

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

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: [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 decision making) 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 measures 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 identified from Medicare inpatient, outpatient, skilled nursing facility, home health, and hospice claims.

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

January 2014 – December 2014 for index discharges
 January 2014 – January 2015 for emergency department encounters
 January 2013 – December 2014 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: (<i>must be consistent with levels entered in item S.26</i>)	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*)

We included all Medicare-certified facilities (n=5,959) that had at least 11 eligible index discharges in 2014. Median facility size was 66 patients as measured by all dialysis patients in the facility on December 31, 2014.

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 266,564 Medicare ESRD dialysis patients with at least one hospitalization during 2014. There were 596,081 hospitalizations for this patient population. After applying the exclusion criteria, these data represent 423,720 index discharges among 201,938 unique patients during 2014.

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

Patient Demographics	Percent of Index Discharges
Age	
Patient Age: 18 to 24	0.7
Patient Age: 25 to 44	11.7
Patient Age: 45 to 59	26.8

	Percent of Index Discharges
Patient Demographics	
Patient Age: 60 to 74	38.7
Patient Age: 75+	21.9
Sex (% female)	48.7
ESRD due to Diabetes	49.1
Medicare coverage	
Medicare primary + Medicaid	44.81
Medicare primary + no Medicaid	38.19
Medicare secondary/HMO	17.0
Time since Start of ESRD	
91 days-6 months	5.9
6 months-1 year	9.4
1-2 years	15.6
2-3 years	13.4
3-5 years	21.1
5+ years	34.6
Employment status 6 months prior to ESRD	
Unemployed	22.9
Employed	18.5
Other/Unknown*	58.5
Race	
White	57.0
Black	36.7
Native American/Alaskan Native	1.1
Asian/Pacific Islander	3.6
Other/Unknown	1.6
Ethnicity	
Hispanic	15.7
Non-Hispanic	82.9
Unknown	1.4

* 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 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 date of index discharge. 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)

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, which uses a resampling scheme to estimate the within facility variation that cannot be directly estimated by ANOVA. Refer to Appendix for a detailed description of this methodology.

The measure calculation only included facilities that had at least 11 eligible index discharges in 2014.

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 = 0.35, which indicates that 35% of the variation in the ED30 can be attributed to the between-facility differences and 65% to within-facility variation.

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 fair. 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 ED30 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 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.

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), dialysis adequacy (Kt/V ≥ 1.2), Standardized Mortality Ratio (SMR), SRR, and the Standardized Emergency Department Ratio (SEDR) which is currently being submitted for endorsement as a companion measure to 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 ED30 and vascular access type, hemodialysis adequacy, Standardized Mortality Ratio (SMR), Standardized Readmission Ratio (SRR), and, SEDR, respectively are presented in Table 2.

Table 2. Spearman Correlation of ED30 and Related Measures, 2014

	Correlation	P-value
Vascular Access: Fistula	-0.03	0.0159
Vascular Access: Catheter >90 days	-0.03	0.0086
Kt/V ≥ 1.2	-0.02	0.0816
SMR	0.08	<0.0001
SRR	-0.06	<0.0001
SEDR	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 ED30 correlates with several dialysis facility processes and outcomes that are commonly thought to be related to quality of care. Higher rates of arteriovenous fistula use are associated with lower emergency department utilization in the 30 days that follow hospital discharge. The magnitude of the association is in the expected direction and is statistically significant, although the strength of the association is weak. The results suggests that facilities with processes of care to provide optimal vascular access may have other processes in place to help coordinate post-hospital care and thus avoid needing the ED for unscheduled acute care. However, a similar result was also seen with long-term dialysis catheter use. It may be that ED providers have a lower threshold to (re)admit patients with a catheter in the post-hospitalization period rather than treat them in the ED outpatient setting. Facilities with higher percentages of patients with Kt/V ≥ 1.2 was only weakly associated with less ED use in the 30 days after hospital discharge however, this was not statistically significant. It may be that patients with poor dialysis adequacy are sicker and more likely to have an inpatient acute care encounter versus an outpatient ED encounter.

Higher ED utilization was weakly associated with higher facility mortality rates, while it was associated with lower readmissions (SRR). ED30 focuses on outpatient use of ED services whereas SRR captures inpatient readmissions and ED use that results in readmission, therefore the ED30 measure likely captures dialysis patients that have a lower acuity of illness than the SRR. The weak association also indicates the competing risks of an outpatient ED visit versus a readmission, as only outpatient ED visits are included in the ED30. A patient that has an ED encounter will either be discharged from the ED (and counted in ED30) or be (re)admitted (and counted in SRR) therefore they can only be counted in one measure, not both.

Lastly, we assessed the correlation between the ED30 and the Standardized Emergency Department Ratio (SEDR) measure, which is also being submitted for consideration of NQF endorsement. The SEDR describes emergency department encounter rates with reference to the totality of patients being treated by a given facility. 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 ED30 measure with and without the hospice exclusion. Additionally, we calculated the number and percentage of excluded discharges for each exclusion criterion.

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

1. Has had ESRD for 90 days or less at time of discharge
2. 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:

3. Actively enrolled in hospice at any time during the calendar month of the discharge or ED encounter

Additionally we exclude hospital discharges that:

4. Do not result in a live discharge
5. Result in a patient dying, being transplanted, discontinuing dialysis, recovering renal function, or becoming lost-to-follow-up within 30 days with no emergency department encounter or hospitalization
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. Result in another hospitalization within four days of discharge

Exclusion criteria 4-9 are aligned with the Standardized Readmission Ratio (NQF 2496) measure. Based on input from the May 2016 TEP, we additionally excluded pediatric patients, hospice patients, and patients in their first 90 days of ESRD treatment. A majority of TEP members proposed excluding pediatric patients due to substantial differences in both the pediatric population comorbidities as well as reasons for seeking care in the ED when compared to the adult population. Hospice patients were excluded to allow for ED care that may be palliative in nature and directed by other providers outside of the dialysis facility. These concerns are relevant in the context of the measure's potential applications, which are to identify poor-performing facilities for quality improvement purposes. The first 90 days of ESRD treatment are excluded to be consistent with the proposed SEDR measure. While the SRR excludes

hospital discharges that are followed by the patient dying within 30 days with no event, for the proposed ED30 measure we exclude these discharges as well as those that are followed by the patient receiving a transplant, discontinuing dialysis, recovering renal function, or becoming lost-to-follow-up within 30 days without an event, and those which result in another hospitalization within four days of discharge. These additional exclusions are consistent with the original exclusion in that they focus the measure on index discharges for which we have complete follow-up for determining whether there is an event within 30 days.

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)*

A total of 172,361 hospital discharges among 64,950 unique patients were excluded. The number and percentage of excluded discharges are as follows:

Discharges for which the patient:

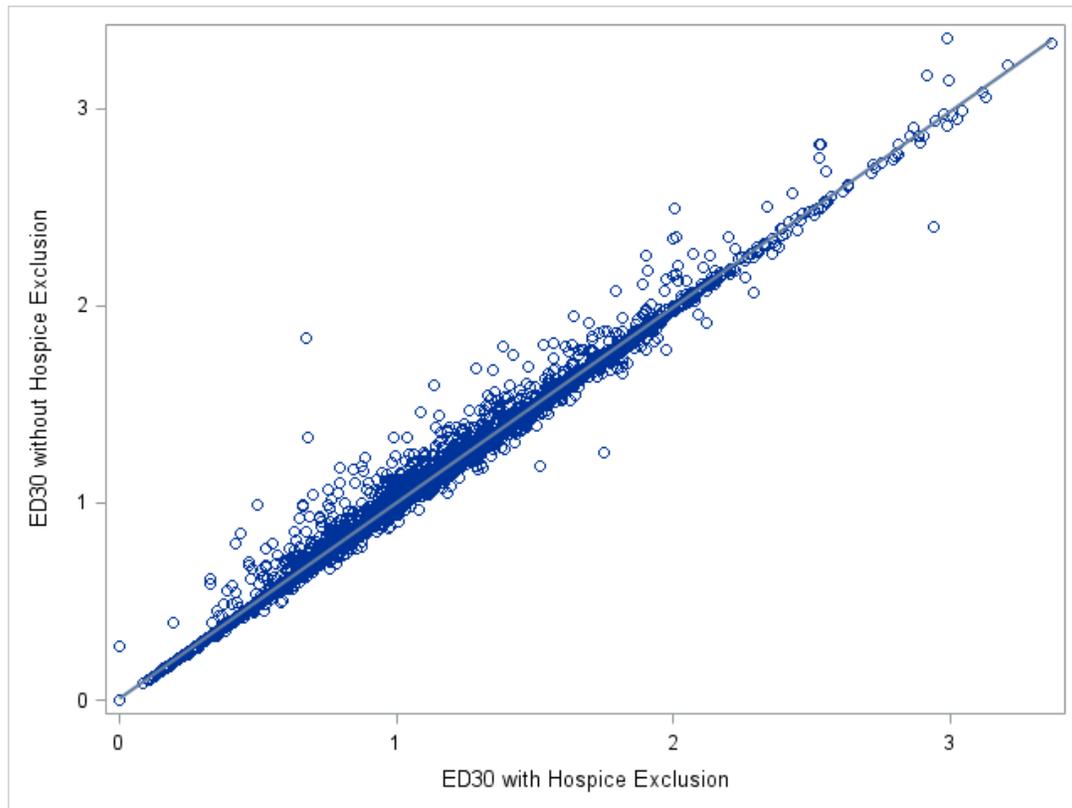
1. Has had ESRD for 90 days or less at time of discharge ($n = 65,681$; 11.0%)
2. Is less than 18 years of age at the time of discharge ($n = 1,203$; 0.2%)
3. Is actively enrolled in hospice at time of discharge ($n = 12,195$; 2.0%)

Additionally, we exclude hospital discharges that:

4. Do not result in a live discharge ($n = 25,968$; 4.4%)
5. Result in a patient dying, receiving a transplant, recovering kidney function, discontinuing dialysis, or becoming lost-to-follow-up within 30 days with no emergency department encounter or hospitalization ($n = 15,323$; 2.6%)
6. Are against medical advice ($n = 9,225$; 1.5%)
7. Include a primary diagnosis for cancer, mental health or rehabilitation ($n = 12,970$; 2.2%)
8. Are from a PPS-exempt cancer hospital ($n = 148$; 0.02%)
9. Result in another hospitalization within four days of discharge ($n=24,237$; 4.1%)

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 $[r] = 0.9939$ p -value <0.0001).

Figure 1. Correlation between ED30 with and without the hospice exclusion (2014)



Overall Correlation=0.99 p-value <0.0001

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 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.

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 88 risk factors
- 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.

We use a two-stage model, the first of which is a double random-effects logistic regression model. In the first stage, both dialysis facilities and hospitals are represented as random effects, and we make regression adjustments for the set of patient-level characteristics listed below. From this model, we obtain the estimated standard deviation of the random effects of hospitals. Next, each facility's expected rate of emergency department encounters is calculated by regressing the probability of an emergency department encounter within 4–30 days on the same set of risk factors.

The second stage is a mixed-effects logistic regression model, in which we include facilities as fixed effects and hospitals as random effects, with the standard deviation specified as equal to its estimates from the first model. The expected number of emergency department encounters within 4-30 days following a hospital discharge 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 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 model.

Risk factors (these are included in both stages of the model):

- Fixed effect for dialysis facility receiving discharged patient
- Random effect for hospital discharging the patient
- Sex: We determine each patient's sex from his/her Medical Evidence Form (CMS-2728).
- Age at index discharge: We determine each patient's age from the date of birth provided in the CROWNWeb and REMIS databases, Medicare claims, SRTR data, and Medical Evidence Form (CMS-2728) and group patients into the following categories: 18-24, 25-44, 45-59, 60-74, 75+
- Years on dialysis as of index discharge: 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 date of discharge.
- Diabetes as cause of ESRD: We determine each patient's primary cause of ESRD from his/her CMS-2728.
- BMI at incidence of ESRD: 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.
- Length (days) of index hospitalization
- Prevalent comorbidities (see Appendix) are determined using the previous 12 months of Medicare claims after the index discharge. 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)

Selection of clinical factors: The list of covariates considered was based on CMS' Standardized Readmission Ratio for Dialysis Facilities (NQF 2496) and separate empirical evaluation of prevalent comorbidities associated with risk of an ED encounter. Therefore, ED30 includes a different set of comorbidities than SRR as the comorbidities associated with high risk of readmission occurred at a very low frequency among patients with an outpatient ED encounter only.

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 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 ED-30 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.

We fit the model adjusting for factors (listed above in 2b4.1.1) that were included in the SRR model, other than discharged with a high-risk condition, and checked for statistical significance.

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

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 (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, 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 a post-acute or acute visit are 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 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, along had 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) 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.

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 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.

Lovasik BP, Zhang R, Hockenberry JM, Schragger 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.

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.

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

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

Table 3. Baseline ED 30 Model Coefficients and Odds Ratios – Data Year 2014

Covariate	Coefficient	Odds Ratio	P-value
Sex			
Female	0.01381	1.01391	0.2499
Male	Reference		
Age			
18-24	0.44076	1.55389	<.0001
25-44	0.39763	1.48829	<.0001
45-59	0.16676	1.18147	<.0001
60-74	Reference		
75+	-0.08072	0.92246	<.0001
Cause of ESRD			
Diabetes	0.01657	1.01670	0.4930
BMI			
Underweight	0.01460	1.01471	0.3267
Normal weight	Reference		
Overweight	-0.02757	0.97281	0.0040
Obese	-0.02183	0.97840	0.0022
Time on ESRD			
91 days-6 months	0.09253	1.09694	0.0003
6 months-1 year	0.01698	1.01712	0.2142
1-2 years	Reference		
2-3 years	0.09127	1.09556	0.0001
3-5 years	0.12796	1.13651	<.0001
5+ years	0.16578	1.18031	<.0001
Length of Hospital Stay (Q1 – Shortest stay)			
Length of Stay (Q1)	Reference		
Length of Stay (Q2)	-0.07195	0.93058	<.0001
Length of Stay (Q3)	-0.05980	0.94195	<.0001
Length of Stay (Q4)	-0.08668	0.91697	<.0001
Prevalent comorbidity groupers			
HIV infection	0.07921	1.08243	0.0133
Hepatitis	0.06151	1.06345	<.0001
Viral infection	-0.01651	0.98363	0.2774
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.01876	1.01893	0.0013
Melanomas of skin; Other non-epithelial cancer of skin	-0.14463	0.86534	<.0001

Covariate	Coefficient	Odds Ratio	P-value
Benign neoplasm of uterus; Other and unspecified benign neoplasm	0.02861	1.02902	0.4209
Diabetes mellitus with or without complication	-0.00133	0.99867	0.3768
Fluid and electrolyte disorders	0.07810	1.08123	0.0633
Encephalitis, Meningitis and other CNS infections	-0.09001	0.91392	0.0442
Epilepsy; convulsions	0.07640	1.07940	<.0001
Headache; including migraine	0.15222	1.16442	<.0001
Otitis, Dizziness, and other ear and sense organ disorders	0.12889	1.13756	<.0001
Neuropathy, pain syndromes, and other neurologic disorders	0.04668	1.04778	<.0001
Essential hypertension	0.25808	1.29444	<.0001
Secondary hypertension and hypertensive complications	0.19064	1.21002	<.0001
Acute myocardial infarction and atherosclerotic heart disease	0.03242	1.03295	0.0015
Nonspecific chest pain	0.23237	1.26159	<.0001
Pulmonary embolism and other pulmonary heart disease	0.00086	1.00086	0.3059
Other and ill-defined heart disease	0.07719	1.08025	<.0001
Conduction disorders; Cardiac dysrhythmias	0.02492	1.02523	0.0105
Other circulatory disease	0.01549	1.01561	0.0086
Phlebitis; thrombophlebitis and thromboembolism	0.01629	1.01642	0.0238
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.10665	1.11254	<.0001
Chronic obstructive pulmonary disease and bronchiectasis; Asthma	0.03394	1.03452	0.0006
Other lower respiratory disease	0.15137	1.16342	<.0001
Other upper respiratory disease	0.05880	1.06057	0.0017
Disorders of teeth, jaw and mouth	0.10666	1.11256	<.0001
Esophageal disorders	0.00988	1.00993	0.0975
Digestive track disorders	0.01691	1.01706	0.0014
Anal and rectal conditions	0.10334	1.10887	0.0029
Peritonitis and intestinal abscess	-0.02597	0.97436	<.0001
Pancreatic disorders (not diabetes)	0.09770	1.10264	<.0001
Gastrointestinal hemorrhage	-0.03236	0.96815	0.0289
Noninfectious gastroenteritis	0.02892	1.02934	<.0001
Other gastrointestinal disorders	0.04794	1.04911	<.0001
Urinary tract infections	0.03780	1.03852	0.0001
Calculus of urinary tract	0.01140	1.01147	0.2497
Other diseases of kidney and ureters (e.g ureteral stricture or reflux; excludes renal calculus)	0.07203	1.07468	<.0001
Prostate hyperplasia, prostatitis and other male genital disorders	0.05566	1.05724	0.1028
Skin disorders: cellulitis, ulcers, inflammatory and others	0.03706	1.03775	0.0144
Infective arthritis and osteomyelitis	-0.05505	0.94644	0.0580
Other non-traumatic joint disorders	0.06313	1.06516	<.0001
Spondylosis; intervertebral disc disorders; other back problems	0.09570	1.10043	<.0001
Osteoporosis	-0.11128	0.89469	<.0001
Other connective tissue disease; Other bone disease and musculoskeletal deformities	0.14536	1.15646	<.0001
Sprains and strains	0.19005	1.20930	<.0001

Covariate	Coefficient	Odds Ratio	P-value
Complication of device; implant or graft	0.11637	1.12341	<.0001
Superficial injury; contusion	0.19419	1.21432	<.0001
Poisoning by medications or nonmedicinal substances	0.03842	1.03917	0.0034
Other injuries and conditions due to external causes	-0.00677	0.99325	0.1470
Syncope	0.09950	1.10462	<.0001
Gangrene	-0.05587	0.94566	0.0011
Shock	-0.08040	0.92275	<.0001
Nausea and vomiting	0.19472	1.21496	<.0001
Abdominal pain	0.24978	1.28375	<.0001
Malaise and fatigue	0.08806	1.09205	<.0001
Allergic reactions	0.04392	1.04490	<.0001
Anxiety disorders	0.05924	1.06103	<.0001
Attention-deficit, conduct, and disruptive behavior disorders	0.13572	1.14536	0.0560
Developmental disorders	0.20095	1.22257	<.0001
Mood disorders	-0.00848	0.99156	0.4986
Personality disorders	0.07231	1.07499	0.0119
Schizophrenia and other psychotic disorders	0.08044	1.08376	0.0002
Alcohol-related disorders	0.15864	1.17191	<.0001
Suicide and intentional self-inflicted injury	0.10595	1.11176	0.0023
Screening and history of mental health and substance abuse codes	0.07782	1.08093	<.0001
Miscellaneous mental health disorders	0.09364	1.09817	0.0076

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)

Using hierarchical binary logistic regression we fit an additional model for ED30 to 2014 hospitalization data, including covariates from the original ED30 model and adding several SES/SDS indicators as well as patients' race, and ethnicity. Table 4 shows effects from these selected additional covariates in the SES/SDS model.

Table 4. Coefficients and Odds Ratios for Baseline Model and Model with Additional SDS/SES Adjustors: 2014

Covariate	Coefficient	Odds Ratio	P-value
Sex			
Female	-0.00122	0.99878	0.1974
Male	Reference		
Age			
18-24	0.44032	1.55321	<.0001
25-44	0.31350	1.36820	<.0001
45-59	0.11957	1.12701	<.0001
60-74	Reference		
75+	-0.01974	0.98046	<.0001
Cause of ESRD			
Diabetes	0.02850	1.02891	0.3002

Covariate	Coefficient	Odds Ratio	P-value
BMI			
Underweight	0.02604	1.02639	0.3683
Normal weight	Reference		
Overweight	-0.04623	0.95482	0.0042
Obese	-0.02384	0.97645	0.0007
Time on ESRD			
91 days-6 months	0.11604	1.12304	<.0001
6 months-1 year	0.06353	1.06560	0.0978
1-2 years	Reference		
2-3 years	0.03715	1.03784	0.0085
3-5 years	0.02576	1.02609	0.0086
5+ years	0.03956	1.04035	0.0023
Length of Hospital Stay (Q1 – Shortest stay)			
Length of Stay (Q1)	Reference		
Length of Stay (Q2)	-0.05593	0.94561	<.0001
Length of Stay (Q3)	-0.06446	0.93757	<.0001
Length of Stay (Q4)	-0.05651	0.94506	<.0001
ADI	-0.09734	0.90725	0.1748
Race			
White	Reference		
Native American/Alaskan Native	-0.10412	0.90112	0.1279
Asian/Pacific Islander	-0.07512	0.92763	0.0952
Black	0.13644	1.14619	<.0001
Other/Unknown	-0.08854	0.91527	0.3767
Ethnicity			
Hispanic	0.08026	1.08357	0.0060
Non-Hispanic	Reference		
Unknown	0.21127	1.23525	0.4389
Medicare coverage*			
Medicare primary + Medicaid	0.12742	1.13589	<.0001
Medicare primary + no Medicaid	Reference		
Medicare secondary/HMO	-1.58880	0.20417	<.0001
Employment status 6 months prior to ESRD			
Employed**	-0.03394	0.96663	0.0509
Unemployed	Reference		
Retired/Other/Unknown***	-0.04535	0.95566	0.0299
Prevalent comorbidity groupers			
HIV infection	0.00281	1.00281	0.2127
Hepatitis	0.01067	1.01073	0.0553
Viral infection	-0.03464	0.96596	0.3984
Other infections; including parasitic; Sexually transmitted infections (not HIV or hepatitis)	0.05344	1.05489	0.0030
Melanomas of skin; Other non-epithelial cancer of skin	-0.13029	0.87784	0.002524
Benign neoplasm of uterus; Other and unspecified benign neoplasm	-0.02175	0.97848	0.314509
Diabetes mellitus with or without complicatio	-0.01658	0.98356	0.479848
Fluid and electrolyte disorders	0.08929	1.09339	0.050783

Covariate	Coefficient	Odds Ratio	P-value
Encephalitis, Meningitis and other CNS infections	-0.02673	0.97363	0.052237
Epilepsy; convulsions	0.07871	1.08189	1.97E-08
Headache; including migraine	0.16174	1.17555	<.0001
Otitis, Dizziness, and other ear and sense organ disorders	0.11284	1.11945	<.0001
Neuropathy, pain syndromes, and other neurologic disorders	0.05923	1.06102	<.0001
Essential hypertension	0.13240	1.14156	<.0001
Secondary hypertension and hypertensive complications	0.22155	1.24801	0.002153
Acute myocardial infarction and atherosclerotic heart disease	0.05028	1.05157	<.0001
Nonspecific chest pain	0.20470	1.22716	<.0001
Pulmonary embolism and other pulmonary heart disease	0.00516	1.00518	0.277761
Other and ill-defined heart disease	0.02701	1.02738	0.000931
Conduction disorders; Cardiac dysrhythmias	0.04424	1.04524	<.0001
Other circulatory disease	0.02251	1.02276	0.186181
Phlebitis; thrombophlebitis and thromboembolism	0.03416	1.03475	0.011428
Acute and chronic tonsillitis; Acute bronchitis; Other upper respiratory infections	0.11152	1.11798	<.0001
Chronic obstructive pulmonary disease and bronchiectasis; Asthma	0.03330	1.03386	0.000137
Other lower respiratory disease	0.09669	1.10151	0
Other upper respiratory disease	0.03474	1.03535	0.000394
Disorders of teeth, jaw and mouth	0.10018	1.10537	<.0001
Esophageal disorders	-0.00302	0.99698	0.085143
Digestive track disorders	0.04376	1.04474	0.000271
Anal and rectal conditions	0.07991	1.08319	0.024825
Peritonitis and intestinal abscess	-0.07429	0.92840	<.0001
Pancreatic disorders (not diabetes)	0.09400	1.09856	<.0001
Gastrointestinal hemorrhage	-0.02054	0.97967	0.018496
Noninfectious gastroenteritis	0.08764	1.09159	<.0001
Other gastrointestinal disorders	0.06681	1.06909	<.0001
Urinary tract infections	0.02591	1.02625	0.000173
Calculus of urinary tract	0.04056	1.04139	0.207938
Other diseases of kidney and ureters (e.g ureteral stricture or reflux; excludes renal calculus)	0.04269	1.04361	0.000238
Prostate hyperplasia, prostatitis and other male genital disorders	0.03525	1.03588	0.007332
Skin disorders: cellulitis, ulcers, inflammatory and others	0.00697	1.00699	0.01011
Infective arthritis and osteomyelitis	-0.01630	0.98383	0.15273
Other non-traumatic joint disorders	0.07207	1.07473	0.013794
Spondylosis; intervertebral disc disorders; other back problems	0.08251	1.08601	<.0001
Osteoporosis	-0.02434	0.97595	0.00029
Other connective tissue disease; Other bone disease and musculoskeletal deformities	0.08567	1.08945	<.0001
Sprains and strains	0.19310	1.21301	<.0001
Complication of device; implant or graft	0.06656	1.06882	<.0001
Superficial injury; contusion	0.12473	1.13284	<.0001
Poisoning by medications or nonmedicinal substances	0.02183	1.02207	0.096101
Other injuries and conditions due to external causes	-0.01566	0.98446	0.260256

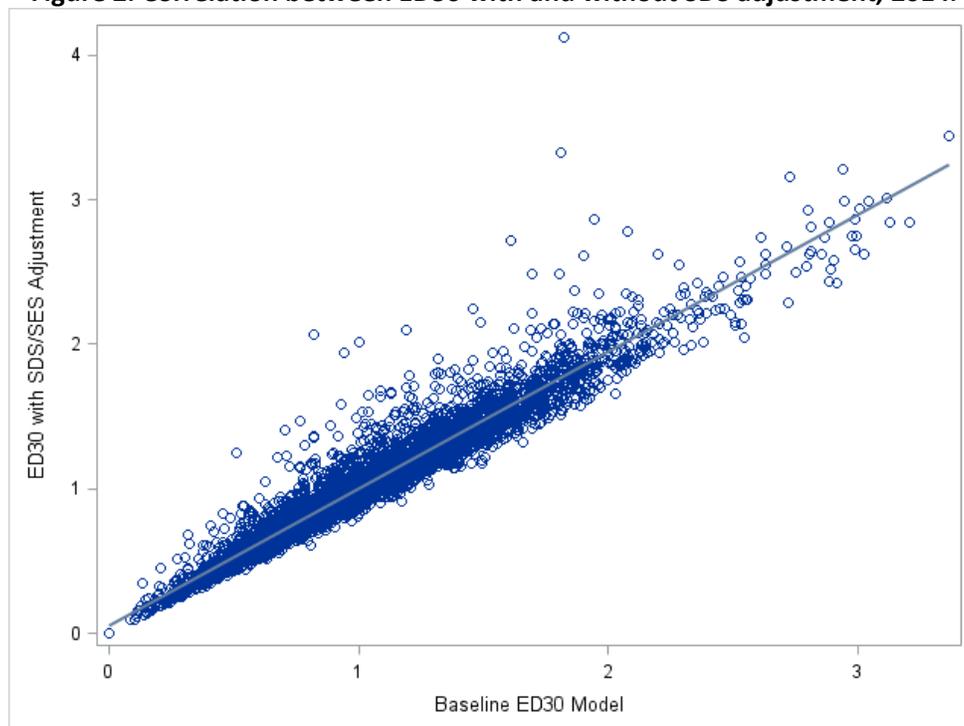
Covariate	Coefficient	Odds Ratio	P-value
Syncope	0.10908	1.11525	<.0001
Gangrene	-0.07197	0.93056	0.01186
Shock	-0.12287	0.88437	<.0001
Nausea and vomiting	0.14154	1.15205	<.0001
Abdominal pain	0.20995	1.23362	<.0001
Malaise and fatigue	0.07528	1.07818	<.0001
Allergic reactions	0.07575	1.07870	<.0001
Anxiety disorders	0.09682	1.10166	<.0001
Attention-deficit, conduct, and disruptive behavior disorders	0.14915	1.16084	0.031569
Developmental disorders	0.17924	1.19631	<.0001
Mood disorders	-0.02082	0.97940	0.147053
Personality disorders	0.10387	1.10946	0.001746
Schizophrenia and other psychotic disorders	0.03170	1.03220	0.001447
Alcohol-related disorders	0.17735	1.19405	<.0001
Suicide and intentional self-inflicted injury	0.14181	1.15236	<.0001
Screening and history of mental health and substance abuse codes	0.06533	1.06751	<.0001
Miscellaneous mental health disorders	0.03896	1.03973	0.013996

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

**Employed includes patients who are full-time employed, part-time employed, or students.

***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.

Figure 2. Correlation between ED30 with and without SDS adjustment, 2014.



Overall Correlation coefficient $\rho=0.97$ ($p < 0.0001$)

We did a sensitivity analysis comparing the baseline ED30 measure to results that included adjustment for SDS/SES factors.

Patient-level SDS: There was no difference between males and females in odds of experiencing an emergency department encounter within 4-30 days of discharge (OR=1.00, p=0.1974). Compared with non-Hispanics, Hispanics had 8% higher odds of an emergency department encounter within 4-30 days of discharge (OR=1.08; p=0.0060). The odds of an ED encounter for Native Americans and Asian/PI patients was slightly reduced compared to whites however these were not statistically significant [Native Americans (OR=0.90, p=0.1279) and Asian/PI patients (OR=0.93, p=0.0952)]. Notably, compared to whites, black patients had 15% higher odds (OR=1.15, p<0.0001) of an emergency department encounter within 4-30 days of discharge. The results for these patient-level SDS factors are consistent with prior studies both in the respective chronic dialysis setting and general population indicating 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 (OR=1.14; p <0.0001) had 14% higher odds of visiting the emergency department within 4-30 days after an inpatient discharge. In striking contrast, patients with Medicare as secondary payer/Medicare HMO (OR=0.20, p <0.0001) had 80% lower odds of having an emergency department encounter within 4-30 days. The result for dually-eligible patients having higher odds 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 had slightly lower odds of having an emergency department encounter within 4-30 days but this was only marginally significant (OR=0.97; 0.0509) compared to unemployed patients. 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.

Area-level SES: While higher area-level deprivation (ADI) reduced the odds of an emergency department encounter (OR=0.91, p=0.1952), the effect was not significant. This could indicate more granular measures of SES may be needed to better assess the impact of SES on ED use.

We also examined how the different modeling approaches changed how facilities were flagged in terms of their expected ED30 performance. As shown in Table 5, the flagging rates changed minimally between the original ED30 measure and the sensitivity model that includes SDS/SES.

Table 5. Flagging rates, baseline ED30 and ED30 adjusted for SES/SDS, 2014

Baseline ED30	ED30 with SDS/SES			Total
	Better than Expected	As Expected	Worse than Expected	
Better than Expected	46	26	0	72 (1.21%)
As Expected	60	5614	37	5711(95.84%)
Worse than Expected	0	62	114	176 (2.95%)
Total	106 (1.78%)	5702 (95.69%)	151 (2.53%)	5959

These results show that facility profiling changes very little with the addition of these selected patient- or area-level SDS/SES factors. Twenty-five fewer facilities are flagged as worse than expected and there are no changes in better than expected in the model adjusting for SDS/SES. This empirical finding demonstrating minimal differences, coupled with the risk of reducing patients' access to high quality care supports the decision to not adjust ED30 for the selected SDS/SES factors.

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.

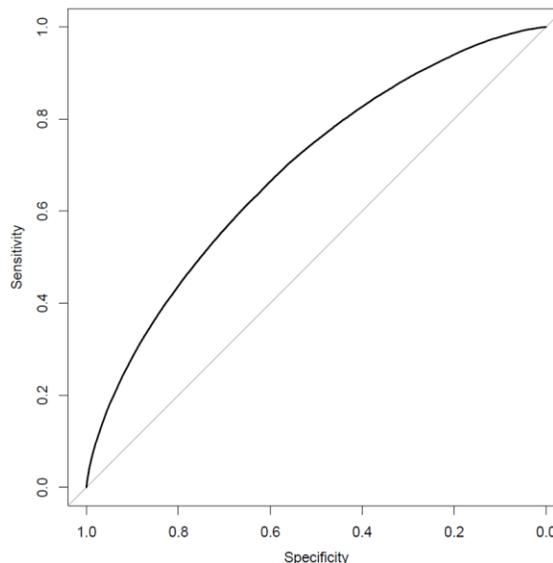
If stratified, skip to 2b4.9

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.

2b4.6. 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 = 0.688.

Figure 3. ROC Curve for Model (2014)

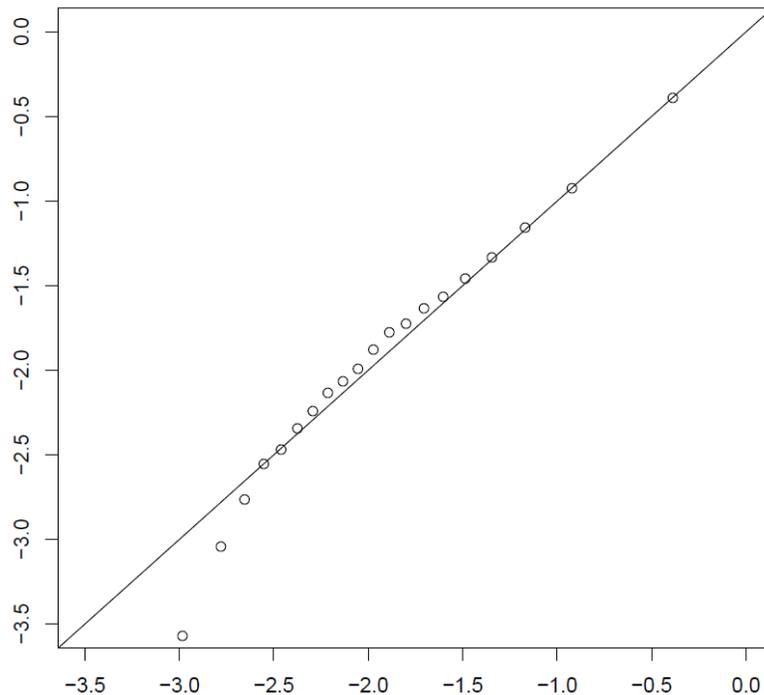


2b4.7. Statistical Risk Model Calibration Statistics (e.g., Hosmer-Lemeshow statistic):

N/A

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

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



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)

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).

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 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 three groups based on numbers of discharges 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. 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.

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 6 shows the number of facilities classified as extreme using the method described in the prior question. We find 72 (1.2%) facilities with ED30 that are better than expected and 176 (3.0%) that are worse than expected.

Table 6. Number and percentage of facilities by classification of ED30, 2014

Better than Expected	As Expected	Worse than Expected	Total Facilities
1.21% (72)	95.84% (5,711)	2.95% (176)	5,959

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 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 (95.8%), while 1.2% are better than expected, and approximately 3.0% 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.

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