Received: 18 March 2022 | Accepted: 20 June 2022

DOI: 10.1002/jhm.12919

ORIGINAL RESEARCH

Journal of Hospital Medicine shm Society of Hospital Medicine

A risk model to identify *Legionella* among patients admitted with community-acquired pneumonia: A retrospective cohort study

Michael B. Rothberg MD, MPH[1] | Peter B. Imrey PhD[2,3] | Ning Guo MS[2] | Abhishek Deshpande MD, PhD[1,4] | Thomas L. Higgins MD, MBA[5] | Peter K. Lindenauer MD, MSc[5,6]

[1]Center for Value-Based Care Research, Community Care, Cleveland Clinic, Cleveland, Ohio, USA

[2]Department of Quantitative Health Sciences, Cleveland Clinic, Cleveland, Ohio, USA

[3]Cleveland Clinic's Medical School, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University, Cleveland, Ohio, USA

[4]Department of Infectious Disease, Respiratory Institute, Cleveland, Ohio, USA

[5]Department of Medicine, University of Massachusetts Medical School-Baystate, Springfield, Massachusetts, USA

[6]Department of Medicine, Institute for Healthcare Delivery and Population Science, University of Massachusetts Medical School-Baystate, Springfield, Massachusetts, USA

Correspondence

Michael B. Rothberg, MD, MPH, Center for Value-Based Care Research, 9500 Euclid Ave, Mail Code G10, Cleveland, OH 44195, USA.
Email: rothbem@ccf.org;
Twitter: @MRothbergMD

Funding information

Agency for Healthcare Research and Quality, Grant/Award Number: R01HS024277

Abstract

Background: Guidelines recommend testing hospitalized patients with community-acquired pneumonia (CAP) for *Legionella pneumophila* only if the infection is severe or risk factors are present. There are no validated models for predicting *Legionella*.

Objective: To derive and externally validate a model to predict a positive Legionella test.

Design, Setting and Participants: Diagnostic study of adult inpatients with pneumonia using data from 177 US hospitals in the Premier Healthcare Database (training and hold-out validation sets) and 12 Cleveland Clinic Health System (CCHS) hospitals (external validation set). We used multiple logistic regression to predict positive *Legionella* tests in the training set, and evaluated performance in both validation sets.

Main Outcome and Measures: The outcome was a positive *Legionella* test. Potential predictors included demographics and co-morbidities, disease severity indicators, season, region, and presence of a local outbreak.

Results: Of 166,689 patients hospitalized for pneumonia, 43,070 were tested for *Legionella* and 642 (1.5%) tested positive. The strongest predictors of a positive test were a local outbreak (odds ratio [OR], 3.4), June–October occurrence (OR, 3.4), hyponatremia (OR, 3.3), smoking (OR, 2.4), and diarrhea (OR, 2.0); prior admission within 6 months (OR, 0.27) and chronic pulmonary disease (OR, 0.49) were associated with a negative test. Model *c*-statistics were 0.79 in the Premier and 0.77 in the CCHS validation samples. High-risk patients were only slightly more likely to have been tested than lower-risk patients. Compared to actual practice, the model-based testing strategy detected twice as many cases.

Conclusions: Although *Legionella* is an uncommon cause of pneumonia, patient characteristics can identify individuals at high risk, allowing for more efficient testing.

INTRODUCTION

Legionella pneumophila is estimated to cause approximately 2% of community-acquired pneumonia (CAP) cases.[1] As the organism is difficult to culture, *Legionella* pneumonia has historically been challenging to diagnose,[2] but polymerase chain reaction (PCR) and urinary antigen testing (UAT) have greatly increased the ability to reliably test for *Legionella*. However, because positivity rates are low,[1] American Thoracic Society/Infectious Disease Society of America (ATA/IDSA) 2019 guidelines recommend against routine UAT, except in those with severe CAP (e.g., requiring mechanical ventilation or vasopressors) or in association with a *Legionella* outbreak or recent travel.[3] In a large retrospective study of US hospitals, we found that 26% of pneumonia patients had *Legionella* testing, few had *Legionella* risk factors, and only 1.5% tested positive.[4] Nevertheless, appropriate testing for *Legionella* is important because delayed treatment is associated with poor outcomes.[5,6]

Legionella is often characterized by hyponatremia, high fever, gastrointestinal symptoms, and history of travel.[7,8] Incidence varies by season and region,[4,9] and cases often occur in clusters. Because of these differentiating attributes, we hypothesized that accurate prediction models for test positivity could be created. Previous models[10–14] have used small samples, lacked validation, and/or required variables not routinely available. To improve testing efficiency, we used data from 177 US hospitals to derive, validate, and refine a prediction model, and further validated it using data from 12 unrelated hospitals. We then estimated the relative screening yields of model-guided and guideline-based testing strategies.

