Electronically Available Patient Claims Data Improve Models for Comparing Antibiotic Use Across Hospitals: Results From 576 US Facilities

Katherine E. Goodman,[1] Lisa Pineles,[1] Laurence S. Magder,[1] Deverick J. Anderson,[2] Elizabeth Dodds Ashley,[2] Ronald E. Polk,[3] Hude Quan,[4] William E. Trick,[5] Keith F. Woeltje,[6] Surbhi Leekha,[1] Sara E. Cosgrove,[7] and Anthony D. Harris[1]

[1]The University of Maryland School of Medicine, Department of Epidemiology and Public Health, Baltimore, Maryland, USA, [2]Duke Center for Antimicrobial Stewardship and Infection Prevention, Duke University School of Medicine, Durham, North Carolina, USA, [3]School of Pharmacy, School of Medicine, Virginia Commonwealth University, Richmond, Virginia, USA, [4]Department of Community Health Sciences, University of Calgary, Calgary, Alberta, Canada, [5]Cook County Health and Rush University Medical Center, Chicago, Illinois, USA, [6]Department of Internal Medicine, Division of Infectious Diseases, Washington University School of Medicine, St. Louis, Missouri, USA, and [7]Division of Infectious Diseases, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA

Background. The Centers for Disease Control and Prevention (CDC) uses standardized antimicrobial administration ratios (SAARs)—that is, observed-to-predicted ratios—to compare antibiotic use across facilities. CDC models adjust for facility characteristics when predicting antibiotic use but do not include patient diagnoses and comorbidities that may also affect utilization. This study aimed to identify comorbidities causally related to appropriate antibiotic use and to compare models that include these comorbidities and other patient-level claims variables to a facility model for risk-adjusting inpatient antibiotic utilization.

Methods. The study included adults discharged from Premier Database hospitals in 2016–2017. For each admission, we extracted facility, claims, and antibiotic data. We evaluated 7 models to predict an admission's antibiotic days of therapy (DOTs): a CDC facility model, models that added patient clinical constructs in varying layers of complexity, and an external validation of a published patient-variable model. We calculated hospital-specific SAARs to quantify effects on hospital rankings. Separately, we used Delphi Consensus methodology to identify Elixhauser comorbidities associated with appropriate antibiotic use.

Results. The study included 11 701 326 admissions across 576 hospitals. Compared to a CDC-facility model, a model that added Delphi-selected comorbidities and a bacterial infection indicator was more accurate for all antibiotic outcomes. For total antibiotic use, it was 24% more accurate (respective mean absolute errors: 3.11 vs 2.35 DOTs), resulting in 31–33% more hospitals moving into bottom or top usage quartiles postadjustment.

Conclusions. Adding electronically available patient claims data to facility models consistently improved antibiotic utilization predictions and yielded substantial movement in hospitals' utilization rankings.

Keywords. antibiotic stewardship; antimicrobial use; benchmarking; risk adjustment.

Reducing inappropriate antibiotic use in inpatient settings is a nationally recognized priority for combating antibiotic resistance [1]. A system for measuring and comparing antibiotic use across hospitals is fundamental to achieving this goal: comparator data help facilities contextualize their use against other institutions and, by identifying outlier hospitals, may uncover targets for intervention to improve antibiotic utilization. However, because hospitals treat different types of patients with different diagnoses and underlying conditions, usage rates

across hospitals may appropriately vary [2]. Although it is impossible to control for every variable that affects rates of antibiotic use, controlling ("risk-adjusting") for the most important factors facilitates fairer and more meaningful interhospital comparisons.

Facilities that submit antibiotic utilization data to the Centers for Disease Control and Prevention (CDC) National Healthcare Safety Network (NHSN) Antimicrobial Use and Resistance Module can receive standardized antimicrobial administration ratios (SAARs) for select units [3]. SAARs compare a hospital unit's observed-to-predicted antibiotic utilization (also historically called "observed-to-expected" ratios [4]). Currently, CDC statistical models adjust for unit type and certain hospital characteristics to calculate predicted antibiotic use [5, 6]. As CDC has itself recognized, however, additional patient factors may also affect antibiotic utilization [6, 7]. For example, some of the potentially relevant clinical constructs that NHSN models do not directly adjust for when predicting antibiotic use include patient diagnoses and comorbidities. To this end, a recent study

by Yu et al (2018) explored a large number of patient-level variables for comparing unit-level antibiotic use across 35 Kaiser Permanente California hospitals and found that many were significantly associated with antibiotic use in their risk-adjustment models [4].

Although SAARs are not currently used for public reporting or reimbursement purposes, they are National Quality Forum (NQF)-endorsed and are currently a quality and efficiency Measure under Consideration by the Centers for Medicare and Medicaid Services (CMS) [6, 8]. Yu et al's findings, coupled with other prior research [9], provide mounting evidence that any candidate SAAR model should account for patient-mix to optimize antibiotic utilization comparisons. However, translating this conclusion into models suitable for wider policy use also raises additional considerations. Models would need to be validated beyond a single hospital network or geographic region and preferably across an entire facility, including the many unit locations that do not currently qualify for NHSN SAARs. Ideally, too, models would only include readily implementable patient variables, that is, those that are easily electronically available, relatively standardized across facilities, and available for all discharges. And most importantly, to mitigate against the risk of unfairly penalizing hospitals, to the fullest extent possible models would predict appropriate antibiotic use [10].

Motivated by the preceding considerations, the objective of the current study was to compare models that incorporate claims data-derived patient variables to the current CDC NHSN facility-variable model for risk-adjusting inpatient antibiotic utilization. To achieve this goal, we (1) convened an expert panel to identify patient comorbidities, derived from electronically available claims data, that are perceived as causally related to appropriate antibiotic use; (2) evaluated, across a large and diverse cohort of US hospitals, whether models that incorporate these comorbidities and other patient diagnoses are more accurate than models that only include facility-level variables; and (3) quantified the impact of adjustment with patient-variable models on hospital antibiotic utilization rankings.

