

Preplanned Studies

Epidemiological and Clinical Features of Patients with Scrub Typhus — Guangzhou City, Guangdong Province, China, 2012–2018

Yuehong Wei^{1,8}; Xiugang Guan^{2,8}; Shixia Zhou^{2,3}; Anran Zhang^{2,4}; Qingbin Lu⁵; Ziwei Zhou²; Jinjin Chen²; Haiyang Zhang²; Yang Ji²; Baogui Jiang²; Yang Yang⁶; Zhicong Yang^{1,7}; Hao Li^{2,7}; Liqun Fang^{2,3,7}; Wei Liu^{2,5,7}

Summary

What is already known about this topic?

Scrub typhus (ST) causes public health challenges in the “tsutsugamushi triangle” in the Asia-Pacific area greater than 13 million square kilometers, affecting an estimated one million people each year.

What is added by this report?

A retrospective study based on 4,501 hospitalized patients with ST in Guangzhou City, China, described the epidemiological and clinical characteristics, laboratory findings of ST, and determined the related factors and a predictive model for severe disease.

What are the implications for public health practice?

The current study provided updated knowledge that might enable public health policymakers to formulate appropriate measures to prevent ST and medical workers to perform targeted management to recognize and treat severe ST patients.

Scrub typhus (ST) is a vector-borne rickettsial zoonosis caused by the organism *Orientia tsutsugamushi*, transmitted to humans by the bite of the larva of trombiculid mites (1–2). Up to the end of 2016 in China, all 31 provincial-level administrative divisions (PLADs) had recorded human cases, with the incidence increased by over 16-fold compared to that of 2006, with dramatic geographic expansions in both rural and urban areas and diversified seasonal patterns (3). Whether these profound epidemiological features might be related to clinical aspects remains obscure. This study described the epidemiological features and clinical outcomes of ST patients and assessed early predictors of severe disease by retrospective medical record review of 4,501 ST patients in 69 hospitals located throughout all 11 districts of Guangzhou City of southern China from January 2012 to December 2018. Severe ST was found to be associated with the decreased levels of albumin (ALB) and platelet (PLT)

count and increased levels of serum creatinine (CREA) and total bilirubin (TBIL) in the blood, as well as the occurrence of dyspnea for ST patients, with estimated relative contributions more than 10% in the final boosted regression trees models, which could be helpful for the recognition and treatment of severe ST in the early clinical management.

According to the *National Scrub Typhus Control and Prevention Guideline (2009)* issued by China CDC, a total of 4,501 patients with clinically diagnosed and laboratory-confirmed ST were included in the study (Supplementary Figure S1, available in <http://weekly.chinacdc.cn/>). Demographic information, medical history, and exposure history were obtained by interviewing patients or their guardians. Clinical data which comprised of date of disease onset (the day when clinical signs or symptoms were noticed), signs and symptoms, laboratory measurements, imaging findings, and treatment regimens, were retrieved from medical records and collected by EpiData software (version 3.1; The EpiData Association; Odense, Denmark).

Patients who had ever developed any severe complications [multiple organ dysfunction syndromes (MODS), shock, or requiring intensive care unit (ICU) admission] during the hospitalization were defined as severe cases, and the remaining patients were defined as mild cases.

Continuous variables were summarized as medians and interquartile-range (IQR). Categorical variables were summarized as frequencies and proportions. Chi-squared test, Fisher’s exact test, or nonparametric test, were used as appropriate to determine the difference between groups. Multivariate logistic regression analysis and boosted regression trees (BRT) model were performed to examine the effect of clinical manifestations and laboratory indicators on illness severity and to attain an early prediction of severe ST. The area under the receiver operating characteristic curve (AUC) was calculated to evaluate the predictive power of the BRT model. The details on variable selection and the modeling analyses of multivariate

logistic regression and BRT model were shown in the Supplementary Material and Supplementary Figure S2 (available in <http://weekly.chinacdc.cn/>). A 2-sided P value of <0.05 is considered as statistically significant. All statistical analyses were performed using R software (version 3.6.2; R Foundation; Vienna, Austria).

A total of 5,354 hospitalized patients with ST from 69 hospitals were included for medical record screening, from which 115 patients had incomplete information, 17 had other infectious diseases, and 721 had vague diagnosis were excluded, which resulted in 4,501 patients with clinically diagnosed ST being included in the final analysis (Supplementary Figure S1), accounting for 76.1% of the total number of cases in Guangzhou City reported to the national surveillance system of China CDC (Figure 1A). The median (IQR) age of the patients was 57 (46–65) years, which gradually increased from 54 (42–62) in 2012 to 59 (49–66) in 2018 ($P<0.001$). A slightly higher proportion of female patients (54.4%) was observed, but no significant difference was shown across the study years ($P=0.882$) (Table 1). Overall, 72.6% of the patients resided in rural areas, with the proportion increased from 60.4% in 2012 to 80.0% in 2018, and farmers and retirees were the main occupations (Table 1). When geographically displayed,

cropland and forest regions were related to higher case incidence, such as in Conghua, Zengcheng, Huadu, and Nansha districts (Figure 1B). Approximately 88.4% of the patients occurred between May and October. The median time from symptom onset to hospital admission was 7 (IQR: 5–9) days and the median length of hospital stay was 7 (IQR: 6–10) days. Almost all patients (93.1%) had received antibiotics treatment (including doxycycline, azithromycin, chloramphenicol, or levofloxacin), in combination with the supportive therapy (1.5%).

Severe illness was determined from 366 patients, with a disease severity rate (DSR) of 8.1% (95% CI: 7.3%–8.9%), which did not differ across the studied years (Table 1, Figure 2A). Overall, 53 of the severe cases died, with a case fatality rate of 1.2% (95% CI: 0.9%–1.5%), which differed during the study years (Table 1). Patients with severe disease comprised 265 (5.9%) who developed MODS, 196 (4.4%) who were admitted to the ICU, and 139 (3.1%) who developed shock. When looking closely, children aged ≤ 14 years had a DSR of 11.0%, higher than that of the next age group of 15–29 years old (2.0%); thereafter, an age-dependent increase of DSR was observed with patients ≥ 80 years having an 8-fold risk of DSR compared with the 15–29 years age group (Figure 2B). The DSR was