METHODS

Derivation sample

We identified adult patients with pneumonia from hospitals that contributed data to the Premier Health Care Database between 2010 and 2015. The database includes date-stamped records of items billed during hospitalization and is widely used for research.[15] A subset of Premier hospitals also report microbiological results, including blood cultures, respiratory cultures, direct fluorescence antibody (DFA) testing, PCR, and UAT. As previously described,[4] we included patients with a principal diagnosis of pneumonia, or principal diagnosis of respiratory failure or sepsis with a secondary diagnosis of pneumonia. To increase the specificity of pneumonia identification, we restricted to patients with chest imaging, antibiotics on each of the first 3 hospital days, and a test for *L. pneumophila*. We defined a positive test as any positive respiratory culture, UAT, DFA, or PCR.

Validation samples

From the initial sample, we reserved a random 20% for model validation. We applied the same inclusion and exclusion criteria to create an external validation sample from the electronic health records (EHR) of Cleveland Clinic Health System (CCHS), including 12 non-Premier hospitals in Northeast Ohio and Florida between January 2017 and August 2020.

Potential predictors

We assessed the following potential predictors based on our work[4] and that of others[1,8,10,11,16]: patient demographics (age, sex, race, marital status), admission source (emergency room, nursing home, clinic, other), clinical findings on admission (hyponatremia, fever, diarrhea), pneumonia severity measures (ICU admission, mechanical ventilation, vasopressor use), comorbid conditions (e.g., chronic pulmonary disease, congestive heart failure, cancer), smoking status, risk factors for multidrug resistant organisms (residence in skilled nursing facility, admission in the past 6 months, hemodialysis, immunosuppression), hospital characteristics (US Census region, size, teaching status), *Legionella* season (June through October), and recent outbreak as indicated by another *Legionella* patient within the same or preceding month (since the deidentified database contains only admission month, this approximates the previous 30 days). As *Legionella* seasonality varies by region, we evaluated the interaction of the region with the season.

Predictor variables for Premier patients were based on ICD-9 and charge codes, whereas those for CCHS included both ICD-10 codes (e.g., comorbidities) and clinical information (e.g., smoking status and hyponatremia). Diagnoses were derived using software from AHRQ, based on the work of Elixhauser.[17] For CCHS data, which includes precise admission dates, we defined the recent outbreak as another *Legionella* case admitted to that hospital within the preceding 30 days. For CCHS patients, we also noted the lowest serum sodium (Na^+), highest lactase dehydrogenase (LDH) and c-reactive protein (CRP), and maximum temperature within 12 h of admission.

Statistical methods

We modeled a positive *Legionella* test based on variables observed at hospital admission in the Premier "training set," using a relaxed standardized group lasso (least absolute shrinkage and selection operator) and regularized logistic regression. This approach fits an ordinary multiple logistic regression model to the subset of variables selected by a group lasso criterion, with the group lasso penalty weight optimized by cross-validation using the "one standard error rule,"[18,19] from which we report ordinary asymptotic Wald-based confidence intervals and *p*-values. Model discrimination was assessed by c-statistic (a.k.a. the area under the receiver operating characteristic curve [AUC], ranging from 0.5 for no to 1.0 for perfect discrimination) in the training sample and by scoring the validation samples by the training sample linear predictor. Model calibration, that is, agreement of predicted with observed *Legionella* rates, was assessed by grouped bar charts of observed and expected positive

tests within deciles of predicted probability. Contributions of predictors were ranked by increases in Akaike's Information Criterion (AIC) and Schwarz's Bayesian Information Criterion (BIC) from removing individual variables, after reweighting the predictors in each validation set.

The *c*-statistic from the hold-out validation sample describes the performance of the initial training sample model without bias. To increase precision and potentially improve future performance, coefficients were re-estimated from the full Premier sample, and the updated model was externally validated using the CCHS sample.

To gauge the potential clinical benefit of our model, we compared the efficiency of testing based on (a) our model and (b) the ATS/IDSA guidelines, which recommend testing only patients with severe pneumonia or epidemiological factors, including association with a *Legionella* outbreak or recent travel. We identified severe pneumonia by invasive mechanical ventilation and/or vasopressors on hospital Day 1. We had no data on travel but defined a recent outbreak as we did for our model. We compared these strategies among the tested patients, comparing the yield of positive tests. For our model, we chose a predicted probability threshold that would produce the same number of tests as the guideline-directed strategy. In this same sample, we then calculated the predicted number of positive tests for each strategy based on our model and compared it to the observed number of positive tests. Next, we compared yields using the entire Premier sample, projecting each strategy's observed to predicted ratio to the untested patients. We repeated this process in the CCHS validation sample.

Finally, in view of the reported associations of serum Na^+, LDH, CRP, and fever with *Legionella*,[10] we evaluated the change in model performance from adding these variables in the CCHS sample. In this augmented model, these were parameterized as zero if within the normal range, and otherwise as the absolute distances above 99.5°F and outside the lower limit of normal Na^+ and upper limits of normal LDH and CRP.

RESULTS

Of 166,689 premier patients, 43,070 (25.8%) were tested for *Legionella*, of whom 642 (1.5%) were positive. Table 1 and the Supporting Information Table show characteristics of patients with positive and negative tests in the derivation and validation sets, respectively. In the derivation set, the median age was 66.7 years, 50.6% were female and 79.6% were White. Compared to patients with negative tests, those with positive tests were more often from the Northeast, presenting between June and October, at hospitals with another *Legionella* admission in the same or previous month. These three local factors (major US Census Bureau region, seasonality, and prior or same month hospitalization of a *Legionella* case), as well as four pre-existing patient risk factors (sex, smoking, chronic pulmonary disease, and hospital admission within the past 6 months), and two signs/symptoms (hyponatremia and diarrhea) were retained as predictors. Recognizing that seasonal factors influencing *Legionella* manifest differently across regions, we ad hoc augmented the model with a region × season interaction term. The interaction was borderline statistically significant ($p = .06$), substantively justified, modestly raised the model c-statistic, and reduced AIC in the validation sample; it was therefore retained.

Odds ratios and confidence intervals for predictors are shown in Table 2 for both the training and full-sample models. *Legionella* outbreak (odds ratio [OR] 3.4), June–October admission (OR 3.4), and hyponatremia (OR 3.3) were most strongly linked to positive *Legionella* tests, as were smoking, diarrhea, and male sex, while chronic pulmonary disease and prior admission within 6 months predicted negative tests.

Receiver operating characteristic curves for the initial model in the training set and both validation sets are superimposed in the Supporting Information Figure. The *c*-statistics for model discrimination were 0.81 (95% confidence interval [CI]: 0.79–0.84) in the training sample, 0.79 (0.74–0.84) in the 20% hold-out validation sample, and 0.77 (0.75–0.80) in the external CCHS sample.

Figure 1 shows the calibration plots for the two validation samples. Each plot is relatively flat across the first seven deciles of predicted risk, with observed rates of *Legionella* of 1.2% or below. Both plots also show substantial increases in test positivity in the top decile. The prevalence of *Legionella* in the top decile was 6.5% in the Premier sample and 8.6% in the CCHS sample.

Importantly, over 75% of patients with positive tests in each sample were in the top three risk deciles of the full Premier sample of tested and untested pneumonia patients. However, physician ordering was relatively insensitive to patient risk (Figure 2). Physicians practicing at hospitals within the Premier network obtained tests on 38.5% of patients in the uppermost versus 18.7% of patients in the lowest risk decile. Physicians in the CCHS were even less sensitive to risk.

Of the 43,070 patients who were tested, 13,366 (31%) were guideline-concordant; 2.79% of these tested positive, which was 86% higher than the 1.50% among all patients tested. A model-based strategy, testing the same number of patients but only those with the highest predicted probabilities of a positive test, produced a positivity rate of 3.73%, which was 33.5% higher than the guidelines-based strategy and 150% higher than actual testing.

In the CCHS data, 20,894 patients were tested and 409 (1.96%) tested positive; 11,618 (56%) of the tests were guideline-concordant, of which 297 patients (2.56%) tested positive, an increase of 31% in test yield. Applying the modeled risk threshold we used in the Premier data (1.19%), only 7902 patients met the criteria for testing, of whom 311 (3.94%) tested positive, an increase in efficiency of 54% relative to the guidelines-based strategy and 101% relative to actual testing.

We then estimated how many positives would be detectable by each strategy among untested patients by using the mean predicted probabilities of all patients in each testing strategy: 1.85% of the 27,552 untested patients satisfying the guidelines-based criteria and 2.99% of the 26,291 patients using the individual risk cutoff from the first analysis (1.19%). Combining these with the tested patients yielded 892/40,918 (2.18%) anticipated positives using the guidelines-based strategy and 1,294/39,657 (3.26%) using the

TABLE 1 Baseline characteristics[a]

	Total (*N* = 166,689)	Positive test (*N* = 642)	Negative test (*N* = 42,428)	Not tested (*N* = 123,619)
Local conditions				
Hospital major US census region, No. (%)				
Midwest	42,496 (25.5)	166 (25.9)	11,823 (27.9)	30,507 (24.7)
Northeast	30,929 (18.6)	248 (38.6)	10,415 (24.5)	20,266 (16.4)
South	70,955 (42.6)	212 (33.0)	18,130 (42.7)	52,613 (42.6)
West	22,309 (13.4)	16 (2.5)	2,060 (4.9)	20,233 (16.4)
Same or preceding month *Legionella* pneumonia admission	19,167 (11.5)	320 (49.8)	7448 (17.6)	11,399 (9.2)
June–October, No. (%)	59,689 (35.8)	446 (69.5)	14,834 (35.0)	44,409 (35.9)
Patient factors				
Age, mean ± SD	69.5 ± 16.2	62.1 ± 14.8	66.7 ± 16.7	70.5 ± 15.9
Gender, No. (%)				
Female	85,616 (51.4)	238 (37.1)	21,567 (50.8)	63,811 (51.6)
Male	81,073 (48.6)	404 (62.9)	20,861 (49.2)	59,808 (48.4)
Race, No. (%)				
White	129,255 (77.5)	462 (72.0)	33,834 (79.7)	94,959 (76.8)
Black	20,604 (12.4)	149 (23.2)	5817 (13.7)	14,638 (11.8)
Hispanic	1066 (0.64)	1 (0.16)	204 (0.48)	861 (0.70)
Others	15,621 (9.4)	29 (4.5)	2529 (6.0)	13,063 (10.6)
Unknown	143 (0.09)	1 (0.16)	44 (0.10)	98 (0.08)
Comorbidities				
Tobacco smoker, No. (%)	30,305 (18.2)	244 (38.0)	9314 (22.0)	20,747 (16.8)
Chronic pulmonary disease, No. (%)	77,456 (46.5)	208 (32.4)	20,416 (48.1)	56,832 (46.0)
Congestive heart failure, No. (%)	47,640 (28.6)	121 (18.8)	11,793 (27.8)	35,726 (28.9)
Dialysis, No. (%)	7810 (4.7)	23 (3.6)	1046 (2.5)	6741 (5.5)
Alcohol abuse, No. (%)	7069 (4.2)	59 (9.2)	2104 (5.0)	4906 (4.0)
Diabetes, No. (%)	54,962 (33.0)	193 (30.1)	13,450 (31.7)	41,319 (33.4)
Liver disease, No. (%)	5860 (3.5)	26 (4.0)	1622 (3.8)	4212 (3.4)
Cancer, No. (%)	14,024 (8.4)	25 (3.9)	3185 (7.5)	10,814 (8.7)
Risk for drug resistant organisms				
Immunosuppressed, No. (%)	26,501 (15.9)	112 (17.4)	7832 (18.5)	18,557 (15.0)
Hospital admission in past 6 months, No. (%)	18,154 (10.9)	16 (2.5)	3599 (8.5)	14,539 (11.8)
Admission from SNF, No (%)	12,868 (7.7)	24 (3.7)	2973 (7.0)	9871 (8.0)
Symptoms				
Diarrhea, No. (%)	5528 (3.3)	57 (8.9)	1739 (4.1)	3732 (3.0)
Hyponatremia, No. (%)	20,404 (12.2)	231 (36.0)	5936 (14.0)	14,237 (11.5)
Fever, No. (%)	2408 (1.4)	10 (1.6)	690 (1.6)	1708 (1.4)

(Continued)

TABLE 1 (Continued)

	Total (*N* = 166,689)	Positive test (*N* = 642)	Negative test (*N* = 42,428)	Not tested (*N* = 123,619)
Severity of illness				
Initial ICU admission, No. (%)	42,338 (25.4)	171 (26.6)	11,510 (27.1)	30,657 (24.8)
Initial IMV, No. (%)	13,307 (8.0)	33 (5.1)	3632 (8.6)	9642 (7.8)
Vasopressor, No. (%)	11,275 (6.8)	33 (5.1)	2892 (6.8)	8350 (6.8)

[a]All comparisons except for fever ($p = .001$) and vasopressor use ($p = .24$) are highly statistically significant ($p < .001$) by Pearson's χ^2, Kruskal–Wallis, or *t*-test as appropriate to the distribution of the variable.

TABLE 2 Mutually adjusted partial associations (adjusted odds ratios: aOR) of factors in multivariable prediction model with positive *Legionella* test, with naïve (unadjusted for variable selection) 95% confidence intervals and *p*-values, from Premier training sample, and reestimated coefficients and odds ratios (ORs) from the Premier full sample

	Training sample				Full sample	
		95% confidence limits				
	aOR	Lower	Upper	*p* Value	aOR	Linear predictor weight
Circumstantial factors						
US Census Bureau Region (reference: Midwest)						
Northeast	1.25	0.81	1.94	.31	1.31	0.27
West	0.82	0.34	1.96	.66	1.11	0.10
South	1.02	0.68	1.52	.93	1.18	0.16
Same or preceding month *Legionella* pneumonia admission	3.35	2.78	4.04	<.001	3.37	1.22
Legionella season (June–Oct)	3.37	2.30	4.94	<.001	3.47	1.24
Interaction between Region and Legionella Season (reference: Midwest)						
Northeast × Season	1.37	0.82	2.28	.23	1.32	0.28
West × Season	0.81	0.25	2.59	.72	0.47	−0.76
South × Season	0.71	0.44	1.17	.18	0.64	−0.45
Patient factors						
Hyponatremia	3.30	2.73	3.99	<.001	3.29	1.19
Smoking	2.41	1.99	2.91	<.001	2.41	0.88
Diarrhea	1.96	1.42	2.69	<.001	1.95	0.67
Male (reference: Female)	1.49	1.24	1.79	<.001	1.56	0.44
Chronic lung disease	0.49	0.41	0.60	<.001	0.50	−0.69
Hospital admission within the past 6 months	0.27	0.14	0.51	<.001	0.35	−1.06

model-based strategy, compared to the 642/43,070 (1.5%) actually performed in clinical practice.

Augmenting the Premier-based model with temperature increased the c-statistic in the CCHS sample from 0.77 to 0.83. LDH and CRP were not related to a positive test result, and adding serum Na^+ had a minimal impact beyond the diagnosis of hyponatremia.

