health outcome, or PRO-PM, or resource use measure, skip to section 2b4.

2b3.1. What method of controlling for differences in case mix is used?

☐ **No risk adjustment or stratification**

☒ **Statistical risk model with 86 risk factors**

☐ **Stratification by** Click here to enter number of categories **risk categories**

☐ **Other,** Click here to enter description

2b3.1.1 If using a statistical risk model, provide detailed risk model specifications, including the risk model method, risk factors, coefficients, equations, codes with descriptors, and definitions.

- Patient age: Age (continuous); Age squared
- Sex
- Diabetes as cause of ESRD
- ESRD duration: categorized as 91 days-6 months, 6 months-1 year, 1-2 years, 2-3 years, 3-5 years, or 5+ years as of the period start date.
- Nursing home status in previous 365 days:

    - No Nursing Home care (0 days)
    - Short-term NH care (1 - 89 days)
    - Long-term NH care (90 - 365 days)
- BMI at incidence of ESRD
    - <18.5
    - 18.5-25
    - 25-30
    - >=30
- Calendar year
- The following incident comorbidities are included. They are taken from the CMS-2728 form. Each comorbidity is included as a separate covariate in the model.
    - Alcohol dependence
    - Atherosclerotic heart disease
    - Cerebrovascular disease
    - Chronic obstructive pulmonary disease
    - Congestive heart failure
    - Diabetes that is not the primary cause of ESRD
    - Drug dependence
    - Inability to ambulate
    - Inability to transfer
    - Malignant neoplasm or cancer
    - Other cardiac disease
    - Peripheral vascular disease
    - Tobacco use (current smoker)
    - No Medical Evidence (CMS-2728) Form
    - At least one of the comorbidities listed
- A set of prevalent comorbidities based on Medicare inpatient claims (individual comorbidities categorized into 66 groups).
    - Includes an adjustment for less than 6 months of Medicare covered months in prior calendar year
- Beside main effects, two-way interaction terms between age, sex, and cause of ESRD are also included:
    - Diabetes as cause of ESRD*Sex
    - Diabetes as cause of ESRD*Age
    - Age*Sex

Prevalent comorbidities (see appendix) are determined using the previous calendar year of CMS claims. We grouped individual comorbidities into clinically related categories. Each comorbidity group is included as a separate covariate in the model. If a patient has less than 6 Medicare covered months in the prior calendar year, we consider prevalent comorbidities to be "missing" for that patient even if there are comorbidities identified in claims.

The modeling process has two stages. At stage I, a stratified model is fitted to the national data with piecewise-constant baseline rates and stratification by facility. Specifically, the model is of the following form

$$Pr(\text{Emergency department encounter on day } t \text{ given covariates } X) = r_{0k}(t)\exp(\beta' X_{ik})$$

where X_{ik} is the vector of covariates for the i^{th} patient in the k^{th} facility and β is the vector of regression coefficients. Time t is measured from the start of ESRD. The baseline rate function $r_{0k}(t)$ is specific to the k^{th} facility, and is assumed to be a step function with break points at 6 months, 1 year, 2 years, 3 years and 5 years since the onset of dialysis. This model allows the baseline emergency department rates to vary between strata

(facilities), but assumes that the regression coefficients are the same across all strata; this approach is robust to possible differences between facilities in the patient mix being treated. The stratification on facilities is important in this phase to avoid bias due to possible confounding between covariates and facility effects.

At stage II, the relative risk estimates from the first stage are used to create offsets and an unstratified model is fitted to obtain estimates of an overall baseline rate function. That is, we estimate a common baseline rate of encounters, $r_0(t)$, across all facilities by considering the model

$$Pr(\text{Emergency department encounter on day } t \text{ given covariates } X) = r_0(t)\, R_{ik,}$$

where $R_{ik} = \exp(\beta' X_{ik})$ is the estimated relative risk for patient i in facility k obtained from the stage I. In our computation, we assume the baseline to be a step function with 6 unknown parameters, $\alpha_1, \ldots, \alpha_6$, to estimate. These estimates are used to compute the expected number of encounters given a patient's characteristics.

Specifically, let t_{iks} represent the number of days that patient i from facility k is under observation in the s^{th} time interval with estimated rate α_s. The corresponding expected number of emergency department encounters in the s^{th} interval for this patient is calculated as

$$E_{iks} = \alpha_s\, t_{iks}\, R_{ik} \ .$$

It should be noted that t_{iks} and hence E_{iks} can be 0 if patient i from facility k is never at risk during the s^{th} time interval. Summing the E_{iks} over all 6 intervals and all N_k patients in facility k gives

$$Exp = \sum_{i=1}^{N_k}\sum_{s=1}^{6} E_{iks} = \sum_{i=1}^{N_k}\sum_{s=1}^{6} \alpha_s t_{iks} R_{ik} \ ,$$

which is the expected number of emergency department encounters during follow-up at that facility.

Let Obs be the observed total number of emergency department encounters at this facility. The SEDR for emergency department encounters is the ratio of the observed total encounters to this expected value, or

$$\text{SEDR} = \text{Obs/Exp}$$

Reference:

Elixhauser A, Steiner C, Palmer L. Clinical Classifications Software (CCS), 2015. U.S. Agency for Healthcare Research and Quality.

Available: http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp

2b3.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b3.3a. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or social risk factors) used in the statistical risk model or for stratification by risk (*e.g.,*

potential factors identified in the literature and/or expert panel; regression analysis; statistical significance of p<0.10; correlation of x or higher; patient factors should be present at the start of care) **Also discuss any "ordering" of risk factor inclusion**; for example, are social risk factors added after all clinical factors?

In this model for SEDR, covariates are taken to act multiplicatively on the ED rate and the adjustment model is fitted with facility defining strata so as to provide valid estimates even if the distribution of adjustment variables differs across facilities [1-6]. All analyses are done using SAS.

In general, adjustment factors for the SEDR were selected based on several considerations. Our starting point was the Standardized Hospitalization Ratio (SHR) (NQF 1463) which is the model on which we developed SEDR. We began with a large set of patient characteristics (listed above), which were first evaluated for face validity by the 2016 TEP. Factors considered appropriate were then investigated with statistical models to determine if they were related to ED encounters.

We identify all unique ICD-9/10 diagnosis codes from each patient's prior year of Medicare claims. We group these diagnosis codes by diagnosis area using the Agency for Healthcare Research and Quality (AHRQ) Clinical Classifications Software (CCS) diagnosis categories. A list of ICD-9/10 codes used for the calculation is provided in the attached data dictionary/code list.

Methodology for prevalent comorbidity selection: We began the selection process with the 283 AHRQ CCS groupers for calendar year 2015. We eliminated the following 32 groupers either due to a possible association with facility care, a reflection of underlying kidney disease, or because they were not appropriate adjusters for our analysis.

AHRQ CCS Groupers Excluded	Description
2	Septicemia
123	Influenza
156	Nephritis / Nephrosis
157	Acute Kidney Failure
158	Chronic Kidney Disease
254	Rehabilitation care; fitting of prostheses; and adjustment of devices
255	Administrative/social admission
256	Medical examination/evaluation
257	Other aftercare
258	Other screening for suspected conditions
259	Residual codes; unclassified
E-Codes	21 Groupers total

Next, five categories of specific ICD-9 codes were removed from the remaining 251 AHRQ CCS groupers. These codes, listed in the Appendix, may be associated with dialysis facility care and include diagnoses such as secondary hyperparathyroidism, fluid overload, hyperkalemia, and vascular access infections. Once these specific ICD-9 codes were excluded, the 251 CCS groupers were consolidated down to a set of 130 nascent groups that we developed by combining similar CCS categories that had specificity beyond what was needed for our risk adjustment.

The selection of prevalent comorbidities was derived using a boosting variable selection method that was applied to the 130 nascent groups to identify a subset of prevalent comorbidities based on their ability to predict outpatient ED encounters. This process is more selective than traditional forward step-wise model building in selecting covariates. The boosting method [7] included the following steps:

1. Use forward stage-wise regression to iteratively detect comorbidities. That is, given the inclusion of some comorbidities, this method identifies additional comorbidity predictors to add to the analysis model.

2. Randomly draw bootstrapped samples and repeatedly apply the boosting procedure on each bootstrapped sample. The variables are ranked based on their selection frequencies.

3. Apply an empirical Bayes false discovery rate (FDR) controlling procedure [8,9] to effectively control the fraction of false discoveries. This procedure is able to control the FDR at a preselected level $0 < q < 1$ (FDR-controlling parameter). For instance, if q = 0:1 and 10 variables are selected with an estimated FDR less than q, at most 1 of these 10 variables would be expected to be a false positive. This is an equivalent process to assessing the statistical significance of the association between the predictor variable and an emergency department encounter.

The boosting method resulted in a set of 67 groups that were predictive of an ED encounter. This list of prevalent comorbidities was presented to the ED TEP in June 2017 and received unanimous support for inclusion in the SEDR and ED30 measures. Since then, due to changes in the CCS groupers, we removed CCS 55 grouper "Fluid and electrolyte disorders", as this condition is likely to be associated with facility care and therefore should not be included as a risk factor since fluid management is under the purview of the facility. The final set of comorbidity groups is 66.

Selected References:

1. Cox D: Regression models and life tables (with discussion). JRStat Soc [SerB]34: 187–220, 1972
2. Cunningham A, Mautner D, Ku B, Scott K, LaNoue M. Frequent emergency department visitors are frequent primary care visitors and report unmet primary care needs. J Eval Clin Pract. 2016 Nov 8. doi: 10.1111/jep.12672. [Epub ahead of print]
3. Cook R, Lawless J. The Statistical Analysis of Recurrent Events, New York, Springer-Verlag, 2007
4. Kalbfleisch J, Prentice R: The statistical analysis of failure time data, Hoboken, New Jersey, John Wiley & Sons, Inc., 2002
5. Lawless J, and Nadeau C. Some Simple Robust Methods for the Analysis of Recurrent Events Technometrics. Vol. 37, No. 2 (May, 1995), pp. 158-168
6. Liu, D., Schaubel, D.E. and Kalbfleisch, J.D. (2012). Computationally efficient marginal models for clustered recurrent event data. Biometrics, 68, 637-647.
7. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232.
8. Benjamini,Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 57, 289-300.
9. Efron, B. (2012). Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction Institute of Mathematical Statistics Monographs, Cambridge University Press.