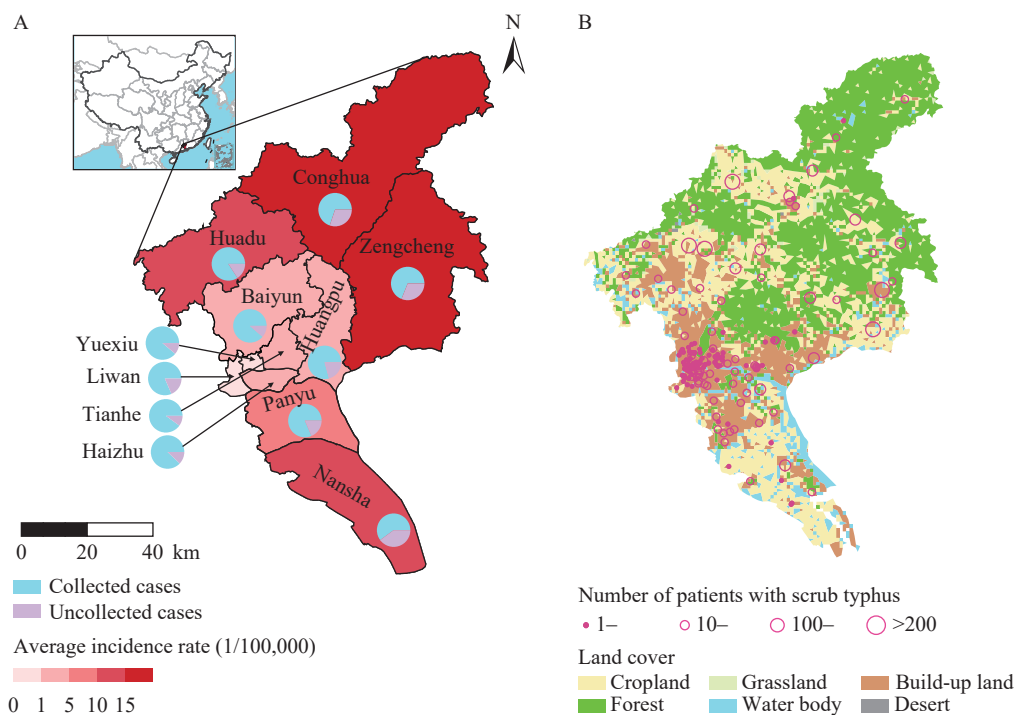


FIGURE 1. The spatial distribution of scrub typhus in Guangzhou, China, 2012–2018. (A) The percentage of currently studied cases among all reported cases in each district of Guangzhou City according to National Surveillance Data, 2012–2018; (B) The overall number of reported patients by National Surveillance Systems of Infectious Diseases for each township, 2012–2018.

TABLE 1. Baseline demographic and clinical characteristics of patients with scrub typhus in Guangzhou, China, 2012–2018.

Variables	Total (n=4,501)	2012 (n=558)	2013 (n=673)	2014 (n=731)	2015 (n=606)	2016 (n=617)	2017 (n=610)	2018 (n=706)	P value*
Sex, n (%)									0.882†
Female	2,450 (54.4)	310 (55.6)	371 (55.1)	394 (53.9)	323 (53.3)	348 (56.4)	323 (53.0)	381 (54.0)	
Male	2,051 (45.6)	248 (44.4)	302 (44.9)	337 (46.1)	283 (46.7)	269 (43.6)	287 (47.0)	325 (46.0)	
Age, years, median (IQR)	57 (46–65)	54 (42–62)	56 (42–63)	56 (45–64)	58 (48–64)	58 (47–65)	59 (49–66)	59 (49–66)	<0.001
Age group, years [§] , n (%)									<0.001†
0–14	209 (4.6)	50 (9.0)	39 (5.8)	35 (4.8)	21 (3.5)	24 (3.9)	16 (2.6)	24 (3.4)	
15–59	2,402 (53.4)	327 (58.6)	376 (55.9)	412 (56.4)	329 (54.3)	315 (51.0)	301 (49.3)	342 (48.4)	
≥60	1,890 (42.0)	181 (32.4)	258 (38.3)	284 (38.8)	256 (42.2)	278 (45.1)	293 (48.0)	340 (48.2)	
Residence, n (%)									<0.001†
Urban	1,232 (27.4)	221 (39.6)	233 (34.6)	197 (26.9)	163 (26.9)	148 (24.0)	129 (21.1)	141 (20.0)	
Rural	3,269 (72.6)	337 (60.4)	440 (65.4)	534 (73.1)	443 (73.1)	469 (76.0)	481 (78.9)	565 (80.0)	
Occupation [§] , n (%)									<0.001†
Farmers	1,093 (24.3)	151 (27.1)	149 (22.1)	150 (20.5)	123 (20.3)	165 (26.7)	170 (27.9)	185 (26.2)	
Retirees	1,242 (27.6)	134 (24.0)	191 (28.4)	216 (29.5)	183 (30.2)	172 (27.9)	155 (25.4)	191 (27.1)	
Children and students	240 (5.3)	52 (9.3)	47 (7.0)	42 (5.7)	24 (4.0)	31 (5.0)	20 (3.3)	24 (3.4)	
Others	848 (18.8)	107 (19.2)	141 (21.0)	136 (18.6)	93 (15.3)	114 (18.5)	128 (21.0)	129 (18.3)	
Unknown	1,078 (24.0)	114 (20.4)	145 (21.5)	187 (25.6)	183 (30.2)	135 (21.9)	137 (22.5)	177 (25.1)	
Time from symptom onset to hospital admission, days, median (IQR)	7 (5–9)	7 (5–10)	7 (5–10)	7 (5–9)	7 (5–9)	7 (4–9)	7 (4–9)	6 (4–9)	<0.001
Length of hospital stay, days, median (IQR)	7 (6–10)	8 (5–10)	8 (6–10)	8 (6–10)	7 (5–10)	7 (6–10)	7 (5–10)	7 (6–9)	0.092
Severe cases, n (%)	366 (8.1)	46 (8.2)	51 (7.6)	70 (9.6)	47 (7.8)	50 (8.1)	39 (6.4)	63 (8.9)	0.481†
Death cases, n (%)	53 (1.2)	8 (1.4)	4 (0.6)	18 (2.5)	6 (1.0)	5 (0.8)	3 (0.5)	9 (1.3)	0.013†

* P value relates the difference among the study years.

† P value calculated by use of χ^2 test.

§ Some columns do not add up to 100% because of rounding.

Abbreviation: IQR=interquartile-range.

comparable between males and females as a whole (8.2% vs. 8.1%, $P=0.894$). The median (IQR) time from symptom onset to hospital admission among severe patients was 8 days (6–10 days), significantly longer than that among mild patients (7 days, 4–9 days, $P<0.001$).

