

NATIONAL QUALITY FORUM—Measure Testing (subcriteria 2a2, 2b2-2b7)

Measure Number (if previously endorsed): [Click here to enter NQF number](#)

Measure Title: Standardized Emergency Department Encounter Ratio for Dialysis Facilities

Date of Submission: [9/29/2017](#)

Type of Measure:

<input checked="" type="checkbox"/> Outcome (including PRO-PM)	<input type="checkbox"/> Composite – STOP – use composite testing form
<input type="checkbox"/> Intermediate Clinical Outcome	<input type="checkbox"/> Cost/resource
<input type="checkbox"/> Process	<input type="checkbox"/> Efficiency
<input type="checkbox"/> Structure	

Instructions

- Measures must be tested for all the data sources and levels of analyses that are specified. **If there is more than one set of data specifications or more than one level of analysis, contact NQF staff** about how to present all the testing information in one form.
- For all measures, sections 1, 2a2, 2b2, 2b3, and 2b5 must be completed.**
- For outcome and resource use measures, section 2b4 also must be completed.**
- If specified for **multiple data sources/sets of specifications** (e.g., claims and EHRs), section **2b6** also must be completed.
- Respond to all questions as instructed with answers immediately following the question. All information on testing to demonstrate meeting the subcriteria for reliability (2a2) and validity (2b2-2b6) must be in this form. An appendix for *supplemental* materials may be submitted, but there is no guarantee it will be reviewed.
- If you are unable to check a box, please highlight or shade the box for your response.
- Maximum of 20 pages (including questions/instructions; minimum font size 11 pt; do not change margins). **Contact NQF staff if more pages are needed.**
- Contact NQF staff regarding questions. Check for resources at [Submitting Standards webpage](#).
- For information on the most updated guidance on how to address sociodemographic variables and testing in this form refer to the release notes for version 6.6 of the Measure Testing Attachment.

Note: The information provided in this form is intended to aid the Steering Committee and other stakeholders in understanding to what degree the testing results for this measure meet NQF’s evaluation criteria for testing.

2a2. Reliability testing ¹⁰ demonstrates 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. For **PRO-PMs and composite performance measures**, reliability should be demonstrated for the computed performance score.

2b2. Validity testing ¹¹ demonstrates that the measure data elements are correct and/or the measure score correctly reflects the quality of care provided, adequately identifying differences in quality. For **PRO-PMs and composite performance measures**, validity should be demonstrated for the computed performance score.

2b3. Exclusions are supported by the clinical evidence; otherwise, they are supported by evidence of sufficient frequency of occurrence so that results are distorted without the exclusion; ¹²

AND

If patient preference (e.g., informed decisionmaking) is a basis for exclusion, there must be evidence that the exclusion impacts performance on the measure; in such cases, the measure must be specified so that the information about patient preference and the effect on the measure is transparent (e.g., numerator category computed separately, denominator exclusion category computed separately). ¹³

2b4. For outcome measures and other measures when indicated (e.g., resource use):

- **an evidence-based risk-adjustment strategy** (e.g., risk models, risk stratification) is specified; is based on patient factors (including clinical and sociodemographic factors) that influence the measured outcome and are present at start of care; ^{14,15} and has demonstrated adequate discrimination and calibration

OR

- rationale/data support no risk adjustment/ stratification.

2b5. Data analysis of computed measure scores demonstrates that methods for scoring and analysis of the specified measure allow for **identification of statistically significant and practically/clinically meaningful** ¹⁶ **differences in performance;**

OR

there is evidence of overall less-than-optimal performance.

2b6. If multiple data sources/methods are specified, there is demonstration they produce comparable results.

2b7. For eMeasures, composites, and PRO-PMs (or other measures susceptible to missing data), analyses 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.

Notes

10. Reliability testing applies to both the data elements and computed measure score. Examples of reliability testing for data elements include, but are not limited to: inter-rater/abstractor or intra-rater/abstractor studies; internal consistency for multi-item scales; test-retest for survey items. Reliability testing of the measure score addresses precision of measurement (e.g., signal-to-noise).

11. Validity testing applies to both the data elements and computed measure score. Validity testing of data elements typically analyzes agreement with another authoritative source of the same information. Examples of validity testing of the measure score include, but are not limited to: testing hypotheses that the measure scores indicate quality of care, e.g., measure scores are different for groups known to have differences in quality assessed by another valid quality measure or method; correlation of measure scores with another valid indicator of quality for the specific topic; or relationship to conceptually related measures (e.g., scores on process measures to scores on outcome measures). Face validity of the measure score as a quality indicator may be adequate if accomplished through a systematic and transparent process, by identified experts, and explicitly addresses whether performance scores resulting from the measure as specified can be used to distinguish good from poor quality.

12. Examples of evidence that an exclusion distorts measure results include, but are not limited to: frequency of occurrence, variability of exclusions across providers, and sensitivity analyses with and without the exclusion.

13. Patient preference is not a clinical exception to eligibility and can be influenced by provider interventions.

14. Risk factors that influence outcomes should not be specified as exclusions

15. With large enough sample sizes, small differences that are statistically significant may or may not be practically or clinically meaningful. The substantive question may be, for example, whether a statistically significant difference of one percentage point in the percentage of patients who received smoking cessation counseling (e.g., 74 percent v. 75 percent) is clinically meaningful; or whether a statistically significant difference of \$25 in cost for an episode of care (e.g., \$5,000 v. \$5,025) is practically meaningful. Measures with overall less-than-optimal performance may not demonstrate much variability across providers.

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: (<i>must be consistent with data sources entered in S.23</i>)	Measure Tested with Data From:
<input type="checkbox"/> abstracted from paper record	<input type="checkbox"/> abstracted from paper record
<input checked="" type="checkbox"/> administrative claims	<input checked="" type="checkbox"/> administrative claims
<input checked="" type="checkbox"/> clinical database/registry	<input checked="" type="checkbox"/> clinical database/registry
<input type="checkbox"/> abstracted from electronic health record	<input type="checkbox"/> abstracted from electronic health record
<input type="checkbox"/> eMeasure (HQMF) implemented in EHRs	<input type="checkbox"/> eMeasure (HQMF) implemented in EHRs
<input type="checkbox"/> other: Click here to describe	<input type="checkbox"/> 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 the CMS Consolidated Renal Operations in a Web-enabled Network (CROWN) system. The CROWN data include the Renal Management Information System (REMIS), 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 data), the historical Standard Information Management System (SIMS) database (formerly maintained by the 18 ESRD Networks until replaced by CROWNWeb in May 2012), the National Vascular Access Improvement Initiative’s Fistula First Catheter Last project (in CROWNWeb since May 2012), Medicare dialysis and hospital payment records, transplant data from the Organ Procurement and Transplant Network (OPTN), the Nursing Home Minimum Dataset, the Quality Improvement Evaluation System (QIES) Workbench, which includes data from the Certification and Survey Provider Enhanced Report System (CASPER), the Dialysis Facility Compare (DFC) and the Social Security Death Master File. The database is comprehensive for Medicare patients. Non-Medicare patients are included in all sources except for the Medicare payment records. CROWNWeb provides tracking by dialysis provider and treatment modality for non-Medicare patients. Information on emergency department visits is obtained from Medicare Outpatient Claims Standard Analysis Files (SAFs). Medicare Inpatient Claims SAFs are used to determine if emergency department visits resulted in an admission. Prevalent comorbidities are obtained using Medicare Physician Supplier, Inpatient, Outpatient, Skilled Nursing, Home Health, and Hospice claims.