2b3.3b. How was the conceptual model of how social risk impacts this outcome developed? Please check all that apply:

☒ **Published literature**

☒ **Internal data analysis**

☐ **Other (please describe)**

SDS/SES factors were evaluated based on appropriateness (whether related to differences in outcomes), empirical association with the outcome, and as supported in published literature.

The relationship among patient-level SDS, socioeconomic disadvantage, access to care, and acute care utilization such as hospitalization and emergency department use is well-established in studies in the general

population and has received considerable attention over the years (AHRQ Reports, 2011; 2012; 2013; 2014; 2015). There is also overlap between patient-level SDS factors such as race, and area-level SES. For example, blacks and other minority races, compared to whites, disproportionately tend to have lower income, experience more neighborhood poverty, residential segregation, levels of educational attainment, and unemployment levels. Together these jointly influence key health outcomes related to morbidity and acute care use (Williams 2006; Williams and Collins, 2001).

Race, insurance status (dual-eligbility), younger age, and SES have been shown to be predictors of emergency department utilization in the general population (Capp et al., 2015; Colligan et al., 2016; LaCalle et al., 2010; Zuckerman and Shen 2004; Hastings et al., 2008). For example, a study by Zuckerman and Shen (2004) reported that black adults had higher odds than whites of being occasional users compared to non-ED users. This difference between blacks and whites was larger when comparing frequent-users to non-users (Zuckerman and Shen, 2004, pg. 178). However, they also found few differences in the likelihood of frequent ED use when comparing patients that have private insurance versus those who are uninsured, while frequent ED use was more likely among those with public insurance (i.e., Medicaid) (Zuckerman and Shen 2004). Those with lower income also had higher odds of being occasional and frequent ED users, while individuals with some college had lower odds of being an occasional or frequent user of the ED, compared to those with no high school diploma. An analysis by Cunningham et al., (2016) of frequent ED use at two urban hospitals found that frequent ED use was associated with younger age, and that frequent users were more likely to be black. However, there was no significant difference in primary care access between infrequent and frequent users, suggesting that access to care did not explain variation in ED utilization. In addition to younger age, another study reported that those who were single/divorced, single-parents, had high school education or less, or had lower income were more likely to be frequent users of the ED (Sun et al., 2003). Among dual-eligible patients that receive care from a Federally Qualified Health Center (FQHC), relative rates of ED use were lower compared to dual-eligibles that did not receive care from an FQHC (Wright et al., 2015), suggesting the importance of access to primary care. Finally, trends in ED use show differences by sex (female), age (45-64), and geography (the Midwest) and in large central metropolitan areas (Skinner et al., 2014, pg 2-3).

In the ESRD population, low health literacy (a proxy of SES) was found to be predictive of ED use in one study (Green et al., 2013), as well as SDS/SES factors of younger age, female sex, black race, and public insurance (Medicaid) while lower ED use was associated with private insurance (Lovasik et al., 2016). ESRD patients discharged from a skilled nursing facility that had a subsequent emergency department encounter within 30 days were more likely to be of black race, have dual Medicare-Medicaid status, and higher comorbidity (Hall et al., 2015). In ESRD patients that received a transplant, higher risk of ED use was associated with younger age, female sex, black race, Hispanic ethnicity, and public insurance (Medicaid) (Schold et al., 2016). Treatment adherence was also found to be a risk factor for emergency department visits (Chan et al., 2014). This suggests that there may be related SDS/SES or community level factors that adversely impact patient treatment adherence.

Area-level factors, typically operating as proxies of patient level factors, have also been found to influence acute care use, such as readmission (Herrin et al., 2015; Kind et al, 2014) as well as ED use (Skinner et al., 2014, pg 2-3). Additionally, area-level SES has been observed to be associated with poor outcomes in ESRD patients (e.g., Almachraki et al 2016).

Given these observed linkages we tested available patient- and area-level SDS/SES variables based on the conceptual relationships described above and demonstrated in the literature, as well as the availability of data for analysis.

In our analyses assessing the impact on facility level emergency department use by ESRD patients, we use the publicly available Area Deprivation Index (ADI) originally developed by Singh and colleagues at the University

of Wisconsin. We applied the updated ADI based on 2009-2013 census data (University of Wisconsin, 2013 v1.5). The ADI reflects a full set of SES characteristics, including measures of income, education, and employment status, measured at the ZIP code level. Singh (2003) has applied the index in a variety of contexts, including analysis of county-level mortality rates. Singh found area differences in mortality associated with low SDS. Over the period studied, mortality differences widened because of slower mortality reductions in more deprived areas. The ADI has also been applied to the calculation of risk-adjusted rates of hospital readmission (Kind et al 2014).

References:

Agency for Healthcare Research and Quality (AHRQ). 2010 National Health Care Disparities Report. Washington, DC: AHRQ; 2011; 2011 National Health Care Disparities Report. Washington, DC: AHRQ; 2012; 2012 National Health Care Disparities Report. Washington, DC: AHRQ; Reports: 2013; 2013 National Health Care Disparities Report. Washington, DC: AHRQ; Reports: 2014; 2014 National Health Care Disparities Report. Washington, DC: AHRQ; 2015.

Almachraki F, Tuffli M, Lee P, Desmarais M, Shih HC, Nissenson A, and Krishnan M. Population Health Management. Volume 19, Number 1, 2016.

Capp R, West DR, Doran K, Sauaia A, Wiler J, Coolman T, Ginde AA. Characteristics of Medicaid-Covered Emergency Department Visits Made by Nonelderly Adults: A National Study. J Emerg Med. 2015 Dec;49(6):984-9.

Chan K, Thadhani R, Maddux F. Adherence barriers to chronic dialysis in the United States. J Am Soc Nephrol. 2014 Nov;25(11):2642-8. Epub 2014 Apr 24.

Colligan E, Pines JM, Colantuoni E, Howell B, Wolff JL. Risk Factors for Persistent Frequent Emergency Department Use in Medicare Beneficiaries. Ann Emerg Med. 2016 Jun;67(6):721-9.

Green JA, Mor MK, Shields AM, Sevick MA, Arnold RM, Palevsky PM, Fine MJ, Weisbord SD. Associations of health literacy with dialysis adherence and health resource utilization in patients receiving maintenance hemodialysis. Am J Kidney Dis. 2013 Jul;62(1):73-80.

Hall R, Toles M, Massing M, Jackson E, Peacock-Hinton S, O'Hare A, Colón-Emeric C. Utilization of acute care among patients with ESRD discharged home from skilled nursing facilities. Clin J Am Soc Nephrol. 2015 Mar 6;10(3):428-34. Epub 2015 Feb 3.

Hastings S, Oddone E, Fillenbaum G, Shane R, and Schmader K. Frequency and predictors of adverse health outcomes in older Medicare beneficiaries discharged from the emergency department. Med Care. 2008 Aug;46(8):771-7

Herrin J, St. Andre J, Kenward K, Joshi MS, Audet AJ, Hines SC. Community Factors and Hospital Readmission Rates. HSR: Health Services Research 50:1 (February 2015).

Kind AJH, Jencks S, Brock J, Yu M, Bartels C, Ehlenbach W, Greenberg C, Smith M. "Neighborhood Socioeconomic Disadvantage and 30 Day Rehospitalizations: An Analysis of Medicare Data ." Ann Intern Med. 2014 Dec 2; 161(11): 765–774. doi: 10.7326/M13-2946 PMCID: PMC4251560.

LaCalle E, Rabin E. Frequent users of emergency departments: the myths, the data, and the policy implications. Ann Emerg Med. 2010 Jul;56(1):42-8.

Lovasik B, Zhang R, Hockenberry JM, Schrager JD, Pastan SO, Mohan S, Patzer RE.Emergency Department Use and Hospital Admissions Among Patients With End-Stage Renal Disease in the United States.JAMA Intern Med. 2016 Oct 1;176(10):1563-1565.

Schold J, Elfadawy N, Buccini L, Goldfarb D, Flechner S, P Phelan M, Poggio E. Emergency Department Visits after Kidney Transplantation. Clin J Am Soc Nephrol. 2016 Apr 7;11(4):674-83.

Singh, GK. Area Deprivation And Widening Inequalities In US Mortality, 1969–1998. Am J Public Health. 2003; 93(7):1137–1143.

Skinner H, Blanchard J, and Elixhauser A. Trends in Emergency Department Visits, 2006–2011. Healthcare Cost And Utilization Project. Statistical Brief #179. September 2014. Pg 2-3

Sun BC, Burstin HR, Brennan TA. Predictors and outcomes of frequent emergency department users. Acad Emerg Med. 2003 Apr;10(4):320-8.

University of Wisconsin School of Medicine Public Health. 2013 Area Deprivation Index v1.5. Downloaded from https://www.neighborhoodatlas.medicine.wisc.edu/ October 31, 2018.

Williams D. Race, Socioeconomic Status, and Health: The Added Effects of Racism and Discrimination. Annals of the New York Academy of Sciences. Volume 896, Issue 1, Article first published online: 6 FEB 2006.

Williams D, and Collins C, Racial Residential Segregation: A Fundamental Cause of Racial Disparities in Health. Public Health Reports / September–October 2001. Volume 116. 404-416.

Wright B, Potter A, and Trivedi A. Federally Qualified Health Center Use Among Dual Eligibles: Rates Of Hospitalizations And Emergency Department Visits Health Affairs, 34, no.7 (2015):1147-1155.