DISCUSSION

In this cross-sectional diagnostic study, we derived and validated a prediction model for *L. pneumophila* among 43,070 patients hospitalized for pneumonia and tested at 177 US hospitals. We identified and combined nine key risk factors, many previously known, into a single

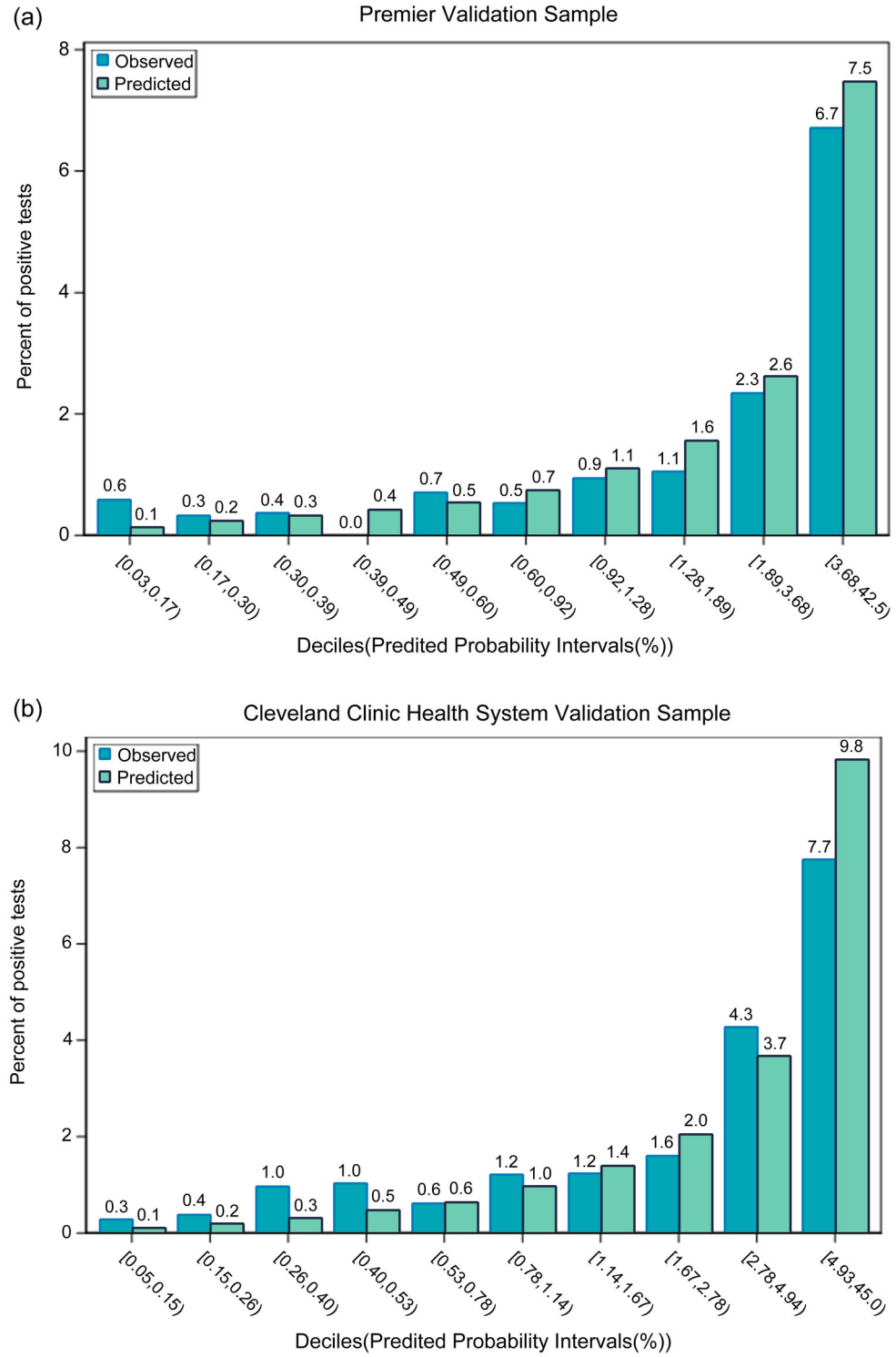

FIGURE 1 Calibration plots for validation samples. Observed and expected test positivity fractions from patients in the Premier and CCHS validation samples, based on a nine-variable prediction model for *Legionella pneumophila* test positivity, by decile of model-predicted test positivity in the respective samples. (a) In the Premier hold-out internal validation sample, based on the model derived from the Premier training sample. (b) In the CCHS external validation sample, based on the revised model derived from the full Premier sample. CCHS, Cleveland Clinic Health System.

prediction model with very good discrimination as measured by the *c*-statistic in two separate validation sets. Based on the observed ordering patterns of this national sample of physicians, the application of this model has the potential to dramatically increase testing efficiency. We estimated that a testing strategy based on the model would have more than doubled test yield compared to observed practice in either the full Premier sample or the subgroup clinicians actually tested, and increased per-test yield by over one-third compared to a guideline-based strategy.

