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 #: 3566

Corresponding Measures:

De.2. Measure Title: Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) for Dialysis Facilities

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

De.3. Brief Description of Measure: The Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge for Dialysis Facilities (ED30) is defined to be the ratio of observed over expected events. The numerator is the observed number of index discharges from acute care hospitals that are followed by an outpatient emergency department encounter within 4-30 days after discharge for eligible adult Medicare dialysis patients treated at a particular dialysis facility. The denominator is the expected number of index discharges followed by an ED encounter within 4-30 days given the discharging hospital's characteristics, characteristics of the dialysis facility's patients, and the national norm for dialysis facilities. Note that in this document, acute care hospital includes critical access hospitals and "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 11 eligible index discharges 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 encounters within 30 days of an index discharge are an important indicator of care coordination, care transitions, 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). This has been demonstrated in the end-stage renal disease (ESRD) population as well with 27% of patients being treated in an ED within 30 days of hospital discharge, most frequently for congestive heart failure (Harel et al., 2015)

Rates of ED visits among end ESRD dialysis patients have increased between 2007 and 2016. As reported by the USRDS, 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), 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 encounters subsequent to a hospital discharge may help dialysis facility efforts to prevent emergent unscheduled care and to help control escalating medical costs, for example through greater care coordination and post-discharge transitional care. Specifically, dialysis facility activities such as evaluation of the patients target weight or medication reconciliation and review may help reduce the risk of ED encounters after hospital discharge. This measure will complement existing measures targeting care coordination (such as the Standardized Readmission Ratio NQF #2496) by identifying impactful events that can be influenced by dialysis facility care.

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

Harel, Z.;Wald, R.;McArthur, E.;Chertow, G. M.;Harel, S.;Gruneir, A.;Fischer, H. D.;Garg, A.

X.;Perl, J.;Nash, D. M.;Silver, S.;Bell, C. M. Rehospitalizations and Emergency Department Visits after Hospital Discharge in Patients Receiving Maintenance Hemodialysis. J Am Soc Nephrol. 2015 26(12):3141-50 doi:10.1681/ASN.2014060614

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

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.

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 index hospital discharges during a year that are followed by an emergency department encounter within 4–30 days of the discharge among eligible adult Medicare patients at a facility.

S.6. Denominator Statement: The expected number of index hospital discharges for eligible adult Medicare ESRD dialysis patients during the two year period that are followed by an emergency department encounter within 4-30 days of the discharge among eligible patients at a facility. The expected value is the result of a risk-adjusted predictive model adjusted for the characteristics of the patients, the dialysis facility, and the discharging hospitals.

S.8. Denominator Exclusions: Index Discharge exclusions that are implicit in the denominator definition include discharges for which the patient:

• Has Medicare Advantage coverage at the time of the index discharge

• Has had ESRD for 90 days or less at time of discharge

• Is less than 18 years of age at the time of discharge

We also exclude discharges and emergency department encounters for which the patient was:

• Actively enrolled in hospice at any time of during the calendar month of the discharge date or ED encounter admit date

Outpatient Medicare claims are the source of ED encounter data, and since outpatient claims are not available for Medicare Advantage (MA) patients, we cannot identify ED encounters for MA patients. Therefore, we exclude index discharges for patients with MA at the time of discharge.

The hospice exclusion is needed because hospice patients are considered to be under the purview of hospice care givers and may have other reasons for Emergency Department use such as pain management.

Additionally we exclude hospital discharges that:

• Do not result in a live discharge

• Are against medical advice

• Include a primary diagnosis for cancer, mental health or rehabilitation (see below for excluded CCSs)

• Are from a PPS-exempt cancer hospital

• Are followed within three days of discharge by the patient being transplanted, discontinuing dialysis, recovering renal function, being lost to follow-up, having another hospitalization, or having an emergency department visit

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?

Criteria 1: Importance to Measure and Report

1a. Evidence

Maintenance measures – less emphasis on evidence unless there is new information or change in evidence since the prior evaluation.

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 developers cited research that showed ESRD dialysis patients had higher rates of hospital readmissions than the national average.
- The developers suggest that several factors can increase the risk of an emergency department (ED) visit including: 1) losing control of target weight from fluid overload, 2) increased infection from poor prevention processes, and 3) inability to maintain electrolyte levels following hospitalization.
- The developers also referenced several interventions that focus on reducing missed dialysis treatments such as increased communication, care coordination, and patient education can reduce the rate of ED readmissions.

Question for the Committee:

- *Is there at least action that the accountable entity can take to achieve a change in the measure results?*

Guidance from the Evidence Algorithm

Measure assess performance on a health outcome: Yes (Box 1) -> Relationship between outcome and one health structure: Yes (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 provided data that evaluated Medicare-certified dialysis facilitates (n=6,728) that treated Medicare dialysis patients (n=271,852) between 2016-2017.
- The developer reports that the median sample sizes of the facilities were 65 patients with measure rates varied between 0.00 to 3.52.The mean value was 1.03 and the SD was 0.37.
- The developer reports deciles of ED30 for 2016-2017:
    - Decile 1, N=672 Min=0.00 Max=0.60
    - Decile 2, N=673 Min=0.60 Max=0.73

- Decile 3, N=673 Min=0.73 Max=0.83
- Decile 4, N=673 Min=0.83 Max=0.92
- Decile 5, N=673 Min=0.92 Max=1.00
- Decile 6, N=673 Min=1.00 Max=1.09
- Decile 7, N=673 Min=1.09 Max=1.18
- Decile 8, N=673 Min=1.18 Max=1.29
- Decile 9, N=673 Min=1.29 Max=1.48
- Decile 10, N=672 Min=1.48 Max=3.52

Disparities

- The developers indicated that differences in age, race, ethnicity, and Medicare-Medicaid status show a statistically higher likelihood of emergency department readmissions within 4-30 days of a recent discharge.
- The developer expressed uncertainty on whether or not these disparities were based on comorbidities or quality of care.

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.

- **Little causal evidence cited and some evidence related generally to ED use and not specifically for ESRD. Much of the evidence is tangential. Little applies to PD patients.EvidenceA systematic study showed mixed results.only tev.**
- **The evidence appears to directly apply to the outcome being measured.**
- **No concerns**
- **meets**
- **yes. There is a at least one intervention that can be made. I wonder though whether there is enough evidence to appropriately attribute post-hospital ED visits to a dialysis unit as opposed to other transitions of care issues or even the specifics of the index hospitalization.**
- **Evidence supports measure**
- **Noted that ESRD dialysis patients had higher rates of hospital readmissions than national average**
- **High**

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?

- **The ESRD data base was used and showed gaps in care including by subgroups showing disparities.**
- **Performance data on the measure was provided and demonstrates an opportunity to improve care across subgroups.**
- **No concerns**
- **There is a wide gap, but the developer states that disparities uncertainty basis of quality or comorbidities.**
- **Large gap between 1st and 10th deciles. Re: disparities,comment on age, race, ethnicity and dual eligibility status**
- **gaps and disparities**
- **Disparities noted in age, race, ethnicity and Medi/Medi status with highlier likelihood of ED readmission within 4-30 days of recent discharge. Not clear if disparities were attributed to comorbidities**
- **Yes, I think so**

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**
  - Several 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 was 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 Acceptability: Preliminary Analysis Form

- Reliability: H-3; M-3; L-2; I-0 (Pass)
- Validity: H-1; M-4; L-2; I-1 (Pass)

Methods Panel Evaluation Summary:

- **Specifications:**
    - No issues
- **Reliability Testing** – *Performance Score Reliability*
    - **Method(s) of reliability testing:**
        - Conducted at facility-level using data from 2015-2018 including more than 6,000 facilities
        - Measure Score reliability testing was conducted at the data source and level of analysis indicated
        - 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".
- SMP reviewers raised concerns with the measure score reliability testing result, which was considered modest/low, and questioned whether the testing approach used, namely the provider inter-unit reliability (PIUR), was intended to demonstrate that the measure was reliable only for the purpose of identifying outliers.

- **Reliability testing results:**
  - The underlying signal to noise ratio (IUR) is 0.45, with PIUR being 0.57.

- **Validity Testing** - *Performance Score Validity (Empirical and Face)*
  - **Method(s) of validity testing:**
    - Face validity and score-level empirical validity testing were conducted
    - The empirical validity testing method that compared equivalent outcome measures stratified by expected performance was appropriate and similar to measure 3565 except for SFR.
  - **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 Emergency Encounter Ratio (SEDR) by classification of facilities as 'better than/as expected' versus 'worse than expected' for ED30 (Table 2).

Table 2. Classification of ED30 and mean facility performance scores for Related Measures, 2016-2017

	Facilities Missing	Better than/As Expected	Worse than expected	As Hypothesized?
SMR	412	1.00	1.05	Yes
STrR	717	0.99	1.21	Yes
SFR	510	63.32	63.64	No
PPPW	341	19.70	14.71	Yes
SRR	346	1.00	1.00	Yes
SEDR	205	0.99	1.49	Yes

- **Exclusions:**
  - No exclusions indicated
- **Risk adjustment Summary:** Method – Statistical Modeling
  - Three stage model
    1. Fixed effects regression
    2. Double random effects
    3. Mixed effects logistic regression model

- The c-stat of the adjustment model is modest -- 0.67
- SDS factors were included in the model
    1. Employment status, ethnicity, duals
    2. Gender is only variable maintained in the final model

- **Meaningful Differences:**
    - Overall, most are flagged as expected (94.10%), while 2.85% are better than expected, and 3.05% are flagged as worse than expected.

Questions for the Committee regarding reliability:

- *Are the IUR values acceptable?*
- *Is the PIUR method appropriate for determining acceptable reliability?*

Questions for the Committee regarding validity:

- *Do you agree with the strength of the associations?*
- *Does this measure identify meaningful differences about quality?*
- *Do you agree with the developer's risk adjustment approach?*

Preliminary rating 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 (rationale that reliability improves as the sample sizes increase, medium and small facilities have lower reliability estimates)

Preliminary rating 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?

- **The measure seems to measure only at the tail of performance begging the question of how much improvement can result from its use.**
- **The algorithm appeared clear to me.**
- **No concerns**
- **meets**
- **ED visits which result in hospitalization are not included--but perhaps quality issues for these units will be picked up in the readmission measure.**
- **Clinically appropriate**
- **The signal to noise ratio is 0.45**
- **Moderate**

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

- **Reliability rating using IUR and PIUR appear to be low and may not differentiate performance reliably.**
- **None.**
- **No concerns**
- **no**
- **signal to noise ratios seem low to me. Not sure about PIUR usability**
- **No concerns**
- **No acceptable**
- **Repeatable definitions**

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

- **Adequate.**
- **None.**
- **No concerns**
- **no**
- **As in the other dialysis ED measure, the vast majority of units "performed as expected"**

- **no concerns**
- **visits to ED should be restricted to dialysis care (e.g. for vascular access, hyperkalemia, heart failure exacerbations etc.**
- **No**

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?

- **Only a small proportion of practices identified as outliers.**
- **Missing data may constituted a threat to the validity of this measure as more broadly applied.**
- **No concerns**
- **if 94% outcomes are expected is there room for change?**
- **Similar to other ED measure 3565. Would have expected that SFR rates would have positively impacted results. Unclear that a measure that results in 94% of units performing "as expected" can demonstrate meaningful differences in quality.**
- **no threat**
- **Meaningful differences and comparabiilty are noted**
- **Unidentified missing data could pose a threat**

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?

- **Approach is logical though believe SES should be included. Also concerned about impact of lack of MA data.**
- **Risk adjustment was tested and the results appear to be acceptable.**
- **No concerns**
- **General question we have disparities in individuals, is there a disparity in facilities rural that would inform the results**
- **Most social risk factors dropped from final model**
- **Appropriate approach**
- **No further comments**
- **It is acceptable**

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 data source for this measure is derived from administrative claims.
- The developer states that the data are collected by healthcare staff during the coordination of care (blood pressure, lab value, diagnosis, depression score), and coded by another healthcare staff member (DRG, ICD-9 codes on claims).

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?

- **No concerns.**
- **No immediate concerns about the data collection strategy.**
- **No concerns**
- **none**
- **Electronic data. No concerns**
- **Feasible**
- **Claims based measure**
- **Not sure if it would work with EHR data.**

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:

- Since this a new measure, the developer states that this measure has not been implemented for public reporting yet.
- The developer states that CMS will consider implementing the measure as part of CMS's Dialysis Facillity public reporting program. This program provides information that can help dialysis patients and caregivers compare the quality of care between different dialysis clinics.

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

Feedback on the measure by those being measured or others [vetting]

- The developer did not provide any information about feedback at this time.

Additional Feedback:

- The developer did not provide any information about feedback at this time.

***Questions for the Committee**:*

- *How have (or can) the performance results be used to further the goal of high-quality, efficient healthcare?*

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.

Improvement results

- The developer did not provide any information.

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).

Unexpected findings (positive or negative) during implementation [unexpected findings]

- The developer did not provide any information about benefits or harms.