1.3. What are the dates of the data used in testing?

January 2012-December 2015

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.26)	Measure Tested at Level of:
<input type="checkbox"/> individual clinician	<input type="checkbox"/> individual clinician
<input type="checkbox"/> group/practice	<input type="checkbox"/> group/practice
<input checked="" type="checkbox"/> hospital/facility/agency	<input checked="" type="checkbox"/> hospital/facility/agency
<input type="checkbox"/> health plan	<input type="checkbox"/> health plan
<input type="checkbox"/> other: Click here to describe	<input type="checkbox"/> 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)

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

Year	Number of Facilities	Median Facility Size (as of 12/31)
2012	5,663	60
2013	5,842	61
2014	6,059	61
2015	6,256	61

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)

Medicare dialysis patients were included in the testing and analysis for each of the four years from 2012-2015 of which there were 394,778; 404,353; 413,602 and 421,570 patients respectively.

Table 2. Descriptives of Patient Characteristics Included in the Measure

Patient Demographics	Percent
Age	
Patient Age: 18-24	0.6
Patient Age: 25-44	10.6
Patient Age: 45-59	25.6
Patient Age: 60-74	39.9
Patient Age: 75+	23.3
Sex (% female)	44.5
ESRD due to Diabetes (%)	46.7
Medicare coverage(%)	
Medicare primary + Medicaid	40.2

Patient Demographics	Percent
Medicare primary + no Medicaid	46.7
Medicare secondary/HMO	13.1
Time since Start of ESRD	
91 days-6 months	11.6
6 months-1 year	13.6
1-2 years	17.1
2-3 years	14.8
3-5 years	18.2
5+ years	24.8
Employment status 6 months prior to ESRD (%)	
Unemployed	22.1
Employed	19.0
Other/Unknown *	59.0
Race (%)	
White	59.7
Black	34.0
Native American/Alaskan Native	1.2
Asian/Pacific Islander	4.8
Other/Unknown	0.3
Ethnicity (%)	
Hispanic	15.8
Non-Hispanic	83.6
Unknown	0.6

* Other/Unknown groups includes Homemaker, Retired due to age/preference, retired due to disability, Medical leave of absence, or missing employment status. Note: Some categories (Time since start of ESRD and Employment) sum to 100.1% due to rounding.

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

N/A

1.8 What were the patient-level sociodemographic (SDS) variables that were available and analyzed in the data or sample used? For example, patient-reported data (e.g., income, education, language), proxy variables when SDS data are not collected from each patient (e.g. census tract), or patient community characteristics (e.g. percent vacant housing, crime rate).

Patient level:

- Employment status 6 months prior to ESRD
- Sex
- Race
- Ethnicity
- Medicare coverage*

**Assessed at the start of time at risk based on calendar year and facility assignment. Medicare coverage in the model was defined as:*

1. Medicare as primary and Medicaid
2. Medicare as primary and NO Medicaid
3. Medicare as secondary or Medicare HMO

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

Proxy/Area level: ZIP code level – Area Deprivation Index (ADI) elements from 2014 Census data:

- Unemployment rate (%)
- Median family income (rescaled as (income-60,000)/10,000)
- Income disparity
- Families below the poverty level (%)
- Single-parent households w/ children <18 (%)
- Home ownership rate (%)
- Median home value (rescaled as (homevalue-200,000)/100,000)
- Median monthly mortgage (rescaled as (mortgage-1,500)/1,000)
- Median gross rent (rescaled as (rent-900)/1,000)
- Population (aged 25+) with <9 years of education (%)
- Population (aged 25+) without high school diploma (%)

2a2. RELIABILITY TESTING

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

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

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

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

2a2.2. For each level checked above, describe the method of reliability testing and what it tests

(describe the steps—do not just name a method; what type of error does it test; what statistical analysis was used)

The reliability of the Standardized Emergency Department Encounter Ratio (SEDR) was assessed using data among Medicare ESRD dialysis patients during 2012-2015. If the measure were a simple average across individuals in the facility, the usual approach for determining measure reliability would be a one-way analysis of variance (ANOVA), in which the between and within facility variation in the measure is determined. The inter-unit reliability (IUR) measures the proportion of the total variation of a measure that is attributable to the between-facility variation. The SEDR, however, is not a simple average and we instead estimate the IUR using a bootstrap approach, which uses a resampling scheme to estimate the

within facility variation that cannot be directly estimated by ANOVA. Refer to Appendix B for a detailed description of this methodology.

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

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, as presented in Table 3, we found that IURs for the one-year SEDRs have a range of 0.65 - 0.72 across the years 2012, 2013, 2014 and 2015, which indicates that approximately 65% to 72% of the variation in the one-year SEDR can be attributed to the between-facility differences and about 28% to 35% to within-facility variation.

Table 3: IUR for one-year SEDR, 2012-2015

	2012		2013		2014		2015	
	IUR	Facilities	IUR	Facilities	IUR	Facilities	IUR	Facilities
Overall	0.69	5675	0.72	5851	0.64	6070	0.65	6267

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 IUR value is considered strong. As described in section 2b5.3 the measure demonstrates it is effective at detecting outlier facilities and statistically meaningful differences in performance scores across measured entities.

2b2. VALIDITY TESTING

2b2.1. 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 performance measure score 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)

2b2.2. For each level of testing checked above, describe the method of validity testing and what it tests (describe the steps—do not just name a method; what was tested, e.g., accuracy of data elements compared to authoritative source, relationship to another measure as expected; what statistical analysis was used)

Face Validity: In May 2016, we presented a preliminary version of the SEDR measure to a CMS Technical Expert Panel (TEP) for clinical validity. The TEP discussed different ED outcomes and recommended limiting an ED encounter measure to visits that do not result in an inpatient admission because ED visits resulting in hospitalization are already captured through the respective NQF endorsed Standardized

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

Validation of performance measure scores: We assessed empirical validity of the measure by calculating Spearman correlations between this measure and vascular access type (fistula use and catheter ≥ 90 days), Kt/V ≥ 1.2 , Standardized Mortality Ratio (SMR), SHR, and the ED30.

2b2.3. What were the statistical results from validity testing? (e.g., correlation; t-test)

Results of the Spearman correlations testing the association between SEDR and vascular access type, Kt/V ≥ 1.2 , Standardized Mortality Ratio (SMR), Standardized Hospitalization Ratio (SHR), and the Standardized Ratio of Emergency Department Encounters Occurring Within 30 Days of Hospital Discharge for Dialysis Facilities (ED30), which is currently being submitted for endorsement as a companion measure to SEDR measure, are presented in Table 4. The correlations below were calculated for each of the calendar years 2012-2015.