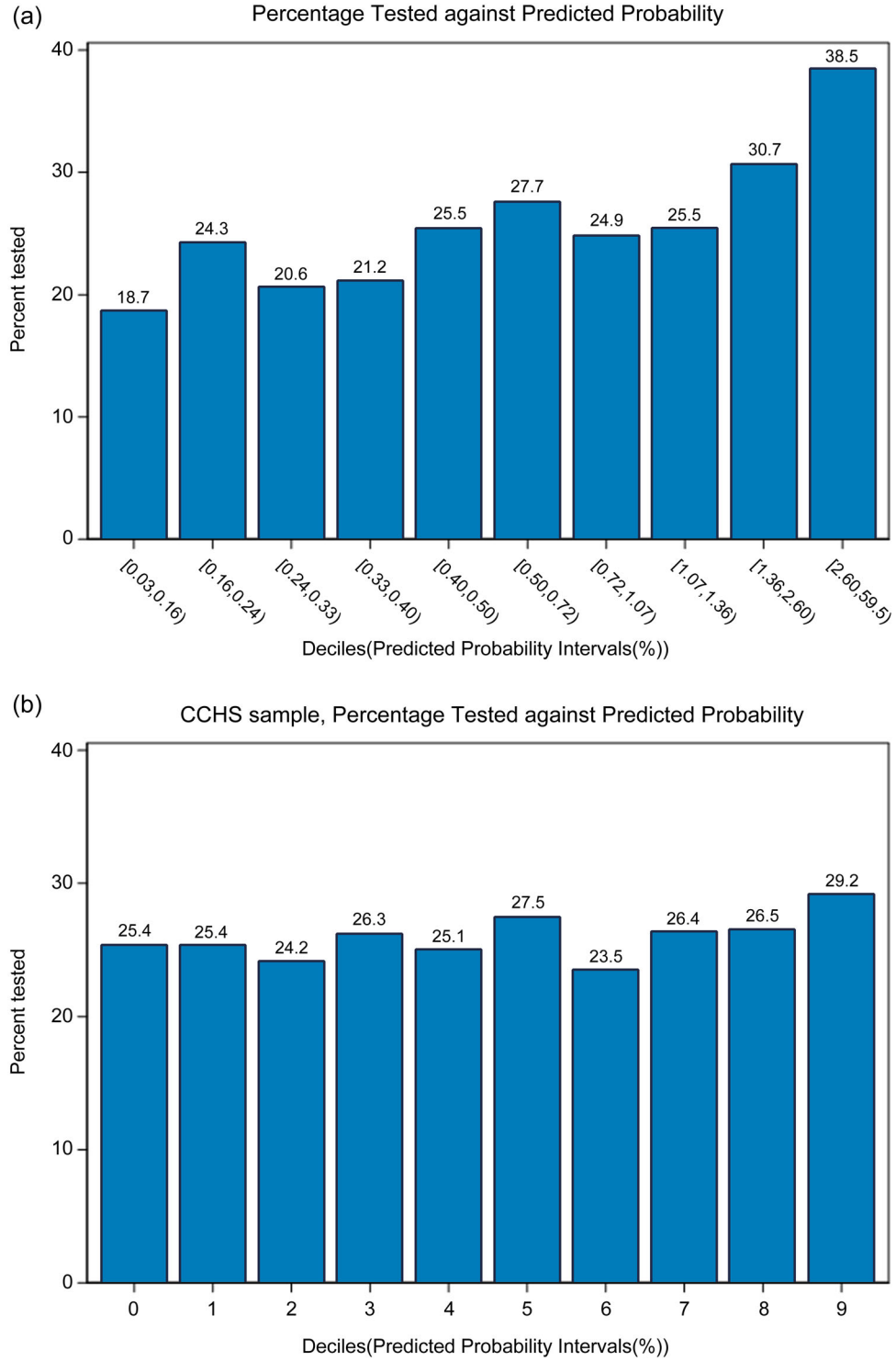

FIGURE 2 Fractions of patients tested for *Legionella pneumophila* by decile of predicted *Legionella* risk. (a) In the full Premier sample. (b) In the CCHS validation sample. CCHS, Cleveland Clinic Health System.

Several groups have attempted to create *Legionella* prediction models. The first point-scoring system was the Winthrop-University Hospital Criteria, published by Cunha[8] in 1998 and validated in 37 veterans administration patients with confirmed *Legionella* and 31 with pneumococcal pneumonia in 2001.[20] Using the "highly probable" cutoff, sensitivity was 78% and specificity was 65%. With these characteristics, in a population with 1.5% *Legionella* prevalence, a "highly probable" score has a positive predictive value (PPV) of only 3.3%. In contrast, our highest risk decile had an observed positivity rate of 6.7% in the Premier validation set and 7.7% in CCHS, and a patient with every risk factor would have a probability of almost 60%. Most recently, Miyashita, et al. developed a scoring system and validated it with 109

patients with *Legionella* and 683 with other causes of pneumonia from 25 Japanese hospitals.[10] Their six predictors (male sex, absence of cough, dyspnea, elevated CRP, high LDH, and hyponatremia) model had an AUC of 0.93 with 93% sensitivity and 75% specificity. However, this model has not been validated in a US population, and two components—CRP and LDH—are not routinely collected for pneumonia patients, limiting its usability.

