Predicting Carbapenem-Resistant Enterobacteriaceae Carriage at the Time of Admission Using a Statewide Hospital Discharge Database

Michael Y. Lin,1 Michael J. Ray,1 Serena Rezny,2 Erica Runningdeer,2 Robert A. Weinstein,1,3 and William E. Trick1,2; for the CDC Prevention Epicenters Program

1Department of Medicine, Rush University Medical Center, Chicago, Illinois, USA; 2Department of Medicine, Cook County Health, Chicago, Illinois, USA; and 3Illinois Department of Public Health, Chicago, Illinois, USA

Background. Timely identification of patients likely to harbor carbapenem-resistant Enterobacteriaceae (CRE) can help health care facilities provide effective infection control and treatment. We evaluated whether a model utilizing prior health care information from a state hospital discharge database could predict a patient's probability of CRE colonization at the time of hospital admission.

Methods. We performed a case–control study using the Illinois hospital discharge database. From a 2014–2015 patient cohort, we defined cases as index adult patient hospital encounters with a positive CRE culture collected within the first 3 days of hospitalization, as reported to the Illinois XDRO registry; controls were all patient admissions from the same hospital and month. We split the data into training (~60%) and validation (~40%) sets and developed a logistic regression model to estimate coefficients for predictors of interest.

Results. We identified 486 index cases and 340,005 controls. Independent risk factors for CRE at the time of admission were age, number of short-term acute care hospital (STACH) hospitalizations in the prior 365 days, mean STACH length of stay, number of long-term acute care hospital (LTACH) hospitalizations in the prior 365 days, mean LTACH length of stay, current admission to LTACH, and prior hospital admission with an infection diagnosis code. When applying the model to the validation data set, the area under the receiver operating characteristic curve was 0.84.

Conclusions. A prediction model utilizing prior health care exposure information could discriminate patients who were likely to harbor CRE at the time of hospital admission.

Keywords. antibiotic resistance; carbapenem resistance; prediction.

Carbapenem-resistant Enterobacteriaceae (CRE) are extensively drug-resistant organisms (XDROs) of public health concern because of their ability to cause life-threatening infections and to spread in health care settings [1]. CRE are often brought into health care facilities by colonized patients, and facilities are encouraged to place CRE-colonized patients on isolation precautions to prevent patient-to-patient transmission. Because CRE colonization is usually asymptomatic, hospitals face the challenge of determining the CRE colonization status of patients at the time of admission [2]. As universal screening to detect colonization is costly, some facilities choose to test only those patients deemed “high risk,” often with limited knowledge of prior risk factors [3–6].

CRE carriage at the time of admission is strongly associated with prior intensity of health care facility exposure (particularly exposure to high-risk facilities such as long-term acute care hospitals) and prior antibiotic exposure [7–9]. However, data related to prior exposures involving external facilities typically reside outside a given hospital’s own information system and thus are inaccessible to hospital providers at the time of admission. State-based hospital discharge databases, which are increasingly available, contain historical patient-level health care exposures and diagnosis codes [10]. Such discharge data sets, which include infection-related diagnosis codes that are associated with prior antibiotic therapy [11], have the potential to inform models predicting risk of CRE carriage.

We hypothesized that the Illinois hospital discharge database could be used to develop a model to predict the probability of CRE colonization at the time of hospital admission. We leveraged the unique availability of the Illinois XDRO registry [12], which contains all reported CRE cases in the state, to train and validate our CRE prediction model.
METHODS

Model Development and Validation With 2014–2015 Patient Cohort
We performed a case–control study using data from 2 state-based databases: the Illinois hospital discharge database and the Illinois XDRO registry. The Illinois hospital database contains comprehensive encounter-level information for all Illinois hospitalizations in short-term acute care hospitals (STACHs) and long-term acute care hospitals (LTACHs) but excludes nonhospital settings such as skilled nursing facilities [10]. Hospital encounter data include patient identifiers, facility identifiers, and encounter characteristics (eg, dates of admission and discharge, diagnosis codes, and procedure codes). We classified hospitals as STACH or LTACH using the Illinois hospital database [13]. The Illinois XDRO registry (xdro.org) is a database of patients reported to the Illinois Department of Public Health as being colonized or infected with CRE; by rule, the first CRE-positive culture from each patient encounter must be reported to the registry. Reported information includes patient identifiers, date of admission, and date of positive CRE culture [12].