Potential harms

- The developer did not provide any information about potential harms.

Additional Feedback:

- None

***Questions for the Committee**:*

- *Are there any potential unintended consequences?*

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

RATIONALE: The developer does not provide information or a rationale to determine an accurate preliminary rating. The Standing Committee should discuss for a final rating on this criterion.

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?

- **No use data available.**
- **It would be helpful to have more information on user feedback on the measure performance and implementation.**
- **No concerns**
- **new measure. meets**
- **Not yet in use.**
- **Not currently being used**
- **Not currently used by CMS is considering implementing as part of CMS's Dialysis Facility public reporting program**
- **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.

- **Measure not in use. Credible rationale provided though weakened by exclusion of MA and lack of consideration of PD patients separately.**
- **Yes, performance results can be used to further the goal of high-quality, efficient healthcare.**
- **No concerns**
- **not much there.**
- **If vast majority of facilities are acceptable per the measure currently, may not be that useful to drive quality**
- **Plan to use in CMS program and benefits outweigh harm**
- **No information provided on usability-the developer should provide this information**
- **Benefits outweight the harm.**

Criterion 5: Related and Competing Measures

Related or competing measures

- 2496 : Standardized Readmission Ratio (SRR) for dialysis facilities
- 2505 : Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health
- 3565: Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities (currently undergoing endorsement review with ED30). Steward: CMS (not NQF-endorsed)

Harmonization

- The developer states that SEDR and ED30 both focus on dialysis facilities's ED use, but SEDR measures overall ED usage and ED30 measures ED after hospitalization. Both ED30 and SEDR use the same target population (adult patients and after 90 days), but SRR includes pediatric patients and less than 90 days.
- The developer states that the index discharge definitions differ between SRR and ED30 measures because: 1) SRR excludes patients who have been admitted 12 or more times in one year, 2) ED30 excludes patients who enrolled in hospice care and Medicare Advantage patients, and 3) ED30 excludes that get hospitalized within 3 days of discharge while SRR includes people who died within 30 days of discharge.
- Additionally, the developer states that ED30 and 2505 differ because: 1) 2505 focuses on ED use with 30 days of home health, 2) target different populatons, 3) risk adjustment factors, and 4) model types.

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?

- **There is a related measure undergoing evaluation.**
- **I am not aware of competing measures.**
- **No concerns**
- **related, but not competing.**
- **3565 among others, does ED use after hospitalization measure some quality difference than simply overall ED use?**
- **measures related but slightly different**
- **Differences are noted**
- **Yes, there are competing measures**

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 Expet Panel Evaluation

Renal Input Form

Measure Number: 3566

Measure Title: Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) 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 #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 care provided by the dialysis facility? Restrict to dialysis-related complications?

TEP Member #3: There is evidence that care coordination activities in a dialysis facility can positively impact the need for emergency or hospital care after an admission. What is does not do which is relevant to this measure is account for an inappropriately early acute hospitalization discharge to which the dialysis facility must deal with the consequences. This pressure to keep length of hospital stay short or to push a 'home' next site of care is a large confounder to this measure.

TEP Member #4: The evidence is credible that ER visits shortly after hospitalization are an indicator of the dialysis care process. While there is not a deep evidence base for interventions to reduce ER visits following hospitalization, there is good evidence for interventions that reduce repeat hospitalizations. It is reasonable to expect these interventions, particularly efforts to reduce missed treatments and address interdialytic weight gain, would impact ER visits following hospitalization. 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 population is clear but the determination of who gets counted is complex and not representative of many patients in a clinic population.

TEP Member #4: The measure population is appropriate and harmonized with Standardized Readmission Ratio (NQF #2496)

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, how is non-ER urgent care being captured? Fourth, the revenue codes for ER visits should be confirmed to ensure that all ER visits are being captured.

I have additional concerns about attribution with this measure including:

1. "The 4-30 day time frame was selected to harmonize with the Standardized Readmission Ratio (NQF #2496) that also uses the same time period after an index hospitalization. This time interval was selected in response to providers and stakeholders concerns that there may be up to 72 hours before a patient is seen at the facility after hospital discharge." I do not think that 1 visit to the outpatient dialysis facility allows it to meaningfully impact care and avoid a repeat ER visit. The dialysis facility does not have responsibility about when dialysis patients are discharged and under what circumstances. Often discharges on done by hospitalists who have very little understanding of outpatient dialysis. I think this measures hospital care and discharge planning more than dialysis facility care.

2. "Index discharges are attributed to the facility of record on the day of discharge for the patient. That is, if the patient transfers dialysis facilities at the time of hospital discharge, it is the new facility that is assigned the index discharge. "Concern #1 is worse in this case.

3. The excluded CCSs for a primary diagnosis for cancer, mental health or rehabilitation are shown below. Unclear why these exceptions are being made?

TEP Member #2: It's appropriate to know the ER visits reencounters. The measure has some issues of not distinguishing what is dialysis related.

TEP Member #3: If the intent of the measure is to identify opportunity to stimulate more care coordination post discharge then the issues above of whether the patient is really ready for discharge and whether the next site of care should in fact be 'home' is in question.

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: Exclusions are extensive and the presumption that if they don't come back in the first 3 days home means that it was a good place and time to discharge a patient is not a premise I can believe given the pressures to discharge that exist in hospitals today. The stated exclusions make sense but are likely incomplete for the myriad situations that individual patients face and their discharging physicians.

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 exclusion of the growing Medicare Advantage population induces a bias to the data set as they should have a higher degree of care coordination available based on the plan benefit design. A

measure like this would be best for a value-based care model and thus the MA exclusion is a poor choice based on the lack of data. I feel sure that induces a level of bias that impacts a facility and its ability to organize its care coordination.

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 is not valid or reliable if it doesn't exclude things that are unrelated to dialysis care.

TEP Member #3: See above, but the MA exclusion and no recognition of whether the patient discharge was appropriate are the two gaps that seem to be glaring in this regard.

TEP Member #4: Exclusions based on advice of previous TEP. 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: I read the comments about reliability. I think the take home message is that this measure if implemented must be carefully explained for interpretation – only for recognizing differences among facility performance at the tails of the distribution.

TEP Member #2: No comment

TEP Member #3: The described correlations are not surprising to me other than I might have expected the SFR to be positively correlated with the better than expected test set. Otherwise I think the measure correlates to organization within the facility to and possible to care coordination. Directionally, I am not concerned. At an individual facility level, I am concerned that the hospitals doing the discharge will influence the result based on choices out of the control of the dialysis facility. That same facility risks losing the patient if they balk at accepting the discharge because they feel it is premature. Neither patient, nephrologist nor facility benefits from that.

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 4 percent of dialysis facilities.

TEP Member #3: The comorbidity adjusters seem reasonable. I continue to be concerned at the developer's avoidance of geographic and race/ethnic adjustment that would make the expected rate meaningful. This is consistent across each measures submitted by the developer but puts facilities at a distinct disadvantage when trying to use the measures for actionable quality improvement.

TEP Member #4: They are relevant and consistent.

Scientific Acceptability: Preliminary Analysis Form

Measure Number: 3566

Measure Title: Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge for Dialysis Facilities (ED30)

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_xxxx" 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 steward is able to assess the day of hospital discharge and the dialysis schedule. An exclusion of post-discharge days 1-3 is logical and certainly supported by dialysis providers, but I am not sure that it is in the best interest of dialysis patients. If we aspire to induce coordination of hospitals and outpatient dialysis providers, then we ought not to automatically abstain from evaluating acute complications during the first 3 days after hospital discharge.

Panel Member #4 No concerns

Panel Member #6 After reading through the numerator and denominatator several times, I am not sure whether they are measuring the hospital or the ESRD facility. Also have a concern about the assignment of Index Discharges to Facilities e.g., attributed to the facility of record on the day of discharge for the patient. From the numerator description, I am assuming that they are counting an ED encounter for ANY reason which would be attributed to the ESRD facility or is it just for ESRD related conditions? The denominator description is somewhat better but that may have been because of looking through the other meausres submitted by CMS. The algorithm helps somewhat, but I think this would confuse and the likelihood of having this measure consistently implemented would be small.

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**

Panel Member #1 The methods and citations provided are compelling. The distinction between overall IUR and PIUR for flagging outliers is interesting and important.

Panel Member #2 The methods used were appropriate – this is one of a large set of measures from CMS set in the context of dialysis care, and the measure developer regularly uses the signal-to-noise ratio method suggested by Adams to assess reliability.

Panel Member #3 (IUR) and profile IUR (PIUR) are estimated.

Panel Member #4 The inter-unit reliability (IUR) using the bootstrap samples and the Profile IUR comparison was appropriate

Panel Member #6 The methods described are apporpaite.

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 (0.50), lower than the last submission (0.59). 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 (.77). Thus, the use of the measure for this specific purpose appears to be supported.

Panel Member #2 The value obtained for the IUR was very modest - .45 or thereabouts. The PIUR value was a little larger and demonstrates that the measure may have moderate reliability for detecting “outlier” facilities. The low reliabilities are not surprising, as the measure reflects a set of dynamics that may not be very much under the control of the dialysis facilities whose performance is being measured.

Panel Member #3 The estimated IUR is equal to 0.451, whereas the estimated PIUR is equal to 0.570.

Panel Member #4 The results demonstrated score measure reliability

Panel Member #6 The results are low reliability especially the PIUR.

Panel Member #8 The underlying signal to noise ratio (IUR) among substandard (.45), with PIUR higher but still not acceptable (0.57).

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 <u>all</u> testing results):

    ☒ **High** (NOTE: Can be HIGH <u>only if</u> 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 IUR = 0.451 and PIUR - 0.570. I consider these values to signfy low signal to noise ratios (see Adams 2009).

Panel Member #2 This is a generous overall rating for reliability, and is only given because CMS and NQF in the past have allowed measures to pass with reliabilities in the .4 range based on what I consider an inappropriate standard derived from an old paper by Landis and Koch having to do with the kappa statistic or inter-rater agreement. I don't believe that this measure has acceptable reliability for use in any application other than identifying high or low outlier facilities, but I can't fail a measure that barely meets a passing standard used frequently in the past.

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 0.6, thus suggesting to me that the measure is relatively unreliable.

Panel Member #4 No concerns

Panel Member #5 Overall, we found that IUR = 0.451. The PIUR is 0.570. As noted above, the PIUR measures reliability in terms of reflagging rates but is placed on the same scale as IUR.

Panel Member #6 Measure description is not very clear and score for IUR and PIUR are low.

Panel Member #8 Overall reliability was substandard.

Panel Member #9 Low rating is based on an IUR estimate of 0.45 and PIUR estimate of 0.57. Subjectively, IUR of 0.45 is a fair amount of noise. I am not 100% sure how to evaluate PIUR but 0.57. Note I am not making a judgment that a measure with low reliability cannot be useful, I'm only saying that this measure's reliability is on the low side.

VALIDITY: ASSESSMENT OF THREATS TO VALIDITY

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

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 #4 No concerns

Panel Member #6 No concerns.

Panel Member #8 Overall reliability was substandard.

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 Roughly 3% of facilities will be flagged as better and 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 #4 No concerns

Panel Member #6 No concerns.

14. **Please describe any concerns you have regarding comparability of results if multiple data sources or methods are specified.**
Panel Member #1 NA
Panel Member #3 This item is not applicable.
Panel Member #6 No concerns.
Panel Member #9 None

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

Panel Member #1 NA

Panel Member #2 No significant concerns

Panel Member #3 I have no specific

Panel Member #4 No concerns

Panel Member #9 None

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 **Panel Member #9** (Depends on definition of social risk factors; race is included) ☐ 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 Panel Member #8 estimated C-statistic of 0.665 is modest

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

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

Panel Member #1 The risk model is adequate. The decision to not include social risk factors is explained but not compelling. The standing panel should decide if the magnitude of category switching in Table 5 of the testing document is concerning. It looks to me like a meaningful percent of the "worse than expected" change to "as expected depending on the inclusion of SES**.**

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. One could debate whether comorbidity should be ascertained only from the index hospitalization, rather than all claims during the year preceding the index hospitalization. I am not convinced that one approach is necessarily superior to the other.

Panel Member #5 Selection of clinical factors: The list of covariates considered was based on CMS' Standardized Readmission Ratio for Dialysis Facilities (NQF 2496). The specific list of prevalent comorbidities included was determined based on an empirical evaluation of prevalent comorbidities associated with risk of an ED encounter. Therefore, ED30 includes a different set of comorbidities than the SRR.

Panel Member #8 Marginal c-statistic with 87 covariates.

For cost/resource use measures ONLY:

17. **Are the specifications in alignment with the stated measure intent?**

    ☐ **Yes** ☐ **Somewhat** ☐ **No (If "Somewhat" or "No", please explain)**

18. **Describe any concerns of threats to validity related to attribution, the costing approach, carve outs, or truncation (approach to outliers):**

VALIDITY: TESTING

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

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

    ☒ **Face validity**

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

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

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

    Submission document: Testing attachment, section 2b2.2

    Panel Member #4 2016 TEP supported measure.Empirical validity testing method comparing equivalent outcome measures stratified by expected performance was appropriate and similar to measure 3565 except for SFR

    Panel Member #1 good

    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 No Concerns

    Panel Member #8 Hypthesis testing; very robust approach.