The frequently seen symptoms and signs of ST patients included fever (98.2%), eschar (74.1%), anorexia (69.9%), headache (51.1%), weakness (43.7%), and cough (41.3%) (Supplementary Table S1, available in <http://weekly.chinacdc.cn/>). Ulcer and skin rash were less frequently seen (among 17.4% and 14.5% of the patients). By multivariate logistic regression analysis, 12 clinical manifestations were significantly related to severe disease, among which dyspnea had the most robust effect (adjusted OR: 13.95, 95% CI: 9.94%–19.58%), followed by confusion (OR: 7.18, 95% CI: 3.29%–15.67%), dysphoria (OR: 6.76, 95% CI: 2.40%–19.10%),

lethargy (OR: 5.80, 95% CI: 2.19%–15.41%), macroscopic hematuria (OR: 5.32, 95% CI: 1.03%–27.39%), and edema (OR: 4.85, 95% CI: 3.25%–7.24%) (Supplementary Table S1).

A total of 28 laboratory parameters (12 hematological and 16 biochemical) were tested on admission, among which the frequently seen abnormalities included increased levels of neutrophil (NEU) percent in the blood, and aspartate aminotransferase (AST), alanine aminotransferase (ALT), lactate dehydrogenase (LDH), γ -glutamyl transpeptidase (GGT), and C-creative protein (CRP) in the serum above the normal range, decreased levels of hematocrit (HCT) in the blood, and total protein (TP), ALB, sodium (Na), and calcium (Ca) in the serum below the normal range (Supplementary Table S2, available in <http://weekly.chinacdc.cn/>). Multivariate logistic regression analysis disclosed laboratory abnormalities of PLT count, ALB, CREA,

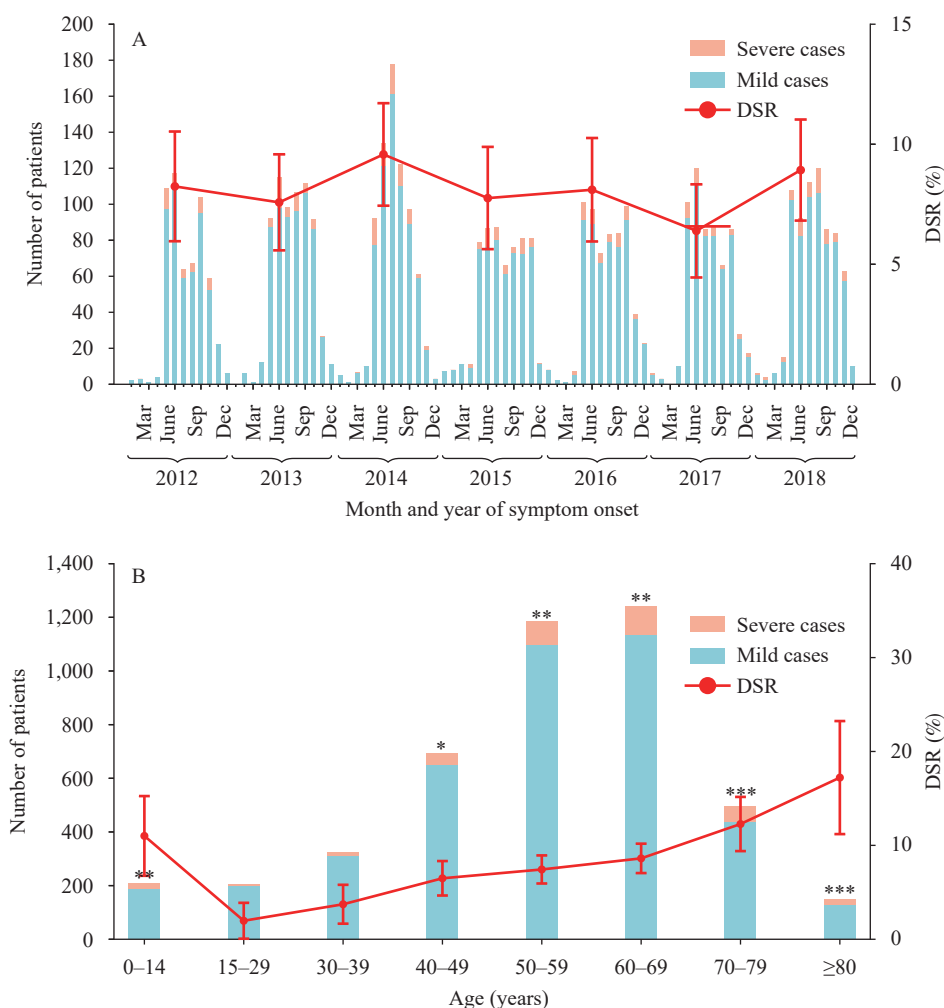


FIGURE 2. Case number and disease severity rate of patients with scrub typhus by time and age in Guangzhou, China, 2012–2018. (A) Number of scrub typhus patients and annual disease severity rate (DSR) by month and year of symptom onset. (B) Number of scrub typhus patients and DSR of patients by age groups. *P*-value indicated association between age groups and severe disease by univariate logistic regression analysis, with 15–29 years group used as reference. *, *P*<0.05; **, *P*<0.01; ***, *P*<0.001.

TBIL, hemoglobin (HGB), and NEU count that were significantly related to severe ST (Supplementary Table S3, available in <http://weekly.chinacdc.cn/>).

The final BRT model that included 10 factors attained a high prediction efficiency, with the mean AUC of 0.92 (95% CI: 0.90–0.93) and an accuracy of 0.88 (95% CI: 0.86–0.89) on testing data. The highest relative contribution (RC) to severe ST was observed for ALB, estimated to be 24.77% (95% CI: 23.70–25.85), followed by dyspnea, PLT count, CREA, and TBIL with their RCs more than 10% (Supplementary Table S4, available in <http://weekly.chinacdc.cn/>).

DISCUSSION

To our knowledge, this study represents the largest

case-cohort study on ST patients. Notably, this cohort also represented properly treated patients by antibiotics, which had a high proportion of mild disease than those studies also including untreated patients in other Asian countries (4–5). The currently determined frequency of severe complications (8.1%) was lower than that of previous cohorts (6–7), although we had exhaustively included all complications, such as MODS, shock, and ICU admission, which are the principal causes of death in patients with severe disease. Even when properly treated, the 0–14 years pediatric patients were related to a higher DSR than its next age group of 15–29 years old. Whether this discrepancy was related to the higher incidence of contact to mites or due to the lack of herd immunity in children remained to be determined. More health education on ST among children should

be implemented from schools and families in endemic areas at high risk of *Orientia tsutsugamushi* infection.

