BMC Infectious Diseases

RESEARCH

Open Access

Epidemiological and clinical characteristics of bacterial co-detection in respiratory syncytial virus-positive children in Wenzhou, China, 2021 to 2023

Xiaoxiao Qu[1], Xuanmei Ye[1], Jian Yu[1], Feifei Zheng[1], Yatian Tang[1], Fangling Yuan[1] and Qipeng Xie[1*]

Abstract

Background Respiratory syncytial virus (RSV) is a leading cause of respiratory tract infections in young children, posing significant health challenges worldwide. This study aims to analyze the etiological, clinical, and imaging characteristics of RSV infection in children from Wenzhou, China, to better understand its epidemiological profile and provide insights for early diagnosis, monitoring, and clinical interventions.

Methods This retrospective descriptive observational study included 1,063 RSV-positive pediatric inpatients (< 14 years) from the Second Affiliated Hospital of Wenzhou Medical University (December, 2021, and April, 2023). Missing data were addressed using multiple imputation. Patients were grouped by co-infection type (viral or bacterial), and differences in demographic, clinical, and imaging features were analyzed. Univariate and multivariate logistic regression identified risk factors for bacterial co-infection in RSV-positive patients. Predictive features were selected via the Boruta algorithm and modeled using XGBoost, with SHAP applied for interpretability. Pathogen distributions were also compared across age groups in both co-infection subtypes.

Results Among 1,063 hospitalized RSV-positive children, RSV primarily affected those under one year of age, with a substantial proportion of cases (47.3%) showing co-infections. The most frequently detected viral co-pathogen was human rhinovirus (HRV, 42.2%), while *Streptococcus pneumoniae* (37.3%) was the most common bacterial pathogen. Some children were found to carry multiple pathogens simultaneously. Age-stratified radiographic analysis revealed that bronchitis was the predominant imaging finding across all age groups. Compared with the viral co-infection group, children with bacterial co-infections had longer hospital stays (6.4 vs. 5.4 days), higher white blood cell and neutrophil counts, and lower IgE levels ($P < 0.05$). Logistic regression analysis identified length of stay, prealbumin (PA), and immunoglobulin E (IgE) as independent predictors of bacterial co-infection. Feature selection using the Boruta algorithm, combined with XGBoost modeling and SHAP interpretation, further confirmed that length of stay, WBC count, neutrophil count, and IgE were the most important predictive variables. Age-specific pathogen analysis showed that influenza A and *Streptococcus pneumoniae* were predominantly found in children aged 3 to < 6 years ($P < 0.05$).

*Correspondence:
Qipeng Xie
pandon2002@163.com

Full list of author information is available at the end of the article

Conclusions RSV infections in Wenzhou predominantly occur in infants under one year of age and frequently involve co-infections with viruses and bacteria, particularly rhinovirus and *Streptococcus pneumoniae*. Early identification of co-infections and tailored interventions are critical to improving patient outcomes.

Trial registration Not applicable.

Keywords Respiratory syncytial virus (RSV), Co-infections, Pathogen analysis, Risk factors, Allergic histories

Background

Respiratory syncytial virus (RSV), a major member of the Pneumovirus genus within the Paramyxoviridae family, is an enveloped virus with a non-segmented, single-stranded, negative-sense RNA genome [1, 2]. RSV is primarily transmitted via respiratory droplets and aerosolized particles, making it a significant agent of respiratory infections [3]. It predominantly affects the pediatric population, especially infants and children under one year of age, and is a leading cause of severe lower respiratory tract infections, including bronchiolitis and pneumonia [4].

Globally, RSV is responsible for approximately 3.2 million hospital admissions annually among children, posing a major public health concern due to its potential to cause long-term pulmonary damage and to exacerbate underlying chronic conditions such as asthma [5, 6]. In China, RSV is also one of the leading viral pathogens causing respiratory tract infections in children aged 0 to 14 years, accounting for over 14% of lower respiratory tract infections among hospitalized pediatric patients [7].

The pathogenesis of RSV involves a complex interplay between viral components and the host immune response [8]. Key viral proteins—including mucosal glycoproteins, fusion glycoproteins, matrix proteins, and non-structural proteins—play pivotal roles in evading the host immune defense and promoting viral replication. These interactions compromise respiratory function, resulting in acute respiratory complications and, in some cases, long-term pulmonary sequelae [9].