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

    Submission document: Testing attachment, section 2b2.3

    Panel Member #1 The hypothesized relationships were generally supported. All of the correlations were statistically significant and in the expected direction, although of modest magnitude.

    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 results are interesting. The positive correlation of the measure with the standardized transfusion ratio is compatible with transfusion administration in the (outpatient) emergency

department during the post-discharge interval, which is typically characterized by anemic status. The absence of any association with the standardized readmission ratio is remarkable. How to interpret this lack of correlation is a matter of discussion for the subgroup and the entire Scientific Methods Panel alike.

Panel Member #4 Both face and empirical validity demonstrated strong validity for the measure.

Panel Member #8 Hypothesis testing indicates measure is valid.

Panel Member #9 Correlations of facility-specific SMR swith related measures were in the expected directions.

23. **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)

24. **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)

25. **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 <u>are</u> threats to validity and/or relevant threats to validity were <u>not assessed OR</u> 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 <u>is required</u>; if not conducted, should rate as INSUFFICIENT.)

26. **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 validity testing supported the hypothesized relationship.

    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. The moderate rating comes mainly from the face validity results; the empirical validity results are weak.

    Panel Member #3 The measure appears to have reasonably good properties, but I do wonder whether the measure ought to be combined with the standardized readmission ratio, rather than exist as a standalone measure. This concern may not be in the scope of the review, but if the Panel is to stand in the way of unnecessary proliferation of measures, then my concern may merit discussion.

    Panel Member #4 No concerns

Panel Member #5 Taken together these results provide validation support for ED30. Performance on key quality measures 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. Also, 2016 TEP.

Panel Member #6 Appropriate approach.

Panel Member #8 Hypothesis testing indicates the measure has good validity, although the c-statistic was modest.

FOR COMPOSITE MEASURES ONLY: Empirical analyses to support composite construction

27. **What is the level of certainty or confidence that the empirical analysis demonstrates that the component measures add value to the composite and that the aggregation and weighting rules are consistent with the quality construct?**

    ☐ **High**

    ☐ **Moderate**

    ☐ **Low**

    ☐ **Insufficient**

28. **Briefly explain rationale for rating of EMPIRICAL ANALYSES TO SUPPORT COMPOSITE CONSTRUCTION**

ADDITIONAL RECOMMENDATIONS

29. **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.

Developer Submission

Additional evaluations and submission materials attachments…

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

ED30_NQF_Evidence.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.

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 Ratio for Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge for Dialysis Facilities (ED30)

IF the measure is a component in a composite performance measure, provide the title of the Composite Measure here: Click here to enter composite measure #/ title

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 occurring within 30 days of hospital discharge 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. Nearly half (46.2%) of ED visits by patients with ESRD result in a hospital admission [1], while the national percentage of ED visits among dialysis patients is 62% as of 2018 (FY2020 Dialysis Facility Report). 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 [2]. The overall aim is to reduce dialysis patients' need for unscheduled acute care in the ED following hospitalization. Post-discharge care by dialysis facilities—and coordination of that care with other providers—has the potential to prevent excessive ED utilization during this time period.

There are numerous dialysis care processes that can influence the likelihood of a patient requiring care in the ED in the 30 days following hospital discharge. These processes include:

(1) Timely evaluation of target weight: 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. This is particularly true in the period immediately following hospitalization where a patient's target weight may have changed abruptly.
(2) Inadequate infection prevention: Inadequate infection prevention processes, including suboptimal management of vascular access, can lead to bacteremia or septicemia, increasing the possibility of the need for ED use.
(3) Management of electrolyte abnormalities: Following hospitalization a patient's electrolyte and nutritional status may change abruptly and failure to maintain processes to ensure adequate dialysis and nutritional counseling can lead to either hypo- or hyperkalemia, increasing the possibility of the need for ED use.

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

****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.

Few interventions to decrease the frequency of ED use in the post-hospitalization period have been examined in the dialysis patient population, however there are effective interventions reported in this population to reduce hospital re-admission [3]. Acknowledging the strong association between ED encounters and subsequent hospitalization, these dialysis facility interventions would likely be effective in preventing outpatient ED encounters as well.

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 adherence to the treatment schedule would be expected to decrease ED utilization, particularly in the post-acute care period. 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].

ESRD patients are often discharged from the hospital to Skilled Nursing Facilities (SNF) before transitioning back to home. After discharge from a SNF back to home, dialysis patients who have visiting home health services are less likely to need acute care in the ED during the subsequent 30 day period [7] Finally, other critical activities in the post-hospitalization period focus on medication reconciliation, appointment scheduling, as well as appraisal of the target weight and volume management. This is particularly important since heart failure has been implicated as one of the most frequent reasons for an ED visit within 30 days of hospital discharge [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. 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.

3. Mathew, A. T.; Strippoli, G. F.;Ruospo, M.;Fishbane, S. Reducing hospital readmissions in patients with end-stage kidney disease**.** *Kidney Int.* 2015 88(6):1250-1260 doi:10.1038/ki.2015.307.

ESKD patients have a large burden of disease, with high rates of readmission to hospital compared with the general population. A readmission after an acute index hospital discharge is either planned or unplanned. A proportion of unplanned readmissions are potentially avoidable, and could have been prevented with optimized transitional care. Readmissions pose financial cost to the health care system and emotional cost to patients and caregivers. In other chronic diseases with high readmission risk, such as congestive heart failure, interventions have improved transitional care and reduced readmission risk. In reviewing the existing literature on readmissions in ESKD, the definition and risk of readmission varied widely by study, with many potentially associated factors including comorbid diseases such as anemia and hypoalbuminemia. An ESKD patient's requisite follow-up in the outpatient dialysis facility provides an opportunity to improve transitional care at the time of discharge. Despite this, our review of existing literature found no studies which have tested interventions to reduce the risk of readmission in ESKD patients. We propose a framework to define the determinants of avoidable readmission in ESKD, and use this framework to define a research agenda. Avoidable readmissions in ESKD patients is a topic prime for in-depth study, given the high-risk nature in this patient population, financial and societal costs, and potential for risk modification through targeted interventions.

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. Hall RK; Toles M; Massing M; Jackson E; Peacokc-Hinton S; O'Hare AM; Colon-Emeric C. Utilization of Acute Care among Patients with ESRD Discharged Home from Skilled Nursing Facilities. Clin J Am Soc Nephrol. 2015 10(3):428-434. doi: 10/2215/CJN.03510414

Background and objectives: Older adults with ESRD often receive care in skilled nursing facilities (SNFs) after an acute hospitalization; however, little is known about acute care use after SNF discharge to home.

Design, setting, participants, & measurements: This study used Medicare claims for North and South Carolina to identify patients with ESRD who were discharged home from a SNF between January 1, 2010 and August 31, 2011. Nursing Home Compare data were used to ascertain SNF characteristics. The primary outcome was time from SNF discharge to first acute care use (hospitalization or emergency department visit) within 30 days. Cox proportional hazards models were used to identify patient and facility characteristics associated with the outcome.

Results: Among 1223 patients with ESRD discharged home from a SNF after an acute hospitalization, 531 (43%) had at least one rehospitalization or emergency department visit within 30 days. The median time to first acute care use was 37 days. Characteristics associated with a shorter time to acute care use were black race (hazard ratio [HR], 1.25; 95% confidence interval [95% CI], 1.04 to 1.51), dual Medicare-Medicaid coverage (HR, 1.24; 95% CI, 1.03 to 1.50), higher Charlson comorbidity score (HR, 1.07; 95% CI, 1.01 to 1.12), number of hospitalizations during the 90 days before SNF admission (HR, 1.12; 95%CI, 1.03 to 1.22), and index hospital discharge diagnoses of cellulitis, abscess, and/or skin ulcer (HR, 2.59; 95% CI, 1.36 to 4.45). Home health use after SNF discharge was associated with a lower rate of acute care use (HR, 0.72; 95%CI, 0.59 to 0.87). There were no statistically significant associations between SNF characteristics and time to first acute care use.

Conclusions: Almost one in every two older adults with ESRD discharged home after a post–acute SNF stay used acute care services within 30 days of discharge. Strategies to reduce acute care utilization in these patients are needed.

8. Harel, Z.;Wald, R.;McArthur, E.;Chertow, G. M.;Harel, S.;Gruneir, A.;Fischer, H. D.;Garg, A. X.;Perl, J.;Nash, D. M.;Silver, S.;Bell, C. M. Rehospitalizations and Emergency Department Visits after Hospital Discharge in Patients Receiving Maintenance Hemodialysis**.** *J Am Soc Nephrol.* 2015 26(12):3141-50 doi:10.1681/ASN.2014060614

   Clinical outcomes after a hospital discharge are poorly defined for patients receiving maintenance in-center (outpatient) hemodialysis. To describe the proportion and characteristics of these patients who are rehospitalized, visit an emergency department, or die within 30 days after discharge from an acute hospitalization, we conducted a population-based study of all adult patients receiving maintenance in-center hemodialysis who were discharged between January 1, 2003, and December 31, 2011, from 157 acute care hospitals in Ontario, Canada. For patients with more than one hospitalization, we randomly selected a single hospitalization as the index hospitalization. Of the 11,177 patients included in the final cohort, 1926 (17%) were rehospitalized, 2971 (27%) were treated in the emergency department, and 840 (7.5%) died within 30 days of discharge. Complications of type 2 diabetes mellitus were the most common reason for rehospitalization, whereas heart failure was the most common reason for an emergency department visit. In multivariable analysis using a cause-specific Cox proportional hazards model, the following characteristics were associated with 30-day rehospitalization: older age, the number of hospital admissions in the preceding 6 months, the number of emergency department visits in the preceding 6 months, higher Charlson comorbidity index score, and the receipt of mechanical ventilation during the index hospitalization. Thus, a large proportion of patients receiving maintenance in-center hemodialysis will be readmitted or visit an emergency room within 30 days of an acute hospitalization. A focus on improving care transitions from the inpatient setting to the outpatient dialysis unit may improve outcomes and reduce healthcare costs.

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**	**N/A**

• **URL**	
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	
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 encounters within 30 days of an index discharge are an important indicator of care coordination, care transitions, 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). This has been demonstrated in the end-stage renal disease (ESRD) population as well with 27% of patients being treated in an ED within 30 days of hospital discharge, most frequently for congestive heart failure (Harel et al., 2015)

Rates of ED visits among end ESRD dialysis patients have increased between 2007 and 2016. As reported by the USRDS, 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), 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 encounters subsequent to a hospital discharge may help dialysis facility efforts to prevent emergent unscheduled care and to help control escalating medical costs, for example through greater care coordination and post-discharge transitional care. Specifically, dialysis facility activities such as evaluation of the patients target weight or medication reconciliation and review may help reduce the risk of ED encounters after hospital discharge. This measure will complement existing measures targeting care

coordination (such as the Standardized Readmission Ratio NQF #2496) by identifying impactful events that can be influenced by dialysis facility care.

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

Harel, Z.;Wald, R.;McArthur, E.;Chertow, G. M.;Harel, S.;Gruneir, A.;Fischer, H. D.;Garg, A.

X.;Perl, J.;Nash, D. M.;Silver, S.;Bell, C. M. Rehospitalizations and Emergency Department Visits after Hospital Discharge in Patients Receiving Maintenance Hemodialysis. J Am Soc Nephrol. 2015 26(12):3141-50 doi:10.1681/ASN.2014060614

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

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.

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 (<u>current and over time</u>) at the specified level of analysis. *(<u>This is required for maintenance of endorsement</u>. 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.*

After applying all exclusion criteria, we evaluated all Medicare-certified dialysis facilities (n=6,728) treating Medicare dialysis patients (n=271,852) that had at least 11 index discharges in 2016-2017. Median facility size was 65 patients. The distribution of ED30 across these facilities is shown in the tables below. ED30 rates vary widely across facilities. For example, for the 6,728 facilities included in 2016-2017, the ED30 varied from 0.00 to 3.52. The mean value was 1.03 and the SD was 0.37.

Deciles of ED30 for 2016-2017:

Decile 1, N=672 Min=0.00 Max=0.60

Decile 2, N=673 Min=0.60 Max=0.73

Decile 3, N=673 Min=0.73 Max=0.83

Decile 4, N=673 Min=0.83 Max=0.92

Decile 5, N=673 Min=0.92 Max=1.00

Decile 6, N=673 Min=1.00 Max=1.09

Decile 7, N=673 Min=1.09 Max=1.18

Decile 8, N=673 Min=1.18 Max=1.29

Decile 9, N=673 Min=1.29 Max=1.48

Decile 10, N=672 Min=1.48 Max=3.52

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.