Uncommonly seen clinical complications, such as neurological manifestations, were present in 2.8% of the patients, which was based on an exhaustive search of the medical records, and were higher than those of a previous study in the Republic of Korea (1.7%) (8), while lower than studies in India and Thailand showing that up to 26% of ST patients had meningitis (9). *Orientia tsutsugamushi* has been suggested as an important cause of central nervous system infections in untreated patients. By contrast, among the properly treated patients with ST, the frequency of neurological can be significantly reduced, as displayed in the current study. The pathophysiological hallmark of ST is disseminated vasculitis, with subsequent vascular injury involved in skin, liver, brain, kidneys, and lungs, etc. Endothelial permeability may increase systemically and lead to the pathologic capillary leak syndrome resulting in tissue edema and intravascular hypovolemia (10). This can explain the prominent role of edema, as well as the increased laboratory abnormalities indicative of kidney and liver injury, i.e., CREA and TBIL. Decreased levels of plasma ALB and PLT count, surrogates of capillary permeability, were associated with severe disease.

Preexisting comorbidities significantly enhanced the odds of developing severe disease, with cerebral infarction showing significant effects according to the multivariate analysis. With no currently approved vaccines, aggressive treatment strategies should be applied and the treatment of a broad range of comorbidities should be advocated in those patients.

This study is subject to several limitations. Only hospitalized patients with ST were included for analysis, hence outpatients who were not admitted to hospitals warranted further investigation. Not all studied patients were laboratory-confirmed and misdiagnosis could not be completely ruled out, which could introduce uncertainty about our results.

In summary, we found that epidemiological and clinical features of ST were changing, with an increasing proportion of ST patients in the age ≥ 60 and in rural and higher DSR in children. Based on this knowledge, persistent surveillance and community health education should be stressed in high-risk populations, while severe ST patients could be recognized and treated in early clinical management.

Conflicts of interest: No conflicts of interest declared.

Funding: The Natural Science Foundation of China (81825019 to WL, 81722041 to HL), the China Mega-project for Infectious Diseases (2018ZX10713002 to WL), the Science and Technology Plan Grant of Guangzhou (202102080035 to YHW), and the Key Research and Development Program of China (2019YFC1200604 to LQF).

doi: 10.46234/ccdcw2021.261

* Corresponding authors: Zhicong Yang, yangzc@gzcdc.org.cn; Hao Li, lihao_1986@126.com; Liqun Fang, fang_lq@163.com; Wei Liu, liuweibmi@bmi.ac.cn.

¹ Guangzhou Center for Disease Control and Prevention, Guangzhou, Guangdong, China; ² State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, Beijing, China; ³ School of Public Health, Anhui Medical University, Hefei, Anhui, China; ⁴ School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China; ⁵ Department of Laboratorial Science and Technology, School of Public Health, Peking University, Beijing, China; ⁶ Department of Biostatistics, College of Public Health and Health Professions, and Emerging Pathogens Institute, University of Florida, Florida, USA.
[‡] Joint first authors.

Submitted: July 31, 2021; Accepted: November 07, 2021

REFERENCES

- Walker DH. Scrub typhus—scientific neglect, ever-widening impact. *N Engl J Med* 2016;375(10):913–5. <http://dx.doi.org/10.1056/NEJMp1608499>.
- Xu G, Walker DH, Jupiter D, Melby PC, Arcari CM. A review of the global epidemiology of scrub typhus. *PLoS Negl Trop Dis* 2017; 11(11):e0006062. <http://dx.doi.org/10.1371/journal.pntd.0006062>.
- Li ZJ, Xin HL, Sun JL, Lai SJ, Zeng LJ, Zheng CJ, et al. Epidemiologic changes of scrub typhus in China, 1952–2016. *Emerg Infect Dis* 2020;26(6):1091–101. <http://dx.doi.org/10.3201/eid2606.191168>.
- Taylor AJ, Paris DH, Newton PN. A systematic review of mortality from untreated scrub typhus (*Orientia tsutsugamushi*). *PLoS Negl Trop Dis* 2015;9(8):e0003971. <http://dx.doi.org/10.1371/journal.pntd.0003971>.
- Bonell A, Lubell Y, Newton PN, Crump JA, Paris DH. Estimating the burden of scrub typhus: a systematic review. *PLoS Negl Trop Dis* 2017;11(9):e0005838. <http://dx.doi.org/10.1371/journal.pntd.0005838>.
- Sivarajan S, Shivalli S, Bhuyan D, Mawlong M, Barman R. Clinical and paraclinical profile, and predictors of outcome in 90 cases of scrub typhus, Meghalaya, India. *Infect Dis Poverty* 2016;5(1):91. <http://dx.doi.org/10.1186/s40249-016-0186-x>.
- Kim DM, Kim SW, Choi SH, Yun NR. Clinical and laboratory findings associated with severe scrub typhus. *BMC Infect Dis* 2010;10:108. <http://dx.doi.org/10.1186/1471-2334-10-108>.
- Kim MH, Kim SH, Choi JH, Wie SH. Clinical and laboratory predictors associated with complicated scrub typhus. *Infect Chemother* 2019;51(2):161–70. <http://dx.doi.org/10.3947/ic.2019.51.2.161>.
- Viswanathan S, Muthu V, Iqbal N, Remalayam B, George T. Scrub typhus meningitis in South India—a retrospective study. *PLoS One* 2013;8(6):e66595. <http://dx.doi.org/10.1371/journal.pone.0066595>.
- Assaly R, Olson D, Hammersley J, Fan PS, Liu J, Shapiro JJ, et al. Initial evidence of endothelial cell apoptosis as a mechanism of systemic capillary leak syndrome. *Chest* 2001;120(4):1301–8. <http://dx.doi.org/10.1378/chest.120.4.1301>.

