Open Forum Infectious Diseases

MAJOR ARTICLE

Derivation and Internal Validation of a Clinical Prediction Model for Diagnosis of Spotted Fever Group Rickettsioses in Northern Tanzania

Robert J. Williams,[1] Ben J. Brintz,[1,2] William L. Nicholson,[3] John A. Crump,[4,5,6,7,8] Ganga Moorthy,[5,9] Venace P. Maro,[7,8] Grace D. Kinabo,[7,8] James Ngocho,[7,8] Wilbrod Saganda,[10,11] Daniel T. Leung,[1,12,a] and Matthew P. Rubach[4,5,7,13,a]

[1]Division of Infectious Diseases, Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA, [2]Division of Epidemiology, Department of Internal Medicine, University of Utah, Salt Lake City, Utah, USA, [3]Rickettsial Zoonoses Branch, Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA, [4]Division of Infectious Diseases and International Health, Department of Medicine, Duke University, Durham, North Carolina, USA, [5]Duke Global Health Institute, Duke University, Durham, North Carolina, USA, [6]Centre for International Health, University of Otago, Dunedin, New Zealand, [7]Department of Internal Medicine, Kilimanjaro Christian Medical Centre, Moshi, Tanzania, [8]Department of Internal Medicine, Kilimanjaro Christian Medical University College, Moshi, Tanzania, [9]Division of Pediatric Infectious Diseases, Department of Pediatrics, Duke University, Durham, North Carolina, USA, [10]Mawenzi Regional Referral Hospital, Moshi, Tanzania, [11]Ministry of Health, Community Development, Gender, Elderly, and Children, Dodoma, Tanzania, [12]Division of Microbiology and Immunology, Department of Pathology, University of Utah, Salt Lake City, Utah, USA, and [13]Programme in Emerging Infectious Diseases, Duke–National University of Singapore Medical School, Singapore

Spotted fever group rickettsioses (SFGR) pose a global threat as emerging zoonotic infectious diseases; however, timely and cost-effective diagnostic tools are currently limited. We used data from 449 patients presenting to 2 hospitals in northern Tanzania between 2007 and 2008, of which 71 (15.8%) met criteria for acute SFGR based on ≥4-fold rise in antibody titers between acute and convalescent serum samples. We fit random forest classifiers incorporating clinical and demographic data from hospitalized febrile participants as well as Earth observation hydrometeorological predictors from the Kilimanjaro Region. In cross-validation, a prediction model with 10 clinical predictors achieved an area under the receiver operating characteristic curve of 0.65 (95% confidence interval, .48–.82). A combined prediction model with clinical, hydrometeorological, and environmental predictors (20 predictors total) did not significantly improve model performance. Novel strategies are needed to improve the diagnosis of acute SFGR, including the identification of diagnostic biomarkers that could enhance clinical prediction models.

Keywords. Spotted fever group rickettsioses; Rickettsia; Tanzania; Clinical Prediction.

Spotted fever group rickettsioses (SFGR) are a group of tick-borne illnesses caused by bacteria from the genus *Rickettsia* that include endemic, new, and emerging zoonotic infectious diseases with a worldwide distribution. In several African countries, SFGR have been identified as the infectious etiology in 5%–22% of febrile hospital admissions [1–3]. Prompt recognition and treatment of SFGR are important as multiple studies have shown that delay in initiation of tetracycline antimicrobials is associated with increased morbidity and mortality [4–6]. However, current diagnostic methods do not allow for timely and accurate diagnosis. The most sensitive reference standard diagnostic, a ≥4-fold rise in immunofluorescent antibody (IFA) titer between paired acute and convalescent serum samples, requires convalescent serum collection, and therefore by definition cannot establish the diagnosis at the time of presentation [7]. In the case of *Rickettsia africae* and *Rickettsia conorii*, SFGR of importance in African countries, seroconversion may not occur until 4 weeks after illness onset [8].

Rickettsiae are intracellular species and do not circulate extensively in the bloodstream, limiting the sensitivity of polymerase chain reaction on blood specimens to around 60% [7]. Laboratory values such as thrombocytopenia, hyponatremia, and elevated transaminases are supportive features, but cannot be relied on to guide early management as they are nonspecific findings and often within normal limits or only slightly above the reference range early in the course of illness [9]. Clinical diagnosis relying on the triad classically associated with SFGR—a history of a tick bite, rash, and fever—only occurs in a third of cases [9]. Incorporating tetracycline therapy into the empiric syndromic management of febrile illness in high-prevalence settings would not only exacerbate antimicrobial resistance, but also subject children to the risks of tetracycline therapy, including bone growth suppression and

Received 19 October 2024; editorial decision 10 February 2025; accepted 15 February 2025; published online 21 February 2025

[a]D. T. L. and M. P. R. contributed equally to this work.

Correspondence: Daniel T. Leung, MD, Division of Infectious Diseases, Department of Internal Medicine, University of Utah, 30 N Mario Capecchi Dr, Salt Lake City, UT 84112 (daniel.leung@utah.edu); Matthew P. Rubach, MD, Division of Infectious Diseases, Department of Medicine, Duke University, 181 Hanes HouseBox 90519, Durham, NC 27710 (matthew.rubach@duke.edu).

Open Forum Infectious Diseases®

https://doi.org/10.1093/ofid/ofaf100

permanent tooth discoloration [10, 11]. Thus, more accurate, timely, and cost-effective tools are needed for diagnosis of SFGR.

Clinical decision support systems (CDSSs) incorporating prediction models have the potential to improve management of infectious diseases. CDSSs have proven effective at enhancing therapeutic management and reducing unnecessary diagnostic tests in both high-income countries [12] and low- and middle-income countries [13–15]. Traditional predictive models generally incorporate clinical information that is obtained solely from the presenting patient. However, as with other zoonoses and tick-borne diseases, the range and incidence of SFGR have been associated with hydrometeorological-related environmental factors [16, 17]. Vegetation indices, including the Normalized Difference Vegetation Index (NDVI), the Enhanced Vegetation Index (EVI), and the Normalized Difference Water Index (NDWI), are hydrometeorological-related environment indices that correlate well with fluctuations in tick populations and have been used to estimate tick-borne disease incidence [18–22]. These indices are measured using Earth observation (EO) satellites, which allow for the monitoring of these environmental conditions over large areas. Incorporating location-specific parameters, such as hydrometeorological and environmental data, into a prediction model may increase diagnostic accuracy [15, 23, 24].

In this study, our primary aim was to derive and internally validate a clinical prediction rule for SFGR, with the goal of developing an accessible and cost-effective CDSS to assist clinicians in diagnosing SFGR. A secondary aim was to examine the impact of the inclusion of hydrometeorological and environmental data on prediction discrimination. To achieve this, we used data from a clinical study of febrile illness in an SFGR-endemic region in northern Tanzania [25], along with EO-hydrometeorological and environmental data.

METHODS

Study Design, Setting, and Data Source

For the derivation and validation of a prediction model, we used de-identified data from a study of participants presenting with febrile illness to 2 hospitals in the Kilimanjaro Region of northern Tanzania in 2007–2008 [26, 27]. Prior to data collection, the Kilimanjaro Region had a population of 1 380 000 in a mostly rural and semirural setting [28, 29]; Moshi, the administrative center of the region, had a population of approximately 144 000 [29]. The climate is characterized by a long rainy period (March–May) and a short rainy period (November–December). Febrile participants presenting to Kilimanjaro Christian Medical Centre (KCMC) or Mawenzi Regional Hospital (MRH) in Moshi, Tanzania, from September 2007 through August 2008 were eligible for enrollment. Complete study methods have been described elsewhere [26, 27]. KCMC is a tertiary care hospital serving several regions in northern Tanzania; at the time of the study KCMC had 458 inpatient beds. MRH, the regional hospital for Kilimanjaro, had 300 beds at the time of the study. Together, KCMC and MRH served as major providers of hospital-based care in the Moshi area.

For pediatric participants aged 2 months to 12 years, inclusion criteria were a history of fever in the past 48 hours or a measured axillary temperature ≥37.5°C or rectal temperature ≥38°C. For adolescent and adult participants (≥13 years of age), the inclusion criterion was an oral temperature ≥38°C on admission to the hospital. All participants required paired sera for inclusion in this analysis. Blood specimens were collected for a complete blood count (CBC) and serologic infectious disease diagnostics. Participants were also tested for human immunodeficiency virus (HIV) and malaria with rapid diagnostic testing. After obtaining informed consent, a trained study team member collected standardized demographic data, clinical history, and physical examination findings. Participants were asked to return 4–6 weeks after enrollment for collection of a convalescent serum sample.

Acute and convalescent serum samples collected for SFGR testing were sent to the Rickettsial Zoonoses Branch of the US Centers for Disease Control and Prevention (CDC). Serum samples were tested for SFGR by immunoglobulin G IFA to *R conorii* (Moroccan strain). SFGR was defined as a ≥4-fold increase in IFA titer to *R conorii* between acute and convalescent serum in a participant. Participants with a <4-fold rise in IFA titer to *R conorii* were considered to have non-SFGR febrile illness.

For each participant, a trained study team member collected standardized demographic data, clinical history, admission vital signs, and physical examination findings. As the presentation of infectious disease can differ between pediatric and adult participants, several of the clinical variables were only obtained for either pediatric or adult participants. For our clinical prediction modeling, we only included variables that were collected for both groups. Consequently, certain predictors associated with SFGR, such as the presence of rash, were excluded from our analysis because they were collected exclusively from adult participants. We included the following predictors in our analysis: age, heart rate, respiration rate, blood pressure, oxygen saturation, height, weight, body mass index, cough, diarrhea, emesis, hematochezia, dyspnea, seizures, crepitations, hepatomegaly, splenomegaly, pallor, lymphadenopathy, oral candidiasis, meningeal signs, HIV rapid diagnostic result, malaria rapid diagnostic result, and if the participant resided in a rural setting. We recorded clinical symptoms and physical examination findings as binary variables and age, vital signs, and CBC results as continuous variables.

Hydrometeorological-Related Environmental Data

We extracted hydrometeorological-related environmental data from the MODIS (Moderate Resolution Imaging Spectroradiometer) sensor, a part of the NASA Earth Observing System, for the Kilimanjaro Region, Tanzania. These data included the NDVI, the EVI, and the NDWI. Respectively, these indices reflect the vegetation health, vegetation structure, and water content in the landscape and have been shown to correlate with tick population abundance and tick-borne disease incidence [18–22, 30, 31]. Likewise, evapotranspiration and surface temperature have been found to be important predictors of active tick questing, tick survival, and pathogen transmission [22, 32–34]. More information regarding these predictors is provided in Supplementary Table S1. NDVI and EVI indicators are based on a 16-day time series composite image at 1 km × 1 km spatial resolution and were obtained from MODIS product MOD13A2. Surface temperatures, acquired from MODIS product MOD11A1, are daily measurements at 1 km × 1 km spatial resolution. Evapotranspiration is based on an 8-day time series composite image at 500 m × 500 m spatial resolution and obtained from MOD16A2GF. Finally, using MODI09A1, a 500 m × 500 m 8-day composite time series, we calculated NDWI from near-infrared (MODIS band 2) and short-wave infrared (MODIS band 6) reflectances [35, 36]. We consolidated all hydrometeorological-related environmental data within a uniform 16-day time series window; for shorter time series, we computed the mean for each 16-day window. For example, data from 1 January through 16 January 2007 constituted 1 window, followed by measurements from 17 January through 2 February 2007 in the subsequent window. To account for recent environmental patterns that may influence SFGR incidence, we lagged each 16-day times series at 1, 2, and 3 months. Finally, we aligned the lagged measurements with the admission dates of study participants. For instance, a participant presenting on 4 April 2007 would have data from the window containing 4 March (1-month lag), 4 February (2-month lag), and 4 January (3-month lag). Details of the exact 16-day time resolution windows used in our study are shown in Supplementary Table S2.

Statistical Analysis and Modeling

To compare acute SFGR versus non-SFGR febrile illness groups on univariate analysis, we used the Wilcoxon rank-sum test due to the nonnormal distribution of age data. We used Pearson χ^2 test and Fisher exact test for categorical variables. We used the random forest algorithm to fit a model to predict risk of participants having acute SFGR versus non-SFGR febrile illness. Random forests are a machine learning algorithm that constructs a multitude of decision trees and averages over them to obtain a prediction robust to nonlinearities and interactions between covariates; random forests algorithms have been widely applied to biomedical sciences for both classification and regression [37, 38].

We excluded predictors with highly skewed binary predictors (predictors with 95% or more values concentrated as either 0 or 1) from analysis. For the remaining predictors, we imputed

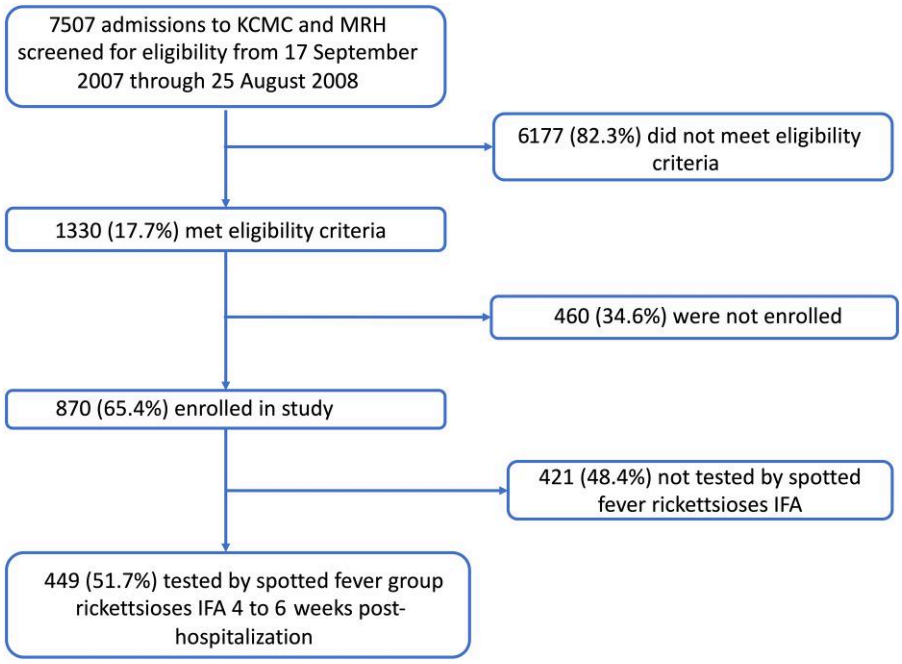

Figure 1. Study flow diagram. Screening and enrollment of patients hospitalized at Kilimanjaro Christian Medical Centre and Mawenzi Regional Hospital. Abbreviations: IFA, immunofluorescence assay; KCMC, Kilimanjaro Christian Medical Centre; MRH, Mawenzi Regional Hospital.

Table 1. Clinical Characteristics and Complete Blood Count Results for Febrile Participants With or Without Spotted Fever Group Rickettsioses, Northern Tanzania, 2007–2008

Characteristic	Overall (n = 449)	Non-SFGR Febrile Illness (n = 378)	Acute SFGR (n = 71)	*P* Value[a]
Age, y	9 (2–32)	8 (1–31)	24 (4–39)	.003
Rural residency	211 (47%)	168 (45%)	42 (59%)	.03
Admission temperature, °C	38.40 (38.00–39.10)	38.40 (38.00–39.00)	38.60 (38.00–39.20)	.4
Heart rate, beats/min	112 (97–131)	112 (98–132)	108 (96–124)	.2
Respiration rate, breaths/min	32 (24–44)	32 (26–45)	28 (24–35)	<.001
Oxygen saturation, %	96.0 (94.0–98.0)	96.0 (94.0–98.0)	97.0 (94.0–98.0)	.5
Systolic blood pressure, mm Hg	108 (100–120)	108 (100–120)	108 (100–117)	.4
Diastolic blood pressure, mm Hg	68 (60–74)	68 (60–74)	67 (60–73)	.7
BMI, kg/m^2	17.2 (14.4–21.6)	17.0 (14.3–21.0)	19.2 (15.6–22.5)	.06
Weight, kg	22 (10–55)	20 (10–55)	49 (14–61)	.01
Height, m	1.25 (0.82–1.62)	1.19 (0.81–1.62)	1.54 (1.05–1.63)	.01
Cough	293 (65%)	252 (67%)	41 (58%)	.1
Oral *Candida*	43 (9.6%)	39 (10%)	4 (5.6%)	.2
Crepitations	196 (44%)	165 (44%)	31 (44%)	>.9
Hepatomegaly	34 (7.6%)	27 (7.2%)	7 (9.9%)	.4
Diarrhea	90 (20%)	77 (20%)	13 (18%)	.7
Emesis	132 (29%)	109 (29%)	23 (32%)	.5
Dyspnea	152 (34%)	132 (35%)	20 (28%)	.3
Seizure	41 (9.2%)	37 (9.8%)	4 (5.6%)	.3
Pallor	32 (7.2%)	27 (7.2%)	5 (7.0%)	>.9
Lymphadenopathy	39 (8.7%)	34 (9.0%)	5 (7.0%)	.6
Rapid malaria diagnostic	19 (4.2%)	16 (4.2%)	3 (4.2%)	>.9
Rapid HIV diagnostic				.13
Positive	103 (23%)	92 (24%)	11 (15%)	
Indeterminant	8 (1.8%)	8 (2.1%)	0 (0%)	
WBC count, K/μL	9 (6–13)	9 (6–13)	7 (5–13)	.1
RBC count, M/μL	4.28 (3.66–4.79)	4.24 (3.59–4.79)	4.43 (3.97–4.82)	.2
Hemoglobin, g/dL	10.80 (9.00–12.40)	10.70 (9.00–12.28)	11.30 (9.95–12.75)	.1
Hematocrit, %	32 (28–37)	32 (27–36)	33 (30–37)	.08
Platelet count, K/μL	267 (161–391)	275 (164–404)	225 (130–360)	.07
Neutrophils, %	62 (42–76)	62 (42–76)	65 (45–79)	.3
Lymphocytes, %	25 (16–43)	26 (16–44)	22 (13–42)	.2
Monocytes, %	8.5 (5.9–11.5)	8.5 (5.9–11.6)	7.8 (5.3–10.5)	.3
Eosinophils, %	0.20 (0.00–0.90)	0.20 (0.00–1.00)	0.20 (0.00–0.70)	.3
Basophils, %	0.80 (0.40–1.20)	0.80 (0.40–1.20)	0.80 (0.45–1.10)	>.9

Data are presented as No. (%) or median (interquartile range).

Abbreviations: BMI, body mass index; HIV, human immunodeficiency virus; RBC, red blood cell; SFGR, spotted fever group rickettsioses; WBC, white blood cell.

[a]Pearson χ^2 test; Wilcoxon rank-sum test; Fisher exact test.

missing data using the "missRanger" package in R. When determining which predictors to include in our analysis, we fit 2 distinct models (1 utilizing solely EO-hydrometeorological and environmental data and another incorporating clinical and demographic data) and assessed each separately. We used the "permimp" package in R to assess the variable importance using permutation-based methods. This method involves systematically shuffling or permuting the values of individual predictors to evaluate their impact on performance. Next, we identified the top 10 predictors from each model based on their respective permuted importance scores. We included these predictors in our final analysis.

To assess predictive performance for each random forest model, we used repeated cross-validation using 80% training/20% testing splits with 100 iterations. In each iteration, we trained models on 80% of the data, made predictions on the 20% test set, and obtained measures of performance. We determined overall model performance by averaging the area under the receiver operating characteristic curve (AUC) and confidence intervals (CIs) across the 100 iterations. To determine statistical significance in the AUC between models, we used a bootstrap method over 100 iterations within cross-validation. The bootstrap method involves resampling the data with replacement multiple times to create bootstrap samples. For each bootstrap sample, we generated the receiver operating characteristic and computed the difference in AUC between the curves. We completed all analyses using R version 4.2.0, and model development/validation was completed in

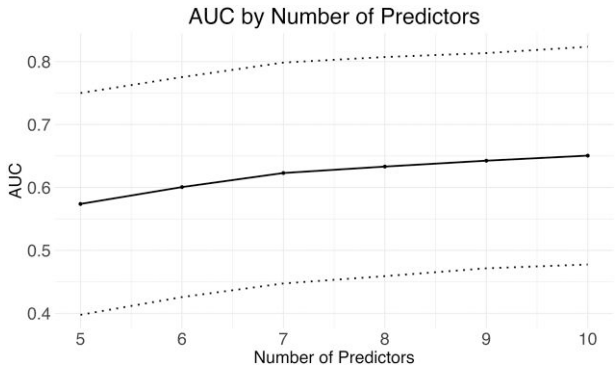

Figure 2. Average area under the receiver operating characteristic curve (AUC; solid line) and 95% confidence intervals (dotted lines) from cross-validation (100 iterations) for each model by number of predictors included in the model.

accordance with the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) checklist (Supplementary Table S3).

Patient Consent Statement

The primary study was approved by the Kilimanjaro Christian Medical University College Health Research Ethics Committee, the Tanzania National Institutes for Medical Research National Health Research Ethics Coordinating Committee, and the institutional review boards (IRBs) of Duke University Medical Center and the CDC. The secondary data analysis was reviewed by the IRB of the University of Utah and determined to be exempt (IRB_00164810). All minors had written informed consent given from a parent or guardian, and all adult participants provided their own written informed consent.

RESULTS

Of the 870 participants enrolled in the study, 449 (51.6%) underwent follow-up for the collection of convalescent serum. Of these 449 participants, 71 (15.8%) met criteria for acute SFGR (Figure 1). To address the potential for ascertainment bias, where individuals living closer to the study sites might be more likely to complete follow-up for convalescent serum collection, we compared the demographic and location characteristics of those who completed follow-up versus those who were lost to follow-up. Our analysis did not reveal statistically significant differences between the 2 groups (Supplementary Table S4). Mapping the locations of participants who completed follow-up versus those who were lost to follow-up showed no discernible patterns (Supplementary Figure S1).

We excluded the highly skewed predictors hematochezia, meningeal signs, and the malaria rapid diagnostic from our analysis. We found statistically significant differences in several clinical variables including vital signs, clinical symptoms, and laboratory results between acute SFGR and non-SGFR febrile illness (Table 1). Overall, acute SFGR participants were older (median age, 24 vs 8 years; $P = .003$) with significantly higher height and weight. Participants with acute SFGR had a lower respiratory rate than non-SFGR febrile illness participants (median, 28 vs 32 breaths per minute; $P < .001$) and were more likely reside in a rural setting (59% vs 45%; $P = .025$). There were no significant differences in the CBC results between the 2 groups, although platelet count and hematocrit had a P value of .07 and .08, respectively.

We also found several significantly different hydrometeorological and environmental predictors between participants with acute SFGR and those with non-SFGR febrile illness (Supplementary Table S5). Acute SFGR was associated with higher recent temperatures—significantly higher nighttime mean (odds ratio [OR], 1.17 [95% CI, 1.03–1.34]; $P = .01$) and nighttime maximum (OR, 1.16 [95% CI, 1.02–1.32]; $P = .02$) temperature at 1-month lag and daytime minimum (OR, 1.08 [95% CI, 1.00–1.17]; $P = .03$) temperature at 1-month lag—as well as lower minimum NDWI (a proxy for plant water stress, where lower values signify increased plant stress) at 1- and 2-month lags (OR, 0.20 [95% CI, .05–.71], $P = .01$; OR, 0.09 [95% CI, .02–.38], $P < .001$, respectively). Additionally, acute SFGR participants had lower minimum evapotranspiration rates at a 2-month lag (OR, 0.78 [95% CI, .64–.96]; $P = .02$) and higher maximum evapotranspiration rates at a 3-month lag (OR, 1.06 [95% CI, 1.01–1.12]; $P = .03$).

Model Performance With Parsimonious Variable Selection

To create a model that would not require overly burdensome input from the clinician, we limited our model assessment to the best-performing 5–10 clinical predictors (measured by change in AUC within cross-validation). Starting with 5 clinical predictors, we used a random forest classifier to analyze the improvement in model performance with each additional clinical variable included (Figure 2). Model performance improved with the inclusion of each predictor; a model with the best-performing 10 clinical predictors had an AUC of 0.65 (95% CI, .48–.82). This model would include the following clinical predictors: respiration rate, admission temperature, oxygen saturation, cough, HIV status, platelet count, hemoglobin, red blood cell count, basophil count, and hematocrit. Using the Youden index, this model had a sensitivity of 91%, a specificity of 32%, a negative predictive value (NPV) of 63%, and a positive predictive value (PPV) of 71%.

The Addition of Hydrometeorological and Environmental Predictors Did Not Significantly Improve Model Performance

We then assessed model performance using only the 10 best-performing hydrometeorological and environmental predictors (AUC, 0.61 [95% CI, .47–.77]). Next, we fit a model within cross-validation using the 10 best-performing clinical and the 10 best-performing hydrometeorological and environmental

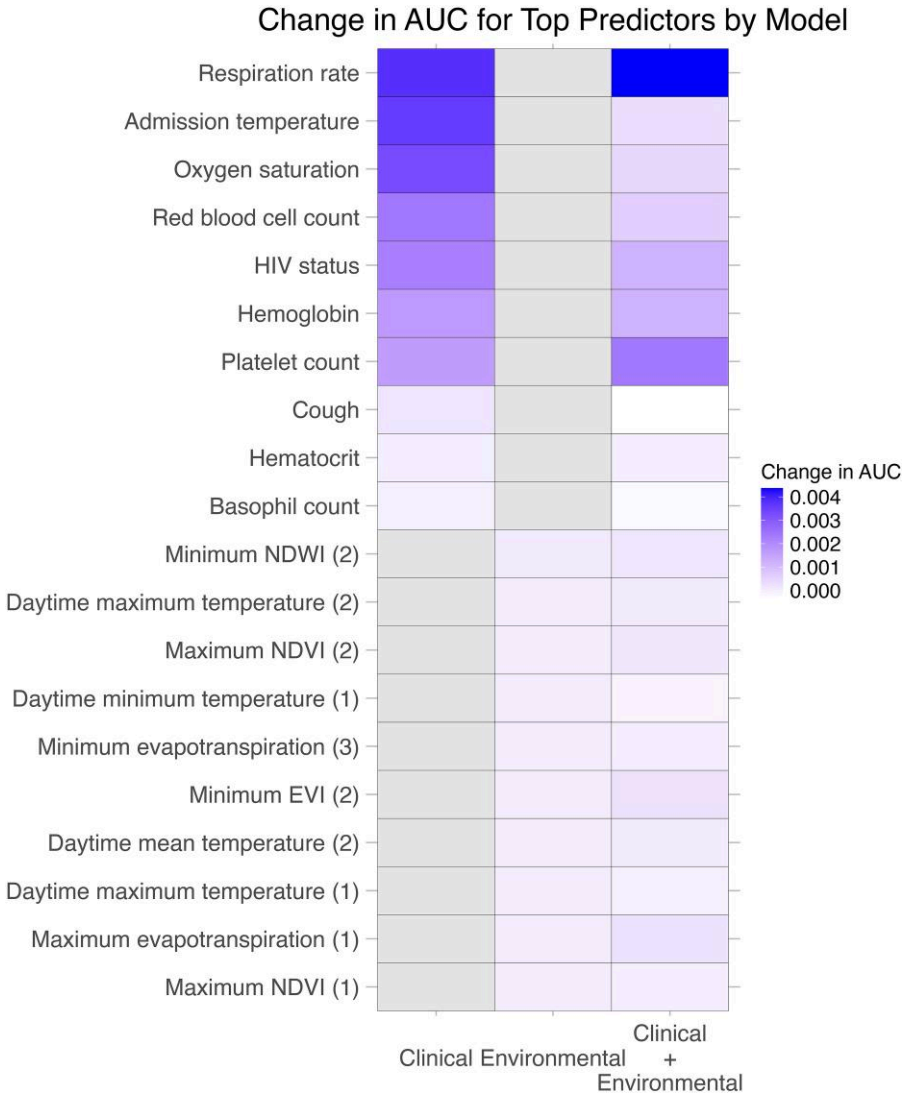

Figure 3. Heatmap of the top-performing predictors for each model. The environmental and clinical models include the 10 best-performing predictors and the clinical + environmental models include the both the top-performing environmental and clinical predictors. Performance was determined by the change in the area under the receiver operating characteristic curve after permutation. Values in parentheses indicate the number of lag-months. Abbreviations: AUC, area under the receiver operating characteristic curve; EVI, enhanced vegetation index; HIV, human immunodeficiency virus; NDVI, normalized difference vegetative index; NDWI, normalized difference water index.

predictors and assessed how this model compared to a model with only the 10 best-performing clinical predictors. By combining clinical, hydrometeorological, and environmental predictors, the AUC improved to 0.72 (95% CI, .57–.86), though this improvement was not statistically significant (median $P = .3$, 12% of P values $<.05$). Figure 3 shows a heat map of the best-performing predictors by change in AUC for each model.

DISCUSSION

Using data from a 2-center clinical study of febrile illness from northern Tanzania, we show the derivation and cross-validation of a diagnostic prediction model for SFGR, a febrile illness lacking an accurate laboratory diagnostic during acute illness. A parsimonious model with 10 clinical predictors including 3 vital signs, 5 results from CBC, 1 symptom, and HIV status had an AUC of 0.65 (95% CI, .48–.82). The optimum threshold for the parsimonious clinical model resulted in a sensitivity of 91%, a specificity of 32%, PPV of 71%, and NPV of 63%. A sensitivity of 91% indicates that this model would be useful for identifying patients who should receive empiric tetracycline therapy, ensuring that the majority of true SFGR cases are promptly treated. Models with higher specificities had much lower sensitivities (ie, high potential for false-negative results); a model with a specificity of 80% and 90% had a sensitivity of 43% and 23%, respectively. While our predictive model offers an improvement over existing clinical prediction models published for SFGR [39], given the suboptimal

performance of these models, there is a critical need for the exploration and validation of specific biomarkers that could enhance diagnostic precision of SFGR clinical prediction models and contribute to more effective management strategies in regions affected by this potentially fatal bacterial disease. We propose assessing candidate biomarkers such as proteins, peptides, and nucleic acids, including routine clinical analytes (eg, fibrinogen) and vetted translational research assays (eg, endothelial activation markers such as angiopoietein-2) that are relevant to SFGR's known pathophysiology of endothelial infection and inflammation.

The use of satellite imagery has been shown to facilitate modeling of population dynamics of ticks [19, 40], the vector for transmission of SFGR. In our prediction modeling, EO-hydrometeorological and environmental predictors did not improve the AUC of our internally validated model in a statistically significant manner. Notably, NDVI, EVI, and NDWI undergo variations influenced by extreme weather events, land use changes, and other anthropogenic impacts [41–43]. Therefore, their dynamic nature would likely impose limitations on their applicability within models spanning several years. In this case, incorporating EO-variables resulted in only a modest and nonsignificant improvement in prediction model performance and their inclusion may not be worth the associated costs (eg, time, funding, computational skill, and infrastructure).

Similar to what has been reported in the literature, we found that a higher body temperature on admission was one of the top clinical predictors associated with acute SFGR infection [44, 45]. CBC results including thrombocytopenia, leukopenia, and lymphopenia have been shown to be significantly different between acute SFGR and non-SFGR febrile illness. In our model, platelet count, hemoglobin, and hematocrit were important CBC predictors; however, lymphopenia and leukopenia were not. Our model also found that respiration rate and oxygen saturation contribute to discrimination between acute SFGR and non-SFGR febrile illness. While our dataset did not have complete information on classic SFGR symptoms including headache, myalgias, and rash [46], these symptoms have been found to occur in similar proportions among acute SFGR and non-SFGR febrile illness [9, 47, 48].

A major limitation of our study is lack of external validation. The lack of external validation, coupled with the fact that we constructed our model using data from a single endemic region in northern Tanzania over a 1-year duration, potentially hinders the model's generalizability to a broader population. Given the intricate interplay between vector and host, the hydrometeorological-related environmental indices that affect SFGR may vary between regions or over time. Furthermore, this model was constructed on data from hospitalized patients, which represent the most severe presentations of disease. Additionally, while we used a robust case definition for acute SFGR based on ≥4-fold rise in antibody titer between acute and convalescent IFA, the requirement for convalescent sample collection could have introduced selection bias to our analysis. However, we did not find any significant differences when comparing the demographic and location characteristics of those who completed follow-up and those who were lost to follow-up. Finally, our model lacks covariates on rash or eschar as well as other laboratory values that have shown to be correlated with SFGR infection (eg, sodium [44, 49], transaminases [44, 49–51], lactic dehydrogenase [49, 51], and fibrinogen [52]); inclusion of these laboratory parameters may have improved model performance.

In conclusion, using an existing dataset of acute SFGR and non-SFGR febrile illness in Tanzania, we demonstrated the derivation and internal validation of a clinical prediction model for SFGR, adding to the paucity of existing diagnostic clinical prediction models for this emerging zoonotic pathogen. Further research should expand upon this analysis by incorporating data from additional febrile cohorts, exploring the inclusion of clinical biomarkers, and assessing the performance of this model in diverse settings endemic for SFGR to ensure its generalizability.

Supplementary Data

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

Notes

Acknowledgments. We acknowledge the Hubert-Yeargan Center for Global Health at Duke University for critical infrastructure support for the Kilimanjaro Christian Medical Centre–Duke University Collaboration.

Disclaimer. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the US Centers for Disease Control and Prevention (CDC). Use of trade names and commercial sources is for identification only and does not imply endorsement by the US Department of Health and Human Services or the CDC.

Financial support. This research was supported by the US National Institutes of Health (grant numbers U01 AI062563 to J. A. C., K24 AI166087 to D. T. L., and R38 HL143605 to R. J. W. through Utah Stimulating Access to Research in Residency [StARR]).