Zuckerman S, Shen Y. Characteristics of occasional and frequent emergency department users: do insurance coverage and access to care matter? Med Care. 2004 Feb;42(2):176-82.

2b3.4a. What were the statistical results of the analyses used to select risk factors?

Table 5 reports the model results.

Table 5. SEDR Model Coefficients, Data Years 2014–2017.

Covariate	Coefficient	P-value^	Hazard Ratio^
Sex			
Female	0.3253	--	--
Male	Reference		

Covariate	Coefficient	P-value^	Hazard Ratio^
Cause of ESRD			
Diabetes	0.3038	--	--
Non-Diabetes	Reference		
BMI			
<18.5	0.0048	0.2817	1.005
18.5-25	Reference		
25-30	-0.0091	<0.0001	0.991
>=30	-0.0280	<0.0001	0.972
Year			
Year 2014	Reference		
Year 2015	0.0198	<0.0001	1.020
Year 2016	0.0210	<0.0001	1.021
Year 2017	0.0397	<0.0001	1.041
Patient Age			
Age (continuous)	-0.0355	--	--
Age Squared	0.0002	--	--
Interaction: Cause of ESRD * Patient Age			
Interaction: Diabetes * Age	-0.0070	<0.0001	0.993
Interaction: Diabetes * Age Squared	0.00004	<0.0001	1.000
Interaction: Sex * Patient Age			
Interaction: Female * Age	-0.0091	<0.0001	0.991
Interaction: Female * Age Squared	0.0001	<0.0001	1.000
Interaction: Cause of ESRD * Sex			
Interaction: Diabetes * Female	0.0192	<0.0001	1.019
Incident Comorbidities			
Atherosclerotic heart disease	0.0218	<0.0001	1.022
Other cardiac disease	0.0149	<0.0001	1.015
Congestive heart failure	0.0329	<0.0001	1.033
Inability to ambulate	0.0108	0.0280	1.011
Chronic obstructive pulmonary disease	0.0352	<0.0001	1.036
Inability to transfer	-0.0325	<0.0001	0.968
Malignant neoplasm, Cancer	-0.0139	0.0004	0.986
Diabetes	0.0402	<0.0001	1.041

Covariate	Coefficient	P-value^	Hazard Ratio^
Peripheral vascular disease	-0.0037	0.1894	0.996
Cerebrovascular disease, CVA, TIA	0.0478	<0.0001	1.049
Tobacco use (current smoker)	0.0653	<0.0001	1.067
Alcohol	0.0262	<0.0001	1.027
Drug dependence	0.1770	<0.0001	1.194
Flag for having at least on incident comorbidity	0.0081	0.0041	1.008
Missing incident comorbidity	0.0293	<0.0001	1.030
Nursing Home previous 365 days			
No Nursing Home	Reference		
Short-term NH care (1 - 89 days)	-0.0166	<0.0001	0.984
Long-term NH care (90 - 365 days)	-0.1901	<0.0001	0.827
Prevalent comorbidity groups			
HIV infection	0.0986	<0.0001	1.104
Hepatitis	0.0453	<0.0001	1.046
Viral infection	0.0324	<0.0001	1.033
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.0426	<0.0001	1.043
Melanomas of skin; Other non-epithelial cancer of skin	-0.1118	<0.0001	0.894
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.0533	<0.0001	0.948
Diabetes mellitus with complications; Diabetes mellitus without complication	0.0407	<0.0001	1.042
Encephalitis (except that caused by tuberculosis or sexually transmitted disease); Meningitis (except that caused by tuberculosis or sexually transmitted disease); Other CNS infection and poliomyelitis	-0.1396	<0.0001	0.870
Conditions associated with dizziness or vertigo; Other ear and sense organ disorders; Otitis media and related conditions	0.0772	<0.0001	1.080
Other nervous system disorders	0.0552	<0.0001	1.057
Essential hypertension	0.0363	<0.0001	1.037
Hypertension with complications and secondary hypertension	0.1143	<0.0001	1.121
Acute myocardial infarction; Coronary atherosclerosis and other heart disease	0.0498	<0.0001	1.051
Nonspecific chest pain	0.1999	<0.0001	1.221
Pulmonary heart disease	0.0235	<0.0001	1.024
Other and ill-defined heart disease	0.0533	<0.0001	1.055
Cardiac dysrhythmias; Conduction disorders	0.0483	<0.0001	1.049

Covariate	Coefficient	P-value^	Hazard Ratio^
Other circulatory disease	0.0243	<0.0001	1.025
Phlebitis; thrombophlebitis and thromboembolism	0.0276	<0.0001	1.028
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.0852	<0.0001	1.089
Asthma; Chronic obstructive pulmonary disease and bronchiectasis	0.0666	<0.0001	1.069
Other lower respiratory disease	0.0997	<0.0001	1.105
Other upper respiratory disease	0.0176	<0.0001	1.018
Diseases of mouth; excluding dental; Disorders of teeth and jaw	0.1150	<0.0001	1.122
Esophageal disorders	0.0138	<0.0001	1.014
Gastroduodenal ulcer (except hemorrhage);Gastritis and duodenitis; Other disorders of stomach and duodenum; Appendicitis and other appendiceal conditions	0.0467	<0.0001	1.048
Anal and rectal conditions	0.0527	<0.0001	1.054
Peritonitis and intestinal abscess	-0.1066	<0.0001	0.899
Pancreatic disorders (not diabetes)	0.1302	<0.0001	1.139
Gastrointestinal hemorrhage	0.0203	<0.0001	1.020
Noninfectious gastroenteritis	0.0948	<0.0001	1.099
Other gastrointestinal disorders	0.0184	<0.0001	1.019
Urinary tract infections	0.0286	<0.0001	1.029
Other diseases of kidney and ureters	0.0093	<0.0001	1.009
Hyperplasia of prostate; Inflammatory conditions of male genital organs; Other male genital disorders	0.0438	<0.0001	1.045
Chronic ulcer of skin; Other inflammatory condition of skin; Other skin disorders; Skin and subcutaneous tissue infections	0.0387	<0.0001	1.040
Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease)	-0.0451	<0.0001	0.956
Sprains and strains	0.1461	<0.0001	1.157
Complication of device; implant or graft	0.0072	0.0001	1.007
Superficial injury; contusion	0.1055	<0.0001	1.111
Poisoning by nonmedicinal substances; Poisoning by other medications and drugs; Poisoning by psychotropic agents	0.0092	0.0055	1.009
Other injuries and conditions due to external causes	0.0358	<0.0001	1.036
Syncope	0.0590	<0.0001	1.061
Gangrene	-0.0703	<0.0001	0.932
Shock	-0.1615	<0.0001	0.851
Nausea and vomiting	0.1466	<0.0001	1.158

Covariate	Coefficient	P-value^	Hazard Ratio^
Abdominal pain	0.1585	<0.0001	1.172
Malaise and fatigue	0.0672	<0.0001	1.070
Allergic reactions	0.0841	<0.0001	1.088
Anxiety disorders	0.0941	<0.0001	1.099
Attention-deficit conduct and disruptive behavior disorders	0.1448	<0.0001	1.156
Developmental disorders	0.0859	<0.0001	1.090
Mood disorders	0.0229	<0.0001	1.023
Personality disorders	0.2004	<0.0001	1.222
Schizophrenia and other psychotic disorders	0.0487	<0.0001	1.050
Alcohol-related disorders; Substance-related disorders	0.1959	<0.0001	1.216
Suicide and intentional self-inflicted injury	0.1372	<0.0001	1.147
Screening and history of mental health and substance abuse codes	0.0847	<0.0001	1.088
Miscellaneous mental health disorders	0.0388	<0.0001	1.040
Epilepsy; convulsions	0.0477	<0.0001	1.049
Headache; including migraine	0.1845	<0.0001	1.203
Calculus of urinary tract	0.0488	<0.0001	1.050
Other non-traumatic joint disorders	0.0450	<0.0001	1.046
Spondylosis; intervertebral disc disorders; other back problems	0.0976	<0.0001	1.103
Osteoporosis	-0.0781	<0.0001	0.925
Other bone disease and musculoskeletal deformities; Other connective tissue disease	0.0657	<0.0001	1.068
Less than 6 Medicare covered months in the prior calendar year	0.9220	<0.0001	2.514

^Interpretation of covariate main effects that are also included in interaction terms is not straightforward. Because of this coefficient p-values and HRs are not reported for the main effect covariates. Interaction terms can be interpreted directly. For example, the interaction between female sex and age means that the effect of female depends on age.

2b3.4b. Describe the analyses and interpretation resulting in the decision to select social risk factors *(e.g. prevalence of the factor across measured entities, empirical association with the outcome, contribution of unique variation in the outcome, assessment of between-unit effects and within-unit effects.)* **Also describe the impact of adjusting for social risk (or not) on providers at high or low extremes of risk.**

Table 6 below shows the parameter estimates from the respective Cox models for the original baseline SEDR and one with patient- and area-level SDS/SES variables added.

Table 6. Coefficients for baseline model and model with additional SDS/SES adjustors, 2014-2017

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Employment status						
Employed	--	--	--	Reference		
Unemployed	--	--	--	0.1300	<0.0001	1.139
Other	--	--	--	0.0919	<0.0001	1.096
Race						
White	--	--	--	Reference		
Black	--	--	--	0.1478	<0.0001	1.159
Asian/Pacific Islander	--	--	--	-0.1665	<0.0001	0.847
Native American	--	--	--	0.0421	<0.0001	1.043
Other	--	--	--	0.0320	0.0564	1.033
Ethnicity						
Ethnicity: non-Hispanic	--	--	--	Reference		
Ethnicity: Hispanic	--	--	--	0.0276	<0.0001	1.028
Dual Eligibility						
Not Dual Eligible	--	--	--	Reference		
Dual Eligible: Medicare and Medicaid	--	--	--	0.2083	<0.0001	1.232
ADI†						
National percentile ADI score	--	--	--	0.0016	<0.0001	1.002
Sex						
Female	0.3253	--	--	0.3202	--	--
Male	Reference			Reference		
Cause of ESRD						
Diabetes	0.3038	--	--	0.2605	--	--
Non-Diabetes	Reference			Reference		
BMI						
<18.5	0.0048	0.2817	1.005	0.0023	0.6115	1.002
18.5-25	Reference			Reference		
25-30	-0.0091	<0.0001	0.991	-0.0094	<0.0001	0.991
>=30	-0.0280	<0.0001	0.972	-0.0284	<0.0001	0.972

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Year						
Year 2014	Reference			Reference		
Year 2015	0.0198	<0.0001	1.020	0.0201	<0.0001	1.020
Year 2016	0.0210	<0.0001	1.021	0.0221	<0.0001	1.022
Year 2017	0.0397	<0.0001	1.041	0.0399	<0.0001	1.041
Patient Age						
Age (continuous)	-0.0355	--	--	-0.0324	--	--
Age Squared	0.0002	--	--	0.0002	--	--
Interaction: Cause of ESRD * Patient Age						
Interaction: Diabetes * Age	-0.0070	<0.0001	0.993	-0.0054	<0.0001	0.995
Interaction: Diabetes * Age Squared	0.00004	<0.0001	1.000	0.00003	<0.0001	1.000
Interaction: Sex * Patient Age						
Interaction: Female * Age	-0.0091	<0.0001	0.991	-0.0092	<0.0001	0.991
Interaction: Female * Age Squared	0.0001	<0.0001	1.000	0.0001	<0.0001	1.000
Interaction: Cause of ESRD * Sex						
Interaction: Diabetes * Female	0.0192	<0.0001	1.019	0.0060	0.0496	1.006
Incident Comorbidities						
Atherosclerotic heart disease	0.0218	<0.0001	1.022	0.0320	<0.0001	1.033
Other cardiac disease	0.0149	<0.0001	1.015	0.0211	<0.0001	1.021
Congestive heart failure	0.0329	<0.0001	1.033	0.0206	<0.0001	1.021
Inability to ambulate	0.0108	0.0280	1.011	0.0002	0.9673	1.000
Chronic obstructive pulmonary disease	0.0352	<0.0001	1.036	0.0366	<0.0001	1.037
Inability to transfer	-0.0325	<0.0001	0.968	-0.0378	<0.0001	0.963

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Malignant neoplasm, Cancer	-0.0139	0.0004	0.986	0.0023	0.5584	1.002
Diabetes	0.0402	<0.0001	1.041	0.0356	<0.0001	1.036
Peripheral vascular disease	-0.0037	0.1894	0.996	-0.0036	0.2081	0.996
Cerebrovascular disease, CVA, TIA	0.0478	<0.0001	1.049	0.0335	<0.0001	1.034
Tobacco use (current smoker)	0.0653	<0.0001	1.067	0.0572	<0.0001	1.059
Alcohol	0.0262	<0.0001	1.027	0.0106	0.1076	1.011
Drug dependence	0.1770	<0.0001	1.194	0.1243	<0.0001	1.132
Flag for having at least on incident comorbidity	0.0081	0.0041	1.008	-0.0019	0.5030	0.998
Missing incident comorbidity	0.0293	<0.0001	1.030	0.0347	0.0538	1.035
Nursing Home previous 365 days						
No Nursing Home	Reference			Reference		
Short-term NH care (1 - 89 days)	-0.0166	<0.0001	0.984	-0.0121	<0.0001	0.988
Long-term NH care (90 - 365 days)	-0.1901	<0.0001	0.827	-0.2276	<0.0001	0.796
Prevalent comorbidity groups						
HIV infection	0.0986	<0.0001	1.104	0.0524	<0.0001	1.054
Hepatitis	0.0453	<0.0001	1.046	0.0237	<0.0001	1.024
Viral infection	0.0324	<0.0001	1.033	0.0387	<0.0001	1.039
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.0426	<0.0001	1.043	0.0384	<0.0001	1.039
Melanomas of skin; Other non-epithelial cancer of skin	-0.1118	<0.0001	0.894	-0.0494	<0.0001	0.952
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.0533	<0.0001	0.948	-0.0466	<0.0001	0.954

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Diabetes mellitus with complications; Diabetes mellitus without complication	0.0407	<0.0001	1.042	0.0333	<0.0001	1.034
Encephalitis (except that caused by tuberculosis or sexually transmitted disease); Meningitis (except that caused by tuberculosis or sexually transmitted disease); Other CNS infection and poliomyelitis	-0.1396	<0.0001	0.870	-0.1320	<0.0001	0.876
Conditions associated with dizziness or vertigo; Other ear and sense organ disorders; Otitis media and related conditions	0.0772	<0.0001	1.080	0.0784	<0.0001	1.082
Other nervous system disorders	0.0552	<0.0001	1.057	0.0508	<0.0001	1.052
Essential hypertension	0.0363	<0.0001	1.037	0.0272	<0.0001	1.028
Hypertension with complications and secondary hypertension	0.1143	<0.0001	1.121	0.1070	<0.0001	1.113
Acute myocardial infarction; Coronary atherosclerosis and other heart disease	0.0498	<0.0001	1.051	0.0559	<0.0001	1.057
Nonspecific chest pain	0.1999	<0.0001	1.221	0.1902	<0.0001	1.210
Pulmonary heart disease	0.0235	<0.0001	1.024	0.0274	<0.0001	1.028
Other and ill-defined heart disease	0.0533	<0.0001	1.055	0.0515	<0.0001	1.053
Cardiac dysrhythmias; Conduction disorders	0.0483	<0.0001	1.049	0.0568	<0.0001	1.058
Other circulatory disease	0.0243	<0.0001	1.025	0.0235	<0.0001	1.024

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Phlebitis; thrombophlebitis and thromboembolism	0.0276	<0.0001	1.028	0.0226	<0.0001	1.023
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.0852	<0.0001	1.089	0.0861	<0.0001	1.090
Asthma; Chronic obstructive pulmonary disease and bronchiectasis	0.0666	<0.0001	1.069	0.0576	<0.0001	1.059
Other lower respiratory disease	0.0997	<0.0001	1.105	0.1020	<0.0001	1.107
Other upper respiratory disease	0.0176	<0.0001	1.018	0.0194	<0.0001	1.020
Diseases of mouth; excluding dental; Disorders of teeth and jaw	0.1150	<0.0001	1.122	0.1056	<0.0001	1.111
Esophageal disorders	0.0138	<0.0001	1.014	0.0084	<0.0001	1.008
Gastroduodenal ulcer (except hemorrhage);Gastritis and duodenitis; Other disorders of stomach and duodenum; Appendicitis and other appendiceal conditions	0.0467	<0.0001	1.048	0.0485	<0.0001	1.050
Anal and rectal conditions	0.0527	<0.0001	1.054	0.0487	<0.0001	1.050
Peritonitis and intestinal abscess	-0.1066	<0.0001	0.899	-0.0830	<0.0001	0.920
Pancreatic disorders (not diabetes)	0.1302	<0.0001	1.139	0.1256	<0.0001	1.134
Gastrointestinal hemorrhage	0.0203	<0.0001	1.020	0.0203	<0.0001	1.020
Noninfectious gastroenteritis	0.0948	<0.0001	1.099	0.0955	<0.0001	1.100
Other gastrointestinal disorders	0.0184	<0.0001	1.019	0.0198	<0.0001	1.020

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Urinary tract infections	0.0286	<0.0001	1.029	0.0337	<0.0001	1.034
Other diseases of kidney and ureters	0.0093	<0.0001	1.009	0.0121	<0.0001	1.012
Hyperplasia of prostate; Inflammatory conditions of male genital organs; Other male genital disorders	0.0438	<0.0001	1.045	0.0440	<0.0001	1.045
Chronic ulcer of skin; Other inflammatory condition of skin; Other skin disorders; Skin and subcutaneous tissue infections	0.0387	<0.0001	1.040	0.0391	<0.0001	1.040
Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease)	-0.0451	<0.0001	0.956	-0.0394	<0.0001	0.961
Sprains and strains	0.1461	<0.0001	1.157	0.1470	<0.0001	1.158
Complication of device; implant or graft	0.0072	0.0001	1.007	0.0028	0.1436	1.003
Superficial injury; contusion	0.1055	<0.0001	1.111	0.1127	<0.0001	1.119
Poisoning by nonmedicinal substances; Poisoning by other medications and drugs; Poisoning by psychotropic agents	0.0092	0.0055	1.009	0.0119	0.0004	1.012
Other injuries and conditions due to external causes	0.0358	<0.0001	1.036	0.0389	<0.0001	1.040
Syncope	0.0590	<0.0001	1.061	0.0579	<0.0001	1.060
Gangrene	-0.0703	<0.0001	0.932	-0.0712	<0.0001	0.931
Shock	-0.1615	<0.0001	0.851	-0.1528	<0.0001	0.858
Nausea and vomiting	0.1466	<0.0001	1.158	0.1421	<0.0001	1.153