Table 4. Spearman Correlation of SEDR and Related Measures (2012-2015)

	2012		2013		2014		2015	
	Corr.	P-value	Corr.	P-value	Corr.	P-value	Corr.	P-value
Vascular Access: Catheter>90 days	-0.04	0.0058	-0.04	0.0017	-0.04	0.0034	-0.02	0.0868
Vascular Access: Fistula	-0.07	<.0001	-0.06	<.0001	-0.05	<.0001	-0.05	<.0001
Kt/V ≥ 1.2	-0.07	<.0001	-0.04	0.0018	-0.05	<.0001	-0.09	<.0001
SHR	-0.09	<.0001	-0.06	<.0001	-0.08	<.0001	-0.1	<.0001
SMR	0.07	<.0001	0.09	<.0001	0.09	<.0001	0.08	<.0001
ED30	---	---	---	---	---	---	0.51	<0.0001

2b2.4. 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?)

The SEDR correlates with dialysis facility processes and outcomes that are commonly thought to be related to quality of care. Higher rates of emergency department visits are associated with suboptimal dialysis adequacy as well as lower rates of arteriovenous fistula use. This suggests that facilities with processes of care to provide optimal small solute clearance and optimal vascular access may have other processes of care to help their patients avoid needing the ED for unscheduled acute care. We found a negative but very weak association between SEDR and having a catheter >90 days for vascular access. It

may be that patients with longer term catheter use are more likely to be admitted (e.g., for catheter associated infections) rather than experience an outpatient ED encounter. This would attenuate the relationship between long-term catheter-based vascular access and outpatient ED utilization.

Higher ED utilization was also associated with lower facility hospitalization rates and higher mortality rates. The correlation with SHR was relatively low, as might be expected, since SEDR focuses on outpatient use of ED services whereas SHR captures ED use that results in hospitalization. Thus, SEDR likely captures dialysis patients that have a lower acuity of illness than the SHR. Higher ED utilization was associated with higher mortality but similar to SHR, the correlation was low.

Lastly, we assessed the correlation between the SEDR and the companion ED30 measure (also being submitted for consideration of NQF endorsement). Since ED encounters that are measured in the ED30 are also captured in the SEDR, these two measures demonstrate a strong degree of correlation while assessing complementary elements of care.

2b3. EXCLUSIONS ANALYSIS

NA no exclusions — skip to section 2b4

2b3.1. Describe the method of testing exclusions and what it tests (*describe the steps—do not just name a method; what was tested, e.g., whether exclusions affect overall performance scores; what statistical analysis was used*)

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

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

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

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

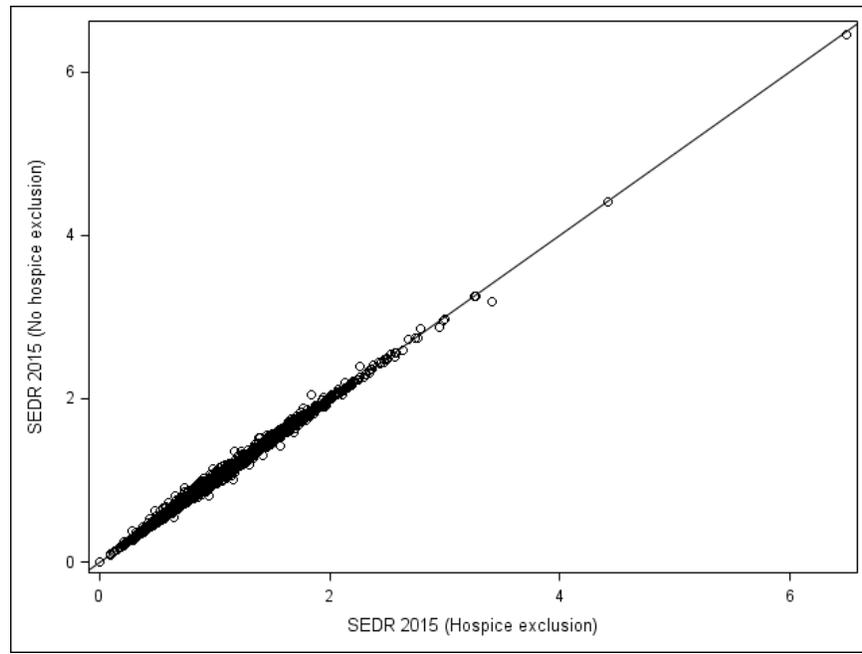
- Actively enrolled in hospice during the calendar month of the ED encounter

2b3.2. What were the statistical results from testing exclusions? (*include overall number and percentage of individuals excluded, frequency distribution of exclusions across measured entities, and impact on performance measure scores*)

There were 2,062 patient years at risk excluded due to active enrollment in hospice, which represents 0.67% of total years in the analysis. This excludes 4,111 (0.90%) ED visits during this time period (2015).

As shown in Figure 1, we compared each facility's SEDR with and without the hospice exclusion and found the two measures to be highly correlated (overall Pearson correlation coefficient $[r] = 0.99875$, $p < 0.0001$).

Figure 1. Correlation between SEDR with and without the hospice exclusion (2015)



2b3.3. 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 exclusion criteria 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.

2b4. 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 2b5.

2b4.1. What method of controlling for differences in case mix is used?

- No risk adjustment or stratification
- Statistical risk model with 86 risk factors (diabetes, sex, age, BMI at incidence, calendar year, nursing home status, 13 comorbidities at incidence, and 67 prevalent comorbidities)
- Stratification by [Click here to enter number of categories](#) risk categories
- Other, [Click here to enter description](#)

2b4.1.1 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 regression model used to compute a facility's "expected" number of emergency department encounters for the Standardized Emergency Department Encounter Ratio measure contains many factors thought to be associated with emergency department encounter rates. Specifically, the model adjusts for patient age, sex, diabetes as cause of ESRD, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, calendar year, and prevalent comorbidities. The stage 1 model allows the baseline emergency department encounter rates to vary between strata, which are defined by facilities, but assumes that the regression coefficients are the same across all strata; this approach is robust to possible differences between facilities in the patient mix being treated. In essence, it avoids a possible confounding between facility effects and patient covariates as can arise, for example, if patients with favorable values of the covariate tend to be treated at facilities with better treatment policies and outcomes. Thus, for example, if patients with diabetes as a cause of ESRD tended to be treated at facilities with higher quality of care, one would underestimate the effect of diabetes unless the model is adjusted for facility. In this model, facility adjustment is done by stratification.