METHODS

Study Population and Collected Data
A description of the study cohort has been published previously [11]. Briefly, adult admissions and associated data were collected from hospitals in the Premier Healthcare Database ("Premier Database"), an all-payer repository of claims and clinical data from more than 870 million inpatient and outpatient US hospital admissions [12]. Although not explicitly nationally representative, Premier Database hospitals cover highly geographically diverse areas across the United States (see Supplementary Materials for further database detail). All admissions with discharge dates on or between 1 January 2016 and 31 December 2017 at hospitals that continuously submitted

data during the study were included. This study did not include personally identifiable information and was exempt from institutional review board review.

For each admission, we extracted (a) facility characteristics; (b) payer and sociodemographic data; (c) location by service-day; (d) daily antibiotic charge data; (e) the Medicare Severity-Diagnosis Related Group (MS-DRG) code, which guides CMS reimbursement; and (f) all ICD-10-CM diagnosis codes, including whether diagnoses were present on admission (POA).

Using publicly available Agency for Healthcare Research and Quality (AHRQ) software, we mapped ICD-10-CM diagnosis codes to 3 clinical constructs: (1) 29 Elixhauser comorbidities [13]; (2) approximately 287 Clinical Classifications Software (CCS) disease categories [14]; and (3) 2 "bacterial infection" indicators based upon (a) any or (b) POA-only bacterial infection-related ICD-10-CM codes [15] (see Supplementary Figure 1).

### Antibiotic Utilization Outcomes
We selected study outcomes to match current CDC NHSN antimicrobial use surveillance practice. We used inpatient days of therapy (DOT) as the primary study metric [3, 9, 16, 17]. If a patient received 2 different antibiotics on the same service day, these events qualified as 2 DOTs [2, 11, 12]. For each admission, we summed a patient's DOTs for 4 antibiotic outcomes, mapped to existing CDC groupings: (1) all antibiotics, (2) broad-spectrum antibiotics predominantly used for hospital-onset infections, (3) broad-spectrum antibiotics predominantly used for community-acquired infections, and (4) antibiotics predominantly used for resistant gram-positive infections (antibiotic appendix available in reference [3]).

Expert Panel Evaluation of Elixhauser Comorbidities Associated With Appropriate Antibiotic Use
We used Delphi methodology [18, 19], a consensus-building technique that has been applied to other infectious disease outcomes [20], to determine which Elixhauser comorbid conditions are causally associated with appropriate antibiotic use as judged by an expert panel. In this context, "associated with appropriate antibiotic use" could mean that the comorbidity includes condition(s) for which antibiotic initiation is generally justified (eg, patients with metastatic cancer are likely to be receiving chemotherapy, and earlier and broader-spectrum antibiotic therapy for suspected infection may be appropriate), or that presence of the comorbidity is on average associated with appropriately greater days of antibiotic therapy compared to an equivalent patient without the comorbidity.

We administered an iterative, 2-round survey with conference call to 8 infectious disease and antimicrobial stewardship experts in the United States (see Supplementary Table 1). Experts were instructed to independently rate each Elixhauser comorbid condition on a Likert scale from 1 (not at all related) to 5 (strongly related), based on its perceived relatedness to

(A) Mean Absolute Error (MAE) for Predicting Days of Therapy (DOTs) at the Patient-Level, by Antibiotic Outcome "Average DOTs Mis-Predicted" n = 5,850,663 (Testing set)		All antibiotics MAE (DOTs)	Broad-spectrum antibiotics for hospital-onset infections (HOI) MAE (DOTs)	Broad-spectrum antibiotics for community-acquired infections (CAI) MAE (DOTs)	Antibiotics for resistant Gram-positive infections (GP) MAE (DOTs)	Percentage Reduction in Mean Absolute Error ("Percentage More Accurate") Compared to the CDC Parallel-Facility Model, Across the Four Antibiotic Outcomes
1	Null	3.153	1.174	1.278	0.857	N/A
2	CDC Parallel-Facility	3.107	1.146	1.259	0.846	N/A
3	Expert Panel Consensus-Driven	3.018	1.128	1.241	0.841	Range: 0.6 – 2.9% • All: 2.86% • HOI: 1.57% • CAI: 1.43% • GP: 0.59%
4	Expert Panel Consensus-Driven + Bacterial Infection	2.348	0.995	1.111	0.762	Range: 9.9 – 24.4% • All: 24.43% • HOI: 13.18% • CAI: 11.76% • GP: 9.93%
5	*Present on Admission* Variant: Expert Panel Consensus-Driven + Bacterial Infection	2.386	1.021	1.118	0.770	Range: 9.0 – 23.2% • All: 23.21% • HOI: 10.91% • CAI: 11.20% • GP: 8.98%
6	Consensus-Driven/Data-Driven Hybrid	2.406	1.380	1.097	0.8406	Range: -20.4 – 22.6% • All: 22.56% • HOI: -20.42% • CAI: 12.87% • GP: 0.64%
7	Yu et al. "Simplified ASP" (External Validation)	2.313	0.926	1.018	0.705	Range: 16.7 – 25.6% • All: 25.56% • HOI: 19.20% • CAI: 19.14% • GP: 16.67%
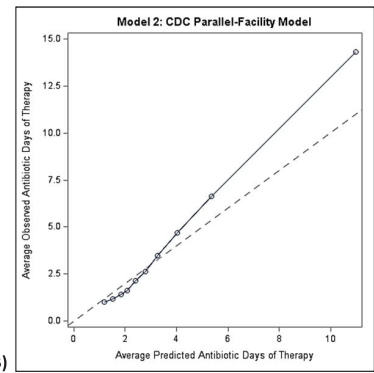
(B)
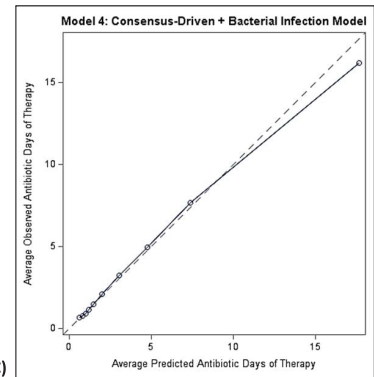
(C)