The most recent ATS/IDSA guidelines for community-acquired pneumonia[3] recommend that testing to detect *Legionella* be reserved for patients with severe pneumonia, exposure to a known *Legionella* outbreak, or recent travel history. We did not find severe pneumonia alone to be associated with a positive test, but testing patients with severe pneumonia may be important if a positive test would influence antimicrobial choice, particularly since antibiotics aimed at multidrug-resistant organisms often do not cover *Legionella*. We did find that a *Legionella* outbreak was strongly associated with a positive test. The current guidelines do not specify what constitutes contact with a known outbreak,[3] but previous guidelines suggest two cases in the same hospital.[7] We considered a similar definition—another case at the same hospital within the past month. We also found strong predictors that the guidelines do not address, including hyponatremia, diarrhea, and seasonality. These have been described by others,[9,16,21] but season has not previously been incorporated into prediction models. Two protective factors, chronic pulmonary disease and previous admission in the past 6 months, are also novel. As such patients are more likely to be admitted for pneumonia due to typical organisms, they may be less likely to be admitted with unusual pathogens such as *Legionella*.

Our model has nine inputs, making it cumbersome to remember. Fortunately, all variables except perhaps diarrhea are readily available in EHRs and could be incorporated into point-of-care clinical decision support tools. Rather than requiring clinicians to always keep rare conditions like *Legionella* in mind, EHRs could alert them to high-risk patients. EHRs are also key because a recent case in the hospital triples the odds of future cases; many clinicians will be unaware of such cases, whereas EHRs can easily scan lab results and incorporate them into clinical decision support. Until such decision support is developed, our model is available as a web-based calculator (https://riskcalc.org/Legionella/). Similar to the ACC/AHA pooled cohort equation for predicting coronary disease, our calculator does not require physicians to remember the risk factors or deal with complex point scoring systems. Clinicians simply check boxes and the calculator will return the probability that the patient will test positive for *Legionella*.

Clinical data could strengthen this model. Other investigators' models have included CRP, LDH, and hyponatremia.[10–12] Although CRP is not routinely collected, LDH often is, and Na^+ and temperature are universally available. When added to our model, only temperature increased predictive capacity substantially (from 0.77 to 0.83), and neither LDH nor CRP was associated with a positive test. Before implementing this augmented version of the model, external validation is warranted.

Legionella accounts for only approximately 1.5% of community cases, making universal *Legionella* testing impractical from a cost standpoint. One study in Texas found that only 0.3% of patients tested positive, at a cost of $12,640 per positive test.[22] Unfortunately, the ATS/IDSA recommendations do not identify high-risk patients either. Moreover, we found that physicians did a poor job of identifying patients at high risk for *Legionella*. In contrast, if they were to use our model to drive decision-making, they could detect many more cases using fewer tests. In the case of the Premier hospitals, we estimated that testing based on our model could have more than doubled the number of *Legionella* cases detected, while simultaneously reducing tests by 10%.

Efficiency could be further improved by limiting testing to patients not expected to receive *Legionella* coverage since there is less need to diagnose *Legionella* in patients already treated for it. The ATS/IDSA guidelines recommend that all patients receive antibiotic regimens that cover *Legionella*, although in practice this may not occur, as many patients, especially patients with severe pneumonia, are treated with vancomycin and piperacillin-tazobactam without a macrolide or quinolone.[23] Failing to test such patients could expose them to harm, as late treatment is associated with worse outcomes.[5,6] Even when patients are treated for *Legionella*, antibiotics aimed at other pathogens could likely be stopped if the *Legionella* test is positive. Thus testing could support antibiotic stewardship.

This study has several limitations. Our model was developed using data only from patients who underwent *Legionella* testing. Applying the same model to all patients could overestimate their risk of *Legionella*. There was a suggestion of such overestimation in our external validation sample, but only in the highest risk decile, and there only modestly. Our model is based on ICD codes, which are applied at discharge. Inaccuracies in coding may have attenuated the relationships between risk factors and the test outcome. Indeed, models incorporating primarily lab values have demonstrated better discrimination.[10]

In conclusion, we derived and externally validated a risk prediction model to identify patients likely to test positive for *Legionella* pneumonia based on local outbreaks, seasonality, comorbidities, and clinical conditions. Using such a model could markedly improve the efficiency of testing for *Legionella* and ensure that patients at high risk receive appropriate empirical antibiotics rather than unnecessarily broad-spectrum treatment. The model is available as an online calculator and could be easily incorporated into EHRs to provide point-of-care clinical decision support.

ACKNOWLEDGMENTS

This work was supported by the Agency for Healthcare Research and Quality [R01HS024277]. All authors except for Thomas Higgins were supported by this grant.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

ETHICS STATEMENT

This study was approved by Cleveland Clinic's Institutional Review Board (IRB #16-1035).