We defined index cases as adult (≥18 years of age) patient hospital encounters from January 1, 2014, to March 31, 2015, with a positive CRE culture collected within the first 3 days of hospitalization, as reported to the XDRO registry; only the first qualifying encounter per patient was analyzed. We defined controls as all patient admissions from the same hospital during the same month–year that an index case was reported; matching of controls by location and by time allowed us to account for confounding due to variation in the geographic distribution of CRE (which is unevenly distributed in Illinois) and potential temporal changes in CRE prevalence (ie, seasonal or secular trends).

For each hospital encounter, we determined the hospital type (STACH or LTACH) and the patient’s age at the time of admission. We determined the following health care exposures in the 365 days before each encounter: number of STACH hospitalizations, number of LTACH hospitalizations, average length of STACH stays in days, and average length of LTACH stays in days. As CRE-contaminated instruments used during endoscopic retrograde cholangiopancreatography (ERCP) have been previously identified as the source of a CRE outbreak in the Chicago region [14], we also determined whether a procedure code (ICD-9: 51.1X) indicating ERCP was present in the 365 days before each encounter. To infer prior antibiotic exposure, we assessed hospital encounters in the prior 365 days for which an infection diagnosis was recorded; prior hospital-associated infection diagnosis codes have been previously shown to be a surrogate for antibiotic exposure, particularly broad-spectrum antibiotics [11].

To develop and validate a prediction model for CRE carriage at the time of admission, we randomly split the data into training (~60%) and validation (~40%) sets. We developed a logistic regression model to estimate coefficients for predictors of interest that were chosen a priori because of their availability in the hospital discharge data set. The full model included the following predictors: age, sex, number of prior STACH and LTACH visits, mean prior STACH and LTACH lengths of stay, current hospital type (STACH vs LTACH), prior admission with infection diagnosis, and prior receipt of ERCP.

To guard against possible overfitting and improve the predictive capabilities of the model, we used a bootstrap aggregation (bagging) procedure [15]. To build the bagged model, we generated 100 random subsamples of cases and their corresponding controls from the training data set and used these subsamples to fit 100 logistic regressions with predictors of interest. We used each logistic regression to produce a set of predicted probabilities for the validation data set and then averaged these predicted probabilities. We used receiver operating characteristic curves based on the averaged predicted probabilities to evaluate the performance of the bagged model.

Although bagging approaches can improve prediction, it may be cumbersome, in practice, for hospitals to utilize a bagging approach and the multiple models entailed to predict CRE carriage at the time of admission. For this reason, we also fitted a single logistic regression model, without bagging, based on all cases and corresponding controls in the training data. The coefficients from this model for predictors of interest can be easily displayed and examined.

Application and Validation of Model in 2016 Cohort
To test the external validity (temporal stability) of our model, we applied the prediction model to a new cohort of adult (≥18 years of age) patients: those with hospitalizations in the Illinois hospital discharge database between January 1, 2016, and December 31, 2016. Diagnosis and procedure codes based on ICD-9 coding used in the original model were translated to ICD-10 codes using General Equivalence Mappings from the Centers for Medicare and Medicaid Services [16]. Using the parameter estimates from the model built, without bagging, on the training data from the first cohort, we generated predicted CRE probabilities for each patient encounter, then calculated each patient’s maximum CRE probability within the 2016 year. We stratified patients into the following CRE risk groups: 0%–5%, >5%–10%, >10%–20%, >20%–30%, >30%–40%, >40%–50%, and >50%. We then assessed how many patients within each risk stratum were reported as CRE-colonized to the XDRO registry during 2016. Because of the large number of patients in the 2 lowest strata for CRE risk, we randomly sampled 50 patients per stratum to estimate the CRE prevalence in those 2 groups. We used binomial methods to construct 95% confidence intervals.

All statistical tests were 2-sided, and P values of <.05 were considered statistically significant. Data were analyzed using SAS 9.4 (Cary, NC, USA) and R statistical software, version 3.1.2 (Vienna, Austria; http://www.R-project.org/).
RESULTS

Model Derivation and Validation Using the 2014–2015 Patient Cohort
In total, we identified 486 index cases (CRE-positive) and 340 005 control (CRE-negative) patient hospital encounters. Three hundred cases were selected for the training data set, and 186 were selected for the validation data set. Only those 143 278 controls in the training data who shared a hospital and month-year with one of the index cases in the training data were used to fit the models.

We compared patient characteristics of cases and their corresponding controls in the training set (Table 1). Cases tended to represent older patients who had more STACH and LTACH hospitalizations in the prior 365 days and a higher mean STACH and LTACH length of stay.