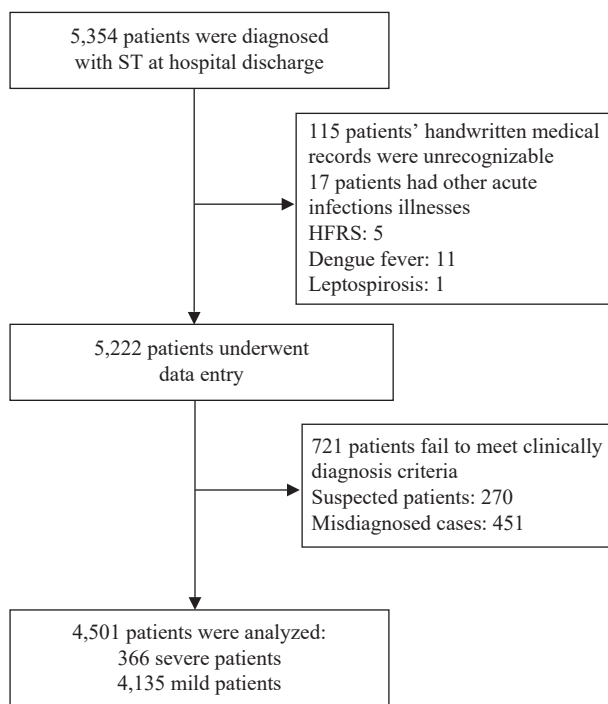
SUPPLEMENTARY METHODS

Steps of Multivariate Logistic Regression Analysis

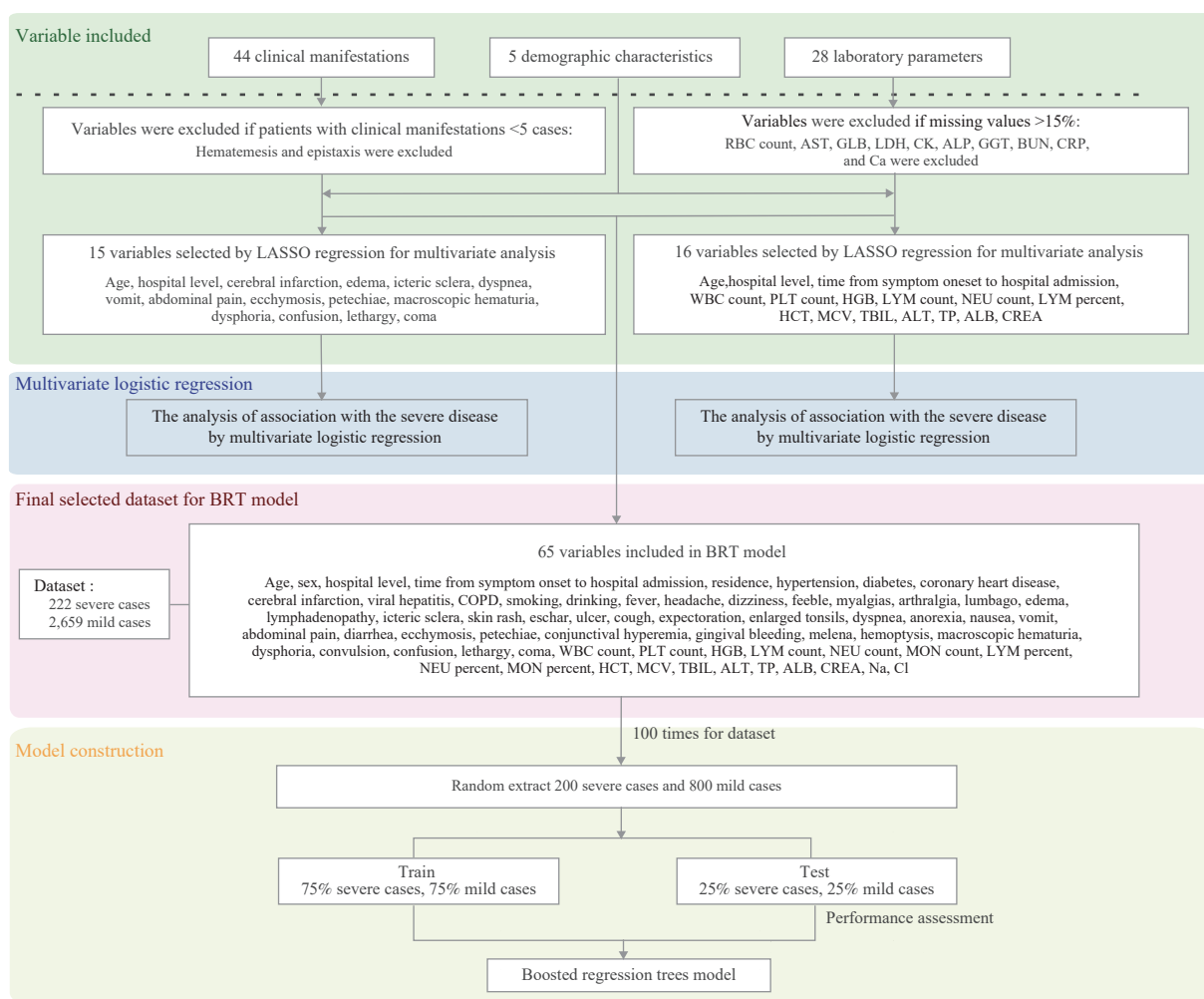
At the first step, to determine the risk factors of clinical manifestations in severe scrub typhus (ST), the presence of clinical manifestations on or before hospital admission was used for statistical analysis. Excluding variables of clinical manifestations was considered if the number of patients with a single clinical manifestation were less than five cases, and hematemesis and epistaxis were excluded. The remaining clinical manifestations, together with 5 demographic characteristics were included in Least Absolute Shrinkage and Selection Operator (LASSO) regression to select the strongest variables. Variables identified by LASSO regression analysis were entered into multivariate logistic regression analysis (Supplementary Figure S2). Meanwhile, to attain the risk factors of laboratory indicators in severe ST, the laboratory indicators result within 24 hours of hospital admission were used for statistical analysis. After excluding the laboratory indicators if the missing value was more than 15%, the laboratory indicators were also examined for their association with development of severe disease in multivariate logistic regression analysis after variable selection of LASSO regression analysis (Supplementary Figure S2). The R library “glmnet” was used for the LASSO regression. Odds ratio (OR) and the 95% confidence interval (95% CI) were estimated using maximum likelihood methods.

Variable Selection for Model Construction of the Boosted Regression Trees (BRT) Modeling

To attain an early prediction of severe ST, the clinical manifestations on or before hospital admission and the laboratory indicators result within 24 hours of hospital admission were used for model analysis. Firstly, for 44 variables of clinical manifestations, we excluded those less than five cases such as hematemesis and epistaxis. Meanwhile, for 28 laboratory parameters, we excluded those with missing values more than 15% of all patients such as RBC count, AST, GLB, LDH, CK, ALP, GGT, BUN, CRP, and Ca (Supplementary Figure S2). The remaining clinical manifestations and laboratory parameters, together with 5 demographic characteristics were included in the BRT models.



SUPPLEMENTARY FIGURE S1. The flowchart of recruited patients with scrub typhus in Guangzhou, China, 2012–2018. Abbreviations: ST=scrub typhus; HFRS=hemorrhagic fever with renal syndrome.



SUPPLEMENTARY FIGURE S2. Variable selection and model construction of the logistic regression and boosted regression trees modeling.

Notes: Demographic characteristics including age, sex, hospital level, time from symptom onset to hospital admission, and residence. Clinical manifestations including hypertension, diabetes, coronary heart disease, cerebral infarction, viral hepatitis, COPD, smoking, drinking, fever, headache, dizziness, weakness, myalgias, arthralgia, lumbago, edema, lymphadenopathy, icteric sclera, skin rash, eschar, ulcer, cough, expectoration, enlarged tonsils, dyspnea, anorexia, nausea, vomit, abdominal pain, diarrhea, ecchymosis, petechiae, conjunctival hyperemia, gingival bleeding, melena, hemoptysis, macroscopic hematuria, dysphoria, convulsion, confusion, lethargy, and coma. Laboratory parameters including WBC count, RBC count, PLT count, HGB, LYM count, NEU count, MON count, LYM percent, NEU percent, MON percent, HCT, MCV, TBIL, AST, ALT, TP, ALB, GLB, LDH, CK, ALP, GGT, CREA, BUN, CRP, Na, Cl, and Ca. Residence, rural or urban were divided according to the county address of patients, in where Huadu, Conghua, Zengcheng, Nansha, and Panyu districts were considered as rural areas, and Tianhe, Yuexiu, Liwan, Haizhu, Baiyun, and Huangpu districts were considered as urban areas (5).

Abbreviations: RBC=red blood cells; AST=aspartate aminotransferase; GLB=globulin; LDH=lactate dehydrogenase; CK=creatine kinase; ALP=alkaline phosphatase; GGT= γ -glutamyl transpeptidase; BUN=blood urea nitrogen; CRP=C-creative protein; Ca=calcium; LASSO=Least Absolute Shrinkage and Selection Operator; WBC=white blood cells; PLT=platelet; HGB=hemoglobin; LYM=lymphocyte; NEU=neutrophil; HCT=hematocrit; MCV=mean corpuscular volume; TBIL=total bilirubin; ALT=alanine aminotransferase; TP=total protein; ALB=albumin; CREA=creatinine; BRT=boosted regression trees; COPD=chronic obstructive pulmonary disease; MON=monocyte; Na=sodium; Cl=chlorine.

Steps of BRT Modeling

Machine learning techniques are increasingly used in developing models for vector predictions with multiple advantages, like considering nonlinearity and interactions and handling different types of predictor variables (1–2).

The BRT modeling was carried out in steps. First, considering the low proportion of severe outcomes, we set a 1-to-4 case-control ratio to adjust every bootstrap data. Severe cases and 4-fold num of mild cases were randomly

SUPPLEMENTARY TABLE S1. Demographic and clinical characteristics associated with severe disease for patients with scrub typhus by multivariate logistic regression.

Variables	Total (n=4,501)	Severe (n=366)	Mild (n=4,135)	Adjusted OR (95% CI)	P value
Sex, female	2,450 (54.4)	198 (54.1)	2,252 (54.5)		
Age, years, median (IQR)	57 (46–65)	60 (51–68)	56 (46–64)	1.01 (1.01–1.02)	0.001
Time from symptom onset to hospital admission, days, median (IQR)	7 (5–9)	8 (6–10)	7 (4–9)	2.76 (1.91–4.00)	<0.001
Length of hospital stay, days, median (IQR)	7 (6–10)	12 (8–17)	7 (5–9)		
Comorbidities					
Hypertension	718 (16.0)	77 (21.0)	641 (15.5)		
Diabetes	372 (8.3)	38 (10.4)	334 (8.1)		
Coronary heart disease	112 (2.5)	8 (2.2)	104 (2.5)		
Cerebral infarction	136 (3.0)	32 (8.7)	104 (2.5)	2.33 (1.37–3.98)	0.002
Viral hepatitis	178 (4.0)	12 (3.3)	166 (4.0)		
COPD	46 (1.0)	6 (1.6)	40 (1.0)		
Lifestyles					
Smoking	731 (16.2)	64 (17.5)	667 (16.1)		
Drinking	384 (8.5)	37 (10.1)	347 (8.4)		
Non-specific					
Fever	4,421 (98.2)	356 (97.3)	4,065 (98.3)		
Headache	2,302 (51.1)	174 (47.5)	2,128 (51.5)		
Dizziness	1,168 (25.9)	90 (24.6)	1,078 (26.1)		
Weakness	1,969 (43.7)	155 (42.3)	1,814 (43.9)		
Myalgias	1,051 (23.4)	82 (22.4)	969 (23.4)		
Arthralgia	136 (3.0)	5 (1.4)	131 (3.2)		
Lumbago	140 (3.1)	9 (2.5)	131 (3.2)		
Edema	178 (4.0)	64 (17.5)	114 (2.8)	4.85 (3.25–7.24)	<0.001
Lymphadenopathy	1,088 (24.2)	92 (25.1)	996 (24.1)		
Icteric sclera	79 (1.8)	33 (9.0)	46 (1.1)	4.67 (2.61–8.33)	<0.001
Skin					
Skin rash	653 (14.5)	61 (16.7)	592 (14.3)		
Eschar/Ulcer	3,921 (87.1)	302 (82.5)	3,619 (87.5)		
Eschar	3,335 (74.1)	260 (71.0)	3,075 (74.4)		
Ulcer	784 (17.4)	60 (16.4)	724 (17.5)		
Respiratory					
Cough	1,858 (41.3)	169 (46.2)	1,689 (40.8)		
Expectoration	1,157 (25.7)	116 (31.7)	1,041 (25.2)		
Enlarged tonsils	426 (9.5)	30 (8.2)	396 (9.6)		
Dyspnea	213 (4.7)	126 (34.4)	87 (2.1)	13.95 (9.94–19.58)	<0.001
Gastrointestinal					
Anorexia	3,148 (69.9)	270 (73.8)	2,878 (69.6)		
Nausea	791 (17.6)	69 (18.9)	722 (17.5)		
Vomit	484 (10.8)	63 (17.2)	421 (10.2)	1.45 (1.02–2.04)	0.036
Abdominal pain	361 (8.0)	64 (17.5)	297 (7.2)	1.90 (1.32–2.73)	0.001
Diarrhea	193 (4.3)	35 (9.6)	158 (3.8)		

TABLE S1. (Continued)