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
Abdominal pain	0.1585	<0.0001	1.172	0.1476	<0.0001	1.159
Malaise and fatigue	0.0672	<0.0001	1.070	0.0673	<0.0001	1.070
Allergic reactions	0.0841	<0.0001	1.088	0.0911	<0.0001	1.095
Anxiety disorders	0.0941	<0.0001	1.099	0.1001	<0.0001	1.105
Attention-deficit conduct and disruptive behavior disorders	0.1448	<0.0001	1.156	0.1458	<0.0001	1.157
Developmental disorders	0.0859	<0.0001	1.090	0.0674	<0.0001	1.070
Mood disorders	0.0229	<0.0001	1.023	0.0277	<0.0001	1.028
Personality disorders	0.2004	<0.0001	1.222	0.1908	<0.0001	1.210
Schizophrenia and other psychotic disorders	0.0487	<0.0001	1.050	0.0423	<0.0001	1.043
Alcohol-related disorders; Substance-related disorders	0.1959	<0.0001	1.216	0.1742	<0.0001	1.190
Suicide and intentional self-inflicted injury	0.1372	<0.0001	1.147	0.1379	<0.0001	1.148
Screening and history of mental health and substance abuse codes	0.0847	<0.0001	1.088	0.0808	<0.0001	1.084
Miscellaneous mental health disorders	0.0388	<0.0001	1.040	0.0356	<0.0001	1.036
Epilepsy; convulsions	0.0477	<0.0001	1.049	0.0398	<0.0001	1.041
Headache; including migraine	0.1845	<0.0001	1.203	0.1777	<0.0001	1.194
Calculus of urinary tract	0.0488	<0.0001	1.050	0.0585	<0.0001	1.060
Other non-traumatic joint disorders	0.0450	<0.0001	1.046	0.0383	<0.0001	1.039
Spondylosis; intervertebral disc disorders; other back problems	0.0976	<0.0001	1.103	0.0916	<0.0001	1.096
Osteoporosis	-0.0781	<0.0001	0.925	-0.0625	<0.0001	0.939
Other bone disease and musculoskeletal deformities; Other	0.0657	<0.0001	1.068	0.0599	<0.0001	1.062

	Baseline SEDR			SDS/SES-adjusted SEDR		
Covariate	**Coefficient**	**P-value^**	**Hazard Ratio^**	**Coefficient**	**P-value^**	**Hazard Ratio^**
connective tissue disease						
Less than 6 Medicare covered months in the prior calendar year	0.9220	<0.0001	2.514	0.9293	<0.0001	2.533

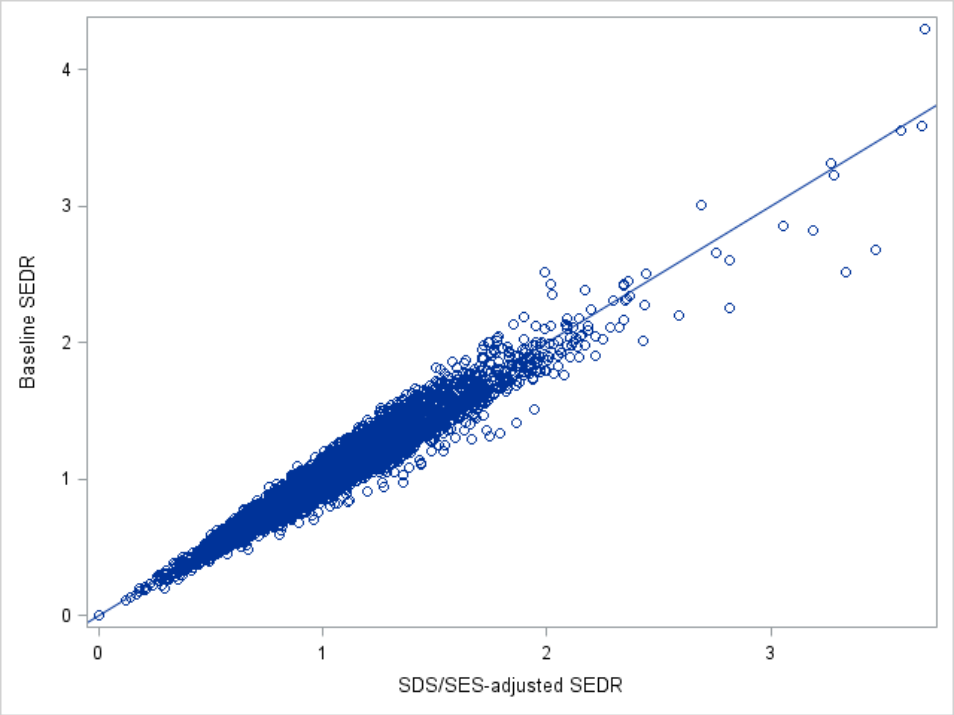
^Interpretation of covariate main effects that are also included in interaction terms is not straightforward. Because of this coefficient p-values and HRs are not reported for the main effect covariates. Interaction terms can be interpreted directly. For example, the interaction between female sex and age means that the effect of female depends on age.

*Patients without Medicare coverage or with unknown coverage type were excluded from the model.

** Other/Unknown includes patients who are on medical leave of absence, retired due to age or disability, homemakers, or those with no employment status information available.

†Area Deprivation Index (ADI) is a national percentile ranking at the block group level from 1 to 100, where 1 is the lowest ADI and 100 is the highest ADI. A block group with a ranking of 1 indicates the lowest level of "disadvantage" within the nation and an ADI with a ranking of 100 indicates the highest level of "disadvantage".

Figure 2. Correlation between SEDR without and with SDS adjustment, 2014-2017



Pearson correlation coefficient rho = 0.9710 (p<0.0001)

Table 7. Flagging rates, baseline SEDR and SEDR adjusted for SDS/SES: 2017

	Baseline SEDR			
SEDR with SDS/SES	**As Expected**	**Better than Expected**	**Worse than Expected**	**Total**

As Expected	6,242	7	61	6,310 (94%)
Better than Expected	18	33	0	51 (1%)
Worse than Expected	53	0	278	331 (6%)
Total	6,313 (94%)	40 (1%)	339 (5%)	6,692

When comparing the baseline SEDR measure with one that includes adjustment for patient and area-level SDS/SES, we observed differences in flagging of facility performance (Figure 2 and Table 7). For example, in the baseline SEDR, 339 facilities are flagged as worse than expected while 331 are flagged as worse than expected in the SEDR adjusted for SDS/SES, resulting in a decrease in the number of facilities flagged for worse than expected performance. Both the baseline SEDR and SEDR adjusted for SDS/SES are highly correlated.

We observed higher risk for female sex, black race, dual eligible status, and higher levels of area deprivation in the model adjusting for SDS/SES. We note that interpretation of the main effect of sex is not straightforward since this SDS covariate is also included as an interaction term with age (in both the baseline and SDS/SES adjusted models).

Race, ethnicity, dual eligible status and area deprivation are not included in the final risk adjusted model. Other studies have reported associations between patient-level race, ethnicity and dual eligible status and acute care utilization, however it is unclear whether these differences are due to underlying biological or other patient factors or represent disparities in care. Adjusting for these patient factors could have the unintended consequence of creating or reinforcing disparities and limiting access to care. The primary goal should be to implement quality measures that result in the highest quality of patient care and equitable access for all patients to that care.

In the final SEDR model we continue to include sex (SDS factor) for risk adjustment. Finally, trends in the general and ESRD population for ED use show differences by sex (female) and age (45-64), (Skinner et al., 2014, pg 2-3; Lovasik et al., 2016; Schold et al., 2016). In the final SEDR model we continue to include sex (SDS factor) for risk adjustment. This approach is consistent with the consensus opinion that adjustment for sex is appropriate based on biologic differences (e.g. genetic, hormonal, metabolic) that may account for higher acute care use (acute care utilization), suggesting a physiologic effect rather than a systematic difference or disparity in care by sex.

2b3.5. Describe the method of testing/analysis used to develop and validate the adequacy of the statistical model or stratification approach (*describe the steps—do not just name a method; what statistical analysis was used*)

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

Risk factors were selected for the final model based on the magnitude of the coefficients, evaluation of their statistical significance, and the model C-statistic. The C-statistic measures the discriminative power of the regression model with considered risk factors.

2b3.6. Statistical Risk Model Discrimination Statistics (*e.g., c-statistic, R-squared*)**:**

The estimate of the C-statistic for the SEDR is 0.61.

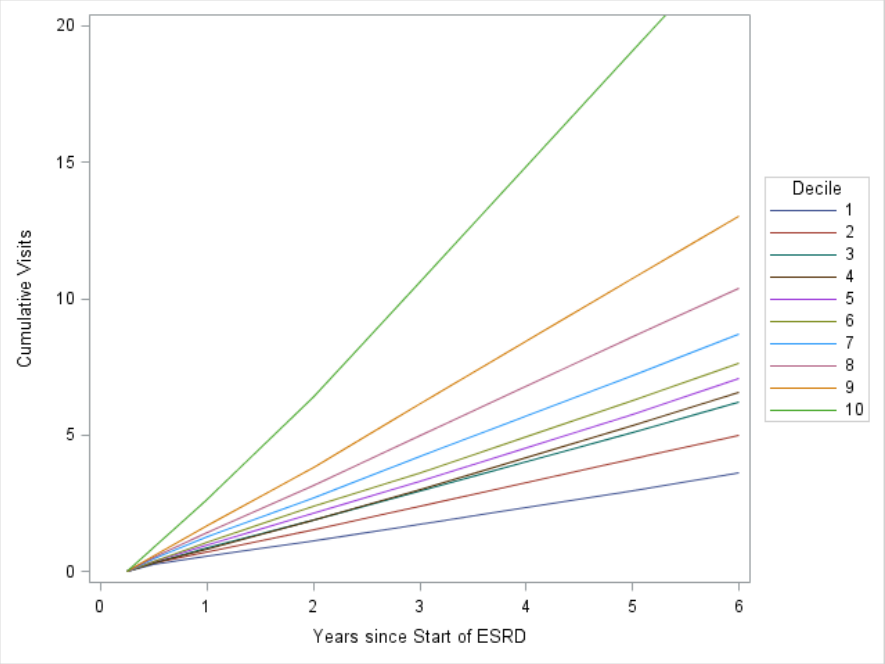
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

N/A

2b3.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

Decile plots showing piecewise linear estimates of the cumulative rates by years since start of ESRD are plotted in Figure 3. This plot creates deciles based on the value of xbeta from the stage 1 model. For each decile we then fit a model with no covariates and pull out the baseline survival curve.