Given the substantial morbidity and broad public health impact of RSV, there is an urgent need to enhance our understanding of its epidemiological patterns, pathogenic mechanisms, and clinical manifestations. This retrospective study analyzed data from 1,063 RSV-positive pediatric patients under 14 years of age, admitted between December, 2021, and April, 2023 to the Second Affiliated Hospital of Wenzhou Medical University. The study aims to elucidate the detailed pathogen landscape and its correlation with clinical outcomes. By doing so, it seeks to provide critical insights into the development of effective management and prevention strategies for RSV while advancing early diagnostic methodologies and improving prognostic assessments in clinical practice.

Methods

Study subjects

This study was a retrospective descriptive observational study that analyzed pediatric patients hospitalized with respiratory infections at the Second Affiliated Hospital of Wenzhou Medical University between December, 2021, and April, 2023. A total of 1,063 patients were included, comprising 661 males and 402 females, with ages ranging from 1 month to 14 years. The study used secondary clinical and laboratory data extracted from hospital records. Eligibility criteria were clearly defined, as detailed below: 1. Pediatric patients aged between 0 and 14 years.2. Pediatric patients meeting the diagnostic criteria for acute respiratory tract infections (ARTI) [10, 11]. 3. Admission to the pediatric department between December 27, 2021, and April 27, 2023.4. Patients with RSV infection confirmed by RT-PCR testing of respiratory samples (nasopharyngeal or oropharyngeal swabs) [12, 13]. Exclusion criteria included: (1) Patients with incomplete clinical or laboratory data. (2) Presence of congenital anomalies, genetic disorders, immunodeficiencies, or malnutrition. (3) Patients with prematurity or conditions involving autoimmune, cardiovascular, endocrine, gastrointestinal, genitourinary, hematologic, hepatobiliary, neuropsychological, respiratory, or solid tumor disorders [14]. A total of 1,063 patients with ARTI met the eligibility criteria for this study, after excluding cases with incomplete data (Fig. 1). For other variables with less than 10% missing data, multiple imputation was performed to minimize bias. The study protocol was approved by the Ethics Review Committee of the Second Affiliated Hospital of Wenzhou Medical University. Given the retrospective nature of the study and the use of anonymized patient data, the requirement for informed consent was waived in accordance with institutional and national ethical guidelines.

Children diagnosed with acute respiratory infections between December, 2021, to April, 2023 were initially screened ($N = 1,153$). A total of 25 patients were excluded due to hematologic diseases ($N = 14$) or incomplete data ($N = 11$), leaving 1,128 patients. Subsequently, 65 additional patients were excluded based on age > 14 years ($N = 39$) or missing laboratory data ($N = 26$). The final cohort included in the analysis comprised 1,063 patients.