Variables	Total (n=4,501)	Severe (n=366)	Mild (n=4,135)	Adjusted OR (95% CI)	P value
Hemorrhagic manifestation*	241 (5.4)	47 (12.8)	194 (4.7)		
Ecchymosis	47 (1.0)	14 (3.8)	33 (0.8)		
Petechiae	22 (0.5)	7 (1.9)	15 (0.4)	3.45 (1.05–11.38)	0.042
Conjunctival hyperemia	108 (2.4)	20 (5.5)	88 (2.1)		
Gingival bleeding	9 (0.2)	2 (0.5)	7 (0.2)		
Melena	59 (1.3)	10 (2.7)	49 (1.2)		
Hemoptysis	10 (0.2)	3 (0.8)	7 (0.2)		
Hematemesis	3 (0.1)	1 (0.3)	2 (<0.1)		
Epistaxis	4 (0.1)	0 (0)	4 (0.1)		
Macroscopic hematuria	8 (0.2)	3 (0.8)	5 (0.1)	5.32 (1.03–27.39)	0.045
Neurological manifestation†	124 (2.8)	70 (19.3)	54 (1.3)		
Dysphoria	29 (0.6)	19 (5.2)	10 (0.2)	6.76 (2.40–19.10)	<0.001
Convulsion	27 (0.6)	15 (4.1)	12 (0.3)		
Confusion	50 (1.1)	34 (9.3)	16 (0.4)	7.18 (3.29–15.67)	<0.001
Lethargy	31 (0.7)	18 (4.9)	13 (0.3)	5.80 (2.19–15.41)	<0.001
Coma	30 (0.7)	18 (4.9)	12 (0.3)	2.90 (1.08–7.79)	0.035
Abnormal image results					
Pericardial effusion	127 (2.8)	37 (10.1)	90 (2.2)		
Pelvic effusion	42 (0.9)	14 (3.8)	28 (0.7)		
Pleural effusion	558 (12.4)	158 (43.2)	400 (9.7)		
Chest radiographic abnormality	1,062 (23.6)	167 (45.6)	895 (21.6)		
Ascites	50 (1.1)	26 (7.1)	24 (0.6)		
Splenomegaly	554 (12.3)	74 (20.2)	480 (11.6)		
Hepatomegaly	228 (5.1)	44 (12.0)	184 (4.4)		

Note: Data are n (%) until otherwise indicated. All clinical symptoms (non-specific, skin, respiratory, gastrointestinal, hemorrhagic, neurological) were reported before or at hospital admission, all abnormal image results were reported after hospital admission.

Abbreviations: COPD=chronic obstructive pulmonary disease. OR=odds ratio. CI=confidence interval; IQR=interquartile-range.

* Hemorrhagic manifestation, patients with one or more hemorrhagic symptoms.

† Neurological manifestation, patients with one or more neurological symptoms.

selected without replacement. Second, for both selected severe and mild cases, 75% of samples of each were used for training the models while the rest 25% were used for model testing. Third, a BRT model was built using the training dataset and validated using the test dataset.

To increase the predictive robustness of model predictions and quantify model uncertainty, we fitted an ensemble of 100 BRT models to separate bootstraps of the data. We adopted a tree complexity of 5, a learning rate of 0.005, and a bag fraction of 75% to identify the optimal trees for each bootstrap data. Herein, we modeled the severe complications as a binary classification. Each of the 100 models predicts severe risk on a continuous scale from 0 to 1, with a final prediction being generated by calculating the mean prediction across all models. To further determine the main contributing factor, we performed BRT model again after removing the factors that showed the relative contributions below 2% (3). The receiver-operating characteristic (ROC) curves were produced for each model based on the average of an ensemble of 100 BRT models, and the area under the receiver operating characteristic curve (AUC) was calculated to evaluate the predictive power of the models. BRT model was performed using R software (version 3.6.2; R Foundation; Vienna, Austria) with the packages of “gbm,” “dismo,” and “ROCR” (4).

SUPPLEMENTARY TABLE S2. Laboratory findings on hospital admission compared between severe patients and mild patients.

Variables	Total (n=4,501)	Severe (n=366)	Mild (n=4,135)	P value
Hematological indicators				
WBC count ($\times 10^9/L$) [4–10]	7.18 (5.23–9.60)	8.99 (6.50–12.92)	7.06 (5.19–9.39)	<0.001
RBC count ($\times 10^{12}/L$) [3.8–5.8]	4.23 (3.87–4.64)	3.95 (3.44–4.31)	4.26 (3.89–4.66)	<0.001
PLT count ($\times 10^9/L$) [100–300]	121.00 (81.00–172.00)	60.00 (40.00–95.50)	125.00 (86.05–178.00)	<0.001
HGB (g/L) [120–165]	121.00 (110.00–132.00)	110.90 (97.00–124.00)	121.00 (111.00–133.00)	<0.001
LYM count ($\times 10^9/L$) [1.1–3.2]	1.42 (0.78–2.57)	1.40 (0.73–2.66)	1.42 (0.79–2.56)	0.672
NEU count ($\times 10^9/L$) [1.8–6.3]	4.67 (3.23–6.55)	6.50 (4.47–9.7)	4.55 (3.17–6.36)	<0.001
MON count ($\times 10^9/L$) [0.1–0.6]	0.44 (0.28–0.69)	0.42 (0.26–0.74)	0.45 (0.28–0.69)	0.719
LYM percent (%) [20–40]	20.50 (12.90–32.30)	15.00 (8.60–26.90)	21.00 (13.30–32.80)	<0.001
NEU percent (%) [50–70] [†]	71.40 (57.50–80.40)	78.40 (63.50–86.30)	71.00 (57.00–79.80)	<0.001
MON percent (%) [3–10]	6.40 (4.40–8.80)	4.90 (3.30–7.50)	6.50 (4.50–8.94)	<0.001
HCT (%) [40–50] [‡]	35.90 (32.50–39.20)	32.80 (28.60–35.70)	36.07 (32.80–39.40)	<0.001
MCV (fL) [82–100]	86.30 (81.70–90.01)	84.65 (79.50–89.00)	86.50 (81.90–90.20)	<0.001
Biochemical indicators				
TBIL ($\mu\text{mol/L}$) [5.1–17.1]	11.50 (8.20–16.50)	18.50 (12.38–48.35)	11.00 (8.00–15.80)	<0.001
AST (U/L) [0–40] [†]	76.00 (49.00–122.00)	136.00 (88.10–218.40)	73.00 (48.00–115.00)	<0.001
ALT (U/L) [0–40] [†]	67.00 (42.00–107.40)	79.50 (54.00–118.00)	65.65 (41.00–106.00)	<0.001
TP (g/L) [65–85] [‡]	61.20 (56.60–66.10)	54.35 (49.20–60.60)	61.80 (57.30–66.40)	<0.001
ALB (g/L) [35–55] [‡]	33.00 (28.90–36.70)	25.90 (22.00–29.70)	33.40 (29.60–37.00)	<0.001
GLB (g/L) [20–40]	28.20 (25.10–31.70)	29.00 (25.30–32.60)	28.11 (25.10–31.60)	0.176
LDH (U/L) [109–245] [†]	459.00 (342.00–693.00)	677.00 (464.00–1,147.00)	446.50 (336.00–658.00)	<0.001
CK (U/L) [25–200]	91.00 (50.00–173.00)	110.00 (59.00–286.00)	90.00 (50.00–167.00)	<0.001
ALP (U/L) [40–150]	94.00 (67.03–143.00)	127.00 (76.00–220.50)	92.00 (67.00–138.00)	<0.001
GGT (U/L) [7–50] [†]	70.00 (33.00–146.00)	98.00 (44.00–189.00)	67.00 (32.00–142.00)	<0.001
CREA ($\mu\text{mol/L}$) [53–106]	78.08 (64.00–98.00)	102.95 (72.00–193.50)	77.85 (63.10–95.10)	<0.001
BUN (mmol/L) [3.2–7.1]	4.34 (3.30–6.04)	9.00 (5.97–16.60)	4.27 (3.24–5.70)	<0.001
CRP (mg/L) [0.07–8.20] [†]	45.00 (14.18–86.36)	84.05 (39.00–140.70)	42.00 (13.31–81.10)	<0.001
Na (mmol/L) [137–147] [‡]	135.70 (132.00–138.60)	133.00 (129.80–137.00)	136.00 (132.40–138.70)	<0.001
Cl (mmol/L) [96–108]	99.70 (96.40–102.90)	100.00 (96.00–103.90)	99.70 (96.40–102.80)	0.272
Ca (mmol/L) [2.10–2.55] [‡]	2.07 (1.96–2.18)	1.90 (1.79–2.03)	2.08 (1.98–2.19)	<0.001