ORCID

Michael B. Rothberg https://orcid.org/0000-0002-2063-1876

Ning Guo https://orcid.org/0000-0002-7875-1491

TWITTER

Michael B. Rothberg @MRothbergMD

REFERENCES

1. Bellew S, Grijalva CG, Williams DJ, et al. Pneumococcal and *Legionella* urinary antigen tests in community-acquired pneumonia: prospective evaluation of indications for testing. *Clin Infect Dis*. 2018;68(12):2026-2033. doi:10.1093/cid/ciy826
2. Harrison T, Uldum S, Alexiou-Daniel S, et al. A multicenter evaluation of the Biotest *Legionella* urinary antigen EIA. *Clin Microbiol Infect*. 1998;4(7):359-365. doi:10.1111/j.1469-0691.1998.tb00079.x
3. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. *Am J Respir Crit Care Med*. 2019;200(7):e45-e67. doi:10.1164/rccm.201908-1581ST
4. Allgaier J, Lagu T, Haessler S, et al. Risk factors, management, and outcomes of *Legionella pneumonia* in a large, nationally representative sample. *Chest*. 2021;159(5):1782-1792. doi:10.1016/j.chest.2020.12.013
5. Heath CH, Grove DI, Looke DFM. Delay in appropriate therapy of *Legionella pneumonia* associated with increased mortality. *Eur J Clin Microbiol Infect Dis*. 1996;15(4):286-290. doi:10.1007/BF01695659
6. Levcovich A, Lazarovitch T, Moran-Gilad J, et al. Complex clinical and microbiological effects on Legionnaires' disease outcome; a retrospective cohort study. *BMC Infect Dis*. 2016;16(1):75. doi:10.1186/S12879-016-1374-9
7. Mandell LA, Wunderink RG, Anzueto A, et al. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. *Clin Infect Dis*. 2007;44(suppl 2):S27-S72. doi:10.1086/511159
8. Cunha BA. Clinical features of legionnaires' disease. *Semin Respir Infect*. 1998;13(2):116-127.
9. Alarcon Falconi TM, Cruz MS, Naumova EN. The shift in seasonality of legionellosis in the USA. *Epidemiol Infect*. 2018;146(14):1824-1833. doi:10.1017/S0950268818002182
10. Miyashita N, Horita N, Higa F, et al. Validation of a diagnostic score model for the prediction of *Legionella pneumophila* pneumonia. *J Infect Chemother*. 2019;25(6):407-412. doi:10.1016/j.jiac.2019.03.009
11. Roed T, Schonheyder HC, Nielsen H. Predictors of positive or negative *Legionella* urinary antigen test in community-acquired pneumonia. *Infect Dis*. 2015;47(7):484-490. doi:10.3109/23744235.2015.1021830
12. Fiumefreddo R, Zaborsky R, Haeuptle J, et al. Clinical predictors for Legionella in patients presenting with community-acquired pneumonia to the emergency department. *BMC Pulm Med*. 2009;9:4. doi:10.1186/1471-2466-9-4
13. Bolliger R, Neeser O, Merker M, et al. Validation of a prediction rule for *Legionella pneumonia* in emergency department patients. *Open Forum Infect Dis*. 2019;6(7). doi:10.1093/OFID/OFZ268
14. Saraya T, Nunokawa H, Ohkuma K, et al. A novel diagnostic scoring system to differentiate between *Legionella pneumophila* pneumonia and *Streptococcus pneumoniae* pneumonia. *Intern Med*. 2018;57(17):2479-2487. doi:10.2169/INTERNALMEDICINE.0491-17
15. Premier Incorporated. Premier Healthcare Database White Paper: data that informs and performs. *Prem Appl Sci*. 2020;2:1-15.
16. Phin N, Parry-Ford F, Harrison T, et al. Epidemiology and clinical management of Legionnaires' disease. *Lancet Infect Dis*. 2014;14(10):1011-1021. doi:10.1016/S1473-3099(14)70713-3
17. Elixhauser A, Steiner C, Harris DR, Coffey RM. Comorbidity measures for use with administrative data. *Med Care*. 1998;36(1):8-27. doi:10.1097/00005650-199801000-00004
18. Hastie T, Tibshirani R, Wainwright M. *Statistical Learning with Sparsity: The Lasso and Generalizations*. 1st ed. Chapman and Hall/CRC; 2015. doi:10.1201/b18401
19. Meinshausen N. Relaxed Lasso. *Comput Stat Data Anal*. 2007;52(1):374-393. doi:10.1016/J.CSDA.2006.12.019
20. Gupta SK, Imperiale TF, Sarosi GA. Evaluation of the Winthrop-University Hospital criteria to identify *Legionella pneumonia*. *Chest*. 2001;120(4):1064-1071. doi:10.1378/chest.120.4.1064
21. Simmering JE, Polgreen LA, Hornick DB, Sewell DK, Polgreen PM. Weather-dependent risk for Legionnaires' disease, United States. *Emerg Infect Dis*. 2017;23(11):1843-1851. doi:10.3201/EID2311.170137
22. Henry C, Boethel C, Copeland LA, Ghamande S, Arroliga AC, White HD. Clinical utility of testing for *Legionella pneumonia* in central Texas. *Ann Am Thorac Soc*. 2017;14(1):65-69. doi:10.1513/ANNALSATS.201606-501BC
23. Haessler S, Guo N, Deshpande A, et al. Etiology, treatments, and outcomes of patients with severe community-acquired pneumonia in a large U.S. sample. *Crit Care Med*. 2022;50:1063-1071. doi:10.1097/CCM.0000000000005498

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Rothberg MB, Imrey PB, Guo N, Deshpande A, Higgins TL, Lindenauer PK. A risk model to identify *Legionella* among patients admitted with community-acquired pneumonia: A retrospective cohort study. *J Hosp Med*. 2022;17:624-632. doi:10.1002/jhm.12919