Figure 3: Decile plot, 2017 data



2b3.9. Results of Risk Stratification Analysis:

N/A

2b3.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.*e., what do the results mean and what are the norms for the test conducted*)

The decile plot shows piecewise linear estimates of the cumulative rates by years since start of ESRD. The plot demonstrates that the risk factors in the model are discriminating well between patients. There is good separation among all 10 groups and the ordering is as predicted by the model (patients predicted to be at lower risk have lower ED visit rates). The absolute differences between the groups is also large with patients predicted to have the highest ED visit rates (line 10) having over a 4 times higher ED visit rate than those predicted to have the lowest rates (line 1).

2b3.11. Optional Additional Testing for Risk Adjustment (*not required, but would provide additional support of adequacy of risk model, e.g., testing of risk model in another data set; sensitivity analysis for missing data; other methods that were assessed*)

N/A

2b4. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b4.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b)*

To adjust for over-dispersion of the data, we compute the p-value for our estimates using the empirical null distribution, a robust approach that takes account of the natural random variation among facilities that is not accounted for in the model (Efron, 2004; Kalbfleisch and Wolfe, 2013). Our algorithm consists of the following concrete steps. First, we fit an over-dispersed Poisson model (e.g., SAS PROC GENMOD with link=log, dist=poisson and scale=dscale) for the number of hospital admissions

$$\log(E[\mathbf{n}_{ik}]) = \log(\mathbf{E}_{ik}) + \boldsymbol{\theta}_k,$$

where $\mathbf{n}_{ik}$ is the observed number of events for patient *i* in facility *k*, $\mathbf{E}_{ik}$ is the expected number of events for patient *i* in facility *k* and $\boldsymbol{\theta}_k$ is the facility-specific intercept. Here, i ranges over the number of patients N_k who are treated in the *k*th facility. The natural log of the SEDR for the *k*th facility is then given by the corresponding estimate of $\boldsymbol{\theta}_k$. The standard error of $\boldsymbol{\theta}_k$ is obtained from the robust estimate of variance arising from the overdispersed Poisson model.

Second, we obtain a z-score for each facility by dividing the natural log of its SEDR by the standard error from the general linear model described above. These z-scores are then grouped into quartiles based on the number of patient years at risk for Medicare patients in each facility. Finally, using robust estimates of location and scale based on the normal curve fitted to the center of the z-scores for the SEDR, we derive the mean and variance of a normal empirical null distribution for each quartile. This empirical null distribution is then used to calculate the p-value for a facility's SEDR. Statistically significant outliers (i.e., better and worse than expected) were determined based on $p<0.05$ level.

References:

Efron B. Large-scale simultaneous hypothesis testing: the choice of a null hypothesis. J Am Stat Assoc. 2004; 99:96–104

Kalbfleisch, J.D. & Wolfe, R.A. On Monitoring Outcomes of Medical Providers. Stat Biosci 2013; 5(2):286-302

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of*

specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator). ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Since many data elements can be obtained from multiple sources, missing data occur rarely. We assessed missing data for BMI and missingness for the CMS 2728 form which is the source of data used for several risk adjusters in the SEDR model.

SEDR relies on Medicare claims and other CMS administrative data for several important components of measure calculation, including identification of outpatient ED encounters from outpatient claims, and for adjustment of prevalent comorbidities (inpatient and outpatient claims). For these reasons, SEDR is restricted to Medicare Fee for Service (FFS) patients.

For several Medicare-only measures developed by UM-KECC, including SEDR, the presence of active Medicare coverage has been defined using a combination of criteria including a defined minimum of paid claims for dialysis services and/or presence of a Medicare inpatient claim during an eligibility period. With the recent increase in Medicare Advantage (MA) coverage for Medicare chronic dialysis patients, and the known systemic issue of unavailable outpatient claims data for MA patients, these criteria have the potential to introduce bias into measure calculations that could affect results for dialysis facilities with either very low or high MA patient populations. More importantly, we are not able to observe outpatient ED encounters for MA patients which introduces bias in the calculation of facility scores because we cannot observe ED encounters from outpatient

claims for MA patients. To demonstrate this we assessed the proportion of years at risk and ED visits excluded among MA patients.

As part of the comprehensive measure review process, we assessed the extent of MA coverage for ESRD dialysis patients and the effect of our historical definition of "active Medicare" status on the measure result. Medicare Advantage patient status was defined using Medicare Enrollment Database (EDB) criteria. Primary FFS coverage was identified using CMS administrative data, and active Medicare status utilized the combination of minimum dialysis paid claims and/or inpatient Medicare hospitalization claims briefly described above. We confirmed the presence of usable ICD diagnosis codes from MA inpatient claims and the nearly complete absence of outpatient Medicare claims data for patients identified as MA in the CMS data used for our measure calculation.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Patient-year-facility Level Analysis of Missing Data for 2014-2017 SEDR model Input Data

Table 9: Percent Missing for Risk Model Data, 2014-2017

Variable	**Missing %**
BMI	2.2
Missing 2728	1.4
Less than 6 Medicare covered months in the prior calendar year*	27.4

*This indicator is used to determine the presence of prevalent comorbidities from Medicare claims.

Table 10: Percent Medicare Advantage Years at Risk and Emergency Room Visits Excluded, 2014-2017

Year	Years At Risk				Emergency Department Visits		
	No Exclude	Yes Exclude	% Excluded		No Exclude	Yes Exclude	% Excluded
2014	352,570	294,543	16%		454,555	454,093	0.1%
2015	362,199	295,558	18%		468,099	467,215	0.2%
2016	374,103	301,048	20%		489,572	488,090	0.3%
2017	381,909	301,357	21%		499,333	497,914	0.3%
Total	1,470,781	1,192,506	19%		1,911,559	1,907,312	0.2%

Summary findings:

1. The percentage of patients with MA coverage receiving chronic dialysis in US dialysis facilities has approximately doubled in the last decade and is approaching 20% based on 2017 data.
2. When applied to MA patients, the historical definition of active Medicare coverage (described earlier) creates systematic bias in the SEDR measure calculation through exclusion of MA patient time at risk in facilities unless the MA patient had one or more hospitalizations in the observation period. MA patients included because of hospitalization are very likely not representative of MA patients as a whole, instead reflecting a sicker subset.
3. We confirmed the presence of usable ICD diagnosis codes from MA inpatient claims and the nearly complete absence of outpatient Medicare claims data for patients identified as MA in the CMS data used for our measure calculation. The time at risk excluded for MA is 19% overall and percentage of ED encounters excluded for MA patients is 0.2% overall, suggesting the unavailability of outpatient claims for MA patients.
4. The percentage of patient months with Less than 6 Medicare covered months in the prior calendar year is 27%.
5. BMI on the 2728 medical evidence form is missing rarely (2.2%), while the percentage missing the 2728 is 1.4%.

Additional analyses demonstrate a variable distribution of Medicare Advantage ESRD dialysis patient proportion following geographic boundaries. For example, the percentage of MA ESRD patient time at risk relative to total Medicare ESRD patient time at risk varies from a low of 2.2% in Wyoming to a high of 44.2% in Puerto Rico.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

There is a very low frequency of patients with missing BMI. In cases where values are missing, patients are assigned to the most common BMI category, which is >=30, to mitigate any potential impact of missing BMI data on performance scores. The frequency of patients missing form 2728 was also low at 1.4% and an adjustment factor (missing form 2728) in the model was used to mitigate any potential impact on performance results.

Based on our results showing very low percentage of ED visits for MA patients as expected due to the unavailability of outpatient ED claims for MA patients, we excluded MA patients from SEDR. As described above we are not able to observe outpatient ED encounters for MA patients. Because we cannot observe ED encounters from outpatient claims for MA patients, facilities with a higher proportion of MA patients versus FFS patients could appear to have a very low overall SEDR, even if the actual overall ED rate at the facility is high; alternatively SEDR could be high if ED visits are concentrated in a few FFS patients relative to the larger MA patient population at the facility, even if the actual overall ED rate at the facility is low.

Less than 6 Medicare covered months in the prior calendar year were considered as having incomplete prevalent comorbidity information but were not excluded from SEDR. While just over one-quarter of patients fell in this category, we acknowledge this is a general limitation of relying on Medicare FFS claims for ascertaining comorbidities. We mitigate this potential bias in measure performance scores through an adjustment which is an indicator for patients with less than 6 Medicare covered months in the prior calendar year. This minimizes risk of biased results at the dialysis facility level and is consistent with a number of other NQF-endorsed measures that are based on Medicare claims data.

2b4.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., *number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined*)

Table 8. Number and percentage of facilities by classification of SEDR, 2017.

Better than expected	As expected	Worse than expected	Total
40 (0.60%)	6,313 (94.35%)	338 (5.05%)	6,691

2b4.3. What is your interpretation of the results in terms of demonstrating the ability to identify statistically significant and/or clinically/practically meaningful differences in performance across measured entities? (i*.e., what do the results mean in terms of statistical and meaningful differences?*)

Each facility is evaluated using the empirical null method to determine whether its actual number of emergency department visits is statistically significantly different from its expected number of emergency department visits in 2017. Specifically, the risk adjustment model to determine the expected number of visits is estimated using risk periods from 2014-2017. Without empirical null methods, a large number of facilities will be flagged. In contrast, the methods based on the empirical null, used here, make appropriate adjustments for overdispersion. Using this method, facilities are flagged if they have outcomes (excessive emergency department encounters) that are extreme when compared to the variation in outcomes for other facilities of a similar size. Overall, most are flagged as expected (about 94%), while <1% are better than expected, and approximately 5% are flagged as worse than expected.

2b5. COMPARABILITY OF PERFORMANCE SCORES WHEN MORE THAN ONE SET OF SPECIFICATIONS

If only one set of specifications, this section can be skipped.