The patient characteristics included in the stage 1 model as covariates are:

- Age: We determine each patient's age for the birth date provided in the CROWNWeb database and group patients into the following categories:
 - 18-24
 - 25-44
 - 45-59
 - 60-74
 - 75+
- Sex: We determine each patient's sex from his/her Medical Evidence Form (CMS-2728) and the CROWNWeb database.
- Diabetes as cause of ESRD: We determine each patient's primary cause of ESRD from his/her CMS-2728.
- ESRD duration: We determine each patient's length of time on dialysis using the first service date from his/her CMS-2728, claims history (all claim types), the CROWNWeb database and the SRTR database and categorize as 91 days-6 months, 6 months-1 year, 1-2 years, 2-3 years, 3-5 years, or 5+ years as of the period start date.
- Nursing home status: Using the Nursing Home Minimum Dataset, we determine if a patient was in a nursing home the previous year.
- BMI: We calculate each patient's BMI as the height and weight provided on his/her CMS 2728. BMI is categorized as underweight, normal weight, overweight, and obese.
- Calendar year
- The following incident comorbidities are included. They are taken from the CMS-2728 form. Each comorbidity is included as a separate covariate in the model.
 - Alcohol dependence
 - Atherosclerotic heart disease
 - Cerebrovascular disease
 - Chronic obstructive pulmonary disease
 - Congestive heart failure
 - Diabetes
 - Drug dependence
 - Inability to ambulate
 - Inability to transfer

- Malignant neoplasm or cancer
- Other cardiac disease
- Peripheral vascular disease
- Tobacco use (current smoker)
- Prevalent comorbidities (see appendix) are determined using the previous 12 months of Medicare claims after the index ED encounter. The fiscal year 2015 Agency for Healthcare Research and Quality Clinical Classification Software (AHRQ CCS) single-level diagnoses groupers were used to define the prevalent comorbidity risk factors. Each comorbidity is included as a separate covariate in the model. If a patient has less than 6 months of claims in the year before the analysis, we consider prevalent comorbidities to be “missing” for that patient even if there are comorbidities identified in claims. (See Appendix for the mapping of the CCS groupers to individual ICD-9 codes)

See Appendix for a more detailed description of the model coefficients, equations, and risk factors.

Reference:

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

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

2b4.2. If an outcome or resource use component measure is not risk adjusted or stratified, provide rationale and analyses to demonstrate that controlling for differences in patient characteristics (case mix) is not needed to achieve fair comparisons across measured entities.

N/A

2b4.3. Describe the conceptual/clinical and statistical methods and criteria used to select patient factors (clinical factors or sociodemographic 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)

Consideration of clinical risk factors: The risk adjustment is based on a Cox (relative risk) model. The adjustment is made for patient age, sex, diabetes as cause of ESRD, duration of ESRD, nursing home status, BMI at incidence, comorbidities at incidence, prevalent comorbidities, and calendar year. In this model for SEDR, covariates are taken to act multiplicatively on the ED rate and the adjustment model is fitted with facility defining strata so as to provide valid estimates even if the distribution of adjustment variables differs across facilities. Relevant references are Cox (1972), Kalbfleisch and Prentice (2002), Lawless and Nadeau (1995), Cook and Lawless (2007) and Liu, Schaubel and Kalbfleisch (2010). All analyses are done using SAS.

In general, adjustment factors for the SEDR were selected based on several considerations. Our starting point was the Standardized Hospitalization Ratio (SHR) (NQF 1463) which is the model on which we developed SEDR. We began with a large set of patient characteristics (listed above), which

were first evaluated for face validity by the 2016 TEP. Factors considered appropriate were then investigated with statistical models to determine if they were related to ED encounters.

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

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

Next, five categories of specific ICD-9 codes were removed from the remaining 251 AHRQ CCS groupers. These codes, listed in the Appendix, may be associated with dialysis facility care and include diagnoses such as secondary hyperparathyroidism, fluid overload, hyperkalemia, and vascular access infections. Once these specific ICD-9 codes were excluded, the 251 CCS groupers were consolidated down to 130 groupers by combining similar 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 AHRQ CCS groupers 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 CCS groupers 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.

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.

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

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

Race, insurance status, younger age, and SES have been shown to be predictors of emergency department utilization in the general population (Capp et al., 2015; Colligan et al., 2016; LaCalle et al., 2010; Zuckerman and Shen 2004; Hastings et al., 2008). For example, a study by Zuckerman and Shen (2004) reported that black adults had higher odds than whites of being occasional users compared to non-ED users. This difference between blacks and whites was larger when comparing frequent-users to non-users (Zuckerman and Shen, 2004, pg. 178). However, they also found few differences in the likelihood of frequent ED use when comparing patients that have private insurance versus those who are uninsured, while frequent ED use was more likely among those with public insurance (i.e., Medicaid) (Zuckerman and Shen 2004). Those with lower income also had higher odds of being occasional and frequent ED users, while individuals with some college had lower odds of being an occasional or frequent user of the ED, compared to those with no high school diploma. An analysis by

Cunningham et al., (2016) of frequent ED use at two urban hospitals found that frequent ED use was associated with younger age, and that frequent users were more likely to be black. However, there was no significant difference in primary care access between infrequent and frequent users, suggesting that access to care did not explain variation in ED utilization. In addition to younger age, another study reported that those who were single/divorced, single-parents, had high school education or less, or had lower income were more likely to be frequent users of the ED (Sun et al., 2003). Among dual-eligible patients that receive care from a Federally Qualified Health Center (FQHC), relative rates of ED use were lower compared to dual-eligibles that did not receive care from an FQHC (Wright et al., 2015), suggesting the importance of access to primary care. Finally, trends in ED use show differences by sex (female), age (45-64), and geography (the Midwest) and in large central metropolitan areas (Skinner et al., 2014, pg 2-3).

In the ESRD population, low health literacy (a proxy of SES) was found to be a predictor of ED use in one study (Green et al., 2013), as well as SDS/SES factors of younger age, female sex, black race, and public insurance (Medicaid) while lower ED use was associated with private insurance (Lovasik et al., 2016). ESRD patients discharged from a skilled nursing facility that had a subsequent emergency department encounter within 30 days were more likely to be of black race, have dual Medicare-Medicaid status, and higher comorbidity (Hall et al., 2015). In ESRD patients that received a transplant, higher risk of ED use was associated with younger age, female sex, black race, Hispanic ethnicity, and public insurance (Medicaid) (Schold et al., 2016). Treatment adherence was also found to be a risk factor for emergency department visits (Chan et al., 2014). This suggests that there may be related SDS/SES or community level factors that adversely impact patient treatment adherence. Area-level factors, typically operating as proxies of patient level factors, have also been found to influence acute care use, such as readmission (Herrin et al., 2015; Kind et al, 2014) as well as ED use (Skinner et al., 2014, pg 2-3). Additionally, area-level SES has been observed to be associated with poor outcomes in ESRD patients (e.g., Almachraki et al 2016).

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

In our analyses we use the publicly available Area Deprivation Index (ADI) developed by Singh and colleagues at the University of Wisconsin. 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. More recently, the ADI has been applied to the calculation of risk-adjusted rates of hospital readmission (Kind et al 2014).

References:

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

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

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

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

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

Cook R, Lawless J. *The Statistical Analysis of Recurrent Events*, New York, Springer-Verlag, 2007

Cox D: Regression models and life tables (with discussion). *JRStat Soc [SerB]*34: 187–220, 1972
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).

Kalbfleisch J, Prentice R: *The statistical analysis of failure time data*, Hoboken, New Jersey, John Wiley & Sons, Inc., 2002

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 PMID: 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.

Lawless J, and Nadeau C. Some Simple Robust Methods for the Analysis of Recurrent Events *Technometrics*. Vol. 37, No. 2 (May, 1995), pp. 158-168

Liu, D., Schaubel, D. and Kalbfleisch, J. Computationally efficient marginal models for clustered recurrent event data, University of Michigan Department of Biostatistics Technical Reports, 2010.

Liu, D., Schaubel, D.E. and Kalbfleisch, J.D. (2012). Computationally efficient marginal models for clustered recurrent event data. *Biometrics*, 68, 637-647.

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

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

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

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

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

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

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

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

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

2b4.4a. What were the statistical results of the analyses used to select risk factors?

Table 5. SEDR Model Coefficients, Data Years 2012–2015.

Covariate	Coefficient	P-value
Comorbidities at start of ESRD		
At least one of the comorbidities listed below	0.00	0.51
Atherosclerotic heart disease	0.02	<.0001
Other cardiac disease	0.00	0.61
Diabetes*	0.03	<.0001
Congestive heart failure	0.03	<.0001
Inability to ambulate	-0.02	0.00

Covariate	Coefficient	P-value
Chronic obstructive pulmonary disease	0.02	<.0001
Inability to transfer	-0.03	0.00
Malignant neoplasm, cancer	-0.03	<.0001
Peripheral vascular disease	-0.01	0.00
Cerebrovascular disease, CVA, TIA	0.03	<.0001
Tobacco use (current smoker)	0.08	<.0001
Alcohol dependence	0.01	0.05
Drug dependence	0.16	<.0001
No Medical Evidence (CMS-2728) Form	0.02	0.01
Cause of ESRD		
Diabetes	0.03	<.001
Sex: Female	0.08	<.0001
Age		
18-24	0.69	<.0001
25-44	0.43	<.0001
45-59	0.19	<.0001
60-74	Reference	
75+	-0.02	<.0001
BMI		
Underweight	0.01	0.04
Normal weight	Reference	
Overweight	-0.02	<.0001
Obese	-0.04	<.0001
Calendar year		
2012	Reference	
2013	0.02	<.0001
2014	0.06	<.0001
2015	0.07	<.0001
In nursing home the previous year	-0.09	<.0001
Diabetes as cause of ESRD X time on ESRD interaction term		
91 days-6 months	Reference	
6 months-1 year	0.03	0.00
1-2 years	0.00	0.95
2-3 years	-0.02	0.01
3-5 years	-0.03	<.0001
5+ years	-0.04	<.0001
Cause of ESRD: diabetes X sex: female interaction term	0.02	<.0001
Age X diabetes as cause of ESRD interaction term		
18-24	0.03	0.37
25-44	0.03	<.0001
45-59	0.03	<.0001

Covariate	Coefficient	P-value
60-74	Reference	
75+	-0.02	<.0001
Age X female sex interaction term		
18-24	0.14	<.0001
25-44	0.06	<.0001
45-59	-0.04	<.0001
60-74	Reference	
75+	0.01	0.26
Prevalent comorbidity groupers		
HIV infection	0.08	<.0001
Hepatitis	0.04	<.0001
Viral infection	0.04	<.0001
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.04	<.0001
Melanomas of skin; Other non-epithelial cancer of skin	-0.09	<.0001
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.05	<.0001
Diabetes mellitus with or without complications	0.04	<.0001
Fluid and electrolyte disorders	0.10	<.0001
Encephalitis, Meningitis and other CNS infections	-0.13	<.0001
Epilepsy; convulsions	0.06	<.0001
Headache; including migraine	0.19	<.0001
Otitis, Dizziness, and other ear and sense organ disorders	0.09	<.0001
Neuropathy, pain syndromes, and other neurologic disorders	0.06	<.0001
Essential hypertension	0.10	<.0001
Secondary hypertension and hypertensive complications	0.08	<.0001
Acute myocardial infarction and atherosclerotic heart disease	0.03	<.0001
Nonspecific chest pain	0.20	<.0001
Pulmonary embolism and other pulmonary heart disease	0.01	<.0001
Other and ill-defined heart disease	0.05	<.0001
Conduction disorders; Cardiac dysrhythmias	0.05	<.0001
Other circulatory disease	0.02	<.0001
Phlebitis; thrombophlebitis and thromboembolism	0.02	<.0001
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.09	<.0001
Chronic obstructive pulmonary disease and bronchiectasis; Asthma	0.06	<.0001
Other lower respiratory disease	0.11	<.0001
Other upper respiratory disease	0.02	<.0001
Disorders of teeth, jaw and mouth	0.12	<.0001
Esophageal disorders	0.01	<.0001
Digestive track disorders (gastritis, gastric ulcers, and other disorders of stomach; appendicitis)	0.05	<.0001
Anal and rectal conditions	0.05	<.0001

Covariate	Coefficient	P-value
Peritonitis and intestinal abscess	-0.10	<.0001
Pancreatic disorders (not diabetes)	0.13	<.0001
Gastrointestinal hemorrhage	0.02	<.0001
Noninfectious gastroenteritis	0.10	<.0001
Other gastrointestinal disorders	0.01	<.0001
Urinary tract infections	0.02	<.0001
Calculus of urinary tract	0.05	<.0001
Other diseases of kidney and ureters (e.g ureteral stricture or reflux; excludes renal calculus)	0.01	<.0001
Prostate hyperplasia, prostatitis and other male genital disorders	0.03	<.0001
Skin disorders: cellulitis, ulcers, inflammatory and others	0.04	<.0001
Infective arthritis and osteomyelitis	-0.07	<.0001
Other non-traumatic joint disorders	0.05	<.0001
Spondylosis; intervertebral disc disorders; other back problems	0.10	<.0001
Osteoporosis	-0.08	<.0001
Other connective tissue disease; Other bone disease and musculoskeletal deformities	0.07	<.0001
Sprains and strains	0.17	<.0001
Complication of device; implant or graft	0.03	<.0001
Superficial injury; contusion	0.11	<.0001
Poisoning by medications or nonmedicinal substances	0.02	<.0001
Other injuries and conditions due to external causes	0.04	<.0001
Syncope	0.05	<.0001
Gangrene	-0.07	<.0001
Shock	-0.16	<.0001
Nausea and vomiting	0.15	<.0001
Abdominal pain	0.17	<.0001
Malaise and fatigue	0.07	<.0001
Allergic reactions	0.08	<.0001
Anxiety disorders	0.10	<.0001
Attention-deficit, conduct, and disruptive behavior disorders	0.09	<.0001
Developmental disorders	0.09	<.0001
Mood disorders	0.01	<.0001
Personality disorders	0.17	<.0001
Schizophrenia and other psychotic disorders	0.02	<.0001
Alcohol-related disorders	0.20	<.0001
Suicide and intentional self-inflicted injury	0.15	<.0001
Screening and history of mental health and substance abuse codes	0.09	<.0001
Miscellaneous mental health disorders	0.05	<.0001
Missing comorbidity flag	0.82	<.0001

*The diabetes indicator includes all diabetes comorbidities on CMS-2728 and diabetes as cause of ESRD

2b4.4b. Describe the analyses and interpretation resulting in the decision to select SDS 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)

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

Table 6. Coefficients for baseline model and model with additional SDS/SES adjustors, 2012-2015