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 disparities in emergency department utilization within 4-30 days of an inpatient discharge. Differences are observed by age (younger age), race (blacks), ethnicity (Hispanic), and dual Medicare-Medicaid status.

Compared to the reference age group, those who were younger had higher odds of an emergency department encounter subsequent to a recent discharge (4-30 days). The odds were highest for 18-<25 year olds, with a monotonic decline in odds for the 25-<44 age group, and the 45-<59 age group. Compared to whites, black patients had 16% higher odds of an emergency department encounter within 4-30 days. Compared to non-Hispanic patients, Hispanic patients had 6 % higher odds of an emergency department encounter within 4-30 days. Finally, those with dual Medicare-Medicaid status had 14% higher odds of an ED encounter within 4-30 days of an index discharge.

While there are notable differences by younger age, black race, and Hispanic ethnicity, as well as Medicare coverage type, it is unclear if these disparities in emergency department encounters following discharge from an inpatient admission 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.

The odds of an ED visit after a discharge are shown below for various patient subgroups.

Age:

For the 18-<24 age group: OR = 1.48, p-value <.0001.

For the 25-<44 age group: OR = 1.31, p-value <.0001.

For the 45-<59 age group: OR = 1.11, p-value <.0001.

The 60-<75 age group was used as the reference group.

For the 75+ age group: OR = 0.97, p-value 0.0036.

Sex:

For Female: OR = 0.98, p-value 0.0283

Male was used as the reference group.

Race:

White was used as the reference group.

For Black: OR = 1.16, p-value <.0001.

For Native American/Alaskan Native: OR = 0.99, p-value = 0.8813.

For Asian/Pacific Islander: OR = 0.98, p-value 0.3490.

For Other race: OR =1.18, p-value = 0.0382.

Ethnicity:

For Hispanic: OR = 1.06, p-value = 0.0001.

Non-Hispanic was used as the reference group.

For Unknown: OR = 0.90, p-value = 0.2405.

Employment Status:

Unemployed was used as the reference group.

For Employed: OR = 0.94, p-value <0.0001.

For Other: OR = 0.95, p-value <0.0001

Medicare Coverage:

Non-dual eligible was used as the reference group.

Dual eligible: OR = 1.14 p-value < 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. Reliability and Validity—Scientific Acceptability of Measure Properties

Extent to which the measure, as specified, produces consistent (reliable) and credible (valid) results about the quality of care when implemented. ***Measures must be judged to meet the sub criteria for both reliability and validity to pass this criterion and be evaluated against the remaining criteria.***

2a.1. Specifications The measure is well defined and precisely specified so it can be implemented consistently within and across organizations and allows for comparability. eMeasures should be specified in the Health Quality Measures Format (HQMF) and the Quality Data Model (QDM).

De.5. Subject/Topic Area *(check all the areas that apply):*

De.6. Non-Condition Specific*(check all the areas that apply):*

De.7. Target Population Category *(Check all the populations for which the measure is specified and tested if any):*

S.1. Measure-specific Web Page *(Provide a URL link to a web page specific for this measure that contains current detailed specifications including code lists, risk model details, and supplemental materials. Do not enter a URL linking to a home page or to general information.)*

N/A

S.2a. If this is an eMeasure, HQMF specifications must be attached. Attach the zipped output from the eMeasure authoring tool (MAT) - if the MAT was not used, contact staff. (Use the specification fields in this online form for the plain-language description of the specifications)

This is not an eMeasure **Attachment:**

S.2b. Data Dictionary, Code Table, or Value Sets *(and risk model codes and coefficients when applicable) must be attached. (Excel or csv file in the suggested format preferred - if not, contact staff)*

Attachment **Attachment:** ED30_Data_Dictionary_Code_Table.xlsx

S.2c. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

No, this is not an instrument-based measure **Attachment:**

S.2d. Is this an instrument-based measure (i.e., data collected via instruments, surveys, tools, questionnaires, scales, etc.)? Attach copy of instrument if available.

Not an instrument-based measure

S.3.1. <u>For maintenance of endorsement:</u> Are there changes to the specifications since the last updates/submission. If yes, update the specifications for S1-2 and S4-22 and explain reasons for the changes in S3.2.

No

S.3.2. <u>For maintenance of endorsement,</u> please briefly describe any important changes to the measure specifications since last measure update and explain the reasons.

S.4. Numerator Statement *(Brief, narrative description of the measure focus or what is being measured about the target population, i.e., cases from the target population with the target process, condition, event, or outcome) DO NOT include the rationale for the measure.*

<u>IF an OUTCOME MEASURE</u>, state the outcome being measured. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

The observed number of index hospital discharges during a year that are followed by an emergency department encounter within 4–30 days of the discharge among eligible adult Medicare patients at a facility.

S.5. Numerator Details *(All information required to identify and calculate the cases from the target population with the target process, condition, event, or outcome such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b)*

<u>IF an OUTCOME MEASURE,</u> describe how the observed outcome is identified/counted. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

Index Discharges

We use Medicare inpatient hospital claims to identify acute hospital discharges. Among these acute hospital discharges, all live discharges of eligible patients in a calendar year are considered eligible for this measure. Those that do not meet one of the index discharge exclusion criteria described in the next section are considered index discharges.

Assignment of Index Discharges to Facilities

Index discharges are attributed to the facility of record on the day of discharge for the patient. That is, if the patient transfers dialysis facilities at the time of hospital discharge, it is the new facility that is assigned the index discharge.

Emergency Department Encounters

Emergency department (ED) encounters are identified from Medicare outpatient claims using revenue center codes that indicate an ED visit (0450, 0451, 0452, 0453, 0454, 0455, 0456, 0457, 0458, 0459, 0981). Note that this means that we include both outpatient ED visits and those that result in an observation stay, but not those that result in a hospital admission. Outpatient ED claims that have overlapping or consecutive dates of service are combined and considered as a single ED encounter. To further ensure that these outpatient ED encounters are distinct from those associated with hospitalizations, we exclude ED encounters where there is an inpatient claim that has dates of service included in any of the same time period covered by the ED encounter.

An ED encounter "follows" the index discharge only if there is no intervening inpatient hospitalization. In other words, if after hospital discharge there is another inpatient hospitalization and then an ED encounter within the time frame the original index discharge is not counted as having been followed by an ED encounter. If eligible, the second hospitalization could become a new index discharge. The measure does not count the number of ED encounters after each index discharge, but instead determines whether or not there is at least one such encounter. If there are multiple ED encounters during days 4-30 after an index discharge, only the

first ED encounter during that time is relevant to determining whether or not the index discharge is counted as having been followed by an ED encounter. ED encounters that occur before the 4th day after index discharge are not considered.

The 4-30 day time frame was selected to harmonize with the Standardized Readmission Ratio (NQF #2496) that also uses the same time period after an index hospitalization. This time interval was selected in response to providers and stakeholders concerns that there may be up to 72 hours before a patient is seen at the facility after hospital discharge.

The time period for the measure calculation is two calendar years, meaning that index discharges must occur during the two calendar year period. The subsequent ED encounters may occur during the calendar years or the first 30 days of the following calendar year.

S.6. Denominator Statement *(Brief, narrative description of the target population being measured)*

The expected number of index hospital discharges for eligible adult Medicare ESRD dialysis patients during the two year period that are followed by an emergency department encounter within 4-30 days of the discharge among eligible patients at a facility. The expected value is the result of a risk-adjusted predictive model adjusted for the characteristics of the patients, the dialysis facility, and the discharging hospitals.

S.7. Denominator Details *(All information required to identify and calculate the target population/denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

IF an OUTCOME MEASURE, describe how the target population is identified. Calculation of the risk-adjusted outcome should be described in the calculation algorithm (S.14).

We use Medicare inpatient hospital claims to identify acute hospital discharges. Among these acute hospital discharges, all live discharges of eligible patients in a calendar year are considered eligible for this measure. See Numerator Details section above for definitions index discharges, patients assignment, and ED encounters.

General Inclusion Criteria for Dialysis Patients

To be eligible for the measure a patient must be an adult (aged 18 or more) Medicare dialysis patient with at least 90 days of ESRD treatment on date of index discharge. The 90 days of ESRD are counted without regard to which facility, or the number of facilities, a patient received their dialysis treatments. The date of index discharge is considered day 0 when identifying ED visits within 4-30 days of discharge.

Expected Calculation

We calculate each dialysis facility's expected number of index hospital discharges during the two year period that are followed by an ED encounter within 4-30 days of the discharge. The expected number is calculated by fitting a model with random effects for discharging hospitals, fixed effects for facilities, and regression adjustments for a set of patient-level characteristics. We compute the expectation for the given facility assuming ED encounter rates corresponding to an "average" facility with the same patient characteristics and same discharging hospitals as this facility. Model details are provided in the testing form.

S.8. Denominator Exclusions *(Brief narrative description of exclusions from the target population)*

Index Discharge exclusions that are implicit in the denominator definition include discharges for which the patient:

- Has Medicare Advantage coverage at the time of the index discharge
- Has had ESRD for 90 days or less at time of discharge
- Is less than 18 years of age at the time of discharge

We also exclude discharges and emergency department encounters for which the patient was:

- Actively enrolled in hospice at any time of during the calendar month of the discharge date or ED encounter admit date

Outpatient Medicare claims are the source of ED encounter data, and since outpatient claims are not available for Medicare Advantage (MA) patients, we cannot identify ED encounters for MA patients. Therefore, we exclude index discharges for patients with MA at the time of discharge.

The hospice exclusion is needed because hospice patients are considered to be under the purview of hospice care givers and may have other reasons for Emergency Department use such as pain management.

Additionally we exclude hospital discharges that:

• Do not result in a live discharge

• Are against medical advice

• Include a primary diagnosis for cancer, mental health or rehabilitation (see below for excluded CCSs)

• Are from a PPS-exempt cancer hospital

• Are followed within three days of discharge by the patient being transplanted, discontinuing dialysis, recovering renal function, being lost to follow-up, having another hospitalization, or having an emergency department visit

S.9. Denominator Exclusion Details *(All information required to identify and calculate exclusions from the denominator such as definitions, time period for data collection, specific data collection items/responses, code/value sets – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format at S.2b.)*

• Death in hospital: We determine a patient's death date from a number of sources: CMS Medicare Enrollment Database, CMS forms 2746 and 2728, OPTN transplant follow-up form, CROWNWeb database, Social Security Death Master File, and Inpatient Claims. In addition, if the discharge status on the index discharge claim indicates death and the death date occurs within 5 days after discharge we consider this a death in the hospital.

• Discharged against medical advice: We determine discharge status from the inpatient claim.

• Certain diagnoses: The primary diagnosis at discharge is available on the inpatient claim; we group these diagnoses into more general categories using AHRQ's Clinical Classification Software (CCS; see http://www.hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp for descriptions of each CCS).

The excluded CCSs for a primary diagnosis for cancer, mental health or rehabilitation are shown below.

o Cancer: 42, 19, 45, 44, 17, 38, 39, 14, 40, 35, 16, 13, 29, 15, 18, 12, 11, 27, 33, 32, 24, 43, 25, 36, 21, 41, 20, 23, 26, 28, 34, 37, 22, 31, 30

o Psychiatric: 657, 659, 651, 670, 654, 650, 658, 652, 656, 655, 662

o Rehab for prosthesis: 254

• PPS-exempt cancer hospitals: The following hospitals are listed as PPS-exempt cancer hospitals in the Federal Register (http://www.gpo.gov/fdsys/pkg/FR-2011-07-18/html/2011-16949.htm): 050146, 050660, 100079, 100271, 220162, 330154, 330354, 360242, 390196, 450076, 500138

• Are followed within three days of discharge by the patient being transplanted, discontinuing dialysis, recovering renal function, being lost to follow-up, having another hospitalization, or having an emergency department visit. We determine transplant status from OPTN, CROWNWeb, and dialysis claims, and discontinuation of dialysis or recovery of renal function from CROWNWeb.

S.10. Stratification Information *(Provide all information required to stratify the measure results, if necessary, including the stratification variables, definitions, specific data collection items/responses, code/value sets, and the risk-model covariates and coefficients for the clinically-adjusted version of the measure when appropriate – Note: lists of individual codes with descriptors that exceed 1 page should be provided in an Excel or csv file in required format with at S.2b.)*

N/A

S.11. Risk Adjustment Type (Select type. Provide specifications for risk stratification in measure testing attachment)

Statistical risk model

If other:

S.12. Type of score:

Ratio

If other:

S.13. Interpretation of Score *(Classifies interpretation of score according to whether better quality is associated with a higher score, a lower score, a score falling within a defined interval, or a passing score)*

Better quality = Lower score

S.14. Calculation Algorithm/Measure Logic (*Diagram or describe the calculation of the measure score as an ordered sequence of steps including identifying the target population; exclusions; cases meeting the target process, condition, event, or outcome; time period for data, aggregating data; risk adjustment; etc.)*

See Flowchart in Appendix.