Figure 1. Results for the 7 evaluated models on the testing set (n = 5 850 663 admissions), with (*A*) model accuracy captured by MAEs. MAEs reflect the average days of antibiotic therapy mis-predicted per admission. Calibration plots in (*B*) and (*C*) reflect the concordance between observed and predicted DOTs by decile for the CDC Parallel-Facility model and the Expert Panel Consensus-Driven + Bacterial Infection model, respectively, for predicting total antibiotic use. Abbreviations: CDC, Centers for Disease Control and Prevention; DOT, day of therapy; MAE, mean absolute error.

appropriate antibiotic use. At the conclusion of the Delphi process, each comorbidity was assigned as "causally related," "indeterminately related," or "not causally related" to appropriate antibiotic use; criteria for determining causal relatedness have been described in detail elsewhere [21], and we deployed the same criteria here. Elixhauser comorbid conditions that qualified as causally related or indeterminately related to appropriate antibiotic use were included as predictors in our Expert Panel Consensus-Driven model.

Evaluated Models

For each outcome, we selected 7 models a priori to predict a patient's antibiotic DOTs during an admission (Table 1). Our guiding objective was to build sequentially on a base model to evaluate the incremental performance gains, if any, achieved by adding additional variables. This process progressed from a model with no predictors to a model with only facility characteristics (approximating the existing CDC NHSN model), and then to models that added the Expert Panel-selected Elixhauser comorbid conditions and other claims data-derived patient variables that represented different clinical constructs. All models and their included variables, apart from a data-driven model, were selected a priori based upon hypothesized

relationships with antibiotic use. In addition, we recreated and externally validated, as closely as our data permitted, a previously published model by Yu et al, the "Simplified ASP" model [4]. Following our primary analyses, we also evaluated a combination model that combined the Yu et al model with our best-performing model. Supplementary Table 2 provides variable operationalization details for all evaluated models.

Statistical Methods, SAAR Calculations, and Evaluating Impact on Hospital Rankings

Descriptive statistics for patient and hospital characteristics were calculated using mean (standard deviation [SD]), median (range or interquartile range [IQR]), or frequency count (percentage). For model evaluation, we randomly divided the dataset into 50/50 training and testing sets. All models used negative binomial regression to predict the DOTs for an admission, for each of the 4 antibiotic outcomes; the model offset equaled the natural log of the admission's service-days.

For each model, we (1) fit the model on the training set and stored model parameters; (2) applied the parameterized model to the held-out testing set to predict each admission's DOTs; (3) calculated the absolute error for each admission by comparing predicted to observed DOTs; and (4) for the entire testing set,

Table 1. Description of Evaluated Models

Number	Model Name	Predictors	Notes
1	Null	None (offset only)	
2	CDC Parallel-Facility	• Facility characteristics • % of patient's encounter days in ICUs • % of patient's encounter days in wards • % of patient's encounter days in stepdown units • % of patient's encounter days in hematology-oncology units	*Variables were selected to parallel existing CDC NHSN SAAR risk-adjustment models. Supplementary Table 2 reflects the full list of facility characteristics.*
3	Expert Panel Consensus-Driven	• [All model #2 variables] • Patient age • Elixhauser comorbidities ranked as causally or indeterminately related to appropriate antibiotic use by the Delphi-consensus process • Total Elixhauser score	
4	Expert Panel Consensus-Driven + Bacterial Infection	• [All model #3 variables] • Patient has an ICD-10-CM code associated with bacterial infection	
5	*POA Variant*: Expert Panel Consensus-Driven + Bacterial Infection	• [All model #2 variables] • Elixhauser comorbidities ranked as causally or indeterminately related to appropriate antibiotic use by the Delphi-consensus process *that were coded as present on admission* • Total Elixhauser score (POA conditions only) • Patient has an ICD-10-CM code associated with bacterial infection *that was coded as present on admission*	*Because hospital-onset infections are not POA, for the broad-spectrum antibiotics predominantly used for hospital-onset infections category, the model did not restrict bacterial infections to POA*
6	Consensus-Driven/Data-Driven Hybrid	• [All model #4 variables] • Clinical Classifications Software (CCS) disease category variables retained in cross-validated lasso regression	*Full details of the variable selection process are available in the Supplementary Materials*
7	Yu et al. "ASP Simplified" Model	• Available in: Yu et al (2018)	*Additional details are provided in Supplementary Table 2*

Abbreviations: CDC, Centers for Disease Control and Prevention; ICU, intensive care unit; NHSN, National Healthcare Safety Network; POA, present on admission; SAAR, standardized antimicrobial administration ratio.

calculated the mean absolute error (MAE) by averaging all admission errors. We used the MAE from the testing set to evaluate model accuracy [22] and compared models using raw MAEs and percentage reductions in MAEs between candidate and reference models $[(\frac{MAE_{Ref} - MAE_{Candidate}}{MAE_{Ref}}) \times 100]$. We evaluated model calibration by dividing predicted DOTs into deciles. For each decile, we calculated the mean observed and predicted DOTs and created calibration plots for visual inspection.