Covariate	Baseline SEDR		SDS/SES-adjusted SEDR	
	Coefficient	P-value	Coefficient	P-value
Medicare coverage*				
Medicare primary + Medicaid	NA	NA	0.19	<.0001
Medicare primary + no Medicaid	NA	NA	Reference	-
Medicare secondary/HMO	NA	NA	-0.91	<.0001
Employment status 6 months prior to ESRD				
Unemployed	NA	NA	Reference	-
Employed	NA	NA	-0.13	<.0001
Other/Unknown **	NA	NA	-0.04	<.0001
Race				
White	NA	NA	Reference	-
Native American/Alaskan Native	NA	NA	0.05	<.0001
Asian/Pacific Islander	NA	NA	-0.19	<.0001
Black	NA	NA	0.15	<.0001
Other/Unknown	NA	NA	0.04	0.01
Ethnicity				
Hispanic	NA	NA	0.04	<.0001
Non-Hispanic	NA	NA	Reference	-
Unknown	NA	NA	0.02	0.20
ADI Index	NA	NA	0.00	<.0001
Comorbidities at start of ESRD				
At least one of the comorbidities listed below	0.00	0.51	0.00	0.32
Atherosclerotic heart disease	0.02	<.0001	0.03	<.0001
Other cardiac disease	0.00	0.61	0.01	<.0001
Diabetes***	0.03	<.0001	0.03	<.0001
Congestive heart failure	0.03	<.0001	0.02	<.0001
Inability to ambulate	-0.02	0.00	-0.03	<.0001
Chronic obstructive pulmonary disease	0.02	<.0001	0.03	<.0001
Inability to transfer	-0.03	0.00	-0.03	<.0001
Malignant neoplasm, cancer	-0.03	<.0001	0.00	0.23
Peripheral vascular disease	-0.01	0.00	-0.01	0.03
Cerebrovascular disease, CVA, TIA	0.03	<.0001	0.02	<.0001
Tobacco use (current smoker)	0.08	<.0001	0.07	<.0001
Alcohol dependence	0.01	0.05	-0.01	0.39
Drug dependence	0.16	<.0001	0.11	<.0001

Covariate	Baseline SEDR		SDS/SES-adjusted SEDR	
	Coefficient	P-value	Coefficient	P-value
No Medical Evidence (CMS-2728) Form	0.02	0.01	-0.02	0.00
Cause of ESRD				
Diabetes	0.03	<.001	0.04	<.0001
Sex: Female	0.08	<.0001	0.05	<.0001
Age				
18-24	0.69	<.0001	0.59	<.0001
25-44	0.43	<.0001	0.34	<.0001
45-59	0.19	<.0001	0.13	<.0001
60-74	Reference	-	Reference	-
75+	-0.02	<.0001	0.02	<.0001
BMI				
Underweight	0.01	0.04	0.01	0.00
Normal weight	Reference	-	Reference	-
Overweight	-0.02	<.0001	-0.02	<.0001
Obese	-0.04	<.0001	-0.04	<.0001
Calendar year				
2012	Reference	-	Reference	-
2013	0.02	<.0001	0.02	<.0001
2014	0.06	<.0001	0.06	<.0001
2015	0.07	<.0001	0.08	<.0001
In nursing home the previous year	-0.09	<.0001	-0.11	<.0001
Diabetes as cause of ESRD X time on ESRD interaction term				
91 days-6 months	Reference	-	Reference	-
6 months-1 year	0.03	0.00	0.03	0.00
1-2 years	0.00	0.95	0.00	0.99
2-3 years	-0.02	0.01	-0.02	0.00
3-5 years	-0.03	<.0001	-0.04	<.0001
5+ years	-0.04	<.0001	-0.05	<.0001
Cause of ESRD: diabetes X sex: female interaction term	0.02	<.0001	0.00	0.23
Age X diabetes as cause of ESRD interaction term				
18-24	0.03	0.37	-0.04	0.32
25-44	0.03	<.0001	0.02	<.0001
45-59	0.03	<.0001	0.03	<.0001
60-74	Reference	-	Reference	-
75+	-0.02	<.0001	-0.04	<.0001
Age X female sex interaction term				
18-24	0.14	<.0001	0.17	<.0001
25-44	0.06	<.0001	0.07	<.0001
45-59	-0.04	<.0001	-0.03	<.0001
60-74	Reference	-	Reference	-
75+	0.01	0.26	0.00	0.73
Prevalent comorbidity groupers				
HIV infection	0.08	<.0001	0.05	<.0001

Covariate	Baseline SEDR		SDS/SES-adjusted SEDR	
	Coefficient	P-value	Coefficient	P-value
Hepatitis	0.04	<.0001	0.01	<.0001
Viral infection	0.04	<.0001	0.05	<.0001
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.04	<.0001	0.04	<.0001
Melanomas of skin; Other non-epithelial cancer of skin	-0.09	<.0001	-0.04	<.0001
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.05	<.0001	-0.05	<.0001
Diabetes mellitus with or without complications	0.04	<.0001	0.03	<.0001
Fluid and electrolyte disorders	0.10	<.0001	0.09	<.0001
Encephalitis, Meningitis and other CNS infections	-0.13	<.0001	-0.13	<.0001
Epilepsy; convulsions	0.06	<.0001	0.05	<.0001
Headache; including migraine	0.19	<.0001	0.18	<.0001
Otitis, Dizziness, and other ear and sense organ disorders	0.09	<.0001	0.08	<.0001
Neuropathy, pain syndromes, and other neurologic disorders	0.06	<.0001	0.06	<.0001
Essential hypertension	0.10	<.0001	0.05	<.0001
Secondary hypertension and hypertensive complications	0.08	<.0001	0.10	<.0001
Acute myocardial infarction and atherosclerotic heart disease	0.03	<.0001	0.04	<.0001
Nonspecific chest pain	0.20	<.0001	0.18	<.0001
Pulmonary embolism and other pulmonary heart disease	0.01	<.0001	0.02	<.0001
Other and ill-defined heart disease	0.05	<.0001	0.05	<.0001
Conduction disorders; Cardiac dysrhythmias	0.05	<.0001	0.06	<.0001
Other circulatory disease	0.02	<.0001	0.02	<.0001
Phlebitis; thrombophlebitis and thromboembolism	0.02	<.0001	0.02	<.0001
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.09	<.0001	0.09	<.0001
Chronic obstructive pulmonary disease and bronchiectasis; Asthma	0.06	<.0001	0.06	<.0001
Other lower respiratory disease	0.11	<.0001	0.10	<.0001
Other upper respiratory disease	0.02	<.0001	0.02	<.0001
Disorders of teeth, jaw and mouth	0.12	<.0001	0.11	<.0001
Esophageal disorders	0.01	<.0001	0.01	<.0001
Digestive track disorders (gastritis, gastric ulcers, and other disorders of stomach; appendicitis)	0.05	<.0001	0.05	<.0001
Anal and rectal conditions	0.05	<.0001	0.05	<.0001
Peritonitis and intestinal abscess	-0.10	<.0001	-0.07	<.0001
Pancreatic disorders (not diabetes)	0.13	<.0001	0.13	<.0001
Gastrointestinal hemorrhage	0.02	<.0001	0.02	<.0001