Note: All measurements within 24 hours of hospital admission were presented as median (interquartile-range).

P value was compared between severe group and mild group by Mann-Whitney U test.

Abbreviations: WBC=white blood cells; RBC=red blood cells; PLT=platelet; HGB=hemoglobin; LYM=lymphocyte; NEU=neutrophil; MON=monocyte; HCT=hematocrit; MCV=mean corpuscular volume; TBIL=total bilirubin; AST=aspartate aminotransferase; ALT=alanine aminotransferase; TP=total protein; ALB=albumin; GLB=globulin; LDH=lactate dehydrogenase; CK=creatinine kinase; ALP=alkaline phosphatase; GGT= γ -glutamyl transpeptidase; CREA=creatinine; BUN=blood urea nitrogen; CRP=C-creative protein; Na=sodium; Cl=chlorine; Ca=calcium.

[†] Variable, with the median value in total population above the normal range.

[‡] Variable, with the median value in total population below the normal range.

SUPPLEMENTARY TABLE S3. Association between laboratory measurements and severe scrub typhus by multivariate logistic regression analysis.

Variables	Adjusted OR (95% CI)	P value
Hospital level	2.26 (1.50–3.43)	<0.001
Time from symptom onset to hospital admission, days	1.06 (1.01–1.10)	0.014
PLT count <100 ($\times 10^9/L$) [100–300]	4.80 (3.39–6.82)	<0.001
HGB <120 (g/L) [120–165]	2.19 (1.59–3.01)	<0.001
NEU count >6.3 ($\times 10^9/L$) [1.8–6.3]	2.03 (1.47–2.79)	<0.001
TBIL >17.1 ($\mu\text{mol/L}$) [5.1–17.1]	2.62 (1.91–3.60)	<0.001
ALB <35 (g/L) [35–55]	3.78 (1.86–7.65)	<0.001
CREA >106 ($\mu\text{mol/L}$) [53–106]	2.91 (2.12–4.00)	<0.001

Abbreviations: OR=odds ratio; CI=confidence interval; PLT=platelet; HGB=hemoglobin; NEU=neutrophil; TBIL=total bilirubin; ALB=albumin; CREA=creatinine.

SUPPLEMENTARY TABLE S4. Relative contribution of predictors estimated for the occurrence of severe complications of scrub typhus by the boosted regression trees model.

Variables	Relative contribution, % (95% CI)
ALB (g/L)	24.77 (23.70–25.85)
Dyspnea	16.63 (15.73–17.53)
PLT count ($\times 10^9/L$)	12.95 (12.16–13.74)
CREA ($\mu\text{mol/L}$)	11.34 (10.58–12.11)
TBIL ($\mu\text{mol/L}$)	10.55 (10.08–11.02)
NEU count ($\times 10^9/L$)	5.71 (5.40–6.02)
TP (g/L)	5.59 (5.20–5.98)
NEU percent (%)	5.40 (5.06–5.75)
HGB (g/L)	4.25 (3.97–4.54)
Age	2.79 (2.59–2.98)

Abbreviations: ALB=albumin; PLT=platelet; CREA=creatinine; TBIL=total bilirubin; NEU=neutrophil; TP=total protein; HGB=hemoglobin; CI=confidence interval.

REFERENCES

1. Sinka ME, Rubio-Palis Y, Manguin S, Patil AP, Temperley WH, Gething PW, et al. The dominant *Anopheles* vectors of human malaria in the Americas: occurrence data, distribution maps and bionomic précis. *Parasit Vectors* 2010;3:72. <http://dx.doi.org/10.1186/1756-3305-3-72>.
2. Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol* 2008;77(4):802–13. <http://dx.doi.org/10.1111/j.1365-2656.2008.01390.x>.
3. Zhao GP, Wang YX, Fan ZW, Ji Y, Liu MJ, Zhang WH, et al. Mapping ticks and tick-borne pathogens in China. *Nat Commun* 2021;12(1):1075. <http://dx.doi.org/10.1038/s41467-021-21375-1>.
4. Araújo MB, New M. Ensemble forecasting of species distributions. *Trends Ecol Evol* 2007;22(1):42–7. <http://dx.doi.org/10.1016/j.tree.2006.09.010>.
5. Wei YH, Huang Y, Luo L, Xiao XC, Liu L, Yang ZC. Rapid increase of scrub typhus: an epidemiology and spatial-temporal cluster analysis in Guangzhou City, Southern China, 2006–2012. *PLoS One* 2014;9(7):e101976. <http://dx.doi.org/10.1371/journal.pone.0101976>.