S.15. Sampling *(If measure is based on a sample, provide instructions for obtaining the sample and guidance on minimum sample size.)*

IF an instrument-based performance measure (e.g., PRO-PM), identify whether (and how) proxy responses are allowed.

N/A

S.16. Survey/Patient-reported data *(If measure is based on a survey or instrument, provide instructions for data collection and guidance on minimum response rate.)*

Specify calculation of response rates to be reported with performance measure results.

N/A

S.17. Data Source *(Check ONLY the sources for which the measure is SPECIFIED AND TESTED).*

If other, please describe in S.18.

Claims, Registry Data

S.18. Data Source or Collection Instrument *(Identify the specific data source/data collection instrument (e.g. name of database, clinical registry, collection instrument, etc., and describe how data are collected.)*

IF instrument-based, identify the specific instrument(s) and standard methods, modes, and languages of administration.

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) and Part B (outpatient) claims.

S.19. Data Source or Collection Instrument *(available at measure-specific Web page URL identified in S.1 OR in attached appendix at A.1)*

No data collection instrument provided

S.20. Level of Analysis *(Check ONLY the levels of analysis for which the measure is SPECIFIED AND TESTED)*

Facility

S.21. Care Setting *(Check ONLY the settings for which the measure is SPECIFIED AND TESTED)*

Other

If other: Dialysis Facility

S.22. <u>COMPOSITE Performance Measure</u> - Additional Specifications *(Use this section as needed for aggregation and weighting rules, or calculation of individual performance measures if not individually endorsed.)*

N/A

2. Validity – See attached Measure Testing Submission Form

ED30_Testing-637139155825464897.docx

2.1 <u>For maintenance of endorsement</u>

Reliability testing: If testing of reliability of the measure score was not presented in prior submission(s), has reliability testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.2 <u>For maintenance of endorsement</u>

Has additional empirical validity testing of the measure score been conducted? If yes, please provide results in the Testing attachment. Please use the most current version of the testing attachment (v7.1). Include information on all testing conducted (prior testing as well as any new testing); use red font to indicate updated testing.

2.3 <u>For maintenance of endorsement</u>

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 Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge for Dialysis Facilities (ED30)
Date of Submission: 1/5/2020

Type of Measure:

☒ **Outcome (*including PRO-PM*)**	☐ **Composite – *STOP – use composite testing form***

☐ Intermediate Clinical Outcome	☐ Cost/resource
☐ Process *(including Appropriate Use)*	☐ Efficiency
☐ Structure	

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? Click here to enter date range

January 2016 – December 2017 for index discharges

January 2016 – January 2018 for emergency department encounters

January 2015 – December 2016 for prior year comorbidities

1.4. What levels of analysis were tested? (*testing must be provided for all 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 measured entities 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*)

We included all Medicare-certified facilities (N= 6728) that had at least 11 eligible index discharges in 2016 and 2017. Median facility size was 65 patients as measured by all dialysis patients in the facility on December 31, 2017.

1.6. How many and which patients 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*)

Before applying the exclusion criteria, we had 304,941 Medicare ESRD dialysis patients with at least one hospitalization during 2016 and 2017. There were 995,788 hospitalizations for this patient population. After applying the exclusion criteria, the data used to calculate the measure included 757,497 index discharges among 271,852 unique patients during 2016 and 2017.

Table 1. Patient Demographics of Index Discharges Included in the Measure

Patient Demographics	**Percent of Index Discharges**
Age	
Patient Age: 18 to 24	0.67
Patient Age: 25 to 44	11.84

Patient Demographics	Percent of Index Discharges
Patient Age: 45 to 59	27.56
Patient Age: 60 to 74	39.74
Patient Age: 75+	20.20
Sex (% female)	47.75
ESRD due to Diabetes	48.80
Medicare coverage	
Dual eligible	44.24
Non-dual eligible	55.76
Time since Start of ESRD	
91 days-6 months	4.95
6 months-1 year	8.34
1-2 years	14.43
2-3 years	12.94
3-5 years	20.86
5+ years	38.49
Employment status 6 months prior to ESRD	
Unemployed	25.11
Employed	19.98
Other/Unknown*	54.91
Race	
White	57.21
Black	36.22
Native American/Alaskan Native	1.27
Asian/Pacific Islander	3.91
Other/Unknown	1.39
Ethnicity	
Hispanic	15.10
Non-Hispanic	83.70
Unknown	1.19

* Other/Unknown groups includes homemaker, retired due to age/preference, retired due to disability, medical leave of absence, or missing employment status.

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.

**Assessed at date of index discharge.*

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*)

If the measure were a simple average across individuals in the facility, the NQF-recommended 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 measure variability that is attributable to the between-facility variance. The ED30, however, is not a simple average and we instead estimate the IUR using a bootstrap approach (see reference 3 below), which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by ANOVA.

Suppose that there are N facilities with at least 11 discharges in the year. Let $T_1,\ldots,T_N$ be the ED30 for these facilities. Within each facility, select at random and with replacement $B = 100$ bootstrap samples. That is, if the ith facility has n_i subjects, randomly draw with replacement n_i subjects from those in the same facility, find their corresponding ED30$_i$ and repeat the process 100 times. Thus, for the ith facility, we have bootstrapped ED30s of $T^*_{i1}\ldots, T^*_{i100}$. 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 ED30, 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 ED30 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 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 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 only included facilities that had at least 11 eligible index discharges in 2016-2017.

To assess more directly the value of ED30 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 statistical 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*)

Overall, we found that IUR = 0.451. The PIUR is 0.570. 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, as described in section 2b5.3, the measure demonstrates it 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 ED30 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 creating a measure of ED use within 30 days of hospital discharge would complement the existing SRR measure while providing a more complete picture of care coordination in the outpatient setting. 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 ED30 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 ED30 we first stratified facilities into the 'better than/as expected' 'and 'worse than expected' categories of the ED30 ratio. Next we calculated mean performance scores for several quality measures: Standardized Transfusion Ratio (STrR), Standardized Fistula Rate (SFR), Percentage of Prevalent Patients Waitlisted (PPPW), Standardized Readmission Ratio (SRR), and Standardized Emergency Department Visit Ratio (SEDR). We then compared mean performance scores across the two strata of 'better than/as expected' 'and 'worse than expected' categories for ED30. 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 ED30 compared to facilities classified as 'worse than expected.' Compared to facilities that perform 'worse than expected', facilities that perform 'better than/as expected' on ED30 are likely to have more successful care coordination and other processes of care in place that may help patients avoid an ED visit in the vulnerable period following a recent discharge:

- 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 after hospitalization may not have care processes in place to support these transitions in care.
- STrR: We expect to observe a lower mean standardized transfusion event ratio for facilities in the 'better than/as expected' category for ED30 compared to facilities classified as 'worse than expected.' Facilities that have a low STrR likely have processes of care in place to support robust anemia management and care transitions for patients recently discharged, compared to facilities with a higher STrR.
- SFR – We expect to observe a higher mean standardized fistula rate for facilities in the 'better than/as expected' category for ED30 compared to facilities classified as 'worse than expected.' Facilities that have higher standardized fistula rates suggest they have more robust processes to coordinate care outside of the dialysis facility. Facilities that do a better job at care coordination and care transitions

reduce the likelihood that patients recently discharged will experience an 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 ED30 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 care transition processes that are expected to also reduce the likelihood that patients recently discharged will experience an acute event resulting in an ED visit.
- SRR: We expect that facilities classified as 'worse than expected' for ED30 will have a standardized readmission ratio that is close to the national norm. ED30 only captures outpatient ED visits that do not result in an (re)admission which, by definition, is a different patient subpopulation than SRR. Patients that require post-acute care from the ED without a readmission likely have lower acuity medical needs that can be handled in an outpatient setting without readmission. Therefore we do not expect ED30 flagging to be related to how facilities perform on SRR.
- SEDR: We expect to observe a lower mean standardized emergency department visit ratio for facilities classified as 'better than/as expected' for ED30 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, SEDR is a companion measure to ED30 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, SRR, and SEDR by classification of facilities as 'better than/as expected' versus 'worse than expected' for ED30 (Table 2).

Table 2. Classification of ED30 and mean facility performance scores for Related Measures, 2016-2017

	Facilities Missing	Better than/As Expected	Worse than expected	As Hypothesized?
SMR	412	1.00	1.05	Yes
STrR	717	0.99	1.21	Yes
SFR	510	63.32	63.64	No
PPPW	341	19.70	14.71	Yes

SRR	346	1.00	1.00	Yes
SEDR	205	0.99	1.49	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 5% 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 ED30.

On average the standardized transfusion event ratio was 21% higher than the national average for facilities classified as 'worse than expected' and 1% lower for facilities classified as 'better than/as expected.' This suggests that facilities which have lower standardized transfusion event ratios may have better processes of care in place to support robust anemia management and care transitions for patients recently discharged.

For SFR, the mean standardized fistula rate was 63.3% in facilities classified as 'better than/as expected' and 63.6% in facilities classified as 'worse than expected.' While we would expect better standardized fistula rates for facilities classified as 'better than/as expected' for ED30 compared to facilities classified as 'worse than expected' this result is consistent with recent national trends in AVF rates that have generally plateaued across many US dialysis facilities. This is also a small difference and it is possible there are other facility level factors that differ between long term vascular access planning and shorter term response after hospital discharge.

The mean standardized percentage of prevalent patients waitlisted (PPPW) in facilities classified as 'better than/as expected' was 19.7% compared to facilities classified as 'worse than expected' 14.7%, suggesting that facilities that have higher rates of prevalent 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 effective care coordination and care transition processes that are expected to help reduce patient utilization of the ED for many post-acute care needs.

SRR: Facilities classified as 'better than/as expected' and those classified as 'worse than expected' for ED30 performed no better or worse on SRR. The mean SRR (1.00) was the same across both flagging categories for ED30 indicating flagging is not related to how facilities perform on SRR. This suggests that both measures are capturing different patient subpopulations and different facets of facility care quality and highlights the need for this measure of ED use after hospital discharge since these patients are not represented in the SRR.

On average facilities classified as 'better than/as expected' for ED30 had a lower standardized emergency department visit ratio compared to the national average, by about 1%, while facilities classified as 'worse than expected' had an SEDR that was 51% higher than the national average. These results reinforce that both ED30 and SEDR assess complementary elements of care. These include 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 many post-acute care needs.

Taken together these results provide validation support for ED30. Performance on key quality measures 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 ED30 measure with and without the hospice exclusion. Additionally, we calculated the number and percentage of excluded discharges for each exclusion criterion 4-9 below.

Exclusions that are implicit in the denominator definition include discharges for which the patient:

1. Has Medicare Advantage (MA) coverage at the time of the index discharge
2. Has had ESRD for 90 days or less at time of discharge
3. Is less than 18 years of age at the time of discharge

We also exclude discharges and emergency department encounters for which the patient was:

4. Actively enrolled in hospice at any time during the calendar month of the discharge or ED encounter
5. Do not result in a live discharge
6. Are against medical advice
7. Include a primary diagnosis for cancer, mental health or rehabilitation
8. Are from a PPS-exempt cancer hospital
9. Discharges followed within 3 days by an ED visit, any hospitalization (at acute care, long-term care, rehabilitation, or psychiatric hospital or unit), transplant, loss to follow-up, withdrawal from dialysis, or recovery of renal function

Based on input from the May 2016 TEP, we exclude 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.

Exclusion criteria 5-8 are aligned with the Standardized Readmission Ratio (NQF 2496) measure. For exclusion 9, the 4-30 day time frame was selected to harmonize with the Standardized Readmission Ratio (NQF #2496) that also uses the same time period after an index hospitalization. This time interval was selected in response to providers and stakeholders concerns that there may be up to 72 hours before a patient is seen at the facility after hospital discharge.