Note: *This item is directed to measures that are risk-adjusted (with or without social risk factors)* ***OR*** *to measures with more than one set of specifications/instructions (e.g., one set of specifications for how to identify and compute the measure from medical record abstraction and a different set of specifications for claims or eMeasures). It does not apply to measures that use more than one source of data in one set of specifications/instructions (e.g., claims data to identify the denominator and medical record abstraction for the numerator).* ***Comparability is not required when comparing performance scores with and without social risk factors in the risk adjustment model. However, if comparability is not demonstrated for measures with more than one set of specifications/instructions, the different specifications (e.g., for medical records vs. claims) should be submitted as separate measures.***

2b5.1. Describe the method of testing conducted to compare performance scores for the same entities across the different data sources/specifications (*describe the steps—do not just name a method; what statistical analysis was used*)

N/A

2b5.2. What were the statistical results from testing comparability of performance scores for the same entities when using different data sources/specifications? (*e.g., correlation, rank order*)

N/A

2b5.3. What is your interpretation of the results in terms of the differences in performance measure scores for the same entities across the different data sources/specifications? (i.*e., what do the results mean and what are the norms for the test conducted*)

N/A

2b6. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b6.1. Describe the method of testing conducted to identify the extent and distribution of missing data (or nonresponse) and demonstrate that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias (*describe the steps—do not just name a method; what statistical analysis was used*)

Since many data elements can be obtained from multiple sources, missing data occur rarely. We assessed missing data for BMI and missingness for the CMS 2728 form which is the source of data used for several risk adjusters in the SEDR model.

SEDR relies on Medicare claims and other CMS administrative data for several important components of measure calculation, including identification of outpatient ED encounters from outpatient claims, and for adjustment of prevalent comorbidities (inpatient and outpatient claims). For these reasons, SEDR is restricted to Medicare Fee for Service (FFS) patients.

For several Medicare-only measures developed by UM-KECC, including SEDR, the presence of active Medicare coverage has been defined using a combination of criteria including a defined minimum of paid claims for dialysis services and/or presence of a Medicare inpatient claim during an eligibility period. With the recent increase in Medicare Advantage (MA) coverage for Medicare chronic dialysis patients, and the known systemic issue of unavailable outpatient claims data for MA patients, these criteria have the potential to introduce bias into measure calculations that could affect results for dialysis facilities with either very low or high MA patient populations. More importantly, we are not able to observe outpatient ED encounters for MA patients which introduces bias in the calculation of facility scores because we cannot observe ED encounters from outpatient claims for MA patients. To demonstrate this we assessed the proportion of years at risk and ED visits excluded among MA patients.

As part of the comprehensive measure review process, we assessed the extent of MA coverage for ESRD dialysis patients and the effect of our historical definition of "active Medicare" status on the measure result. Medicare Advantage patient status was defined using Medicare Enrollment Database (EDB) criteria. Primary FFS coverage was identified using CMS administrative data, and active Medicare status utilized the combination of minimum dialysis paid claims and/or inpatient Medicare hospitalization claims briefly described above. We confirmed the presence of usable ICD diagnosis codes from MA inpatient claims and the nearly complete absence of outpatient Medicare claims data for patients identified as MA in the CMS data used for our measure calculation.

2b6.2. What is the overall frequency of missing data, the distribution of missing data across providers, and the results from testing related to missing data? (*e.g., results of sensitivity analysis of the effect of various rules for missing data/nonresponse; if no empirical sensitivity analysis, identify the approaches for handling missing data that were considered and pros and cons of each*)

Patient-year-facility Level Analysis of Missing Data for 2014-2017 SEDR model Input Data

Table 9: Percent Missing for Risk Model Data, 2014-2017

Variable	**Missing %**
BMI	2.2
Missing 2728	1.4
Less than 6 Medicare covered months in the prior calendar year*	27.4

*This indicator is used to determine the presence of prevalent comorbidities from Medicare claims.

Table 10: Percent Medicare Advantage Years at Risk and Emergency Room Visits Excluded, 2014-2017

Year	Years At Risk				Emergency Department Visits		
	No Exclude	Yes Exclude	% Excluded		No Exclude	Yes Exclude	% Excluded
2014	352,570	294,543	16%		454,555	454,093	0.1%
2015	362,199	295,558	18%		468,099	467,215	0.2%
2016	374,103	301,048	20%		489,572	488,090	0.3%
2017	381,909	301,357	21%		499,333	497,914	0.3%
Total	1,470,781	1,192,506	19%		1,911,559	1,907,312	0.2%

Summary findings:

1. The percentage of patients with MA coverage receiving chronic dialysis in US dialysis facilities has approximately doubled in the last decade and is approaching 20% based on 2017 data.
2. When applied to MA patients, the historical definition of active Medicare coverage (described earlier) creates systematic bias in the SEDR measure calculation through exclusion of MA patient time at risk in facilities unless the MA patient had one or more hospitalizations in the observation period. MA patients included because of hospitalization are very likely not representative of MA patients as a whole, instead reflecting a sicker subset.
3. We confirmed the presence of usable ICD diagnosis codes from MA inpatient claims and the nearly complete absence of outpatient Medicare claims data for patients identified as MA in the CMS data used for our measure calculation. The time at risk excluded for MA is 19% overall and percentage of ED encounters excluded for MA patients is 0.2% overall, suggesting the unavailability of outpatient claims for MA patients.
4. The percentage of patient months with Less than 6 Medicare covered months in the prior calendar year is 27%.
5. BMI on the 2728 medical evidence form is missing rarely (2.2%), while the percentage missing the 2728 is 1.4%.

Additional analyses demonstrate a variable distribution of Medicare Advantage ESRD dialysis patient proportion following geographic boundaries. For example, the percentage of MA ESRD patient time at risk relative to total Medicare ESRD patient time at risk varies from a low of 2.2% in Wyoming to a high of 44.2% in Puerto Rico.

2b6.3. What is your interpretation of the results in terms of demonstrating that performance results are not biased due to systematic missing data (or differences between responders and nonresponders) and how the specified handling of missing data minimizes bias**?** (i.*e., what do the results mean in terms of supporting the selected approach for missing data and what are the norms for the test conducted; if no empirical analysis, provide rationale for the selected approach for missing data*)

There is a very low frequency of patients with missing BMI. In cases where values are missing, patients are assigned to the most common BMI category, which is >=30, to mitigate any potential impact of missing BMI data on performance scores. The frequency of patients missing form 2728 was also low at 1.4% and an adjustment factor (missing form 2728) in the model was used to mitigate any potential impact on performance results.

Based on our results showing very low percentage of ED visits for MA patients as expected due to the unavailability of outpatient ED claims for MA patients, we excluded MA patients from SEDR. As described above we are not able to observe outpatient ED encounters for MA patients. Because we cannot observe ED encounters from outpatient claims for MA patients, facilities with a higher proportion of MA patients versus FFS patients could appear to have a very low overall SEDR, even if the actual overall ED rate at the facility is high; alternatively SEDR could be high if ED visits are concentrated in a few FFS patients relative to the larger MA patient population at the facility, even if the actual overall ED rate at the facility is low.

Less than 6 Medicare covered months in the prior calendar year were considered as having incomplete prevalent comorbidity information but were not excluded from SEDR. While just over one-quarter of patients fell in this category, we acknowledge this is a general limitation of relying on Medicare FFS claims for ascertaining comorbidities. We mitigate this potential bias in measure performance scores through an adjustment which is an indicator for patients with less than 6 Medicare covered months in the prior calendar year. This minimizes risk of biased results at the dialysis facility level and is consistent with a number of other NQF-endorsed measures that are based on Medicare claims data.

3. Feasibility

Extent to which the specifications including measure logic, require data that are readily available or could be captured without undue burden and can be implemented for performance measurement.

3a. Byproduct of Care Processes

For clinical measures, the required data elements are routinely generated and used during care delivery (e.g., blood pressure, lab test, diagnosis, medication order).

3a.1. Data Elements Generated as Byproduct of Care Processes.

Generated or collected by and used by healthcare personnel during the provision of care (e.g., blood pressure, lab value, diagnosis, depression score), Coded by someone other than person obtaining original information (e.g., DRG, ICD-9 codes on claims).

If other:

3b. Electronic Sources

The required data elements are available in electronic health records or other electronic sources. If the required data are not in electronic health records or existing electronic sources, a credible, near-term path to electronic collection is specified.

3b.1. To what extent are the specified data elements available electronically in defined fields (*i.e., data elements that are needed to compute the performance measure score are in defined, computer-readable fields*) Update this field for **maintenance of endorsement**.

ALL data elements are in defined fields in a combination of electronic sources

3b.2. If ALL the data elements needed to compute the performance measure score are not from electronic sources, specify a credible, near-term path to electronic capture, OR provide a rationale for using other than electronic sources. For **maintenance of endorsement**, if this measure is not an eMeasure (eCQM), please describe any efforts to develop an eMeasure (eCQM).

3b.3. If this is an eMeasure, provide a summary of the feasibility assessment in an attached file or make available at a measure-specific URL. Please also complete and attach the NQF Feasibility Score Card.

Attachment:

3c. Data Collection Strategy

Demonstration that the data collection strategy (e.g., source, timing, frequency, sampling, patient confidentiality, costs associated with fees/licensing of proprietary measures) can be implemented (e.g., already in operational use, or testing demonstrates that it is ready to put into operational use). For eMeasures, a feasibility assessment addresses the data elements and measure logic and demonstrates the eMeasure can be implemented or feasibility concerns can be adequately addressed.

3c.1. Required for maintenance of endorsement. Describe difficulties (as a result of testing and/or operational use of the measure) regarding data collection, availability of data, missing data, timing and frequency of data collection, sampling, patient confidentiality, time and cost of data collection, other feasibility/implementation issues.

IF instrument-based, consider implications for both individuals providing data (patients, service recipients, respondents) and those whose performance is being measured.

N/A

3c.2. Describe any fees, licensing, or other requirements to use any aspect of the measure as specified *(e.g., value/code set, risk model, programming code, algorithm)*.

N/A

4. Usability and Use

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

4a. Accountability and Transparency

Performance results are used in at least one accountability application within three years after initial endorsement and are publicly reported within six years after initial endorsement (or the data on performance results are available). If not in use at the time of initial endorsement, then a credible plan for implementation within the specified timeframes is provided.

4.1. Current and Planned Use

NQF-endorsed measures are expected to be used in at least one accountability application within 3 years and publicly reported within 6 years of initial endorsement in addition to performance improvement.

Specific Plan for Use	Current Use (for current use provide URL)
Not in use	

4a1.1 For each CURRENT use, checked above (update for maintenance of endorsement), provide:

- Name of program and sponsor
- Purpose
- Geographic area and number and percentage of accountable entities and patients included
- Level of measurement and setting

N/A

4a1.2. If not currently publicly reported OR used in at least one other accountability application (e.g., payment program, certification, licensing) what are the reasons? (*e.g., Do policies or actions of the developer/steward or accountable entities restrict access to performance results or impede implementation?*)

As the measure is undergoing endorsement review it has not been implemented in public reporting or for use in another accountability application

4a1.3. If not currently publicly reported OR used in at least one other accountability application, provide a credible plan for implementation within the expected timeframes -- any accountability application within 3 years and publicly reported within 6 years of initial endorsement. (*Credible plan includes the specific program, purpose, intended audience, and timeline for implementing the measure within the specified timeframes. A plan for accountability applications addresses mechanisms for data aggregation and reporting.*)

Public Reporting: CMS will consider implementing the SEDR measure as part of CMS' Dialysis Facility public reporting program, of which the purpose is to help dialysis patients and their caregivers understand the quality of care provided by dialysis facilities and to be able to compare selected aspects of care between dialysis facilities. All Medicare-certified dialysis facilities that treat dialysis patients in the U.S. are reported on CMS's Dialysis Facility public reporting program.

4a2.1.1. Describe how performance results, data, and assistance with interpretation have been provided to those being measured or other users during development or implementation.

How many and which types of measured entities and/or others were included? If only a sample of measured entities were included, describe the full population and how the sample was selected.

N/A

4a2.1.2. Describe the process(es) involved, including when/how often results were provided, what data were provided, what educational/explanatory efforts were made, etc.

N/A

4a2.2.1. Summarize the feedback on measure performance and implementation from the measured entities and others described in 4d.1.

Describe how feedback was obtained.

N/A

4a2.2.2. Summarize the feedback obtained from those being measured.

N/A

4a2.2.3. Summarize the feedback obtained from other users

N/A

4a2.3. Describe how the feedback described in 4a2.2.1 has been considered when developing or revising the measure specifications or implementation, including whether the measure was modified and why or why not.

N/A

Improvement

Progress toward achieving the goal of high-quality, efficient healthcare for individuals or populations is demonstrated. If not in use for performance improvement at the time of initial endorsement, then a credible rationale describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

4b1. Refer to data provided in 1b but do not repeat here. Discuss any progress on improvement (trends in performance results, number and percentage of people receiving high-quality healthcare; Geographic area and number and percentage of accountable entities and patients included.)

If no improvement was demonstrated, what are the reasons? If not in use for performance improvement at the time of initial endorsement, provide a credible rationale that describes how the performance results could be used to further the goal of high-quality, efficient healthcare for individuals or populations.

N/A

4b2. Unintended Consequences

The benefits of the performance measure in facilitating progress toward achieving high-quality, efficient healthcare for individuals or populations outweigh evidence of unintended negative consequences to individuals or populations (if such evidence exists).

4b2.1. Please explain any unexpected findings (positive or negative) during implementation of this measure including unintended impacts on patients.

N/A

4b2.2. Please explain any unexpected benefits from implementation of this measure.

N/A

5. Comparison to Related or Competing Measures

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

5. Relation to Other NQF-endorsed Measures

Are there related measures (conceptually, either same measure focus or target population) or competing measures (conceptually both the same measure focus and same target population)? If yes, list the NQF # and title of all related and/or competing measures.

Yes

5.1a. List of related or competing measures (selected from NQF-endorsed measures)

1463 : Standardized Hospitalization Ratio for Dialysis Facilities (SHR)

2505 : Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health

5.1b. If related or competing measures are not NQF endorsed please indicate measure title and steward.

Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) (currently undergoing endorsement review with SEDR). Steward: CMS

5a. Harmonization of Related Measures

The measure specifications are harmonized with related measures;

OR

The differences in specifications are justified

5a.1. If this measure conceptually addresses EITHER the same measure focus OR the same target population as NQF-endorsed measure(s):
Are the measure specifications harmonized to the extent possible?

Yes

5a.2. If the measure specifications are not completely harmonized, identify the differences, rationale, and impact on interpretability and data collection burden.

These measures are not completely harmonized. Each measure assesses different outcomes and/or target populations as reflected in the measure specifications. The proposed Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities and Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) for Dialysis Facilities measures both the dialysis facilities' ED use but they measure different aspects of ED use. The SEDR measures the overall rate of ED use while the ED30 focuses on ED utilization closely following a hospitalization. Both SEDR and ED30 apply to the same target population - adult Medicare-covered dialysis patients who have had ESRD for more than 90 days. The SEDR and SHR are both intended to encourage appropriate management of acute conditions but measure two different acute care outcomes. SEDR measures overall outpatient acute care services while SHR measure inpatient acute care services. SEDR is harmonized with SHR and ED30 in several aspects. All are harmonized to the population they measure (eligible Medicare-covered ESRD patients); however SHR also includes pediatric patients. All three measures have risk adjustment for prevalent comorbidities while only SEDR and SHR also adjust for incident comorbidities taken from CMS form 2728. Exclusions: 1) Only SEDR and ED30 exclude hospice patients; 2) ED30 includes additional exclusions based on discharge type, that are not part of SEDR or SHR; 3) ED30 adjusts for discharging hospital, acknowledging that for ED encounters after a hospital discharge, that hospitals also bear accountability for properly coordinating care with the dialysis facility; and 4) both SEDR and ED30 exclude patient time at risk, or index discharges, respectively, that are covered by Medicare Advantage. We do this because Medicare Advantage outpatient encounter data (i.e., ED encounters) are not available from CMS claims therefore we are not able to capture data for the measure outcome. SEDR and NQF measure 2505: Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health have the same focus (emergency department encounters). Differences: 1) Home Health is focused on emergency department use within the first 30 days of home health; 2) each measure has distinct target populations; 3) risk adjustment factors; and 4) model type (2-stage Cox model vs multinomial logistic model). For example, the Home Health 30 measure adjusts for over 400 covariates that were statistically significantly predictive of acute care hospitalization or emergency use (without admission). SEDR currently adjusts for a set of comorbidities present at ESRD incidence and for a set of prevalent comorbidities. Because of the different care settings and comorbidity profile of Home Health patients, different risk adjustment approaches are justified.

5b. Competing Measures

The measure is superior to competing measures (e.g., is a more valid or efficient way to measure);

OR

Multiple measures are justified.

5b.1. If this measure conceptually addresses both the same measure focus and the same target population as NQF-endorsed measure(s): Describe why this measure is superior to competing measures (e.g., a more valid or efficient way to measure quality); OR provide a rationale for the additive value of endorsing an additional measure. (Provide analyses when possible.)

Appendix

A.1 Supplemental materials may be provided in an appendix. All supplemental materials (such as data collection instrument or methodology reports) should be organized in one file with a table of contents or bookmarks. If material pertains to a specific submission form number, that should be indicated. Requested information should be provided in the submission form and required attachments. There is no guarantee that supplemental materials will be reviewed.

Attachment **Attachment:** SEDR_Flowchart.pdf

Contact Information

Co.1 Measure Steward (Intellectual Property Owner): Centers for Medicare and Medicaid Services

Co.2 Point of Contact: Kimberly, Rawlings

Co.3 Measure Developer if different from Measure Steward: University of Michigan Kidney Epidemiology and Cost Center

Co.4 Point of Contact: Jennifer, Sardone, jmsto@med.umich.edu

Additional Information

Ad.1 Workgroup/Expert Panel involved in measure development

Provide a list of sponsoring organizations and workgroup/panel members' names and organizations. Describe the members' role in measure development.

List the workgroup/panel members´ names and organizations. Describe the members´ role in measure development.

Amy Williams, MD

Medical Director of Hospital Operations

Division of Nephrology and Hypertension

Rochester, MN

Terry Ketchersid, MD, MBA

Senior Vice President and Chief Medical Officer

Integrated Care Division, Fresenius Medical Care North America

Waltham, MA

Sarah Swartz, MD

Medical Director of Dialysis,

Texas Children´s Hospital

Baylor College of Medicine

Houston, TX

Michael Phelan, MD, JD, RDMS, FACEP

Medical Director of the Quality and Patient Safety Institute

Cleveland Clinic

Cleveland, OH

Arjun Venkatesh, MD, MBA, MHS

Assistant Professor

Department of Emergency Medicine

Yale University School of Medicine

Yale New Haven Hospital

New Haven, CT

Alexis Chettiar, RN, MSN, ACNP-BC

Acute Care Nurse Practitioner

East Bay Nephrology Medical Group

Oakland, CA

Julie Crandall

Board Member

Dialysis Patient Citizens (DPC) Board of Directors

Hurricane, UT

Maggie Carey

Kidney Patient Advisory Council (KPAC) Chair

Forum of ESRD Networks

Richard Knight, MBA

Vice President/Chair of Public Policy

American Association of Kidney Patients (AAKP)

New Carrollton, MD

Measure Developer/Steward Updates and Ongoing Maintenance

Ad.2 Year the measure was first released: 2020

Ad.3 Month and Year of most recent revision: 04, 2020

Ad.4 What is your frequency for review/update of this measure? Annually

Ad.5 When is the next scheduled review/update for this measure? 04, 2021

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: