Using Clinical History Factors to Identify Bacterial Infections in Young Febrile Infants

Jeffrey P. Yaeger, MD, MPH[1,2], Jeremiah Jones, MA[3], Ashkan Ertefaie, PhD[3], Mary T. Caserta, MD[1], Edwin van Wijngaarden, PhD[2], and Kevin Fiscella, MD, MPH[4]

Objective To develop a novel predictive model using primarily clinical history factors and compare performance to the widely used Rochester Low Risk (RLR) model.

Study design In this cross-sectional study, we identified infants brought to one pediatric emergency department from January 2014 to December 2016. We included infants age 0-90 days, with temperature ≥38°C, and documented gestational age and illness duration. The primary outcome was bacterial infection. We used 10 predictors to develop regression and ensemble machine learning models, which we trained and tested using 10-fold cross-validation. We compared areas under the curve (AUCs), sensitivities, and specificities of the RLR, regression, and ensemble models.

Results Of 877 infants, 67 had a bacterial infection (7.6%). The AUCs of the RLR, regression, and ensemble models were 0.776 (95% CI 0.746, 0.807), 0.945 (0.913, 0.977), and 0.956 (0.935, 0.975), respectively. Using a bacterial infection risk threshold of .01, the sensitivity and specificity of the regression model was 94.6% (87.4%, 100%) and 74.5% (62.4%, 85.4%), compared with 95.5% (87.5%, 99.1%) and 59.6% (56.2%, 63.0%) using the RLR model.

Conclusions Compared with the RLR model, sensitivities of the novel predictive models were similar whereas AUCs and specificities were significantly greater. If externally validated, these models, by producing an individualized bacterial infection risk estimate, may offer a targeted approach to young febrile infants that is noninvasive and inexpensive. *(J Pediatr 2021;232:192-9)*.

Each year, approximately 500 000 young febrile infants, 0-90 days old, are brought to medical attention.[1,2] Of these, at least one-half will be hospitalized because they will be stratified as high risk for a bacterial infection, such as urinary tract infection (UTI), bacteremia, and meningitis, but only 6%-10% will actually have a bacterial infection.[3] Because a high risk predictive model has remained elusive, efforts have focused on using a combination of categorical clinical characteristics (eg, full term/premature) and serum biomarkers to identify infants at low risk of bacterial infection. Only when each characteristic is satisfied as low risk will an infant be stratified as low risk. These models have high sensitivity (ie, they identify almost all infants with bacterial infections as high risk) but have low specificity (ie, many infants without bacterial infections are classified as high risk and subsequently hospitalized).[4-10] This is problematic because a proportion of febrile infants without bacterial infections may experience adverse events with costly financial and psychosocial effects for the family due to unnecessary hospitalizations.[11,12]

Low-risk predictive models are limited in two ways. First, failure to meet any variable threshold would result in high risk stratification, prompting subsequent hospitalization. Only 1 risk factor is evaluated at a time, without a comprehensive evaluation of the infant, as clinicians are trained to do. Second, low risk models often rely on sophisticated biomarkers, which are invasive, expensive, and can be difficult to obtain in some settings. Furthermore, despite the addition of novel biomarkers, diagnostic characteristics of recent models have shown only marginal improvements.[8,10]

Most medical diagnoses are made based on an individual's history.[13,14] Inexpensive and noninvasive elements of the personal history, such as maximum temperature and duration of illness, are associated with bacterial infections in

AUC	Area under the curve
ED	Emergency department
NNH	Number needed to hospitalize
NPV	Negative predictive value
PPV	Positive predictive value
PROS	Pediatric Research in the Outpatient Setting
RLR	Rochester Low Risk
UTI	Urinary tract infection

From the [1]Department of Pediatrics, University of Rochester School of Medicine and Dentistry, Rochester, NY; Departments of [2]Public Health Sciences, and [3]Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY; and [4]Department of Family Medicine, University of Rochester School of Medicine and Dentistry, Rochester, NY

The authors declare no conflicts of interest.

Portions of this study were presented at the Pediatric Academic Societies annual meeting, April 24-May 1, 2019, Baltimore, Maryland; and at the Pediatric Hospital Medicine National Conference, July 25-July 28, 2019, Seattle, Washington.

https://doi.org/10.1016/j.jpeds.2020.12.079

infants and children,[4,5,8,9,15-19] but are not considered in low risk models. Machine learning methods represent an advanced analytic approach that can identify nonlinear associations between predictors and outcomes and may enhance diagnostic capabilities. Machine learning methods have been used to improve risk stratification of acute coronary syndrome,[20] predict mortality from myocardial infarction,[21] and detect acute kidney injury.[22] To address this gap and the limitations of current models while optimizing state-of-the-art analytic capabilities, our objective was to use primarily clinical history factors to develop 2 novel bacterial infection predictive models using regression and machine learning methods that produce an individualized bacterial infection risk estimate, and compare their performance with the Rochester Low Risk (RLR) model.

Methods

We performed a cross-sectional study from a single, large urban pediatric emergency department (ED). The Drexel University Institutional Review Board approved this study.

Study Design and RLR Model

To identify subjects, 3 members of the study team used a standardized data abstraction tool to review manually the medical records of all infants who presented to the ED from January 1, 2014 to December 31, 2016. Inclusion criteria consisted of (1) age 0-90 days old; (2) fever, defined as a temperature ≥38 °C, within 6 hours of arriving to the ED or reported by the caregiver from a measurement prior to arrival, but during the current illness; (3) documented gestational age at birth; and (4) documented duration of illness (any symptom). Study personnel then used an abstraction instrument to review manually the electronic health record for variables of interest. We chose to use the RLR model (**Table I**; available at www.jpeds.com)[4] as the standard because it is widely used and diagnostic characteristics are similar to other models.[5,7,8,10] Based on our clinical experience, we considered missing values for variables as low risk for purposes of RLR stratification and tested this assumption by comparing bacterial infection rates in infants with complete data to infants with missing data. One author reviewed the abstracted data to assign risk status.

Novel Predictive Models

Informed by the literature,[4,5,8,9,15-19] we selected 10 predictor variables a priori to develop regression and machine learning models. Undocumented or missing values for predictor variables may reflect the clinician's perception of bacterial infection risk. For example, a clinician may not order a urinalysis when their concern for a bacterial infection is low. To address this issue and more completely capture this risk differentiation, we included an additional value for categorical variables, labeled "not documented" or "not ordered." With this in mind, categorical variables included sex (male/female), insurance type (public/private), presence of a chronic medical condition (yes/no), appearance (well/ill/not documented), and cough status (yes/no/not documented). Insurance type was included as a marker of socioeconomic status as poverty may be associated with bacterial infections.[23] In an independent and blinded fashion, 2 authors manually reviewed all records to dichotomize chronic medical conditions based on the likelihood that the condition may be related to a bacterial infection. We discussed discrepant designations until we reached consensus regarding the classification of the condition. Examples of chronic medical conditions include vesicoureteral reflux, hypospadias, and gastroschisis. Infants were recorded as ill-appearing if they were described as toxic, limp, inconsolable, ill-appearing, listless, lethargic, irritable, or unresponsive.[24] A "Review of Systems" template, which noted the presence/absence of cough, was consistently completed in the ED per standard practice with >99% compliance. Continuous variables included age (days), caregiver report of gestational age at birth (weeks), maximum temperature (Celsius), and duration of illness in days (any symptom). For some infants, gestational age was recorded only as "full term" so we used 37.5 weeks for analysis. The only laboratory study included in the novel predictive models was urinary tract inflammation (yes/no/not ordered), defined as ≥5 white blood cells/high power field of unspun urine or positive (≥trace) leukocyte esterase.[4,8,10,25]

Outcome

The primary outcome was bacterial infection (ie, UTI, bacteremia, or meningitis). As previously described, we defined UTI as (1) a catheterized urine specimen that grew ≥10 000 colony forming units/mL of a pathogenic organism; (2) evidence of urinary tract inflammation; and (3) clinical management as a pathogen.[26] We defined bacteremia and meningitis as growth of a single pathogenic organism from blood and/or cerebrospinal fluid cultures that was treated clinically as a pathogen.

Statistical Analyses

Using the 10 predictor variables, we developed 2 novel predictive models: 1 using regression and 1 using an ensemble machine learning method. We considered both regression and super learner analyses because each method possesses inherent advantages and disadvantages. Regression analyses are easily understood and interpretable, however, can be limited in their ability to detect nonlinear associations. Conversely, machine learning algorithms are capable of identifying complex, nonlinear relationships between predictors and outcomes and may offer improvements in diagnostic capabilities. However, the nature of these relationships are often opaque and poorly understood,[27] and therefore, clinicians may appropriately be skeptical of using them in clinical settings. Second, reproducibility is a challenge because there are a substantial number of variables that must be considered and established.[28,29] In addition, studies suggest that machine learning methods may not be better than traditional

methods.[30] To address these possible concerns, we have estimated the importance of each predictor variable. For the regression model, we calculated the coefficients and 95% CIs for each predictor. For the super learner model, we used the VIMP package in R to estimate the importance of each variable by removing 1 variable at a time and then calculating the difference in the observed area under the curve (AUC).[31] We have also provided the data and code we used to perform the analysis, found here (https://zenodo.org/record/4081821#.YA1n_elKhR4).

To fit the regression model, we used a generalized linear model with logit link function. For the ensemble learning approach, we used a super learner model,[32] which combines a set of machine learning algorithms to optimally produce a predictive model which is superior to each individual algorithm. A super learner model first builds a predictive model for each algorithm and then uses cross-validation to find the optimal weighted combination of the predicted values as a final output. We used the SuperLearner R package that includes random forest, earth, generalized additive models with default settings, and generalized linear model. To avoid overfitting, we used 10-fold cross-validation in which we used 9 folds to train and tune the models and one fold to estimate performance.[33]

We compared the RLR model to the regression and super learner models using AUC, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), likelihood ratios, and number needed to hospitalize (NNH). Assuming that high-risk infants are hospitalized, NNH represents the number of infants needed to hospitalize to identify 1 bacterial infection. NNH was calculated using the inverse of the PPV (1/PPV). The output of the 2 novel predictive models is an estimate of the probability, or risk, of a bacterial infection for an individual infant. To demonstrate their features, we considered 3 prespecified probabilities of bacterial infection (.01, .03, and .05) and used them as cut-points for the estimated bacterial infection risks. We calculated sensitivity, specificity, PPV, NPV, likelihood ratios, and NNH for each model based on these cut-points and compared the resulting diagnostic characteristics with the RLR model. Last, we identified infants who were misclassified as disease-free after cross-validation.

To construct CIs, we used a percentile bootstrap technique where we resampled the entire set 500 times and performed 10-fold cross-validation on each bootstrapped sample. We created the cross-validation folds to ensure all of the resampled individuals were grouped together in the same fold and used the out-of-fold predictions to create bootstrapped estimates of sensitivity, specificity, PPV, and NPV. We then used the 0.025 and 0.975 percentiles of these estimates to form the corresponding 95% CIs.[34] We used 10-fold cross-validation in which we used 9 folds to train and tune the models and 1 fold to estimate performance. We used a bootstrap hypothesis test to examine statistically significant differences in cross-validated AUCs, sensitivities, and specificities between the RLR, regression, and super learner models. In addition, we performed 10-fold cross-validation to identify infants with bacterial infections who were misclassified using the novel predictive models. To illustrate the clinical application of this novel bacterial infection predictive model, we used the regression model to create a preliminary web-based risk calculator that could be used at the point-of-care, found here (https://www.urmc.rochester.edu/sites/biriskcalculator/).

Clinicians may be skeptical of using clinical history factors obtained retrospectively because they may be recorded inconsistently and/or inaccurately, potentially introducing a source of bias. To address this issue, we performed a sensitivity analysis by removing the 4 subjective predictor variables (gestational age at birth, appearance, cough status, and duration of illness) that relied on caregiver's report or clinician's assessment. We then repeated the cross-validation procedure for both models with the remaining 6 objective variables (sex, insurance, chronic medical condition, age, maximum temperature, and urinary tract inflammation) that were either reported as a distinct number (eg, temperature) or were documented as a discrete value (eg, urine laboratory study). For the sensitivity analysis, we again used a bootstrap hypothesis test to assess for statistically significant differences in sensitivities and specificities between the RLR, regression, and super learner models.

Results

Of 9177 infants, 877 met inclusion criteria (**Figure**; available at www.jpeds.com). The median age was 57 days, a majority were male, one-half were Hispanic, and almost all infants had public health insurance. Over 90% of infants were full-term, almost two-thirds of whom were documented as "full-term" without a discrete gestational age (**Table II**). We observed moderate agreement in classifying infants with chronic medical conditions (n = 29; 3.3%) and resolved any differences through consensus. Using the RLR model, 390 infants (44%) were high risk (**Table II**). Sixty-seven infants had a bacterial infection (7.6%), 17 of whom had bacteremia or meningitis (1.9%). The most common

Table II. Sample characteristics

Subject characteristics	N = 877 (%)
Demographic characteristics	
Median age, d (SD)	57 (22.3)
Male	500 (57)
Public insurance	801 (91)
Race/ethnicity	
White/other	109 (12)
Black	326 (37)
Hispanic	424 (48)
Unknown race/ethnicity	30 (3)
Clinical characteristics	
Full term	801 (91)
Chronic medical condition	29 (3)
Cough present	439 (50)
Duration of symptoms, mean; median (d)	1.89; 0
Rochester high risk	390 (44)
Outcome	
Bacterial infection	67 (7.6)

Table III. Etiology and type of bacterial infection categorized by age

Type of bacterial infection	0-30 d	31-60 d	61-90 d	Total
Febrile infants, n	147	351	379	877
Infants with any bacterial infection, n (%)	21 (14.3%)	25 (7.1%)	21 (5.5%)	67 (7.6%)
Infants with bacteremia or meningitis, n (%)	4 (2.7%)	9 (2.6%)	4 (1.1%)	17 (1.9%)
Infants with UTI only, n	17 (11.6%)	16 (4.6%)	17 (4.5%)	50 (5.7%)
Organism recovered from urine only (n)	*Escherichia coli* (12) *Klebsiella pneumoniae (3)* *Enterococcus faecalis (1)* Group B Streptococcus (1)	*Escherichia coli* (13) *Enterococcus faecalis* (1) *Citrobacter koseri* (2)	*Escherichia coli* (16) *Enterococcus faecalis (1)*	–
Infants with UTI and bacteremia, n (%)	0	2 (0.6%)	2 (0.5%)	4 (0.5%)
Organism recovered from urine and blood	—	*Escherichia coli (2)*	*Escherichia coli (1)* Group B Streptococcus (1)	—
Infants with bacteremia only, n (%)	3 (2%)	5 (1.4%)	1 (0.3%)	9 (1%)
Organism recovered from blood only (n)	Group B Streptococcus (3)	Group B Streptococcus (2) Salmonella species (1) *Klebsiella pneumoniae (1)* *Staphylococcus aureus (1)*	*Neisseria meningitidis,* Group B (1)	–
Infants with meningitis only, n (%)	1 (0.7%)	0	0	1 (0.1%)
Organism recovered from cerebrospinal fluid only (n)	*Escherichia coli* (1)	–	–	–
Infants with bacteremia and meningitis, n (%)	0	2 (0.6%)	1 (0.3%)	3 (0.3%)
Organism recovered from blood and cerebrospinal fluid (n)	–	Group B Streptococcus (2)	Salmonella species (1)	–

organism identified was *Escherichia coli* (69%) followed by Group B streptococcus (13%) (**Table III**). The RLR model misclassified 3 infants with bacterial infections as low risk: a 22-day old infant with Group B streptococcus bacteremia, a 27-day old infant with *Escherichia coli* meningitis, and a 49-day old infant with *Staphylococcus aureus* bacteremia.

The AUCs of the RLR, regression, and super learner models were 0.776 (95% CI 0.746, 0.807), 0.945 (95% CI 0.913, 0.977), and 0.956 (95% CI 0.935, 0.975), respectively. The bootstrap hypothesis test showed that the cross-validated AUCs of the regression and super learner models were significantly greater than that of the RLR model by 15.5% and 16.6%, respectively. The sensitivity, specificity, and NNH for the RLR model were 95.5% (95% CI 87.5%, 99.1%), 59.6% (95% CI 56.2%, 63.0%), and 6.1, respectively. Using a bacterial infection risk threshold of .01, the sensitivity, specificity, and NNH for the regression model was 94.6% (95% CI 87.4%, 100%), 74.5% (95% CI 62.4%, 85.4%), and 4.2 compared with 95.6% (95% CI 89.6%, 100%), 73.6% (95% CI 66.1%, 81.7%), and 4.3 for the super learner model (**Table IV**). Diagnostic characteristics for the remaining risk thresholds are shown in **Table IV**. At both the .01 and .03 risk thresholds, the sensitivities of the novel predictive models were the same as the RLR model (regression: $P = .418$, $P = .126$; super learner: $P = .718$, $P = .230$) and the specificities were significantly greater (regression: $P < .012$, $P < .002$; super learner: $P < .002$, $P < .002$).

Using a bacterial infection risk threshold of .01, the cross-validated regression and super learner models misclassified 3 infants and 2 infants, respectively. Both models misclassified a 27-day old infant with *Escherichia coli* meningitis and a 22–day old infant with Group B streptococcus bacteremia, as did the RLR model. The regression model also misclassified a 64-day old infant with *Neisseria meningitidis* group B bacteremia. Presence of urinary tract inflammation was most strongly associated with bacterial infections for both models. In addition, female sex, age, ill appearance, maximum temperature, and presence of cough were significantly associated with bacterial infections for the regression model whereas duration of illness was a significant predictor for the super learner model (**Table V**; available at www.jpeds.com). For the sensitivity analysis that was restricted to 6 factors measured objectively and at a risk threshold of .01, the sensitivity and specificity of the regression model was 94.1% (95% CI 86.1%, 98.8%) and 58.2% (95% CI 35.4%, 77.2%) compared with 91.0% (95% CI 84.1%, 96.9%) and

Table IV. Test characteristics for the RLR, regression, and super learner models at selected risk estimates

Predictive model	Sensitivity	Specificity	PPV	NPV	+ LR	− LR	NNH
Rochester risk	0.955 (0.875, 0.991)	0.596 (0.562, 0.630)	0.164 (0.151, 0.178)	0.994 (0.982, 0.998)	2.37 (2.14, 2.61)	0.08 (0.02, 0.23)	6.1
Regression							
Risk = .01	0.946 (0.874, 1)	0.745 (0.624, 0.854)	0.239 (0.162, 0.339)	0.994 (0.986, 1)	3.94 (2.46, 6.59)	0.07 (0.00, 0.18)	4.2
Risk = .03	0.921 (0.842, 0.973)	0.847 (0.792, 0.900)	0.333 (0.254, 0.429)	0.993 (0.985, 0.997)	6.25 (4.31, 9.42)	0.09 (0.03, 0.19)	3
Risk = .05	0.904 (0.822, 0.970)	0.880 (0.842, 0.916)	0.383 (0.301, 0.475)	0.991 (0.984, 0.997)	7.75 (5.48, 10.93)	0.11 (0.03, 0.20)	2.6
Super learner							
Risk = .01	0.956 (0.896, 1)	0.736 (0.661, 0.817)	0.231 (0.173, 0.307)	0.995 (0.987, 1)	3.73 (2.76, 5.39)	0.06 (0.00, 0.14)	4.3
Risk = .03	0.930 (0.861, 0.986)	0.827 (0.778, 0.872)	0.307 (0.238, 0.387)	0.993 (0.986, 0.999)	5.51 (3.99, 7.63)	0.09 (0.02, 0.17)	3.3
Risk = .05	0.912 (0.833, 0.974)	0.860 (0.821, 0.897)	0.349 (0.274, 0.435)	0.992 (0.984, 0.997)	6.66 (4.94, 9.23)	0.10 (0.03, 0.20)	2.9

− LR, negative likelihood ratio; *+ LR*, positive likelihood ratio.
NNH to evaluate and empirically treat 1 infant with a bacterial infection.
Numbers in parentheses represent 95% CIs.

70.2% (95% CI 57.8%, 81.9%) for the super learner model. At this risk threshold, both novel predictive models were as sensitive and specific as the RLR model (regression: $P = .398$, $P = .532$; super learner: $P = .148$, $P = .066$).

For RLR stratification, 13% of values were missing, almost one-half of which occurred among infants >60 days old and consisted mostly of clinical history regarding the nursery stay. Approximately 75% of infants with a missing laboratory study were ≥60 days old. Compared with infants with complete data, the OR of a bacterial infection for infants with missing data was 0.72 (95% CI 0.43, 1.18). For the novel predictive models, 9% of values were either not documented or not ordered, over two-thirds of which were for infants >60 days old. Of infants with at least 1 missing value (n = 273), there were 1 bacterial infections (0.7%), both of whom had bacteremia. The OR of a bacterial infection for infants with at least 1 value that was not documented or not ordered was significantly lower compared with infants with complete data (OR 0.06; 95% CI 0.01, 0.25). No infants were diagnosed with a bacterial infection within 7 days following their initial encounter.

Discussion

As in previous studies,[30] our findings demonstrate that the regression and super learner models performed similarly. Compared with the RLR model and using a conservative bacterial infection risk estimate threshold of .01, the point estimates of the novel predictive models indicate that 25% fewer infants could be hospitalized while misclassifying the same number of infants with bacterial infection. At this lower bacterial infection risk estimate threshold of .01, the RLR, regression and super learner models misclassified 3 infants, 3 infants, and 2 infants with bacterial infections, respectively. Notably, all 3 models failed to identify an infant with *Escherichia coli* meningitis and Group B streptococcus bacteremia. This finding exemplifies the central challenge with evaluating young febrile infants that has been observed in each iteration of low-risk models. Even highly sensitive predictive models will occasionally miss an infant with a bacterial infection. Using the regression model and a more lenient risk threshold of .03, a clinician could expect to misclassify the same number of infants with bacterial infections compared with the RLR tool, but may hospitalize more than 2 times fewer infants. Our findings show that the sensitivities for the regression and super learner models are similar to the RLR model, and the specificities are significantly greater. If confirmed in validation studies, implementation of these models may safely decrease unnecessary hospitalizations, resulting in lower costs and fewer invasive tests while mitigating other harmful effects.[35,36]

Reliance on clinical history factors, rather than serum or cerebrospinal fluid biomarkers, distinguishes our work from other studies and represents a paradigm shift in how clinicians conceptualize bacterial infection risk in young febrile infants. Our approach reflects the importance of the clinical history and illustrates how inexpensive, noninvasive markers of disease can be used to identify infants with bacterial infections in clinical practice. In addition, this approach is relevant from a practical standpoint for 3 reasons. First, we wanted to operationalize a clinician's reasoning skills to evaluate simultaneously relevant clinical information, coupling principles of illness scripts and disease patterns with current evidence. Second, 70% of children in the US are hospitalized in general/community hospitals,[37] many with limited resources to perform on-site laboratory studies. Procalcitonin, for example, is associated with bacterial infections,[8,38] but is not readily available in all settings, limiting point-of-care utilization. By focusing on clinical history factors, we strived to develop an accurate predictive model that could be widely implemented, regardless of available resources. Last, up to 80% of positive blood cultures are due to contaminants,[39-41] the treatment of which can result in unnecessary care and complications while awaiting final determination.[42] One effective approach to reduce the rates of contaminated blood cultures is to refrain from collecting them if they are unlikely to result in growth of a pathogenic organism.[43] By omitting serum biomarkers as predictor variables, clinicians may be able to avoid obtaining blood cultures in a subset of infants found to be at extremely low risk of bacterial infections and whose risk of contamination may be substantially greater than the risk of bacteremia. We envision that this approach could have at least 2 benefits. First, as the decision to collect a blood culture is largely made by individual clinicians, there are subsets of infants who do not receive a blood culture, resulting in extensive practice variation.[41,44] By quantifying this risk, clinicians may be better equipped to embrace a standardized approach to blood culture collection in extremely low-risk infants, thereby reducing practice variation. Second, by identifying infants whose bacterial infection risk is negligible, thousands of blood culture contaminants each year could be avoided, thereby reducing the likelihood that infants with contaminants would be unnecessarily hospitalized and treated. This novel predictive model and web-based bacterial infection risk calculator may offer a path forward that is as conservative as current models but is less invasive with fewer costs. This approach requires rigorous investigation with proper evaluation and external validation before clinicians should forego the initial collection of serum biomarkers and blood cultures for some young febrile infants.

Although 2 infants may be stratified into the same risk-tier using a low-risk model, it is unlikely that they actually have the same bacterial infection risk. The novel predictive models may provide an accurate estimate of this risk. Recognizing that the balance of risk tolerance/risk aversion may vary widely between clinicians, an individualized bacterial infection risk estimate allows clinicians to identify their preferred risk estimate threshold (eg, .01 vs .03). As demonstrated in other clinical decision support tools,[45,46] the practical application of the regression model as a web-based risk calculator is appealing because it can be integrated into a clinician's workflow, facilitating point-of-care utilization with mobilization of resources based on estimated bacterial infection risk. If externally validated, we expect that a web-based risk

calculator could be useful for emergency and hospital medicine clinicians. Clinicians could either manually enter data into a risk calculator or the model could be embedded within an electronic medical record system to automatically calculate risk. In addition, office-based clinicians, who may have limited access to on-site laboratory studies, may be able to utilize the risk calculator to help decide which young febrile infants warrant further diagnostic evaluation, thereby safely avoiding ED visits for a subset of infants.

At least 2 elements of the study design may have contributed to the improved performance of our models compared with the RLR model. First, to avoid sacrificing information that may affect model performance, we used continuous variables, when appropriate, rather than categorizing variables and introducing possibly arbitrary thresholds. Second, in contrast to other studies,[4,8,9,16,41] we included premature infants to avoid potentially introducing selection bias as prematurity is associated with race[47-49] and poverty,[50] which are correlated with each other and may also be associated with bacterial infections.[23]

The sensitivity analysis showed that a subset of 6 objective clinical history factors performed at least as well as the RLR model in terms of sensitivity and specificity. Although external validation studies are needed, this finding suggests there may indeed be value to identifying infants with bacterial infections by using primarily clinical history factors and should assuage concerns that results are due solely to the use of variables obtained from caregiver reports or subjective assessments that may be inaccurate or incomplete. Notably, excluding the infant's appearance in the sensitivity analysis minimally affected model performance. Perhaps appearance may be more important for younger infants or when solely invasive bacterial infections, such as bacteremia and meningitis, are considered. Future work should focus on using separate datasets to identify the combination of clinical history factors for each model that produce the best predictive value.

The Pediatric Research in the Outpatient Setting (PROS) study[16] showed clinicians could safely decrease testing and hospitalization of young febrile infants without strict adherence to a low-risk model. Some clinicians have interpreted this as a justification to rely more heavily on clinical judgment. Given the infrequent occurrence of bacterial infections, others have voiced concern that these findings justify the systematic use of low-risk models to avoid missing bacterial infections. It has also been noted that the PROS study was conducted in largely suburban practices, with experienced clinicians, low rates of public insurance, and established family relationships to ensure appropriate follow-up.[51] This clinical environment contrasts markedly with the urban EDs where predictive models were first developed and was the setting for our study. Now 16 years later, the core question remains: how can clinicians systematically identify infants truly at high risk of a bacterial infection without missing bacterial infections? We propose that the clinical history factors operationalized in this study may overlap with those considered in the PROS study. These clinical history factors, together with a more systematic approach to determine a family's ability to follow-up, may help answer this core question and require further investigation.

This study has several limitations. First, this is a retrospective study from a single site with high rates of public insurance. Documentation and performance of tests were completed at the discretion of the treating clinician, resulting in missing data. However, for the RLR model, the OR of a bacterial infection for infants with missing data was lower relative to infants with complete data, suggesting our approach of labeling missing data as low risk was reasonable. For the novel predictive models, the OR of a bacterial infection for infants with missing data was significantly lower compared with infants with complete data, indicating that missing data did not create a major source of bias in assessing sensitivity for bacterial infections. In addition, the sensitivity analysis affirms there is value in collecting clinical history factors retrospectively to predict bacterial infections. This potential issue requires further investigation with a prospective study design. Second, the novel predictive models are not externally validated, limiting their generalizability. We did, however, use 10-fold cross-validation as an internal validation technique to avoid overfitting. One of the perceived advantages of machine learning models is their ability to consider a large number of predictors, however, the novel predictive models only considered 10 predictors. We did so because we sought to include evidence-based predictors that were consistently documented, thereby limiting the amount of missing data and ensuring that the models made intuitive sense. It is possible that additional data, such as circumcision status, may enhance the diagnostic capabilities of the models. In addition, because we chose predictor variables a priori, it is possible that some variables, such as insurance status, may not substantially contribute to the prediction of bacterial infections. Although model reduction analyses may improve clinical utilization by removing variables that are not strongly associated with an outcome, it is problematic to re-select variables and retrain models with the same dataset because this approach can result in overfitting, thereby limiting reproducibility.[52] Last, given the small sample size, we did not evaluate the outcome of invasive bacterial infections. Future studies, in separate, larger populations, should address these limitations by identifying factors that are most strongly associated with bacterial infections to refine and validate an optimal predictive model.

In this study, we used primarily clinical history factors to develop novel predictive models that estimate bacterial infection risk based on each infant's unique clinical profile. If externally validated, implementation may result in fewer hospitalizations and invasive procedures than the RLR model, without missing infants with bacterial infections. We employed a targeted approach, creating a unique risk estimate for each infant that would allow clinicians to determine their preferred risk thresholds. By developing a preliminary web-based risk calculator, we demonstrated how this technology

can be applied to clinical decision-making at the point-of-care. Although external validation and refinement in variable selection are needed, our results are promising as they may represent a more individualized and value-based approach to manage young febrile infants. ■

Submitted for publication Jul 20, 2020; last revision received Dec 30, 2020; accepted Dec 31, 2020.

Reprint requests: Jeffrey P. Yaeger, MD, MPH, Department of Pediatrics, University of Rochester School of Medicine and Dentistry, 601 Elmwood Ave, Rochester, NY 14642. E-mail: Jeffrey_Yaeger@urmc.rochester.edu

References

1. McCaig LF, Nawar EW. National Hospital Ambulatory Medical Care Survey: 2004 emergency department summary. Adv Data 2006;372:1-29.
2. Woll C, Neuman MI, Aronson PL. Management of the febrile young infant: update for the 21st century. Pediatr Emerg Care 2017;33:748-53.
3. Pantell RH, Roberts KB, Greenhow TL, Pantell MS. Advances in the diagnosis and management of febrile infants: challenging tradition. Adv Pediatr 2018;65:173-208.
4. Dagan R, Powell KR, Hall CB, Menegus MA. Identification of infants unlikely to have serious bacterial infection although hospitalized for suspected sepsis. J Pediatr 1985;107:855-60.
5. Baker MD, Bell LM, Avner JR. Outpatient management without antibiotics of fever in selected infants. N Engl J Med 1993;329:1437-41.
6. Baskin MN, O'Rourke EJ, Fleisher GR. Outpatient treatment of febrile infants 28 to 89 days of age with intramuscular administration of ceftriaxone. J Pediatr 1992;120:22-7.
7. Herr SM, Wald ER, Pitetti RD, Choi SS. Enhanced urinalysis improves identification of febrile infants ages 60 days and younger at low risk for serious bacterial illness. Pediatrics 2001;108:866-71.
8. Gomez B, Mintegi S, Bressan S, Da Dalt L, Gervaix A, Lacroix L. Validation of the "step-by-step" approach in the management of young febrile infants. Pediatrics 2016;138:e20154381.
9. Bachur RG, Harper MB. Predictive model for serious bacterial infections among infants younger than 3 months of age. Pediatrics 2001;108:311-6.
10. Kuppermann N, Dayan PS, Levine DA, Vitale M, Tzimenatos L, Tunik MG, et al. A Clinical prediction rule to identify febrile infants 60 days and younger at low risk for serious bacterial infections. JAMA Pediatr 2019;173:342-51.
11. DeAngelis C, Joffe A, Wilson M, Willis E. Iatrogenic risks and financial costs of hospitalizing febrile infants. Am J Dis Child 1983;137:1146-9.
12. Paxton RD, Byington CL. An examination of the unintended consequences of the rule-out sepsis evaluation: a parental perspective. Clin Pediatr (Phila) 2001;40:71-7.
13. Hampton JR, Harrison MJ, Mitchell JR, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. Br Med 1975;2:486-9.
14. Summerton N. The medical history as a diagnostic technology. Br J Gen Pract 2008;58:273-6.
15. Gorelick MH, Hoberman A, Kearney D, Wald E, Shaw KN. Validation of a decision rule identifying febrile young girls at high risk for urinary tract infection. Pediatr Emerg Care 2003;19:162-4.
16. Pantell RH, Newman TB, Bernzweig J, Bergman DA, Takayama JI, Segal M, et al. Management and outcomes of care of fever in early infancy. JAMA 2004;291:1203-12.
17. Bonadio WA, Hennes H, Smith D, Ruffing R, Melzer-Lange M, Lye P, et al. Reliability of observation variables in distinguishing infectious outcome of febrile young infants. Pediatr Infect Dis J 1993;12:111-4.
18. Roberts KB, Downs SM, Finnell SME, Hellerstein S, Shortliffe LD, Wald ER, et al. Urinary tract infection: clinical practice guideline for the diagnosis and management of the initial UTI in febrile infants and children 2 to 24 months. Pediatrics 2011;128:595-610.
19. Klouda TM, Wang H, Yaeger JP. Association of cough status with bacterial infections in febrile infants. Hospital Pediatr 2020;10:185-9.
20. Hu D, Huang Z, Chan TM, Dong W, Lu X, Duan H. Acute Coronary syndrome risk prediction based on GRACE risk score. Stud Health Technol Inform 2017;245:398-402.
21. Li X, Liu H, Yang J, Xie G, Xu M, Yang Y. Using machine learning models to predict in-hospital mortality for st-elevation myocardial infarction patients. Stud Health Technol Inform 2017;245:476-80.
22. Koyner JL, Carey KA, Edelson DP, Churpek MM. The development of a machine learning inpatient acute kidney injury prediction model. Crit Care Med 2018;46:1070-7.
23. Yaeger JP, Moore KA, Melly SJ, Lovasi GS. Associations of neighborhood-level social determinants of health with bacterial infections in young, febrile Infants. J Pediatr 2018;203:336-44.
24. Aronson PL, Wang ME, Nigrovic LE, Shah SS, Desai S, Pruitt CM, et al. Time to Pathogen detection for non-ill versus ill-appearing infants ≤60 days old with bacteremia and meningitis. Hosp Pediatr 2018;8:379-84.
25. Mintegi S, Bressan S, Gomez B, Da Dalt L, Blazquez D, Olaciregui I, et al. Accuracy of a sequential approach to identify young febrile infants at low risk for invasive bacterial infection. EMJ 2014;31:e19-24.
26. Greenhow TL, Hung YY, Herz AM. Changing epidemiology of bacteremia in infants aged 1 week to 3 months. Pediatrics 2012;129:e590-6.
27. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Machine Intell 2019;1:206-15.
28. Beam AL, Manrai AK, Ghassemi M. Challenges to the reproducibility of machine learning models in health care. JAMA 2020;323:305-6.
29. Liu Y, Chen PC, Krause J, Peng L. How to Read articles that use machine learning: users' guides to the medical literature. JAMA 2019;322:1806-16.
30. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 2019;110:12-22.
31. Williamson B, Simon N, Carone M. vimp: Perform Inference on Algorithm-Agnostic Variable Importance. 2020. R Package; 2020. 2.1.0 ed, Accessed February 10, 2021. https://CRAN.R-project.org/package=vimp
32. van der Laan MJ, Polley EC, Hubbard AE. Super learner. Stat Appl Genet Mol Biol 2007;6. Article25.
33. Hastie T, Tibshirani R, Friedman J. Model assessment and selection. the elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, NY: Springer; 2009. p. 219-59.
34. DiCiccio TJ, Efron B. Bootstrap confidence intervals. Statist Sci 1996;11:189-228.
35. Albenberg L, Kelsen J. Advances in gut microbiome research and relevance to pediatric diseases. J Pediatr 2016;178:16-23.
36. Mbakwa CA, Scheres L, Penders J, Mommers M, Thijs C, Arts IC. Early life antibiotic exposure and weight development in children. J Pediatr 2016;176:105-13.e2.
37. Leyenaar JK, Ralston SL, Shieh MS, Pekow PS, Mangione-Smith R, Lindenauer PK. Epidemiology of pediatric hospitalizations at general hospitals and freestanding children's hospitals in the United States. J Hosp Med 2016;11:743-9.
38. Milcent K, Faesch S, Gras-Le Guen C, Dubos F, Poulalhon C, Badier I, et al. Use of procalcitonin assays to predict serious bacterial infection in young febrile infants. JAMA Pediatr 2016;170:62-9.
39. Richter SS, Beekmann SE, Croco JL, Diekema DJ, Koontz FP, Pfaller MA, et al. Minimizing the workup of blood culture contaminants: implementation and evaluation of a laboratory-based algorithm. J Clin Microbiol 2002;40:2437-44.
40. Weinstein MP, Towns ML, Quartey SM, Mirrett S, Reimer LG, Parmigiani G, et al. The clinical significance of positive blood cultures in the 1990s: a prospective comprehensive evaluation of the microbiology, epidemiology, and outcome of bacteremia and fungemia in adults. Clin Infect Dis 1997;24:584-602.

41. Greenhow TL, Hung YY, Pantell RH. Management and outcomes of previously healthy, full-term, febrile infants ages 7 to 90 days. Pediatrics 2016;138:e20160270.
42. Bates DW, Goldman L, Lee TH. Contaminant blood cultures and resource utilization. The true consequences of false-positive results. JAMA 1991;265:365-9.
43. Hall KK, Lyman JA. Updated review of blood culture contamination. Clin Microbiol Rev 2006;19:788-802.
44. Aronson PL, Thurm C, Alpern ER, Alessandrini EA, Williams DJ, Shah SS, et al. Variation in care of the febrile young infant <90 days in US pediatric emergency departments. Pediatrics 2014;134:667-77.
45. Kuzniewicz MW, Puopolo KM, Fischer A, Walsh EM, Li S, Newman TB, et al. A Quantitative, risk-based approach to the management of neonatal early-onset sepsis. JAMA Pediatr 2017;171:365-71.
46. D'Agostino RB Sr, Vasan RS, Pencina MJ, Wolf PA, Cobain M, Massaro JM, et al. General cardiovascular risk profile for use in primary care: the Framingham Heart Study. Circulation 2008;117:743-53.
47. Mendez DD, Hogan VK, Culhane JF. Institutional racism, neighborhood factors, stress, and preterm birth. Ethn Health 2014;19:479-99.
48. Kramer MR, Hogue CJ, Dunlop AL, Menon R. Preconceptional stress and racial disparities in preterm birth: an overview. Acta Obstet Gynecol Scand 2011;90:1307-16.
49. Menon R, Dunlop AL, Kramer MR, Fortunato SJ, Hogue CJ. An overview of racial disparities in preterm birth rates: caused by infection or inflammatory response? Acta Obstet Gynecol Scand 2011;90:1325-31.
50. DeFranco EA, Lian M, Muglia LA, Schootman M. Area-level poverty and preterm birth risk: a population-based multilevel analysis. BMC Public Health 2008;8:316.
51. Roberts KB. Young, febrile infants: a 30-year odyssey ends where it started. JAMA 2004;291:1261-2.
52. Kuchibhotla A, Brown D, Buja A, Cai J, George E, Zhao L. Valid post-selection inference in model-free linear regression. Ann Stat 2020;48:2953-81.

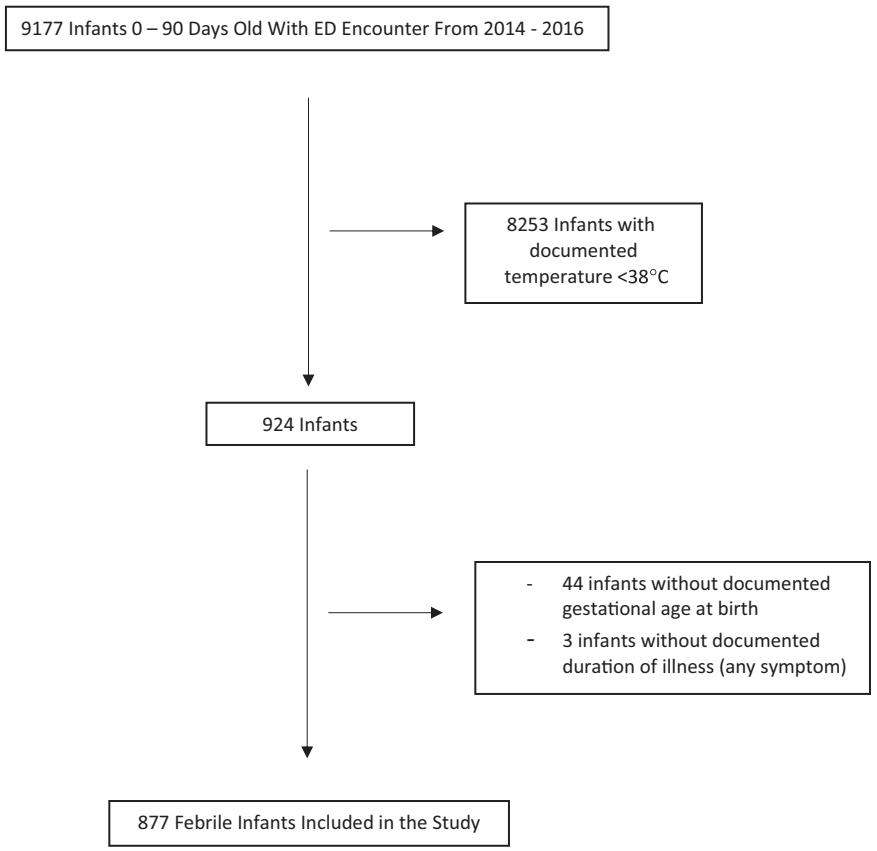

Figure. Flow diagram to indicate the included and excluded infants.

Table I. Factors used to classify febrile infants as low risk

History
Gestational age ≥37 wk
No prolonged nursery stay/NICU stay
No hospitalizations in the past 72 h
No history of perinatal antibiotics
No reported social concerns
No chronic disease
Examination
Well appearing
Laboratory
Peripheral WBC count >5000/mm^3 and <15 000/mm^3
Absolute band count ≤1500
<5 WBC/high power field in urine (unspun)
Negative leukocyte esterase in urine

NICU, neonatal intensive care unit; *WBC*, white blood cell.
All factors and thresholds must be met to satisfy low-risk designation.

Table V. Importance of variable

A. Regression model	Coefficients (95% CI)
Sex (female)	1.07 (0.33, 1.86)
Insurance (public)	−0.18 (−1.58, 1.44)
Chronic medical condition (yes)	1.10 (−0.43, 2.65)
Age	−.02 (−.04, −.01)
Gestational age	.01 (−0.24, 0.25)
Appearance (Ill)	2.85 (1.75, 4.05)
Maximum temperature	1.28 (0.69, 1.91)
Duration of illness	0.13 (−.03, 0.27)
Cough status (present)	−1.46 (−2.35, −0.63)
Urinary tract inflammation (present)	4.76 (3.8, 5.94)
B. Super learner model	**Improvement in AUC (95% CI)**
Sex	0 (0, .07)
Insurance	.01 (0, .05)
Chronic medical condition	0 (0, 0.05)
Age	0 (0, .05)
Gestational age	.04 (0, .08)
Appearance	.06 (0, .13)
Maximum temperature	0 (0, .04)
Duration of illness	.08 (.02, .14)
Cough status	0 (0, .04)
Urinary tract inflammation	.23 (.14, 0.32)