Medicare Advantage patients are excluded from ED30 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*)

A total of 238,291 hospital discharges among 33,089 unique patients were excluded. The number and percentage of excluded discharges are as follows:

Exclusions that are implicit in the denominator definition include discharges for which the patient:

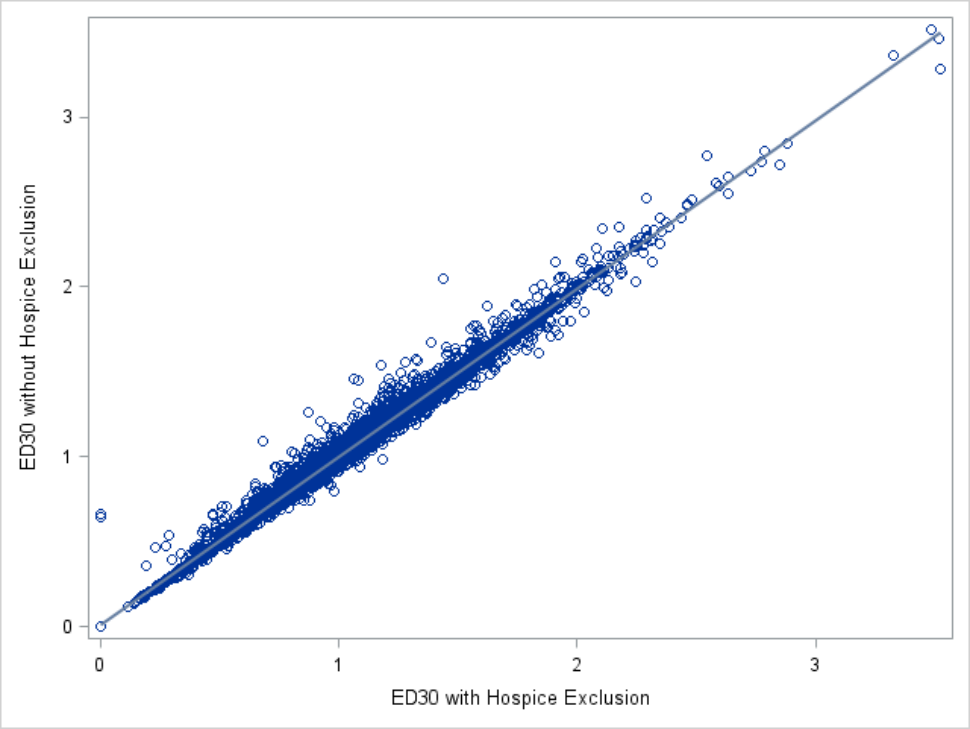
1. Has Medicare Advantage coverage at the time of discharge
2. Has had ESRD for 90 days or less at time of discharge
3. Is less than 18 years of age at the time of discharge

We also exclude the following discharges:

4. Patient actively enrolled in hospice at time of discharge (*n* = 21,550; 2.16%)
5. Do not result in a live discharge (*n* =45,826; 4.60%)
6. Are against medical advice (*n* =20,263; 2.03%)
7. Include a primary diagnosis for cancer, mental health or rehabilitation (*n* =10,784; 1.08%)
8. Are from a PPS-exempt cancer hospital (*n* = 515; 0.05%)
9. Result in another hospitalization, emergency department visit, transplant within three days of discharge; or loss to follow-up, withdrawal from dialysis, or recovery of renal function (n=135,939; 13.65%)
10. Facility has less than 11 index discharges (n = 3,414; 0.34%)

As shown in Figure 1, we compared each facility's ED30 with and without the hospice exclusion and found the two measures to be highly correlated (overall Pearson correlation coefficient [r] = 0.9933, p-value <0.0001).

Figure 1. Correlation between ED30 with and without the hospice exclusion (2016-2017)



Overall Correlation=0.9933, p-value <0.0001

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 is highly correlated suggesting the overall impact on the measure's validity is 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 87 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.

The model accounts for a set of patient-level characteristics:

- Sex
- Age
- Years on dialysis
- Diabetes as cause of ESRD
- Nursing Home status in previous 365 days
    - No Nursing Home care (0 days)
    - Short-term Nursing Home care (1 - 89 days)
    - Long-term Nursing Home care (90 - 365 days)
- BMI at incidence of ESRD
    - <18.5
    - 18.5-25
    - 25-30
    - >=30
- Length (days) of index hospitalization
- A set of prevalent comorbidities based on Medicare inpatient claims (individual comorbidities categorized into 66 groups).

Prevalent comorbidities are determined using the previous 12 months of CMS claims after the index discharge. We grouped individual comorbidities into clinically related categories. Each comorbidity group is included as a separate covariate in the model.

To estimate the probability of 30-day emergency department encounter, we use a three-stage model, the first of which is a fixed-effects logistic regression model. In this first stage, facility-hospital combinations are included as fixed effects, adjusting for a set of patient-level characteristics. The results of this step are estimates of the regression coefficients of patient-level characteristics in the logistic regression model. These regression coefficients are then used as an offset variable in the second stage model. The results from this model are unbiased regardless of correlation between hospital effects and patient-case mix.

The second stage of the model is a double random-effects logistic regression model. In this stage of the model, both dialysis facilities and hospitals are represented as random effects, and the sum of regression adjustments multiplied by estimated parameters obtained from the first stage are included as the offset variable. From this model, we obtain the estimated standard deviation of the random effects of hospitals (Diggle, et. al., 2002).

The third stage of the model is a mixed-effects logistic regression model, in which dialysis facilities are modeled as fixed effects and hospitals are modeled as random effects, with the standard deviation specified as equal to its estimates from the second stage and the estimated parameters obtained in the first stage are included as the offset variable. The expected number of emergency department encounters for each facility is estimated as the summation of the probabilities of emergency department encounters of all patients in this facility and assuming the national norm (i.e., the median) for facility effect. This model accounts for a given facility's case mix using the same set of patient-level characteristics as those in the first stage model.

The equations used in the measure calculation are as follows:

- To estimate the probability of 30-day emergency department encounter, we use a three-stage approach. The main model, which produces the estimates used to calculate ED30, takes the form:

$$\log \frac{p_{ijk}}{1-p_{ijk}} = \gamma_i + \alpha_j + \beta^T Z_{ijk}\,, \qquad (1)$$

where p_{ijk} represents the probability of an emergency department encounter for the k^{th} discharge among patients from the i^{th} facility who are discharged from j^{th} hospital, and Z_{ijk} represents the set of patient-level characteristics. Here, $\gamma_i$is the fixed effect for facility and α_j is the random effect for hospital j. It is assumed that the α_js arise as independent normal variables (i.e., $\alpha_j \sim N(0, \sigma^2)$).

- We then use the estimates from this model to calculate each facility's ED30:

$$ED30_i = \frac{O_i}{E_i} = \frac{O_i}{\sum_{j \in H(i)} \sum_{k=1}^{n_{ij}} \tilde{p}_{ijk}}, \quad (2)$$

where, for the i^{th} facility, O_i is the number of observed emergency department encounter, E_i is the expected number of emergency department encounter for discharges, $H(i)$ is the collection of indices of hospitals from which patients are discharged, and $\tilde{p}_{ijk}$ is the predicted probability of emergency department encounter under the national norm for each discharge. Specifically, $\tilde{p}_{ijk}$ takes the form

$$\tilde{p}_{ijk} = \frac{\exp(\widehat{\gamma_M} + \widehat{\alpha_j} + \hat{\beta}^T Z_{ijk})}{1 + \exp(\widehat{\gamma_M} + \widehat{\alpha_j} + \hat{\beta}^T Z_{ijk})}, \quad (3)$$

which estimates the probability that a discharge from hospital *j* of an individual in facility *i* with characteristics Z_{ijk} would result in an emergency department encounter if the facility effect corresponded to the median of national facility effects, denoted by $\widehat{\gamma_M}$. Here, $\widehat{\alpha_j}$ and $\hat{\beta}$ are estimates from model (1). The sum of these probabilities is the expected number of emergency department encounter E_i at facility *i*; e.g., the number of emergency department encounter that would have been expected in facility *i* had they progressed to the emergency department encounter at the same rate as the national population of dialysis patients. If a facility has less than 11 discharges, they are excluded from the measure for the purposes of modeling.

Reference:

Diggle PJ, Heagerty P, Liang K-Y, Zeger SL. Analysis of Longitudinal Data. 2 New York: Oxford Univ. Press; 2002.

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 <u>not risk adjusted or stratified</u>, provide <u>rationale and analyses</u> 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 <u>and</u> 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?

Selection of clinical factors: The list of covariates considered was based on CMS' Standardized Readmission Ratio for Dialysis Facilities (NQF 2496). The specific list of prevalent comorbidities included was determined

based on an empirical evaluation of prevalent comorbidities associated with risk of an ED encounter. Therefore, ED30 includes a different set of comorbidities than the SRR.

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, can 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 [1] 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 [2,3] 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. Friedman, J.H. (2001). Greedy function approximation: A gradient boosting machine. Annals of Statistics, 29(5), 1189-1232.

2. 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.

3. Efron, B. (2012). Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction Institute of Mathematical Statistics Monographs, Cambridge University Press.

As described in 2b3.1.1,we fit the model adjusting for patient demographic factors and comorbidities identified in the above process and checked for statistical significance.

We conducted all analyses in R and SAS. The analyses presented here are based on ICD-9 and ICD-10 codes.

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 (ED visits within 30 days of a hospital discharge), 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 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, race may interact with lower income, neighborhood poverty, residential segregation, levels of educational attainment, and unemployment levels that jointly influence key health outcomes related to morbidity and acute care use (Williams 2006; Williams and Collins, 2001).

Race, insurance status (dual-eligibility), 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 patient that are privately insured versus 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 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, and 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).

Emergency department utilization after an acute visit is associated with age and insurance type. For example, Hastings et al., report that Medicare beneficiaries that had a return ED visit or other acute care encounter were associated with older age, and Medicaid status, along with higher chronic health burden (Hastings et al, 2008). Chu and Pei (1999, pg. 220) found that in addition to clinical risk factors, socioeconomic characteristics of patient were predictive of early emergency readmission among elderly patient population.

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, along with 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 adherence to dialysis treatment.

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 have 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 as 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.

Chu L, Pei C. Risk factors for early emergency hospital readmission in elderly medical patients. Gerontology. 1999 Jul-Aug;45(4):220-6.

Colligan EM, 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.

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]

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 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.

Schold JD, Elfadawy N, Buccini LD, Goldfarb DA, Flechner SM, P Phelan M, Poggio ED. 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 YC. 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?

In Table 3 below, we list results from the adjusted baseline model described above.

Table 3. Baseline ED 30 Model Coefficients and Odds Ratios – Data Years 2016-2017

Covariate	Coefficient	Odds Ratio	P-value
Sex			
Female	0.0034	1.0034	0.7061
Male	**Reference**		
Age			
18-24	0.4727	1.6042	<0.0001
25-44	0.3295	1.3903	<0.0001
45-59	0.1372	1.1470	<0.0001
60-74	**Reference**		
75+	-0.0521	0.9492	<0.0001
BMI			
<18.5	-0.0290	0.9714	0.1928
18.5-25	**Reference**		
25-30	-0.0118	0.9883	0.2554
>=30	-0.0173	0.9828	0.0699
Cause of ESRD			
Diabetes	0.0107	1.0108	0.2193
Time on ESRD			
91 days - 6 months	0.0124	1.0125	0.5253
6 months - 1 year	-0.0103	0.9898	0.5195
1-2 years	**Reference**		
2-3 years	0.0228	1.0230	0.1044
3-5 years	0.0028	1.0028	0.8288
5+ years	0.0172	1.0174	0.1601
Length of Hospital Stay (Q1 - Shortest stay)			

Covariate	Coefficient	Odds Ratio	P-value
Length of Stay (Q1)	**Reference**		
Length of Stay (Q2)	-0.0634	0.9386	<0.0001
Length of Stay (Q3)	-0.0881	0.9156	<0.0001
Length of Stay (Q4)	-0.1113	0.8947	<0.0001
Nursing Home (NH) previous 365 days			
No NH care (0 days)	**Reference**		
Short-term NH care (1 - 89 days)	-0.0737	0.9290	<0.0001
Long-term NH care (90 - 365 days)	-0.1874	0.8291	<0.0001
Year			
Year 2016	-0.0099	0.9901	0.1750
Year 2017	**Reference**		
Prevalent comorbidity groups			
HIV infection	0.0596	1.0614	0.0091
Hepatitis	0.0471	1.0482	0.0001
Viral infection	0.0269	1.0273	0.0241
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.0201	1.0203	0.0655
Melanomas of skin; Other non-epithelial cancer of skin	-0.0818	0.9215	<0.0001
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.0384	0.9623	<0.0001
Diabetes mellitus with complications; Diabetes mellitus without complication	0.0321	1.0326	0.0033
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	-0.0008	0.9992	0.9751
Conditions associated with dizziness or vertigo; Other ear and sense organ disorders; Otitis media and related conditions	0.0730	1.0757	<0.0001
Other nervous system disorders	0.0720	1.0747	<0.0001
Essential hypertension	0.2109	1.2348	<0.0001
Hypertension with complications and secondary hypertension	0.0021	1.0021	0.9708
Acute myocardial infarction; Coronary atherosclerosis and other heart disease	0.0387	1.0394	<0.0001
Nonspecific chest pain	0.1795	1.1966	<0.0001
Pulmonary heart disease	0.0135	1.0135	0.1034
Other and ill-defined heart disease	0.0695	1.0720	<0.0001
Cardiac dysrhythmias; Conduction disorders	0.0396	1.0404	<0.0001
Other circulatory disease	0.0563	1.0579	<0.0001