We calculated each hospital's SAARs by summing its admissions' observed and predicted DOTs, similar to current NHSN methodology but adapted to predict at the hospital, rather than unit, level (see Equation 1 in the Supplementary Materials). A SAAR >1 indicates the hospital reported higher antibiotic use than predicted, whereas a SAAR <1 indicates lower use than predicted.

We used model-specific SAARs to quantify effects of adjustment on hospital rankings of antibiotic use. Given the large number of evaluated hospitals, we used quartile rankings. We ranked hospitals by quartile based upon their unadjusted (observed) rates of antibiotic DOTs per 1000 patient-days. We then reranked hospitals by their SAARs, and unadjusted and SAAR rankings were compared. To assess the practical impact of adding patient-level variables to risk-adjustment models, we evaluated the number of hospitals that changed in the bottom and top quartiles of use when adjusting with models that included patient-level variables versus when adjusting with a model that only included facility-level variables. Analyses were performed using SAS version 9.4 (SAS Institute Inc.).

RESULTS

During the 24-month study period, there were 11 701 326 admissions (64 064 632 patient-days) across 576 US hospitals. Hospital and patient characteristics are presented in Table 2. Overall, 65% of patients received at least 1 antibiotic during their hospitalization. Across all admissions, the DOT distribution for each antibiotic outcome was the following [(mean), 25th, 50th, 75th, 95th percentiles]: all antibiotics: (3.91), 0, 2, 5, 15; antibiotics for hospital-onset infections: (0.89), 0, 0, 0, 6;

Table 2. Description of Patient and Facility Characteristics Among US Adult Inpatient Admissions in the Premier Healthcare Database, 2016–2017

Encounters	n = 11 701 326
No. of admissions by year	
2016	5 834 810
2017	5 866 516
Total patient-days	64 064 632
Patient characteristics	**n = 11 701 326 (%)**
Age, median (IQR)	62 (42–75)
Male	4 834 283 (41)
Race	
White	8 690 211 (75)
Black	1 651 263 (14)
Other	1 135 218 (10)
Unknown	224 634 (2)
Payer	
Medicare	5 829 127 (50)
Medicaid	1 976 689 (17)
Private	3 065 584 (26)
Other	829 926 (7)
Length of stay in days, median (IQR)	4 (3–6)
Died	258 668 (2)

Table 2. Continued

Encounters	n = 11 701 326
Top 5 MS-DRGs[a]	
Vaginal delivery w/o complicating diagnosis	828 478 (7)
Septicemia or severe sepsis w/o MV >96 hours w/ MCC	509 476 (4)
Major joint replacement or reattachment of lower extremity w/o MCC	483 684 (4)
Cesarean section w/o CC/MCC	284 173 (2)
Heart failure and shock w/ MCC	258 083 (2)
Top 5 Elixhauser comorbidities[a,b]	
Hypertension	4 830 425 (41)
Fluid and electrolyte disorders	3 500 288 (30)
Chronic pulmonary disease	2 783 243 (24)
Deficiency anemias	2 330 685 (20)
Congestive heart failure	2 055 661 (18)
Elixhauser comorbidity score, median (IQR)[c]	3 (1–5)
Top 5 CCS disease categories[a,c]	
Essential hypertension	4 334 551 (37)
Disorders of lipid metabolism	4 191 814 (26)
Fluid and electrolyte disorders	3 499 631 (30)
Other nutritional- endocrine- and metabolic disorders	3 277 471 (28)
Coronary atherosclerosis and other heart disease	2 796 670 (24)
Facility characteristics	**n = 576 (%)**
Urban[d]	432 (75)
Teaching	170 (30)
Bed size	
0–99	126 (22)
100–199	143 (25)
200–299	102 (18)
300–399	82 (14)
400–499	41 (7)
500+	82 (14)
US Census Region and Division	
Northeast	76 (13)
Mid-Atlantic	63 (11)
New England	13 (2)
South	260 (45)
East South Central	37 (6)
West South Central	62 (11)
South Atlantic	161 (28)
Midwest	147 (26)
West North Central	46 (8)
East North Central	101 (18)
West	93 (16)
Mountain	25 (4)
Pacific	68 (12)

Abbreviations: CC, complication or comorbidity; CCS, Clinical Classifications Software (maintained by the Agency for Healthcare Research and Quality [AHRQ]); IQR, interquartile range; MCC, major complication or comorbidity; MS-DRG, Medicare Severity-Diagnosis Related Group; MV, mechanical ventilation; w/ and w/o, with and without.

[a]Each encounter receives 1, and only 1, MS-DRG assignment. Patients can have multiple Elixhauser comorbidities and CCS diseases per encounter.

[b]Elixhauser comorbidity classifications modified to also include primary diagnoses. Patient Elixhauser scores represent unweighted Elixhauser comorbidity sums (1 point per comorbidity).

[c]Two most common CCS categories excluded from this listing because they do not represent disease categories: "Residual codes—unclassified" (54%) and "Other aftercare" (38%).

[d]Designation provided by Premier, based upon American Hospital Association Annual Survey response.

antibiotics for community-acquired infections: (0.96), 0, 0, 1, 5; and antibiotics for resistant gram-positive infections: (0.63), 0, 0, 0, 4. In other words, taking total antibiotic use as an example, 50% of all admissions had ≤2 days of therapy, and 95% of admissions had ≤15 DOTs.

Of the 29 Elixhauser comorbidities, 14 were ranked as causally related to appropriate antibiotic use at the conclusion of the Delphi-consensus process, and a further 6 were ranked as indeterminately related (Table 3). In total, these 20 variables, plus patient age and total Elixhauser comorbidity score, were included in our Expert Panel Consensus-Driven model.