Potential conflicts of interest. All authors: No potential conflicts.

References

1. Crump JA, Morrissey AB, Nicholson WL, et al. Etiology of severe non-malaria febrile illness in northern Tanzania: a prospective cohort study. PLoS Negl Trop Dis **2013**; 7:e2324.
2. Ndip LM, Fokam EB, Bouyer DH, et al. Detection of *Rickettsia africae* in patients and ticks along the coastal region of Cameroon. Am J Trop Med Hyg **2004**; 71: 363–6.
3. Maina AN, Farris CM, Odhiambo A, et al. Q fever, scrub typhus, and rickettsial diseases in children, Kenya, 2011–2012. Emerg Infect Dis **2016**; 22:883–6.
4. Dalton MJ, Clarke MJ, Holman RC, et al. National surveillance for Rocky Mountain spotted fever, 1981–1992: epidemiologic summary and evaluation of risk factors for fatal outcome. Am J Trop Med Hyg **1995**; 52:405–13.
5. Kirkland KB, Wilkinson WE, Sexton DJ. Therapeutic delay and mortality in cases of Rocky Mountain spotted fever. Clin Infect Dis **1995**; 20:1118–21.

6. Regan JJ, Traeger MS, Humpherys D, et al. Risk factors for fatal outcome from Rocky Mountain spotted fever in a highly endemic area—Arizona, 2002–2011. Clin Infect Dis **2015**; 60:1659–66.
7. Biggs HM, Behravesh CB, Bradley KK, et al. Diagnosis and management of tick-borne rickettsial diseases: Rocky Mountain spotted fever and other spotted fever group rickettsioses, ehrlichioses, and anaplasmosis—United States. MMWR Recomm Rep **2016**; 65:1–44.
8. Fournier PE, Jensenius M, Laferl H, Vene S, Raoult D. Kinetics of antibody responses in *Rickettsia africae* and *Rickettsia conorii* infections. Clin Diagn Lab Immunol **2002**; 9:324–8.
9. Traeger MS, Regan JJ, Humpherys D, et al. Rocky Mountain spotted fever characterization and comparison to similar illnesses in a highly endemic area—Arizona, 2002–2011. Clin Infect Dis **2015**; 60:1650–8.
10. Cross R, Ling C, Day NP, McGready R, Paris DH. Revisiting doxycycline in pregnancy and early childhood–time to rebuild its reputation? Expert Opin Drug Saf **2016**; 15:367–82.
11. Wormser GP, Wormser RP, Strle F, Myers R, Cunha BA. How safe is doxycycline for young children or for pregnant or breastfeeding women? Diagn Microbiol Infect Dis **2019**; 93:238–42.
12. Bright TJ, Wong A, Dhurjati R, et al. Effect of clinical decision-support systems. Ann Intern Med **2012**; 157:29–43.
13. Bilal S, Nelson E, Meisner L, et al. Evaluation of standard and mobile health–supported clinical diagnostic tools for assessing dehydration in patients with diarrhea in rural Bangladesh. Am J Trop Med Hyg **2018**; 99:171–9.
14. Tuon FF, Gasparetto J, Wollmann LC, Moraes TP. Mobile health application to assist doctors in antibiotic prescription—an approach for antibiotic stewardship. Braz J Infect Dis **2017**; 21:660–4.
15. Garbern SC, Nelson EJ, Nasrin S, et al. External validation of a mobile clinical decision support system for diarrhea etiology prediction in children: a multicenter study in Bangladesh and Mali. Elife **2022**; 11:e72294.
16. Kerins JL, Dorevitch S, Dworkin MS. Spotted fever group rickettsioses (SFGR): weather and incidence in Illinois. Epidemiol Infect **2017**; 145:2466–72.
17. Zhang YY, Sun YQ, Chen JJ, et al. Mapping the global distribution of spotted fever group rickettsiae: a systematic review with modelling analysis. Lancet Digit Health **2023**; 5:e5–15.
18. Estrada-Peña A. Distribution, abundance, and habitat preferences of *Ixodes ricinus* (Acari: Ixodidae) in northern Spain. J Med Entomol **2001**; 38:361–70.
19. Randolph SE. Ticks and tick-borne disease systems in space and from space. Adv Parasitol **2000**; 47:217–43.
20. Estrada-Peña A. Increasing habitat suitability in the United States for the tick that transmits Lyme disease: a remote sensing approach. Environ Health Perspect **2002**; 110:635–40.
21. Polo G, Labruna MB, Ferreira F. Satellite hyperspectral imagery to support tick-borne infectious diseases surveillance. PLoS One **2015**; 10:e0143736.
22. Garcia-Martí I, Zurita-Milla R, van Vliet AJH, Takken W. Modelling and mapping tick dynamics using volunteered observations. Int J Health Geogr **2017**; 16:41.
23. Fine AM, Brownstein JS, Nigrovic LE, et al. Integrating spatial epidemiology into a decision model for evaluation of facial palsy in children. Arch Pediatr Adolesc Med **2011**; 165:61–7.
24. Nelson EJ, Khan AI, Keita AM, et al. Improving antibiotic stewardship for diarrheal disease with probability-based electronic clinical decision support: a randomized crossover trial. JAMA Pediatr **2022**; 176:973–9.
25. Pisharody S, Rubach MP, Carugati M, et al. Incidence estimates of acute Q fever and spotted fever group rickettsioses, Kilimanjaro, Tanzania, from 2007 to 2008 and from 2012 to 2014. Am J Trop Med Hyg **2021**; 106:494–503.
26. Crump JA, Ramadhani HO, Morrissey AB, et al. Invasive bacterial and fungal infections among hospitalized HIV-infected and HIV-uninfected adults and adolescents in northern Tanzania. Clin Infect Dis **2011**; 52:341–8.
27. Crump JA, Ramadhani HO, Morrissey AB, et al. Invasive bacterial and fungal infections among hospitalized HIV-infected and HIV-uninfected children and infants in northern Tanzania. Trop Med Int Health **2011**; 16:830–7.