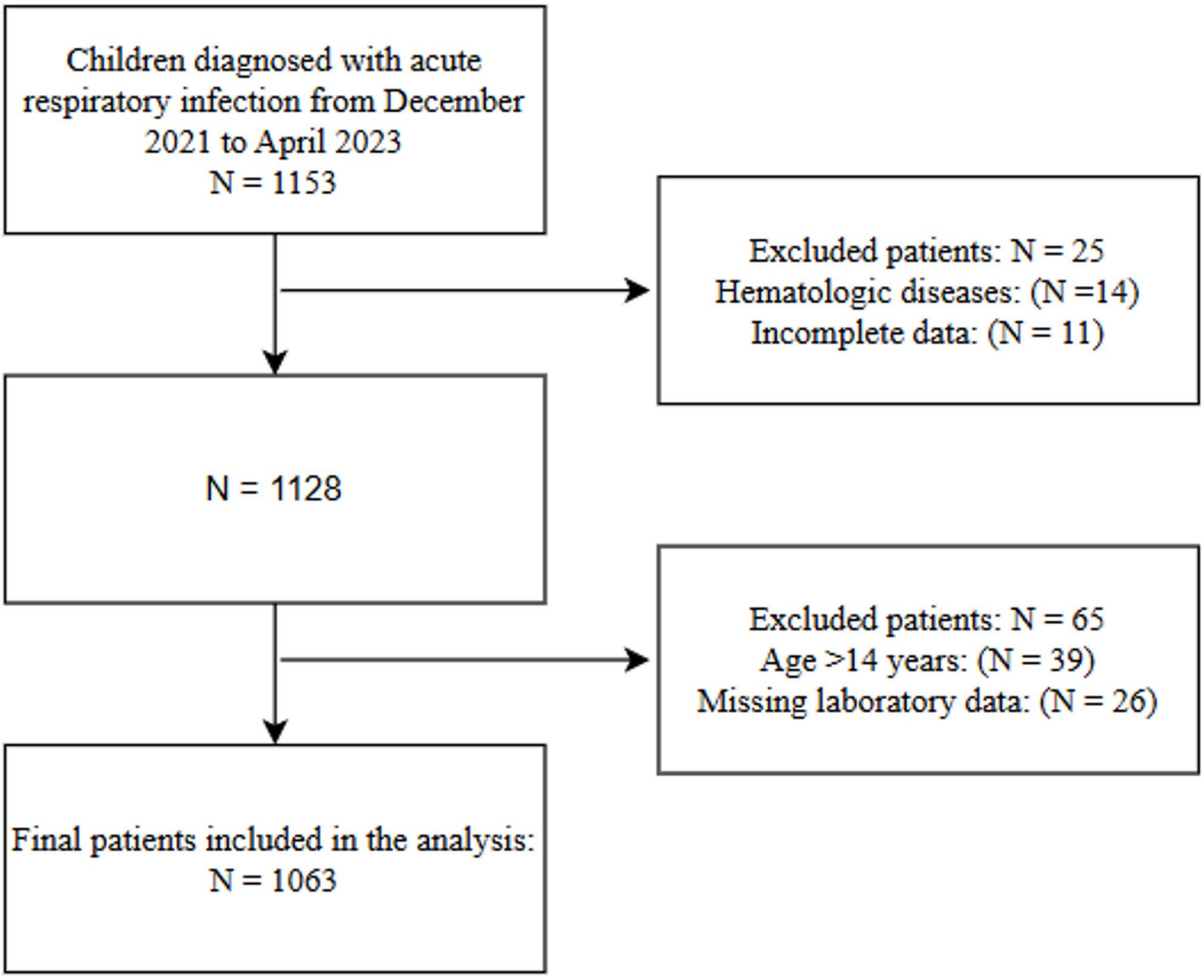

Fig. 1 Study flowchart depicting patient inclusion and exclusion criteria

Data collection

Patient data were collected from the electronic medical record system, including baseline demographic information and laboratory indicators. Biomarkers measured included a comprehensive range of hematological, biochemical, and inflammatory parameters to assess overall health, organ function, and immune status.

The laboratory measurements comprised: (1) Hematological Parameters: White blood cell count (WBC), neutrophil count (Neut), lymphocyte count (Lymph), hemoglobin (Hb), red blood cell count (RBC), hematocrit (HCT), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), Platelet count (PLT). (2) Biochemical Parameters: Prealbumin (PA), aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TBIL), total protein (TP), albumin (ALB), blood urea nitrogen (BUN), creatinine (CREA), calcium (Ca), magnesium (Mg), sodium (Na), chloride (Cl)0.3. Inflammatory and Immune Markers: C-reactive protein (CRP), procalcitonin (PCT), and immunoglobulin E (IgE).

Hematological data were obtained using the Mindray BC-5300 analyzer, while biochemical results were measured with the Siemens ADVIA 2400 analyzer. All laboratory tests were performed within 24 h of patient admission for ARTI, prior to the initiation of any clinical treatment.

These parameters are routinely measured in clinical practice and provide essential insights into liver and kidney function, anemia status, inflammatory responses, and electrolyte balance, supporting the comprehensive assessment of patient health.

Disease definition

The diagnostic criteria for RSV infection were based on the 2020 Expert Consensus on the Diagnosis, Treatment, and Prevention of Respiratory Syncytial Virus (RSV) Infection in Children [15]. These criteria included clinical

manifestations of respiratory tract infections, such as cough, wheezing, and increased respiratory effort, along with laboratory confirmation via reverse transcription polymerase chain reaction (RT-PCR).

In this study, mixed infection was defined as the simultaneous detection of Respiratory Syncytial Virus (RSV) and one or more other pathogens (including viruses and bacteria) from respiratory specimens (nasopharyngeal swabs) within 24 h after RSV confirmation, using either PCR or culture methods [16].

The diagnosis of bacterial co-infection was based on a combination of clinical signs—such as persistent fever and elevated inflammatory markers—and microbiological findings, including positive bacterial cultures from lower respiratory tract specimens. Standard techniques for bacterial isolation and identification were employed to confirm the presence of co-infecting bacterial pathogens.