Model Performance for Predicting Antibiotic Utilization (DOTs)

Models' mean absolute errors (average days of antibiotic therapy mispredicted per admission), and calibration plots for select models, are reflected in Figure 1. Overall, models were most accurate at predicting use of antibiotics for resistant gram-positive infections and least accurate at predicting total antibiotic use.

Building upon a null model with no variables ("Null" model), we added facility-level variables that were selected to match current CDC NHSN models as closely as possible (our "CDC Parallel-Facility" model). The CDC Parallel-Facility model improved predictions over the Null model by 1.3–2.4%. Adding Expert Panel-selected Elixhauser comorbidities, patient age,

Table 3. Relationship Between Elixhauser Comorbidities and Appropriate Antibiotic Use, as Rated by an Expert Panel[a] Using Delphi Consensus Methodology

Causally Related, n = 14	Indeterminately Related, n = 6	Not Causally Related, n = 9
Valvular disease	Congestive heart failure	Hypertension
Peripheral vascular disease	Pulmonary circulation disorders	Hypothyroidism
Paralysis	Other neurological disorders	Coagulation deficiencies
Chronic pulmonary disease	Diabetes without chronic complications	Solid tumor without metastasis
Diabetes with chronic complications	Metastatic cancer	Fluid and electrolyte disorders
Renal failure	Weight loss	Blood loss anemia
Liver disease		Deficiency anemias
Chronic peptic ulcer disease		Psychoses
HIV and AIDS		Depression
Lymphoma		
Rheumatoid arthritis/ collagen vascular diseases		
Obesity		
Alcohol abuse		
Drug abuse		

Abbreviation: HIV, human immunodeficiency virus.

[a]Consisting of 8 infectious disease and antimicrobial stewardship experts in the United States; details available in the Supplementary Materials.

and total Elixhauser score to the CDC Parallel-Facility model (together with our "Expert Panel Consensus-Driven" model) improved predictions over the CDC Parallel-Facility model by a further 0.6–2.9% (Figure 1). However, adding an additional variable for bacterial infection, as derived from ICD-10-CM codes, achieved the most significant prediction improvements. For predicting total antibiotic use, a model incorporating these facility- and patient-level variables (our "Expert Panel Consensus-Driven + Bacterial Infection" model) was 24% more accurate than the CDC Parallel-Facility model, equating to a reduction in average antibiotic DOTs mis-predicted per admission from 3.107 DOTs to 2.348 DOTs. This model was also very well calibrated for all patients except those with the highest 90th percentile of antibiotic use (Figure 1). Our present on admission-variant of the "Expert Panel Consensus-Driven + Bacterial Infection" model restricted to Expert Panel-selected Elixhauser comorbidities and bacterial infections that were coded as present on admission, in order to reduce the risk of adjusting for conditions that were a consequence, rather than a cause, of antibiotic initiation. This model performed similarly to the main Expert Panel Consensus-Driven + Bacterial Infection model (see Supplementary Figure 1 and Figure 2).

Yu et al's "Simplified ASP" model was substantially similar to our Expert Panel Consensus-Driven + Bacterial Infection

model for predicting total antibiotic use (<1.5-percentage point difference) but was somewhat more accurate for the other antibiotic outcomes (6–7 percentage points). We also combined the Yu et al model with our best-performing model, the Expert Panel Consensus-Driven + Bacterial Infection model, which marginally improved predictions for total antibiotic use and antibiotics for community-acquired infections but did not improve performance for the remaining two antibiotic outcomes (see Supplementary Table 3).

Impact of Risk-adjustment Using Patient-level Variables on Hospital Rankings

Figure 2 describes the impact of adjustment using the CDC Parallel-Facility model and select patient-variable models on hospitals' rankings of use by quartile. We ranked hospitals from highest to lowest crude rates of antibiotic use for each of the 4 outcomes, split these rankings into quartiles of 144 hospitals each, and then reranked hospitals by their SAARs. The bottom or "worst" quartile corresponded to hospitals with the highest crude usage rates or, after adjustment, the highest SAARs; the converse applied to hospitals in the top or "best" quartile [23, 24].

For total antibiotic use, when risk-adjusting using the CDC Parallel-Facility model, 46 (32%) and 42 (29%) of the

	RISK-ADJUSTMENT USING THE:	Number (%) of the 144 hospitals in the bottom ("worst") quartile of unadjusted antibiotic use that change after risk-adjustment:	Number (%) of the 144 hospitals in the top ("best") quartile of unadjusted antibiotic use that change after risk-adjustment:
ALL (TOTAL) ANTIBIOTICS	CDC Parallel-Facility Model	46 (32%)	42 (29%)
	Expert Panel Consensus-Driven + Bacterial Infection Model (A)	61 (42%)	55 (38%)
	POA Variant: Expert Panel Consensus-Driven + Bacterial Infection Model (B)	62 (43%); 52*	59 (41%); 54*
	Increase in Hospital Movement Into/Out of Quartile When Adjusting with a Patient-Level Model Compared to Adjustment with the CDC-Parallel Facility Model (%)[1]	33% (A), 35% (B)	31% (A), 40% (B)
ANTIBIOTICS FOR HOSPITAL-ONSET INFECTIONS	CDC Parallel-Facility Model	41 (28%)	30 (21%)
	Expert Panel Consensus-Driven + Bacterial Infection Model (A)	53 (37%)	40 (28%)
	POA Variant: Expert Panel Consensus-Driven + Bacterial Infection Model (B)	55 (38%); 47*	43 (30%); 35*
	Increase in Hospital Movement Into/Out of Quartile When Adjusting with a Patient-Level Model Compared to Adjustment with the CDC-Parallel Facility Model (%)[1]	29% (A), 34% (B)	33% (A), 43% (B)
ANTIBIOTICS FOR COMMUNITY-ACQUIRED INFECTIONS	CDC Parallel-Facility Model	56 (39%)	52 (36%)
	Expert Panel Consensus-Driven + Bacterial Infection Model (A)	65 (45%)	58 (40%)
	POA Variant: Expert Panel Consensus-Driven + Bacterial Infection Model (B)	66 (46%); 47*	61 (42%); 35*
	Increase in Hospital Movement Into/Out of Quartile When Adjusting with a Patient-Level Model Compared to Adjustment with the CDC-Parallel Facility Model (%)[1]	16% (A), 18% (B)	12% (A), 17% (B)
ANTIBIOTICS FOR RESISTANT GRAM-POSITIVE INFECTIONS	CDC Parallel-Facility Model	34 (24%)	30 (21%)
	Expert Panel Consensus-Driven + Bacterial Infection Model (A)	52 (36%)	39 (27%)
	POA Variant: Expert Panel Consensus-Driven + Bacterial Infection Model (B)	52 (36%); 44*	40 (28%); 36*
	Increase in Hospital Movement Into/Out of Quartile When Adjusting with a Patient-Level Model Compared to Adjustment with the CDC-Parallel Facility Model (%)[1]	53% (Both A and B)	30% (A), 33% (B)