Covariate	Coefficient	Odds Ratio	P-value
Phlebitis; thrombophlebitis and thromboembolism	0.0485	1.0497	<0.0001
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.0362	1.0368	<0.0001
Asthma; Chronic obstructive pulmonary disease and bronchiectasis	0.0296	1.0301	0.0003
Other lower respiratory disease	0.1002	1.1054	<0.0001
Other upper respiratory disease	0.0045	1.0045	0.6151
Diseases of mouth; excluding dental; Disorders of teeth and jaw	0.0518	1.0531	0.0001
Esophageal disorders	0.0202	1.0204	0.0162
Gastroduodenal ulcer (except hemorrhage); Gastritis and duodenitis; Other disorders of stomach and duodenum; Appendicitis and other appendiceal conditions	0.0353	1.0360	0.0001
Anal and rectal conditions	0.0489	1.0501	0.0025
Peritonitis and intestinal abscess	-0.0901	0.9138	<0.0001
Pancreatic disorders (not diabetes)	0.0778	1.0809	<0.0001
Gastrointestinal hemorrhage	-0.0267	0.9737	0.0036
Noninfectious gastroenteritis	0.0550	1.0565	<0.0001
Other gastrointestinal disorders	0.0488	1.0500	<0.0001
Urinary tract infections	0.0302	1.0307	0.0005
Other diseases of kidney and ureters	0.0165	1.0166	0.0373
Hyperplasia of prostate; Inflammatory conditions of male genital organs; Other male genital disorders	0.0279	1.0283	0.0120
Chronic ulcer of skin; Other inflammatory condition of skin; Other skin disorders; Skin and subcutaneous tissue infections	0.0422	1.0431	<0.0001
Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease)	-0.0442	0.9568	0.0004
Sprains and strains	0.1222	1.1300	<0.0001
Complication of device; implant or graft	0.0386	1.0394	<0.0001
Superficial injury; contusion	0.0800	1.0833	<0.0001
Poisoning by nonmedicinal substances; Poisoning by other medications and drugs; Poisoning by psychotropic agents	0.0013	1.0013	0.9262
Other injuries and conditions due to external causes	0.0274	1.0278	0.0010
Syncope	0.0607	1.0626	<0.0001
Gangrene	-0.0493	0.9519	0.0005
Shock	-0.1556	0.8559	<0.0001
Nausea and vomiting	0.1188	1.1262	<0.0001
Abdominal pain	0.1352	1.1448	<0.0001
Malaise and fatigue	0.0683	1.0707	<0.0001
Allergic reactions	0.0487	1.0499	<0.0001

Covariate	Coefficient	Odds Ratio	P-value
Anxiety disorders	0.0783	1.0815	<0.0001
Attention-deficit conduct and disruptive behavior disorders	0.0780	1.0811	0.0019
Developmental disorders	0.1440	1.1548	<0.0001
Mood disorders	0.0208	1.0210	0.0163
Personality disorders	0.0987	1.1037	0.0013
Schizophrenia and other psychotic disorders	0.0750	1.0778	<0.0001
Alcohol-related disorders; Substance-related disorders	0.1311	1.1401	<0.0001
Suicide and intentional self-inflicted injury	0.1390	1.1491	<0.0001
Screening and history of mental health and substance abuse codes	0.0296	1.0301	0.0001
Miscellaneous mental health disorders	0.0407	1.0416	0.0019
Epilepsy; convulsions	0.0289	1.0293	0.0037
Headache; including migraine	0.1412	1.1517	<0.0001
Calculus of urinary tract	0.0241	1.0244	0.0467
Other non-traumatic joint disorders	0.0384	1.0391	<0.0001
Spondylosis; intervertebral disc disorders; other back problems	0.0704	1.0729	<0.0001
Osteoporosis	-0.0727	0.9299	<0.0001
Other bone disease and musculoskeletal deformities; Other connective tissue disease	0.0541	1.0556	<0.0001

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.**

Using hierarchical binary logistic regression, we fit an additional model for ED30 to 2016 and 2017 hospitalization data, including covariates from the original ED30 model and adding several SES/SDS indicators (dual-eligible insurance status, employment status at ESRD incidence, area deprivation index) as well as patient race and ethnicity. Table 4 shows the associations from these selected additional covariates in the SES/SDS adjusted model.

Table 4. Coefficients and Odds Ratios for Model with Additional SDS/SES Adjustors: 2016-2017

Covariate	Coefficient	Odds Ratio	P-value
Sex			
Female	-0.0195	0.9807	0.0283
Male	**Reference**		
Age			
18-24	0.3929	1.4813	<0.0001

Covariate	Coefficient	Odds Ratio	P-value
25-44	0.2699	1.3099	<0.0001
45-59	0.1051	1.1108	<0.0001
60-74	**Reference**		
75+	-0.0326	0.9679	0.0036
BMI			
<18.5	-0.0309	0.9696	0.1619
18.5-25	**Reference**		
25-30	-0.0097	0.9903	0.3450
>=30	-0.0110	0.9891	0.2489
Cause of ESRD			
Diabetes	0.0080	1.0081	0.3576
Time on ESRD			
91 days - 6 months	0.0381	1.0388	0.0490
6 months - 1 year	0.0020	1.0020	0.8974
1-2 years	**Reference**		
2-3 years	0.0196	1.0198	0.1577
3-5 years	-0.0026	0.9974	0.8356
5+ years	0.0096	1.0096	0.4149
Length of Hospital Stay (Q1 - Shortest stay)			
Length of Stay (Q1)	**Reference**		
Length of Stay (Q2)	-0.0594	0.9424	<0.0001
Length of Stay (Q3)	-0.0919	0.9122	<0.0001
Length of Stay (Q4)	-0.1085	0.8972	<0.0001
Nursing Home (NH) previous 365 days			
No NH care (0 days)	**Reference**		
Short-term NH care (1 - 89 days)	-0.0731	0.9295	<0.0001
Long-term NH care (90 - 365 days)	-0.2124	0.8086	<0.0001
Year of Index Discharge			
Year 2016	-0.0089	0.9912	0.2225
Year 2017	**Reference**		
Area Deprivation Index (ADI)[†]	0.1438	1.1547	<0.0001
Race			
White	**Reference**		
Native American/Alaskan Native	-0.0061	0.9939	0.8813
Asian/Pacific Islander	-0.0223	0.9779	0.3490
Black	0.1468	1.1581	<0.0001
Other/Unknown	0.1691	1.1842	0.0382
Ethnicity			
Hispanic	0.0544	1.0559	0.0001
Non-Hispanic	Reference		
Unknown	-0.1038	0.9014	0.2405
Medicare coverage			
Dual eligible	0.1345	1.1440	<0.0001
Non-dual eligible	Reference		

Covariate	Coefficient	Odds Ratio	P-value
Employment status 6 months prior to ESRD			
Employed	-0.0628	0.9391	<0.0001
Unemployed	Reference		
Retired/Other/Unknown	-0.0531	0.9483	<0.0001
Prevalent comorbidity groups			
HIV infection	0.0394	1.0402	0.0812
Hepatitis	0.0293	1.0298	0.0141
Viral infection	0.0313	1.0318	0.0083
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.0206	1.0208	0.0583
Melanomas of skin; Other non-epithelial cancer of skin	-0.0416	0.9592	0.0168
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.0268	0.9736	0.0041
Diabetes mellitus with complications; Diabetes mellitus without complication	0.0126	1.0127	0.2451
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	-0.0058	0.9943	0.8289
Conditions associated with dizziness or vertigo; Other ear and sense organ disorders; Otitis media and related conditions	0.0742	1.0770	<0.0001
Other nervous system disorders	0.0620	1.0640	<0.0001
Essential hypertension	0.2011	1.2227	<0.0001
Hypertension with complications and secondary hypertension	-0.0014	0.9986	0.9807
Acute myocardial infarction; Coronary atherosclerosis and other heart disease	0.0409	1.0417	<0.0001
Nonspecific chest pain	0.1774	1.1941	<0.0001
Pulmonary heart disease	0.0069	1.0069	0.4008
Other and ill-defined heart disease	0.0535	1.0549	<0.0001
Cardiac dysrhythmias; Conduction disorders	0.0418	1.0427	<0.0001
Other circulatory disease	0.0529	1.0543	<0.0001
Phlebitis; thrombophlebitis and thromboembolism	0.0471	1.0482	<0.0001
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.0422	1.0431	<0.0001
Asthma; Chronic obstructive pulmonary disease and bronchiectasis	0.0346	1.0352	<0.0001
Other lower respiratory disease	0.1028	1.1083	<0.0001
Other upper respiratory disease	0.0104	1.0104	0.2405
Diseases of mouth; excluding dental; Disorders of teeth and jaw	0.0451	1.0461	0.0005
Esophageal disorders	0.0220	1.0222	0.0086

Covariate	Coefficient	Odds Ratio	P-value
Gastroduodenal ulcer (except hemorrhage); Gastritis and duodenitis; Other disorders of stomach and duodenum; Appendicitis and other appendiceal conditions	0.0379	1.0386	<0.0001
Anal and rectal conditions	0.0385	1.0392	0.0166
Peritonitis and intestinal abscess	-0.0769	0.9260	<0.0001
Pancreatic disorders (not diabetes)	0.0717	1.0744	<0.0001
Gastrointestinal hemorrhage	-0.0238	0.9765	0.0088
Noninfectious gastroenteritis	0.0492	1.0504	<0.0001
Other gastrointestinal disorders	0.0391	1.0398	<0.0001
Urinary tract infections	0.0375	1.0382	<0.0001
Other diseases of kidney and ureters	0.0132	1.0133	0.0928
Hyperplasia of prostate; Inflammatory conditions of male genital organs; Other male genital disorders	0.0227	1.0230	0.0394
Chronic ulcer of skin; Other inflammatory condition of skin; Other skin disorders; Skin and subcutaneous tissue infections	0.0441	1.0451	<0.0001
Infective arthritis and osteomyelitis (except that caused by tuberculosis or sexually transmitted disease)	-0.0418	0.9591	0.0007
Sprains and strains	0.1334	1.1427	<0.0001
Complication of device; implant or graft	0.0302	1.0307	0.0006
Superficial injury; contusion	0.0979	1.1028	<0.0001
Poisoning by nonmedicinal substances; Poisoning by other medications and drugs; Poisoning by psychotropic agents	0.0196	1.0198	0.1706
Other injuries and conditions due to external causes	0.0381	1.0389	<0.0001
Syncope	0.0614	1.0633	<0.0001
Gangrene	-0.0606	0.9412	<0.0001
Shock	-0.1516	0.8593	<0.0001
Nausea and vomiting	0.1176	1.1248	<0.0001
Abdominal pain	0.1274	1.1358	<0.0001
Malaise and fatigue	0.0696	1.0721	<0.0001
Allergic reactions	0.0590	1.0608	<0.0001
Anxiety disorders	0.0969	1.1018	<0.0001
Attention-deficit conduct and disruptive behavior disorders	0.0628	1.0649	0.0120
Developmental disorders	0.1213	1.1289	<0.0001
Mood disorders	0.0229	1.0232	0.0078
Personality disorders	0.0787	1.0819	0.0096
Schizophrenia and other psychotic disorders	0.0745	1.0774	<0.0001
Alcohol-related disorders; Substance-related disorders	0.1071	1.1130	<0.0001
Suicide and intentional self-inflicted injury	0.1279	1.1364	<0.0001

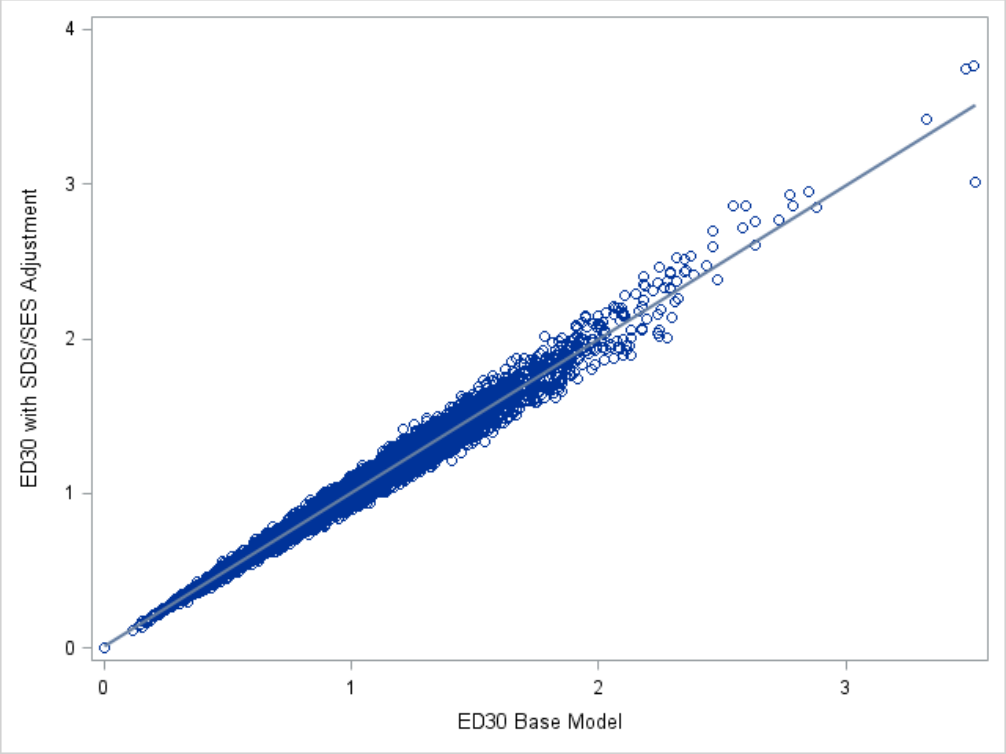
Covariate	Coefficient	Odds Ratio	P-value
Screening and history of mental health and substance abuse codes	0.0275	1.0279	0.0004
Miscellaneous mental health disorders	0.0349	1.0355	0.0073
Epilepsy; convulsions	0.0302	1.0307	0.0022
Headache; including migraine	0.1372	1.1471	<0.0001
Calculus of urinary tract	0.0423	1.0432	0.0004
Other non-traumatic joint disorders	0.0378	1.0385	<0.0001
Spondylosis; intervertebral disc disorders; other back problems	0.0683	1.0707	<0.0001
Osteoporosis	-0.0506	0.9507	0.0002
Other bone disease and musculoskeletal deformities; Other connective tissue disease	0.0531	1.0546	<0.0001

*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.

