# SUPPLEMENTARY MATERIALS

## 1. Supplementary Information on CRE Surrogate Model Selection using Multivariable Cox Model

The primary model outcome is the first subsequent inpatient hospitalization for any cause within 12 months since the index date. We used a Cox regression equation to evaluaterisk factors for hospital readmission among CRE case-patients within 365 days of discharge from the index hospitalization (defined as a patient’s first hospitalization with a CRE-positive culture) for inpatients or since the collection date of the first CRE-positive culture for outpatients. We evaluated an initial set of 41 predictors likely to influence the risk of CRE acquisitions and readmissions from *a priori* knowledge in literature. These included patient demographics, primary insurance, underlying conditions, sepsis during the index hospitalization, and previous healthcare exposures 15,40,41. We extracted 17 underlying conditions that composed the Charlson-Deyo Comorbidity Index (DCCI) from the International Classification of Diseases (ICD)-9 and ICD-10 diagnostic codes in HDDS discharge diagnoses before the first specimen collection date 42,43Each underlying condition was evaluated as a potential predictor in the model, in addition to contributing to the DCCI score.

Previous SNF stays were identified from three sources: identification of SNF as the ordering facility in the culture laboratory reports submitted to the TDH surveillance data, patient address matching a registered SNF address, or discharge to SNF in HDDS within 12 months preceding their first specimen collection. Healthcare exposures within 12 months before the first specimen collection date, including emergency department visits, inpatient hospitalizations, the total length of stay (LOS) at Short-term ACHs, and total LOS at LTACHs, were extracted from HDDS. Specific healthcare procedures were extracted from ICD-10-Procedure Coding System (PCS) and ICD-9 or ICD-10 discharge diagnoses (Supplementary Material, Appendix 1).

We used multivariable Cox regression models to obtain a final set of predictors most strongly associated with the subsequent hospitalization among CRE-infected patients. Microbiology information, such as culture and antibiotic susceptibility test results, were not included as potential predictors because they are not recorded in HDDS. Candidate predictors were selected into the multivariable model through a five-step selection process. First, we reviewed the frequency of all categorical predictors in all general population hospitalizations from 2016–2019 HDDS. Underlying conditions reported in less than 4% of the general population in 2016–2019 HDDS were excluded from the model. All healthcare exposures and continuous predictors were included in the next step. Next, we evaluated highly-correlated predictor pairs (Pearson coefficient >=0.70) for possible multicollinearity. For a highly correlated pair of predictors, we considered excluding one with narrow distribution or a large proportion of missing data and providing less information than the other predictor.

We evaluated whether the risk factor required stratification by index hospitalization status by plotting the log of the negative log of the Kaplan-Meier survival estimates against the log of time to observe the departure from proportional hazards (PH) assumption violation. We evaluated the model for the fit with the proportional hazard (PH) assumption using its Schoenfeld residuals. A p-value of <0.05 of Schoenfeld residual goodness-of-fit tests for the overall model and each predictor was deemed evidence of PH assumption violation. Additionally, we hypothesized that the baseline hazard differs between inpatients and outpatients.

We used a relaxed criterion of p-value < 0.20 in univariate regressions to consider including the covariate in the next iteration of covariate selection into the multivariable model. The final, parsimonious set of predictors was selected from the candidate set based on *a priori* knowledge of their association with CRE risks. Bootstrap cross-validation with replacement and repeated backward selection process with 1,500 iterations were used to identify the model with the lowest Akaike information criterion (AIC). While Cox regression can estimate the survival rate at any point during the follow-up period, we evaluated the predictive ability of the model to assess re-admission risks at 365 days (end of follow-up period). At 365 days, the model had a 0.70 Area Under the Curve (AUC) with an optimal risk score cut-off of -0.22. The sensitivity of this risk score to discriminate hospitalizations within 365 days was 77% (95% CI:75%–80%), and the specificity was 56% (95% CI:54%–59%).