*: Number of hospitals that newly entered the quartile that were the same when adjusting with the POA Variant Model (B) and with the Expert Panel Consensus-Driven + Bacterial Infection Model (A). For example, for risk-adjusting total antibiotic use, 52 of the hospitals that entered the bottom quartile after risk-adjustment would have been the same whether you had risk-adjusted using the Expert Panel Consensus-Driven + Bacterial Infection model (A) or using the POA Variant of this model (B).

1 – Percentages calculated as follows: [(# of Hospitals that Entered/Exited Quartile after Adjustment with Model A or B − # of Hospitals that Entered/Exited Quartile After Adjustment with the CDC-Parallel Facility Model) / (# of Hospitals that Entered/Exited Quartile After Adjustment with the CDC-Parallel Facility Model)]*100. For example, for calculating the increase in movement into or out of the bottom quartile of total antibiotic use when adjusting with the Expert Panel Consensus-Driven + Bacterial Infection Model compared to when adjusting with the CDC Parallel-Facility Model, the calculation was [(61-46)/46]*100 = 33%. Because quartile sizes are fixed at 25% of total hospitals (144/576), the number of hospitals that enter each quartile after risk-adjustment is offset by an equivalent number of hospitals that exit.

Figure 2. Changes in number of hospitals ranked in the top or bottom quartiles of use, compared to rankings by unadjusted usage rates (DOTs/1000 patient-days), after risk-adjustment with: the CDC Parallel-Facility model, the Expert Panel Consensus-Driven + Bacterial Infection model, and the POA) variant of the Expert Panel Consensus-Driven + Bacterial Infection model. Abbreviations: CDC, Centers for Disease Control and Prevention; DOT, day of therapy; POA, present on admission.

144 hospitals in the bottom and top quartiles, respectively, changed after adjustment (Figure 2). By comparison, when risk-adjusting using the Expert Panel Consensus-Driven + Bacterial Infection model, 61 (42%) and 55 (38%) hospitals changed in these respective quartiles, representing 31–33% more hospital movement compared to CDC Parallel-Facility model adjustment. Hospital movement was even more substantial under the present-on-admission variant of our Expert Panel Consensus-Driven + Bacterial Infection model (Figure 2). Relative to adjustment with the CDC Parallel-Facility model, large quartile movements were also observed for antibiotics for hospital-onset infections and for resistant gram-positive infections when risk-adjusting with patient-variable models (29–53% more movement into bottom or top quartiles of use when adjusting with patient-variable models); antibiotics for community-acquired infections yielded the smallest differences between adjustment with the CDC Parallel-Facility model and adjustment with patient-variable models (12–18%, Figure 2).

DISCUSSION

Across a large and diverse cohort of US hospitals, and nearly 12 million admissions, this study found that adding patient-level data to existing, facility-variable risk-adjustment models consistently improved predictions for inpatient antibiotic utilization. More accurate predictions produce more accurate observed-to-predicted use ratios (eg, SAARs), the current bedrock of antibiotic utilization comparison used by hospitals and the CDC and the primary method for identifying outlier prescribing. Importantly, these accuracy improvements were achieved using variables (1) derived from patient ICD-10-CM diagnosis codes, which are readily electronically available and mandated for all patients discharged from HIPAA-compliant US hospitals [25]; and (2) Elixhauser comorbidities expert-rated as associated with appropriate antibiotic use. If SAARs are eventually deployed for quality assessment or reimbursement purposes, these attributes lend practical and policy value. Including these patient data also yielded large downstream effects on hospital rankings. For example, for ranking total antibiotic use, approximately 30% more hospitals moved into the bottom or top quartiles after adjusting with a model that included patient variables, compared to adjusting with a model that only included facility characteristics.

For all antibiotic outcomes, the "Expert Panel Consensus-Driven + Bacterial Infection" was our best-performing model. The Expert Panel-selected Elixhauser comorbidities included in this model are a distinguishing feature of this study and helped to focus predictions more squarely on appropriate antibiotic use. Predicting appropriate antibiotic use is a goal shared by CDC, antibiotic stewardship programs, and other researchers [1, 10, 26]. To be clear, however, we do not intend to suggest (and indeed, would not expect) that antibiotic use in patients with these comorbid conditions is always appropriate. We further recognize that bacterial infections, the other principal variable in this model, may be misdiagnosed or treated with antibiotics inappropriately. But on average across facilities, models that include these expert-selected Elixhauser conditions should yield fairer, more accurate SAARS that protect hospitals that treat more patients with these comorbidities from being penalized for resultantly higher rates of antibiotic use.

The prior work that most closely relates to our current study is a 2018 paper by Yu et al, which also compared patient-variable models to a facility-variable model for predicting antibiotic utilization [4]. We recreated and externally validated their primary model ("Simplified ASP") and found that its accuracy was similar to, and in some cases a few percentage points better than, our best-performing model. This information is new, because although the Yu et al study established high concordance between their Simplified ASP model and a "Complex" model that they also derived, their study did not quantify model accuracy outright. Interestingly, both the Yu et al model and our best-performing model were most accurate at predicting use of antibiotics for resistant gram-positive infections. We attribute this finding not to a stronger relationship between predictors and outcome per se but rather to low use: with 75% of admissions having 0 DOTs for these antibiotics, even our null negative binomial model with no predictors fit the data relatively well. Conversely, outcomes with higher DOTs, such as total antibiotics, demonstrated higher absolute errors across all tested models but also proportionally greater error reductions as patient-level variables were added.

More broadly, our findings and our external validation of the Yu et al model reinforce the benefits of adding patient data to risk-adjustment models and suggest that the accuracy improvements achieved by our and the Yu et al models are likely real and reproducible. Of note, the Yu et al model's most important predictor was an admission's MS-DRG, which is a CMS reimbursement code assigned at discharge based upon the principal diagnosis, as well as certain secondary diagnoses and procedures [27]. Although we had MS-DRG code data available for all admissions because Premier calculates them, we intentionally did not evaluate this variable as a predictor because some commercial insurers do not use them, and CMS has specifically urged caution when applying MS-DRGs to non-Medicare populations [28]. Importantly, by incorporating clinical constructs generated exclusively from ICD-10-CM diagnosis codes, our models should be executable regardless of payer and without requiring additional extraction of procedure data.

Our study has several limitations. First, it did not include pediatric patients. NHSN has developed pediatric care location-specific SAARs [6] and validating our models against existing NHSN facility-variable models in pediatric patients would be an important area of future study. Second, we derived our clinical constructs from patient claims data, specifically

ICD-10-CM diagnosis codes. Although ICD-10-CM codes have imperfect sensitivity and specificity and will not fully capture preadmission healthcare information, from a practical perspective, claims data-based constructs are a significant study strength, because they are available across all US hospitals and payers. However, because ICD-10-CM codes are assigned at discharge, ascertaining whether diagnoses preceded or followed antibiotic initiation is challenging. To address this uncertainty, we evaluated a model variant that restricted to bacterial infections and comorbidities coded as present on admission; this model performed similarly to the main model. As a matter of policy, however, national variability in POA coding and the potential for POA code misuse [29, 30] would require further consideration, standardization and study. Third, we calculated antibiotic DOTs using charge data. Many large studies have used claims data to measure antibiotic utilization [11, 16, 17], and research has demonstrated strong agreement between pharmacy charge and administration records for antibiotics [31], but some residual discordance between these data sources remains possible. Fourth, we predicted antibiotic use at the admission level because, beyond increasing study power, it accounted for patients who move among units during their admission. Consequently, we could closely, but not exactly, recreate unit-level prediction models (eg, CDC NHSN models, the Yu et al model), and we therefore caution against direct comparisons across studies.

Overall, our study found that compared to only adjusting for facility characteristics and patient care location, adding patient comorbidities and bacterial infection diagnoses consistently improved models for predicting inpatient antibiotic utilization. As such, this study in a large, diverse cohort of 576 US hospitals adds to a growing body of evidence from prior, smaller studies: accounting for patient-mix is necessary to optimize antibiotic use comparisons. Importantly, our model variables were easily derived from standard claims data and were selected to correlate more closely with appropriate antibiotic use. Moreover, the improvements they yielded were material—a nearly 25% reduction in the average error rate for predicting total antibiotic use and substantial movement in hospitals' utilization rankings. Whether, and when, SAARs are ready for reimbursement or public reporting purposes is a subjective policy question for CDC, CMS, and professional organizations; we hope that this study provides informative evidence to guide these deliberations. We encourage continued investigation of other, electronically available patient data that may further improve antibiotic utilization comparisons.

Supplementary Data

Supplementary materials are available at *Clinical 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.