The logistic regression model fit without bagging found the following variables to be independent risk factors for CRE at the time of admission (Table 1): age, number of STACH hospitalizations in the prior 365 days, mean STACH length of stay, number of LTACH hospitalizations in the prior 365 days, mean LTACH length of stay, current admission to LTACH, and prior hospital admission with an infection diagnosis. Applying the bagged logistic regression model to the validation data, we calculated the area under the receiver operating characteristic curve (AUC) to be 0.84 (Supplementary Appendix, Supplementary Table 1, and Supplementary Figure 1). When we restricted the training and validation data to STACH admissions only and fitted another bagged model, we calculated the AUC to be 0.81.

Model Application to 2016 Cohort
The 2016 cohort from the Illinois Hospital Discharge Database included 1 229 158 hospital visits by 816 500 unique adult patients; 38% of the patient hospital encounters were preceded by a prior hospitalization within 365 days. The full model predicted 1141 visits (745 patients) associated with a CRE risk >5%–10%, 572 visits (427 patients) with a >10%–20% risk, 144 visits (118 patients) with a >20%–30% risk, 61 visits (65 patients) with a >30%–40% risk, 53 visits (45 patients) with a >40%–50% risk, and 96 visits (63 patients) with a >50% risk.

Within each stratum, we estimated the actual CRE prevalence using XDRO registry reporting (Figure 1). In general, actual CRE prevalence increased in parallel with the predicted CRE risk. There appeared to be a ceiling effect of actual CRE prevalence; the predicted CRE prevalence was significantly greater than the actual CRE prevalence for the >50% stratum.

DISCUSSION

We found that a model based on the information contained in a state hospital discharge database could discriminate patients who were likely to harbor CRE at the time of hospital admission. In particular, variables from the discharge database that captured prior frequency and duration of health care exposure and prior infection treatment (as a surrogate for antibiotic receipt) were found to be independent predictors of CRE carriage. Applying the prediction model could effectively risk-stratify patients and, if incorporated into an automated alerting system, could efficiently guide infection prevention efforts, such as active surveillance and preemptive isolation precautions.

The current strategies to identify patients at high risk of CRE or other multidrug-resistant organism carriage are not efficient. For example, hospitals may choose to screen patients directly transferred from other hospitals including LTACHs [3, 7], but such a strategy would miss patients with indirect health care exposures (ie, exposed to an LTACH but discharged to home before current admission). Another strategy that has been used to assess the risk of multidrug-resistant organism carriage is patient self-report of prior health care exposure [17–20], but this strategy requires relatively simplistic questions, accurate recall by patients, and hospital staff time to ask questions. Although electronic prediction models have been developed to identify patients at risk for certain multidrug-resistant organisms, data for such models are typically limited to the medical records of the admitting hospital system [21, 22]. Our current prediction

Table 1. Adjusted Predictors of Carbapenem-Resistant Enterobacteriaceae Carriage on Admission, 2014–2015 Cohort

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Case (n = 300)</th>
<th>Control (n = 143 278)</th>
<th>aOR</th>
<th>95% CI</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td>65</td>
<td>57</td>
<td>1.02</td>
<td>1.01–1.03</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Male sex, %</td>
<td>50</td>
<td>42</td>
<td>1.07</td>
<td>0.85–1.35</td>
<td>.58</td>
</tr>
<tr>
<td>STACH hospitalizations in prior 365 d, No.</td>
<td>3.7</td>
<td>1.4</td>
<td>1.03</td>
<td>1.01–1.06</td>
<td>.02</td>
</tr>
<tr>
<td>Mean STACH length of stay, d</td>
<td>8.9</td>
<td>2.5</td>
<td>1.04</td>
<td>1.03–1.06</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>LTACH hospitalizations in prior 365 d, No.</td>
<td>0.5</td>
<td>0.02</td>
<td>2.32</td>
<td>1.94–2.78</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Mean LTACH length of stay, d</td>
<td>11.1</td>
<td>0.3</td>
<td>1.02</td>
<td>1.02–1.03</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Current facility is LTACH, %</td>
<td>20.3</td>
<td>1.0</td>
<td>5.80</td>
<td>4.15–8.12</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Prior infection diagnosis, %</td>
<td>74</td>
<td>27</td>
<td>3.03</td>
<td>2.23–4.12</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Prior ERCP, %</td>
<td>1.7</td>
<td>0.5</td>
<td>1.72</td>
<td>0.69–4.27</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

Unless otherwise specified, case and control group means are presented. All predictors were obtained from data in the 365 days before a given patient’s admission, except for “Current facility is LTACH.” Odds ratios and P values are from a logistic regression model that included all covariates listed and an intercept.