## 2. Supplementary Tables and Figures

| **Table S1. Characteristics of Unique Patients with Positive CRE Specimens, Tennessee July 2015 – September 2019** | | | | |
| --- | --- | --- | --- | --- |
| Patient Characteristics | Inpatients | Outpatients | p-value\* | Overall |
| n | 1506 | 1297 |  | 2803 |
| Female sex (%) | 849 (56.4) | 881 (67.9) | <0.001 | 1730 (61.7) |
| Age (median [IQR]) | 66 [55, 76] | 68 [52, 78] | 0.238 | 67 [54, 77] |
| Race and Ethnicity (%) |  |  | <0.001 |  |
| Non-Hispanic White | 1027 (68.2) | 813 (62.7) |  | 1840 (65.6) |
| Non-Hispanic Black | 403 (26.8) | 157 (12.1) |  | 560 (20.0) |
| Hispanic/Other | 73 (4.8) | 34 (2.6) |  | 107 (3.8) |
| Missing | 3 (0.2) | 293 (22.6) |  | 296 (10.6) |
| Primary insurance (%) |  |  | <0.001 |  |
| Medicare/ Tenncare | 1187 (78.8) | 697 (53.7) |  | 1884 (67.2) |
| Commercial | 249 (16.5) | 224 (17.3) |  | 473 (16.9) |
| Uninsured | 70 (4.6) | 54 (4.2) |  | 124 (4.4) |
| Missing | 0 (0.0) | 322 (24.8) |  | 322 (11.5) |
| Had specimen(s) positive for Carbapenemase-Producing CRE (%) | 528 (35.1) | 186 (14.3) | <0.001 | 714 (25.5) |
| Sepsis during Index Hospitalization (%) | 627 (41.6) | 0 (0.0) | <0.001 | 627 (22.4) |
| Died Within 365 Days (%) | 606 (40.2) | 136 (10.5) | <0.001 | 742 (26.5) |
| Hospitalized within 365 Days (%) | 825 (54.8) | 344 (26.5) | <0.001 | 1169 (41.7) |
| **Healthcare Utilization within the Previous 365 days** | | | | |
| Inpatient Hospitalization(s) (%) | 1083 (71.9) | 467 (36.0) | <0.001 | 1550 (55.3) |
| LOS at Short-termACH, days (median [IQR])$ | 25 [10, 52] | 14 [5, 30] | <0.001 | 21 [8, 45] |
| Previous LTACH Stays (%) | 104 (6.9) | 29 (2.2) | <0.001 | 133 (4.7) |
| LOS at LTACH, days (median [IQR])$ | 64 [47, 101] | 70 [45, 112] | 0.605 | 65 [46, 105] |
| Emergency Department Visits (%) | 903 (60.0) | 470 (36.2) | <0.001 | 1373 (49.0) |
| SNF Stays (%) | 630 (41.8) | 228 (17.6) | <0.001 | 858 (30.6) |
| **Characteristics Among Patients >=1 Admission During the Study Period** | | | | |
| n | 1506 | 1012 |  | 2518 |
| **Healthcare Procedures within the Previous 365 days** | | | | |
| Dialysis (%) | 145 (9.6) | 25 (2.5) | <0.001 | 170 (6.8) |
| Urinary Catheters (%) | 169 (11.2) | 72 (7.1) | 0.001 | 241 (9.6) |
| Central Venous Access (%) | 287 (19.1) | 73 (7.2) | <0.001 | 360 (14.3) |
| Any Mechanical Ventilation (%) | 209 (13.9) | 52 (5.1) | <0.001 | 261 (10.4) |
| GI Endoscopy (%) | 175 (11.6) | 64 (6.3) | <0.001 | 239 (9.5) |
| **Underlying Conditions** | | | | |
| Congestive Heart Failure (%) | 621 (41.2) | 234 (23.1) | <0.001 | 855 (34.0) |
| Dementia (%) | 221 (14.7) | 103 (10.2) | 0.001 | 324 (12.9) |
| Chronic Pulmonary Disease (%) | 657 (43.6) | 282 (27.9) | <0.001 | 939 (37.3) |
| Paralysis (%) | 242 (16.1) | 78 (7.7) | <0.001 | 320 (12.7) |
| Renal Disease (%) | 658 (43.7) | 224 (22.1) | <0.001 | 882 (35.0) |
| HIV/AIDS (%) | 10 (0.7) | 2 (0.2) | 0.170 | 12 (0.5) |
| Diabetes (%) | 576 (38.2) | 273 (27.0) | <0.001 | 849 (33.7) |
| Any Malignancy (%) | 261 (17.3) | 101 (10.0) | <0.001 | 362 (14.4) |
| Charlson Comorbidity Index (median [IQR]) | 4 [2, 7] | 2 [0, 4] | <0.001 | 3 [1, 6] |

**Abbreviations:** IQR, interquartile range; CP-CRE, carbapenemase-producing CRE; LOS, length of stay; Short-term ACHs, short-term acute care hospitals; LTACHs, long-term acute ate hospitals; SNF, skilled nursing facilities.

\*p-values were calculated using Kruskal-Wallis test for continuous variables and Chi-square for categorical variable.

$ The displayed median total length of stay at LTACHs and SHORT-TERM ACHs was calculated among patients with previous LTACH and SHORT-TERM ACH stays only, respectively.

**Table S2. Multivariable Cox Regression Model Estimating the Readmissions of CRE cases within 365 days (Risk Factor Model) July 2015 – September 2019, Tennesssee**

| **Patient Characteristics** | **Univariable Model** | | | **Multivariable Imputed Model** | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Unadjusted HR** | **(95% CI)** | **p-value** | **Adjusted HR** | **(95% CI)** | **p-value** |
| Deyo-Charlson Comorbidity Index | 1.118 | (1.1, 1.135) | <.0001 | 1.062 | (1.042, 1.083) | <.0001 |
| Chronic Lung Diseases | 1.637 | (1.457, 1.841) | <.0001 | 1.197 | (1.056, 1.356) | <.01 |
| Sepsis at Index Hospitalization | 1.260 | (1.095, 1.45) | <.01 | 1.219 | (1.056, 1.406) | <.01 |
| Primary Insurance, Ref: Medicare/TennCare | | | | | | |
| Commercial Insurance | 0.616 | (0.522, 0.728) | <.0001 | 0.773 | (0.652, 0.918) | <.01 |
| Uninsured | 0.757 | (0.572, 1.001) | 0.0504 | 1.065 | (0.795, 1.426) | 0.6733 |
| **Prior Healthcare Exposures within 365 days** | | | | | | |
| Inpatient Hospitalization | 3.266 | (2.837, 3.761) | <.0001 | 1.919 | (1.646, 2.238) | <.0001 |
| Total Length of SHORT-TERM ACH Stays | 1.006 | (1.005, 1.006) | <.0001 | 1.003 | (1.002, 1.004) | <.0001 |
| Total Length of LTACH Stays | 1.002 | (1, 1.005) | 0.0653 | 0.998 | (0.995, 1) | 0.0768 |
| Urinary Catheter | 1.822 | (1.543, 2.151) | <.0001 | 1.260 | (1.062, 1.495) | <.01 |

Abbreviations: HR, hazard ratio; 95% CI (95% confidence interval); Ref, reference group; LOS, length of stay; SHORT-TERM ACH, short-term acute care hospitals; LTACH, long-term acute care hospitals.

**Supplementary Material 3. Directed Acyclic Graph of the Causal Association between Centrality Measures and CRE Prevalence.**