References

1. The White House. National strategy for combating antibiotic-resistant bacteria. **2014**. Available at: https://www.cdc.gov/drugresistance/pdf/carb_national_strategy.pdf. Accessed 24 October 2019.
2. Ibrahim OM, Polk RE. Benchmarking antimicrobial drug use in hospitals. Expert Rev Anti Infect Ther **2012**; 10:445–57.
3. Centers for Disease Control and Prevention (CDC). Antimicrobial use and resistance (AUR) module. **2020**. Available at: https://www.cdc.gov/nhsn/PDFs/pscManual/11pscAURcurrent.pdf. Accessed 1 May 2020.
4. Yu KC, Moisan E, Tartof SY, et al. Benchmarking inpatient antimicrobial use: a comparison of risk-adjusted observed-to-expected ratios. Clin Infect Dis **2018**; 67:1677–85.
5. O'Leary E for C. Standardized antimicrobial administration ratio (SAAR). **2019**. Available at: https://www.cdc.gov/nhsn/pdfs/training/2019/saar-508.pdf. Accessed 1 May 2020.
6. van Santen KL, Edwards JR, Webb AK, et al. The standardized antimicrobial administration ratio: a new metric for measuring and comparing antibiotic use. Clin Infect Dis **2018**; 67:179–85.
7. O'Leary EN, Edwards JR, Srinivasan A, et al. National Healthcare Safety Network standardized antimicrobial administration ratios (SAARs): a progress report and risk modeling update using 2017 data. Clin Infect Dis **2020**.
8. CMS. CMS measures inventory tool: National Healthcare Safety Network (NHSN) antimicrobial use measure; MUC15-531. **2020**.
9. Polk RE, Hohmann SF, Medvedev S, Ibrahim O. Benchmarking risk-adjusted adult antibacterial drug use in 70 US academic medical center hospitals. Clin Infect Dis **2011**; 53:1100–10.
10. Spivak ES, Cosgrove SE, Srinivasan A. Measuring appropriate antimicrobial use: attempts at opening the black box. Clin Infect Dis **2016**; 63:1639–44.
11. Goodman KE, Cosgrove SE, Pineles L, et al. Significant regional differences in antibiotic use across 576 US hospitals and 11°701°326 million adult admissions, 2016–2017 (Accepted). Clin Infect Dis **2020**.
12. Premier Applied Sciences. Premier Healthcare Database: data that informs and performs. **2018**. Available at: https://products.premierinc.com/downloads/PremierHealthcareDatabaseWhitepaper.pdf. Accessed 1 May 2020.
13. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. Med Care **1998**; 36:8–27.
14. Agency for Healthcare Research and Quality. Clinical Classifications Software (CCS) for ICD-10-CM diagnoses - Beta. Rockville, MD: **2018**. Available at: www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp.
15. Agency for Healthcare Research and Quality. Patient safety indicators: appendix F (Infection Diagnosis Codes). **2019**. Available at: https://www.qualityindicators.ahrq.gov/Downloads/Modules/PSI/V2019/TechSpecs/PSI_Appendix_F.pdf. Accessed 1 May 2020.
16. Baggs J, Fridkin SK, Pollack LA, Srinivasan A, Jernigan JA. Estimating national trends in inpatient antibiotic use among US hospitals from 2006 to 2012. JAMA Intern Med **2016**; 176:1639–48.
17. Kazakova SV, Baggs J, McDonald LC, et al. Association between antibiotic use and hospital-onset *Clostridioides difficile* infection in US acute care hospitals, 2006–2012: an ecologic analysis. Clin Infect Dis **2019**; 70:11–18.
18. Dalkey N, Helmer O. An experimental application of the DELPHI method to the use of experts. Manag Sci **1963**; 9:458–67.
19. Powell C. The Delphi technique: myths and realities. J Adv Nurs **2003**; 41:376–82.

20. Jackson SS, Leekha S, Magder LS, et al. Electronically available comorbidities should be used in surgical site infection risk adjustment. Clin Infect Dis 2017; 65:803–10.
21. Harris AD, Pineles L, Anderson D, et al. Which comorbid conditions should we be analyzing as risk factors for healthcare-associated infections? Infect Control Hosp Epidemiol 2017; 38:449–54.
22. Reich NG, Lessler J, Sakrejda K, Lauer SA, Iamsirithaworn S, Cummings DA. Case study in evaluating time series prediction models using the relative mean absolute error. Am Stat 2016; 70:285–92.
23. Calderwood MS, Kleinman K, Huang SS, Murphy MV, Yokoe DS, Platt R. Surgical site infections: volume-outcome relationship and year-to-year stability of performance rankings. Med Care 2017; 55:79–85.
24. Yokoe DS, Avery TR, Platt R, Kleinman K, Huang SS. Ranking hospitals based on colon surgery and abdominal hysterectomy surgical site infection outcomes: impact of limiting surveillance to the operative hospital. Clin Infect Dis 2018; 67:1096–102.
25. Department of Health and Human Services. HIPAA administrative simplification: modifications to medical data code set standards to adopt ICD–10–CM and ICD–10–PCS. 2009. Available at: https://www.govinfo.gov/content/pkg/FR-2009-01-16/pdf/E9-743.pdf. Accessed 1 May 2020.
26. Hood G, Hand KS, Cramp E, Howard P, Hopkins S, Ashiru-Oredope D. Measuring appropriate antibiotic prescribing in acute hospitals: development of a national audit tool through a Delphi consensus. Antibiotics 2019; 8:1–11. Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6627925/. Accessed 28 March 2020.
27. CMS. Defining the Medicare severity diagnosis related groups (MS-DRGs), version 37.0. 2019. Available at: https://www.cms.gov/icd10m/version37-fullcode-cms/fullcode_cms/Defining_the_Medicare_Severity_Diagnosis_Related_Groups_(MS-DRGs).pdf. Accessed 1 May 2020.
28. CMS. Medicare Program: changes to the hospital inpatient prospective payment systems and fiscal year 2008 rates; final rule. 2007.
29. Calderwood MS, Kawai AT, Jin R, Lee GM. Centers for Medicare and Medicaid services hospital-acquired conditions policy for central line-associated bloodstream infection (CLABSI) and catheter-associated urinary tract infection (CAUTI) shows minimal impact on hospital reimbursement. Infect Control Hosp Epidemiol 2018; 39:897–901.
30. Kawai AT, Calderwood MS, Jin R, et al. Impact of the centers for Medicare and Medicaid services hospital-acquired conditions policy on billing rates for 2 targeted healthcare-associated infections. Infect Control Hosp Epidemiol 2015; 36:871–7.
31. Chan KH, Moser EA, Cain M, Carroll A, Benneyworth BD, Bell T. Validation of antibiotic charges in administrative data for outpatient pediatric urologic procedures. J Pediatr Urol 2017; 13:185–6.
32. Polk RE, Fox C, Mahoney A, Letcavage J, MacDougall C. Measurement of adult antibacterial drug use in 130 US hospitals: comparison of defined daily dose and days of therapy. Clin Infect Dis 2007; 44:664–70.