Abbreviations: aOR, adjusted odds ratio; CI, confidence interval; ERCP, endoscopic retrograde cholangiopancreatography; LTACH, long-term acute care hospital; STACH = short-term acute care hospital.
model makes use of medical information in the year before admission from a statewide hospital discharge database and could be automated, such that infection control personnel could receive an electronic alert at the time of a patient’s admission.

An automated infrastructure for alerting health care facilities regarding the admission of patients who have laboratory-confirmed CRE history previously reported to the XDRO registry already exists in Illinois within the XDRO registry framework [12]. Such information-sharing is permitted under the Health Insurance Portability and Accountability Act’s Privacy Rule exemptions for public health to control the spread of a communicable disease [23]. Although there may be uncertain precedent for public health authorities to share patient health care information that is driven by risk-based rather than confirmed disease status, hospitals already make risk-based infection control decisions based on prior health care exposures, such as implementing screening or preemptive isolation precautions based on prior post–acute care facility exposure or exposure to high-risk geographic regions [3, 24–26].

Our model’s current prediction capability could be used to guide not only infection control measures, but also treatment. For example, a patient admitted with a high CRE risk score and diagnosed with suspected bacterial sepsis could receive empiric antibiotic treatment that would be active against CRE. Furthermore, although the current prediction model is calibrated for CRE, it may also provide risk stratification for other multidrug-resistant organisms that share similar risk factors. Prior exposure to health care and antibiotics are common risk factors for carriage of methicillin-resistant Staphylococcus aureus, vancomycin-resistant Enterococcus, and extended-spectrum beta-lactamase–producing organisms, and co-colonization is common [8, 22, 27–31]. The ability of the current prediction model to predict carriage of non-CRE multidrug-resistant pathogens warrants further investigation.

There are limitations to this study. First, this model was calibrated on patients who were reported to the Illinois XDRO registry as CRE-positive. Because asymptomatic and unrecognized CRE carriage occurs and because the XDRO registry is a passive surveillance system, under-reporting of CRE-colonized patients is possible. The presence of patients who were CRE-colonized but unreported would lead to misclassification bias toward the null and could explain, in part, the ceiling effect seen in the 2016 cohort validation, where predicted CRE risk diverged from actual CRE reports. Second, hospital discharge databases are comprehensive but rely on hospital report and lag by several months in Illinois. Our retrospective analyses used nonlagged health care exposure information at the time of discharge, but prospective implementation of such a model would be compromised by missing data or would require improvement in the timeliness of the data source. Third, the hospital discharge database does not contain skilled nursing facility exposures. Such exposures have been associated with carriage of multidrug-resistant organisms, including CRE [32, 33]. Lastly, our model is calibrated to the epidemiology of CRE in Illinois and needs validation in other geographic regions. As we began with an already limited set of available covariates with biological

---

**Figure 1.** Predicted vs actual reported CRE colonization among Illinois patients in the 2016 validation cohort. Whiskers represent 95% confidence intervals. “Actual CRE-Positive” status refers to whether the patient was reported as CRE-positive to the XDRO registry. Abbreviations: CI, confidence interval; CRE, carbapenem-resistant Enterobacteriaceae.
plausibility for predicting CRE carriage, we favored retaining the full model for our validation; however, in other geographic regions, some predictors such as prior ERCP may not be needed to maintain model performance.

Despite these limitations, our study has several strengths. The Illinois XDRO registry represents one of the largest cohorts of CRE carriers available for modeling. The modeling parameters estimated in this study have biologic plausibility and are consistent with risks of CRE colonization that have been described in other geographic regions with CRE transmission [26, 34–39].

In summary, we demonstrated that a CRE prediction model informed by routinely available health care exposure data in a state hospital discharge database can effectively predict the CRE status of patients at the time of admission. Such a model could provide timely, actionable information to front-line providers to improve the care of patients at risk of CRE infection.

Supplementary Data
Supplementary materials are available at Open Forum Infectious Diseases online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Acknowledgments
We thank Dejan Jovanov for his assistance with data extraction and data preparation for analysis.

Financial support. This work was supported by the Centers for Disease Control and Prevention (grant numbers 5U54CK000161 and 5U54CK000161-05S1 to R.A.W.; grant number 1U54CK000481 to W.E.T.). Potential conflicts of interest. M.Y.L. has received research support in the form of a contributed product from OpGen and Sage Products (now part of Stryker Corporation) and has received an investigator-initiated grant from CareFusion Foundation (now part of BD). R.A.W. has received research support in the form of a contributed product from Clorox, Medline, Molnlycke, and Sage Products. W.E.T. has received an investigator-initiated grant from CareFusion Foundation. All other authors report no potential conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References