† 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 ED30 with and without SDS/SES adjustment, 2016-2017.



Overall Correlation coefficient rho=0.99 (p <0.0001)

Table 5. Flagging rates, baseline ED30 and ED30 adjusted for SDS/SES, 2016-2017

Baseline ED30	**ED30 with SDS/SES**			**Total**
	Better than Expected	As Expected	Worse than Expected	
Better than Expected	164	28	0	**2.85%(192)**
As Expected	29	6265	37	**94.10%(6331)**
Worse than Expected	0	36	169	**3.05%(205)**
Total	**2.87%(193)**	**94.07%(6329)**	**3.06%(206)**	6728

These results show that there was virtually no difference in the overall flagging rates between the baseline model and the model that adjusts for SES/SDS factors. However, of the facilities that were flagged as worse than expected in the baseline model, about 18% improve to 'as expected' after SDS/SES adjustment. Similarly, of the facilities that were flagged as 'better than expected' in the baseline model, about 15% were reclassified as 'as expected' in the SDS/SES model. About 21% of facilities improve from worse to as expected in the model adjusted for SDS/SES.

Along with changes in facility flagging, higher odds of an ED visit within 30 days of discharge were associated with black race, Hispanic ethnicity, dual eligible status, unemployment, and higher levels of area deprivation in the model adjusting for SDS/SES. Race, ethnicity, dual eligible status, employment status, and area deprivation are not included in the final risk adjusted model. Other studies have reported associations between patient-level race, ethnicity, dual eligible status, neighborhood deprivation 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 social risk 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.

In the baseline model female sex does not have an impact on ED30 however we continue to include sex (SDS factor) for risk adjustment in ED30. 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 (and hospital utilization), suggesting a physiologic effect rather than a systematic difference or disparity in care by sex. Our adjustment for sex in ED30 also aligns with the SRR.

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*)

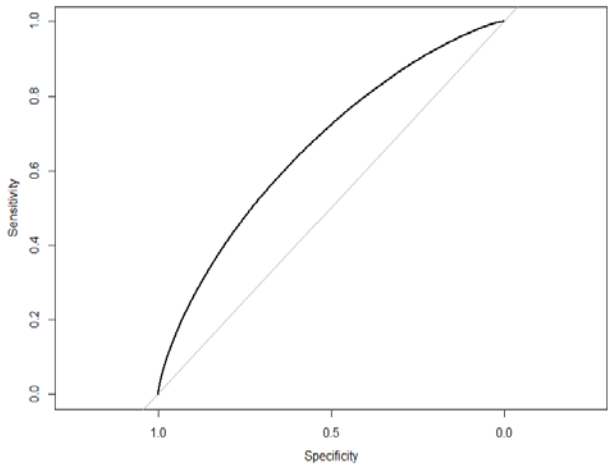
Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.
If stratified, skip to 2b3.9

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 C-statistic measures the discriminative power of the regression model with considered risk factors. As the ROC curve demonstrates, the model's accuracy is good (Figure 3); C-statistic = 0.665.

Figure 3. ROC Curve for Model (2016-2017)



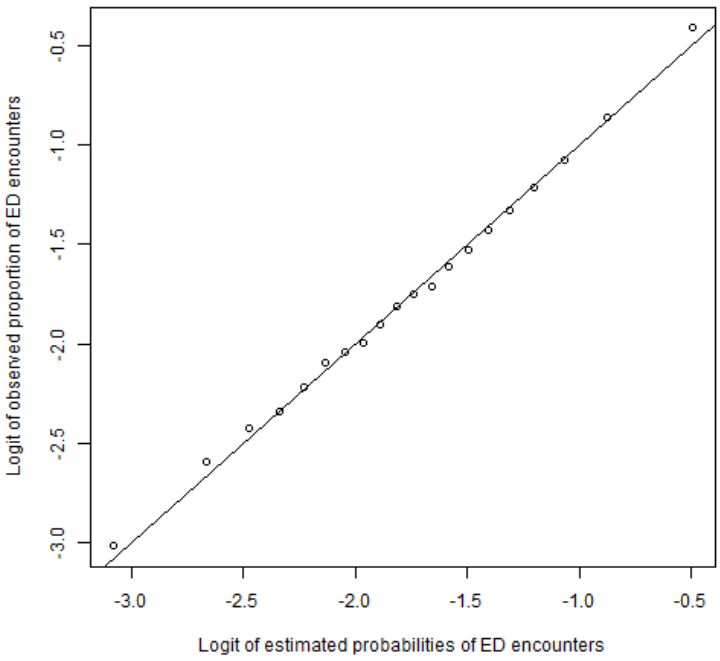
2b3.7. Statistical Risk Model Calibration Statistics (*e.g., Hosmer-Lemeshow statistic*):

N/A

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

The model's fit is demonstrated in Figure 4 below, which compares the observed rates with the model-based predictions. We bin all observations into 20 groups based on their model-based predicted values and compute the observed emergency department encounter proportion for each group. We then apply the logit transformation to each group's observed emergency department encounter proportion and plot it against the same group's average linear prediction. The 45-degree line would represent a perfect match between the observed values and the model-based predictions. In general, the closer the observed values are to this line the better the model fit.

Figure 4. Logit of the observed proportion of ED encounters Against the model estimated probabilities.



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*)

As Figure 4 shows, the observed values are spaced fairly equally and lie very close to the 45-degree line. This suggests that the model fit is reasonably good and therefore adequately adjusts for patient characteristics (case mix).

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 test the null hypothesis that the ED30 for a given facility is statistically different from the national average, we use a simulation method to calculate the nominal p-value as the probability that the observed number of emergency department encounters should be at least as extreme as that expected. This calculation is based on the supposition that, having adjusted for case mix, this facility has a true ED30 rate corresponding to the average facility. Our approach captures the most important aspects of the variability in the ED30. It also avoids difficulties with more traditional methods based on estimates and standard errors. Methods are described in detail in He et al. (2013).

To address the problem of simultaneously monitoring a large number of facilities and to take account of the intrinsic unexplained variation among facilities, we used the approach described in Kalbfleisch and Wolfe (2013). This method is based on the empirical null as described in Efron (2004, 2007). The p-value for each facility is converted to a Z-score, stratified into four groups based on patient-years within each facility. The empirical null corresponds to a normal curve that is fitted to the center of each Z-score histograms using a robust M-estimation method. The standard deviation of empirical null distribution is then used for a reference distribution (with mean 0) to identify outlier facilities (i.e. with the empirical null based p-value less than 0.05). This method aims to separate underlying intrinsic variation in facility outcomes from variation that might be attributed to poor (or excellent) care. Without empirical null methods, a large number of facilities will be flagged, including many larger facilities with a relatively small difference between the rates of emergency department encounters. In contrast, the methods based on the empirical null make appropriate adjustments for overdispersion. Using this method, facilities are flagged if they have outcomes that are extreme when compared to the variation in outcomes for other facilities of a similar size.

References:

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

Efron B. (2007). Size, power and false discovery rates. Ann. Statist. 35(4):1351-1377.

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 6 shows the number of facilities classified as extreme using the method described in the prior sub-section. We find 192 (2.85%) facilities with ED30 that are better than expected and 205 (3.05%) that are worse than expected.

Table 6. Percentage and Number of facilities by classification of ED30, 2016-2017

Better than Expected	As Expected	Worse than Expected	Total Facilities
192 (2.85%)	6331 (94.10%)	205 (3.05%)	6,728

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?*)

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 in the 4-30 days after hospital discharge) that are extreme when compared to the variation in outcomes for other facilities of a similar size. Overall, most are flagged as expected (94.10%), while 2.85% are better than expected, and 3.05% are flagged as worse than expected. This analysis demonstrates both practical and statistically significant differences in performance across facilities based on their proportion of patients who are seen in the ED within 30 days after hospital discharge.

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*)

ED30 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, ED30 is restricted to Medicare Fee for Service (FFS) patients.

For ED30, the population is defined by index discharges (inpatient claim). Inpatient claims are available for Medicare Advantage (MA) patients however outpatient claims are not available for this population. 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, inclusion of MA patients with an index discharge in ED30 introduces bias into the measure calculation that could affect results for dialysis facilities with either very low or high MA patient populations. That is we are not able to observe outpatient ED encounters for MA patients because we cannot observe ED encounters from outpatient claims for MA patients. To demonstrate this we assessed the proportion of observed ED visits following index discharges covered by Medicare FFS and MA, respectively.

Many data elements can be obtained from multiple sources and missing data occur rarely for covariates included in ED30. We assessed missing data for BMI which comes from form CMS 2728 which is the source of data used for the BMI risk adjustment in the ED30 model.

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*)

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. As shown in the table below, over 99% of index discharges covered by MA at the time of discharge had, as expected, no observed ED visits. In contrast about 18% of index discharges covered by Medicare FFS at the time of discharge had a subsequent ED visit, which is a rate that generally corresponds with the broader Medicare FFS chronic dialysis population.
3. The number of index discharges with BMI reported was 736,677 (97.25%); those with missing BMI was 20,820 (2.75%).

ED visit within 4-30 days	Medicare Advantage status at the time of discharge	
	Non MA	MA
No	893770 (81.81%)	246226 (99.53%)
Yes	198713 (18.19%)	1164 (0.47%)

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*)

Because outpatient ED claims are not available for MA patients, we excluded MA patients from ED30. This exclusion approach 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. 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 ED30 value, even if the actual number of ED visits within 30 days of a discharge is high; alternatively ED30 could be high if ED visits within 30 days are concentrated in a few FFS patients relative to the larger MA patient population at the facility, even if the actual overall number at the facility is low for these types of ED visits.

There is a very low frequency of 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 the measure score.

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.

Because data are derived from administrative databases, there is no additional data collection required.

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 ED30 measure as part of CMS's Dialysis Facility public reporting program. The program provides information that can help dialysis patients and their caregivers and other consumers 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)

2496 : Standardized Readmission Ratio (SRR) for dialysis facilities

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.

Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities (currently undergoing endorsement review with ED30). 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?

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 certain differences across the measure specifications. The proposed Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge (ED30) for Dialysis Facilities and Standardized Emergency Department Encounter Ratio (SEDR) for Dialysis Facilities measures both focus on 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 use closely following a hospitalization. The ED30 and SRR are both intended to encourage care coordination for patients recently discharged from an inpatient admission, but measure two different outcomes after discharge. The ED30 applies to the same target population as SEDR—adult Medicare-covered dialysis patients who have had ESRD for more than 90 days. The target population for CMS's Standardized Readmission Ratio (SRR) for dialysis facilities (NQF #2496) is similar but also includes pediatric patients and the first 90 days of ESRD treatment. ED30, SRR, and SEDR adjust for a similar set of patient characteristics. All three measures adjust for prior-year comorbidities although the SRR set of comorbidity risk factors is different than that for the ED30 and SEDR. Only the SEDR also includes adjustment for comorbidities at ESRD incidence. The ED30 and SRR adjust for a number of factors related to the index discharge that are not included in the SEDR model because index discharges are not relevant in that context. The definition of index discharges is very similar for SRR and ED30 but there are some differences: 1) SRR excludes index discharges that follow a patient's 12th admission in the year; 2) ED30 excludes index discharges that occur in a calendar month in which the patient was enrolled in hospice; and patients with Medicare Advantage at the time of the index discharge 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; 3) ED30 excludes index discharges that result in another hospitalization, emergency department visit, or transplant within three days of discharge; or loss to follow-up, withdrawal from dialysis, or recovery of renal function while SRR further excludes only those that result in a patient dying within 30 days with no readmission. ED30 and Emergency Department Use without Hospital Readmission During the First 30 Days of Home Health also 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 different target populations; 3) risk adjustment factors; and 4) model type (logistic 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). 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: ED30_Appendix-637227313803254097.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? Annual

Ad.5 When is the next scheduled review/update for this measure? 04, 2020

Ad.6 Copyright statement:

Ad.7 Disclaimers:

Ad.8 Additional Information/Comments: