4.2.3 Reliability Testing Results

Encounter-Level Reliability Testing Results

Table 13. Inter-rater reliability testing of data elements used in the proposed chest-imaging confirmed diagnosis of pneumonia eCQM.

Among hospitalizations with pneumonia identified from emergency department or inpatient stay during admission or discharge by ICD-9 and ICD-10 codes or NLP algorithm for pneumonia (random sample of 100 hospitalizations from denominator/initial population (50 at each site).			
**		2-Reviewer Inter-rater Reliability (kappa)	
**	Prevalence	VA	Univ. of Utah
Hospital discharge diagnosis of pneumonia based on presence of ICD-10 code	VA: 26% UU: 24% UM: 99.9%	**0.88**	**0.91**
Receipt of antimicrobials (antibiotics or antivirals)	VA: 99% UU: 58% UM: 99.5%	n/a*	n/a*
Chest imaging consistent with pneumonia (NLP document level reliability)	VA: 64% UU: 50% UM: 99.9%	**0.87**	**0.87**
Random sample of 100 records per site among hospitalizations with pneumonia diagnosed in the emergency department or as an inpatient that were initially identified by ICD code or NLP, the hospitalizations with a discharge diagnosis of pneumonia and treatment with antimicrobials (N=89,767 within VA and 3,030 at U. of Utah).			
Discharge diagnosis and treatment for pneumonia confirmed by chest imaging	VA: 90% UU: 92%	n/a*	n/a*
Prior work has previously established excellent data capture of receipt antimicrobials, thus this data element was not re-validated.[73] ****Cells intentionally left blank.**			

Accountable Entity-Level Reliability Testing Results

Table 14. Accountable entity level reliability testing of one year of eCQM performance among 100 VA acute care hospitals, stratified by decile of denominator size (where denominator is number of hospitalizations with a diagnosis of pneumonia and treatment with antimicrobials), 2021 (N=8,253 hospitalizations). Reliability estimates using both Adams and empirical Bayes methods are shown.

**	Min	Decile 1	Decile 2	Decile 3	Decile 4	Decile 5	Decile 6	Decile 7	Decile 8	Decile 9	Decile 10	Max	Overall
Average N patients per facility	6	24	36	47	56	66	75	89	110	129	194	283	82.5
Mean reliability (Adams)	1.9	31.3	19.8	20.7	31.5	39.1	39.4	46.6	44.0	52.6	55.2	54.5	38.0
Mean reliability (empirical Bayes)	35.0	59.9	54.0	54.1	58.7	63.6	62.4	66.1	64.4	68.9	69.8	68.1	62.2
Mean Performance Score	83.3	92.0 (6.2)	91.8 (3.7)	89.7 (4.4)	93.1 (3.3)	92.6 (5.9)	92.8 (4.0)	94.0 (3.6)	92.2 (3.9)	94.1 (2.1)	92.0 (2.6)	87.6	92.4 (4.1)
N Hospitals	1	10	10	10	10	10	10	10	10	10	10	1	100
N Encounters	6	236	356	467	564	657	748	893	1098	1290	1944	283	8,253

*Note: mean performance score here differs from the scores reported in the performance gap results as here deciles are grouped by denominator/target population size rather than by performance score. **Cell intentionally left blank.

Table 15. Accountable entity level reliability testing of one year of eCQM performance among 100 VA acute care hospitals, stratified by decile of reliability level and estimation approach, 2021 (N=8,253 hospitalizations).

**	Mean Reliability (Adams Approach)	Mean Reliability (Empirical Bayes Approach)	Mean Performance Score*
Mean (SD)	38.0 (11.7)	62.2 (10.4)	92
Minimum	1.9	35.0	83
Decile 1	11.6	48.1	86
Decile 2	17.1	52.7	90
Decile 3	22.9	55.3	92
Decile 4	27.5	57.4	90
Decile 5	33.7	60.0	94
Decile 6	39.4	62.6	94
Decile 7	44.2	64.6	93
Decile 8	50.1	67.3	94
Decile 9	57.0	70.8	94
Decile 10	76.7	83.2	98
Maximum	100	100	100
Interquartile range	26.5	12.0	-

*Note: the mean performance score here differs from the scores reported in the performance gap results as here deciles are grouped by reliability (Adams method) rather than by performance score.