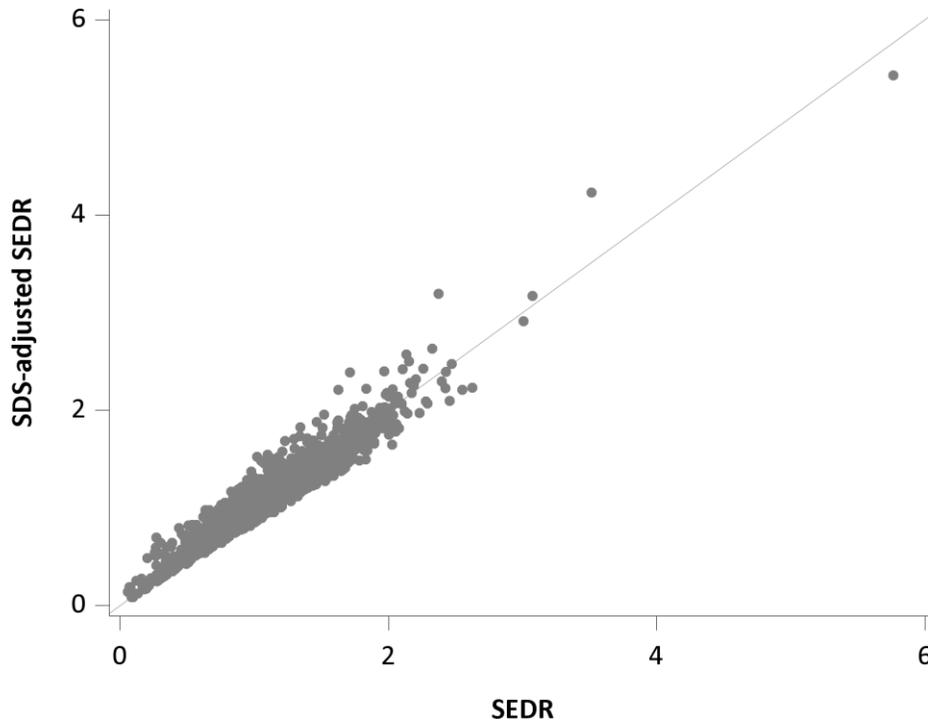
Covariate	Baseline SEDR		SDS/SES-adjusted SEDR	
	Coefficient	P-value	Coefficient	P-value
Noninfectious gastroenteritis	0.10	<.0001	0.10	<.0001
Other gastrointestinal disorders	0.01	<.0001	0.02	<.0001
Urinary tract infections	0.02	<.0001	0.03	<.0001
Calculus of urinary tract	0.05	<.0001	0.05	<.0001
Other diseases of kidney and ureters (e.g ureteral stricture or reflux; excludes renal calculus)	0.01	<.0001	0.02	<.0001
Prostate hyperplasia, prostatitis and other male genital disorders	0.03	<.0001	0.03	<.0001
Skin disorders: cellulitis, ulcers, inflammatory and others	0.04	<.0001	0.04	<.0001
Infective arthritis and osteomyelitis	-0.07	<.0001	-0.06	<.0001
Other non-traumatic joint disorders	0.05	<.0001	0.03	<.0001
Spondylosis; intervertebral disc disorders; other back problems	0.10	<.0001	0.09	<.0001
Osteoporosis	-0.08	<.0001	-0.07	<.0001
Other connective tissue disease; Other bone disease and musculoskeletal deformities	0.07	<.0001	0.06	<.0001
Sprains and strains	0.17	<.0001	0.16	<.0001
Complication of device; implant or graft	0.03	<.0001	0.01	<.0001
Superficial injury; contusion	0.11	<.0001	0.12	<.0001
Poisoning by medications or nonmedicinal substances	0.02	<.0001	0.02	<.0001
Other injuries and conditions due to external causes	0.04	<.0001	0.04	<.0001
Syncope	0.05	<.0001	0.05	<.0001
Gangrene	-0.07	<.0001	-0.07	<.0001
Shock	-0.16	<.0001	-0.15	<.0001
Nausea and vomiting	0.15	<.0001	0.14	<.0001
Abdominal pain	0.17	<.0001	0.15	<.0001
Malaise and fatigue	0.07	<.0001	0.07	<.0001
Allergic reactions	0.08	<.0001	0.09	<.0001
Anxiety disorders	0.10	<.0001	0.11	<.0001
Attention-deficit, conduct, and disruptive behavior disorders	0.09	<.0001	0.10	<.0001
Developmental disorders	0.09	<.0001	0.07	<.0001
Mood disorders	0.01	<.0001	0.02	<.0001
Personality disorders	0.17	<.0001	0.17	<.0001
Schizophrenia and other psychotic disorders	0.02	<.0001	0.01	0.03
Alcohol-related disorders	0.20	<.0001	0.18	<.0001
Suicide and intentional self-inflicted injury	0.15	<.0001	0.15	<.0001
Screening and history of mental health and substance abuse codes	0.09	<.0001	0.09	<.0001
Miscellaneous mental health disorders	0.05	<.0001	0.05	<.0001
Missing comorbidity flag	0.82	<.0001	0.92	<.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.

***The diabetes indicator includes all diabetes comorbidities on CMS-2728 and diabetes as cause of ESRD.

Figure 2. Correlation between SEDR without and with SDS adjustment, 2012-2015



Pearson correlation coefficient $\rho = 0.96$ ($p < 0.0001$)

Patient-level SDS: Compared with males, females were 5% more likely to experience an emergency department encounter (HR=1.05; $p < 0.0001$). Hispanics had a slightly higher risk of having an emergency department encounter (HR=1.04; $p < 0.0001$) than non-Hispanics. Compared with white patients, Asian/PI (HR=0.83, $p < 0.0001$) patients were almost 20% less likely to have an emergency department encounter, while Native Americans were slightly more likely (HR=1.05, $p < 0.0001$). Notably, compared to whites, black patients had a 17% higher risk (HR=1.17, $p < 0.0001$) of having an emergency department encounter. Patients in the youngest age group (18-24) had almost two-times higher risk of an emergency department encounter (HR=1.18; $p < 0.0001$) compared with the reference group (60-74). The effect shows a negative gradient moving from younger to older age categories. The results for these SDS factors are consistent with prior studies both in the respective chronic dialysis setting and general population indicating younger age, black race and female sex as potential SDS risk factors for ED use.

Patient-level SES: Compared with Medicare-only patients, dually-eligible patients with both Medicare and Medicaid (HR=1.21; $p < 0.0001$) were around 20% more likely to have an emergency department encounter. However, patients with Medicare as secondary payer/Medicare HMO (HR=0.40, $p < 0.0001$) were 60% less likely to visit the emergency department. The result for dually-eligible patients having higher risk of an emergency department encounter is consistent with prior studies demonstrating that this insurance category, on average, represents an at-risk group.