28. Biggs HM, Hertz JT, Munishi OM, et al. Estimating leptospirosis incidence using hospital-based surveillance and a population-based health care utilization survey in Tanzania. PLoS Negl Trop Dis **2013**; 7:e2589.
29. Tanzania National Bureau of Statistics. Tanzania census 2002: analytical report. Dar es Salaam: Tanzania National Bureau of Statistics, **2006**.
30. Spare M, Boorgula GD, Thomson D, et al. Surveillance of host-seeking ticks in the Flint Hills region (USA) and associations with environmental determinants. Parasitologia **2021**; 1:137–47.
31. Westra S, Goldberg MS, Didan K. The association between the incidence of Lyme disease in the USA and indicators of greenness and land cover. Curr Res Parasitol Vector Borne Dis **2023**; 4:100132.
32. Ogden NH, Ben Beard C, Ginsberg HS, Tsao JI. Possible effects of climate change on ixodid ticks and the pathogens they transmit: predictions and observations. J Med Entomol **2021**; 58:1536–45.
33. Bouchard C, Dibernardo A, Koffi J, Wood H, Leighton PA, Lindsay LR. Increased risk of tick-borne diseases with climate and environmental changes. Can Commun Dis Rep **2019**; 45:83–9.
34. Ruiz-Fons F, Fernández-de-Mera IG, Acevedo P, Gortázar C, de la Fuente J. Factors driving the abundance of *Ixodes ricinus* ticks and the prevalence of zoonotic *I. ricinus*–borne pathogens in natural foci. Appl Environ Microbiol **2012**; 78:2669–76.
35. Chen D, Huang J, Jackson TJ. Vegetation water content estimation for corn and soybeans using spectral indices derived from MODIS near- and short-wave infrared bands. Remote Sens Environ **2005**; 98:225–36.
36. Jackson TJ, Chen D, Cosh M, et al. Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. Remote Sens Environ **2004**; 92:475–82.
37. Peng SY, Chuang YC, Kang TW, Tseng KH. Random forest can predict 30-day mortality of spontaneous intracerebral hemorrhage with remarkable discrimination. Eur J Neurol **2010**; 17:945–50.
38. Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review. Front Aging Neurosci **2017**; 9:329.
39. Lopez DM, de Mello FL, Giordano Dias CM, et al. Evaluating the surveillance system for spotted fever in Brazil using machine-learning techniques. Front Public Health **2017**; 5:323.
40. Estrada-Peña A. Geostatistics and remote sensing using NOAA-AVHRR satellite imagery as predictive tools in tick distribution and habitat suitability estimations for *Boophilus microplus* (Acari: Ixodidae) in South America. National Oceanographic and Atmosphere Administration–advanced very high resolution radiometer. Vet Parasitol **1999**; 81:73–82.
41. Aburas MM, Abdullah SH, Ramli MF, Ash'aari ZH. Measuring land cover change in Seremban, Malaysia using NDVI index. Procedia Environ Sci **2015**; 30:238–43.
42. Lunetta RS, Knight JF, Ediriwickrema J, Lyon JG, Worthy LD. Land-cover change detection using multi-temporal MODIS NDVI data, eds. Geospatial information handbook for water resources and watershed management, volume II. Boca Raton, FL: CRC Press, **2022**:65–88.
43. Wang G, Peng W, Zhang L, Zhang J, Xiang J. Vegetation EVI changes and response to natural factors and human activities based on geographically and temporally weighted regression. Glob Ecol Conserv **2023**; 45:e02531.
44. Buckingham SC, Marshall GS, Schutze GE, et al. Clinical and laboratory features, hospital course, and outcome of Rocky Mountain spotted fever in children. J Pediatr **2007**; 150:180–4.e1.
45. Silva-Ramos CR, Hidalgo M, Faccini-Martínez ÁA. Clinical, epidemiological, and laboratory features of *Rickettsia parkeri* rickettsiosis: a systematic review. Ticks Tick Borne Dis **2021**; 12:101734.
46. Cohen R, Finn T, Babushkin F, et al. Spotted fever group rickettsioses in Israel, 2010–2019. Emerg Infect Dis **2021**; 27:2117–26.
47. Prabhu M, Nicholson WL, Roche AJ, et al. Q fever, spotted fever group, and typhus group rickettsioses among hospitalized febrile patients in northern Tanzania. Clin Infect Dis **2011**; 53:e8–15.
48. Faruque LI, Zaman RU, Gurley ES, et al. Prevalence and clinical presentation of *Rickettsia*, *Coxiella*, *Leptospira*, *Bartonella* and chikungunya virus infections among hospital-based febrile patients from December 2008 to November 2009 in Bangladesh. BMC Infect Dis **2017**; 17:141.
49. Antón E, Font B, Muñoz T, Sanfeliu I, Segura F. Clinical and laboratory characteristics of 144 patients with Mediterranean spotted fever. Eur J Clin Microbiol Infect Dis **2003**; 22:126–8.
50. Mahara F. Japanese spotted fever: report of 31 cases and review of the literature. Emerg Infect Dis **1997**; 3:105–11.
51. Miyashima Y, Iwamuro M, Shibata M, et al. Prediction of disseminated intravascular coagulation by liver function tests in patients with Japanese spotted fever. Intern Med **2018**; 57:197–202.
52. Rao AK, Schapira M, Clements ML, et al. A prospective study of platelets and plasma proteolytic systems during the early stages of Rocky Mountain spotted fever. N Engl J Med **1988**; 318:1021–8.