**Cells intentionally left blank

Table 16a. Facility and Patient Characteristics of Facilities with Low Reliability (Decile 1) vs. High Reliability (Decile 10) on Proposed Pneumonia Chest Imaging eCQM. Reliability estimates used 1/1/2021–12/31/2021 data among 100 VA facilities. VA facilities were grouped into deciles by reliability level using RAND (Adams) approach.

Variable	Decile 1 Reliability N = 334 patients[1]	Decile 10 Reliability N = 986 patients[1]	p-value[2]
Hospital ED size/complexity[3]			<0.001
High Complexity	54%	93%	
Medium/Low Complexity	46%	7.3%	
ED daily patient load (number of patients seen in ED)	27 (18, 40)	57 (42, 70)	<0.001
Hospital daily census (number of patient in hospital)	25 (17, 42)	80 (46, 98)	<0.001
Hospital daily patient census (% hospitaliz. per qtrly avg)	1.05 (0.93, 1.18)	1.03 (0.93, 1.12)	0.030
Age	74 (68, 80)	73 (68, 79)	0.8
Female sex	5.7%	4.1%	0.2
Race			0.5
Non-white	16%	18%	
White	84%	82%	
Rural residence	57%	41%	<0.001
Congestive Heart Failure	24%	31%	0.015
COPD	51%	51%	0.8
Charlson Comorbidity Index	2 (1, 5)	3 (1, 6)	0.057
ePSI*	99 (79, 120)	102 (83, 122)	0.2
Receipt of guideline-concordant antibiotics in first 24 hrs	68%	63%	0.061
Receipt of broad-spectrum antibiotics in first 24 hours	25%	33%	0.006
ED length of stay > 8 hours	6.0%	9.1%	0.073
CT chest Obtained	37%	37%	>0.9
Admission to the ICU	12%	10%	0.4
Ward-to-ICU transfer in first 72 hours	3.4%	6.2%	0.067
Length of Stay (days)	5 (3, 8)	6 (3, 11)	0.019
30-day Mortality	7.5%	11%	0.045
30-day Readmission	16%	15%	0.5

[1]Median (IQR); %
[2]Wilcoxon rank sum test; Pearson's Chi-squared test; Fisher's exact test
[3]Complexity Score is an operations project where facilities are differentiated by measuring the complexities of the services provided at each administrative parent facility within the VA. Complexity score is made up of resource allocation, intensive care unit level, operative complexity based on type of procedures that can be done at a facility, number of residents (GME and Adjusted AH), research dollars, complex clinical programs (SCI, Blind Rehab, Cardiac Surgery, Invasive Cardiac Catheterization, Neurosurgery, Transplant, Radiation Oncology, and polytrauma program), also care in the community coordination and the mental health program (domiciliary, acute psych, acute MH/PTSD). The model used to generate the Complexity Score is reviewed and updated every 3 years across the VA.

Table 16b. Comparison of Facility-level Characteristics between Facilities with Low Reliability (Decile 1) and High Reliability (Decile 10) on Proposed Pneumonia Chest Imaging eCQM. Reliability estimates based on 1/1/2021 through 12/31/2021 data among 100 VA facilities. VA facilities were grouped into deciles by reliability level using RAND (Adams et al) approach.

Facility Factor	Decile 1 facilities N=10 (%)	Decile 10 facilities N= 10 (%)
Census Region and Division		
Northeast	2 (20%)	2 (20%)
New England	2 (20%)	0 (0%)
Middle Atlantic	0 (0%)	2 (20%)
Midwest	3 (30%)	1 (10%)
East North Central	*1 (10%)*	*1 (10%)*
West North Central	*2 (20%)*	*0 (0%)*
South	2 (20%)	6 (60%)
South Atlantic	2 (20%)	2 (20%)
East South Central	0 (0%)	3 (30%)
West South Central	0 (0%)	1 (10%)
West	3 (30%)	1 (10%)
Mountain	2 (20%)	1 (10%)
Pacific	1 (10%)	0 (0%)
Bed size (maximum monthly average)		
≤50	0 (0%)	0 (0%)
51-100	1 (10%)	3 (30%)
101-200	9 (90%)	6 (60%)
>200	0 (0%)	1 (10%)
VHA VSSC Facility Complexity Composite Score for ED		
1a	2 (20%)	2 (20%)
1b	0 (0%)	3 (30%)
1c	0 (0%)	3 (30%)
2	8 (80%)	2 (20%)
3	0 (0%)	0 (0%)