Patients who were employed prior to ESRD incidence were 11% less likely to have an emergency department encounter (HR=0.88; p<0.0001) compared to unemployed patients. This difference could reflect that patients still able to work may have potentially lower comorbidity burden and have fewer acute care encounters. However, employment information is obtained only at ESRD incidence, therefore we are unable to capture changes to patients' employment status over time and whether that corresponds with changes in emergency department use. Note that for employment categories, the "Other/Unknown" category also had a slightly lower risk of having an emergency department encounter (about 4%). We note this likely represents a diverse mix of patients with regard to SES, such as homemakers and those who are retired. The lower risk of emergency department visits may be associated with unmeasured characteristics of this heterogeneous group.

Area-level SES: The Area Deprivation Index had no impact on the risk of emergency department encounters (HR = 1.00; p<0.0001), suggesting the level of area-SES is not predictive of outpatient ED utilization.

Table 7. Flagging rates, baseline SEDR and SEDR adjusted for SDS/SES: 2012-2015

Baseline SEDR	SEDR with SDS/SES			Total
	Better than Expected	As Expected	Worse than Expected	
Better than Expected	56	16	0	72 (1.11%)
As Expected	18	6041	58	6117 (94.15%)
Worse than Expected	0	61	305	308 (4.74%)
Total	74 (1.14%)	6118 (94.17%)	305 (4.69%)	6,497

Several patient-level SDS/SES factors were predictive of higher emergency department encounter use, however when comparing the baseline SEDR measure with one that includes adjustment for patient and area-level SDS/SES, we observed very small differences in flagging of facility performance (Figure 2 and Table 7). For example, in the baseline SEDR, 308 facilities are flagged as worse than expected while 305 are flagged as worse than expected in the SEDR adjusted for SDS/SES, resulting in a negligible decrease in the number of facilities flagged for worse than expected performance. Additionally, both the baseline SEDR and SEDR adjusted for SDS/SES are highly correlated ($\rho = 0.96$ ($p < 0.0001$)). For these reasons and the lack of definitive evidence indicating that differences are primarily attributable to patient or area-level SDS/SES factors versus facility practices, no additional risk adjustment is made for patient race, ethnicity, or patient and area-level SES.

2b4.5. 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*)

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.

Provide the statistical results from testing the approach to controlling for differences in patient characteristics (case mix) below.

If stratified, skip to 2b4.9

2b4.6. Statistical Risk Model Discrimination Statistics (e.g., c-statistic, R-squared):

The estimate of the C-statistic for the SEDR is 0.67.

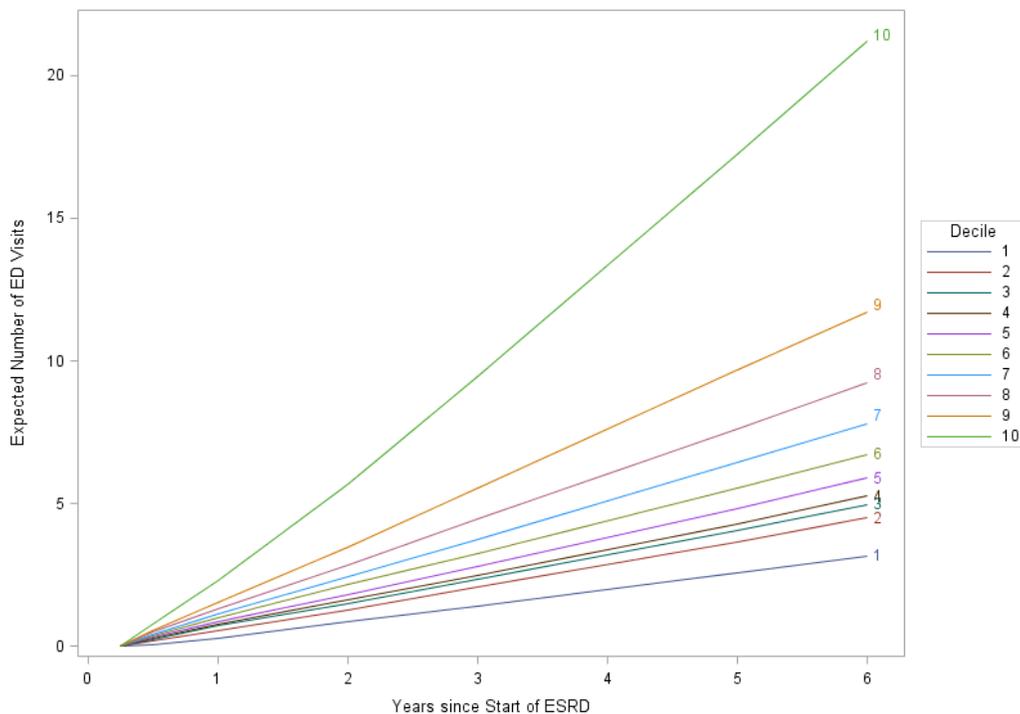
2b4.7. Statistical Risk Model Calibration Statistics (e.g., Hosmer-Lemeshow statistic):

N/A

2b4.8. Statistical Risk Model Calibration – Risk decile plots or calibration curves:

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

Figure 3. Decile Plot for SEDR (2012-2015 data)



2b4.9. Results of Risk Stratification Analysis:

N/A

2b4.10. What is your interpretation of the results in terms of demonstrating adequacy of controlling for differences in patient characteristics (case mix)? (i.e., what do the results mean and what are the norms for the test conducted)

The decile plot (Figure 3) shows that the risk factors in the model are discriminating well between patients. There is good separation among all 10 groups, and the ordering is as predicted by the model (i.e. patients predicted to be at lower risk have lower emergency department rates). The absolute differences between the groups is also large, with patients predicted to have the highest emergency department rates (line 10) having about 7 times higher emergency department rates than those predicted to have the lowest rates (line 1). This means that the model fit is good and therefore adequately adjusts for patient characteristics (case mix).

2b4.11. 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

2b5. IDENTIFICATION OF STATISTICALLY SIGNIFICANT & MEANINGFUL DIFFERENCES IN PERFORMANCE

2b5.1. Describe the method for determining if statistically significant and clinically/practically meaningful differences in performance measure scores among the measured entities can be identified (*describe the steps—do not just name a method; what statistical analysis was used? Do not just repeat the information provided related to performance gap in 1b*)

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

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

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

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

References:

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

2b5.2. What were the statistical results from testing the ability to identify statistically significant and/or clinically/practically meaningful differences in performance measure scores across measured entities? (e.g., number and percentage of entities with scores that were statistically significantly different from mean or some benchmark, different from expected; how was meaningful difference defined)

Table 8. Number and percentage of facilities by classification of SEDR, 2015.

Better than expected	As expected	Worse than expected	Total
0.64% (40)	93.86% (5,872)	5.50% (344)	6,256

2b5.3. 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) that are extreme when compared to the variation in outcomes for other facilities of a similar size. Overall, most are flagged as expected (about 94%), while <1% are better than expected, and approximately 6% are flagged as worse than expected.

2b6. 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 SDS 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 SDS 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.**

2b6.1. 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

2b6.2. 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

2b6.3. 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

2b7. MISSING DATA ANALYSIS AND MINIMIZING BIAS

2b7.1. 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*)

N/A

2b7.2. 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)

N/A

2b7.3. 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*)

N/A