Reverse transcription polymerase chain reaction (RT-PCR) detection

Nasopharyngeal swab samples were collected from pediatric patients using sterile synthetic fiber swabs. After collection, each swab was placed in viral transport medium (VTM), stored at 2–8 °C, and transferred to the laboratory within 24 h. Upon arrival, total nucleic acids were extracted using either the Haier A-96 automatic nucleic acid extraction instrument (Ningbo Health Gene Technologies) with corresponding reagent kits or the GeneRotex automatic magnetic bead extractor (Xi'an Tianlong Technology, qEx-DNA/RNA kits), according to the manufacturer's protocols. For Haier A-96, extraction took approximately 38 min under standard mode, and for GeneRotex, fast mode took approximately 13 min. The extracted nucleic acids were then used as templates for downstream PCR analysis. Pathogen detection was conducted using a one-step reverse transcription polymerase chain reaction (RT-PCR) combined with capillary electrophoresis. Highly conserved genetic sequences were targeted using specific primers, enabling amplification within a single tube. The amplified products were separated by capillary electrophoresis based on fragment size, facilitating the identification of specific pathogens.

Pathogen detection was conducted using a one-step reverse transcription polymerase chain reaction (RT-PCR) combined with capillary electrophoresis. Highly conserved genetic sequences were targeted using specific primers, enabling amplification within a single tube. The amplified products were separated by capillary electrophoresis based on fragment size, facilitating the identification of specific pathogens.

To ensure data quality, human RNA and DNA were detected in each sample to confirm sample integrity. An internal reference standard, included in the RT-PCR kit, monitored the entire process, including nucleic acid extraction, RT-PCR amplification, and capillary electrophoresis.

Multiplex RT-PCR was employed to simultaneously identify RSV and other common respiratory pathogens, including: Viruses: Human Rhinovirus (HRV), Human Parainfluenza Virus (HPIV), Human Coronavirus (HCOV), Human Metapneumovirus (HMPV), Influenza A Virus (INFA), and Human Bocavirus (HBoV). Bacteria: Mycoplasma pneumoniae (MP).

Bacterial strain identification and antimicrobial susceptibility testing

For neonates and young infants unable to expectorate spontaneously, sputum specimens were collected via nasopharyngeal or oropharyngeal suctioning by trained medical staff under aseptic conditions. Samples were inoculated onto Columbia blood agar plates and incubated at 37 °C for 24 h to promote bacterial growth, following the protocols outlined in the Clinical Laboratory Standards Institute (CLSI) [17].

Bacterial identification was performed using the MALDI Biotyper Automatic Rapid Microbiological Mass Spectrometer, and antimicrobial susceptibility testing was conducted according to Clinical and Laboratory Standards Institute (CLSI) guidelines to ensure standardized and reproducible results. In addition, bacterial culture and identification were supported by the Bact/ALERT 3D Automatic Microbial Culture System, utilizing GP identification and ASTGN13 susceptibility cards, alongside Columbia blood agar medium.

Hematological analyses were carried out using Myriad BC-5390 and Myriad 7500 automatic blood cell analyzers, with reagents and quality control materials provided by Shenzhen Myriad Electronics Co. Biochemical parameters were measured using Siemens automatic biochemistry analyzers, ensuring consistency and accuracy across assays.

Outcome

The primary outcome of the study was the occurrence of pathogen co-infections associated with RSV. The study included patients with laboratory-confirmed RSV infection. Based on the presence of additional viral or bacterial pathogens, cases were classified into two groups: (1) RSV with viral co-infection, and (2) RSV with bacterial co-infection. This classification enabled the analysis of co-infection patterns and their potential clinical implications.

Data analyses and effect measures

The Shapiro-Wilk test was used to evaluate the normality of continuous variables. Continuous variables with a normal distribution were expressed as mean ± standard

deviation (SD) and compared using one-way analysis of variance (ANOVA). Skewed continuous variables were presented as median values with interquartile ranges (IQR) and analyzed using the Kruskal-Wallis H test. Categorical or dichotomous variables were reported as absolute counts (percentages) and compared using chi-square (χ^2) tests. Missing values exceeding 1% were addressed through multiple imputations (five iterations).

Univariate and multivariate logistic regression models were employed to evaluate associations between clinical variables and RSV infection, as well as bacterial co-infections. Results were reported as odds ratios (OR) with 95% confidence intervals (CI).

The Boruta algorithm was utilized to identify key predictive features by comparing the Z-scores of true features to those of "shadow features," which are randomly shuffled duplicates of the original features. Features with Z-scores consistently exceeding the maximum Z-score of shadow features across multiple tests were classified as important (acceptable variables), while those failing to meet this criterion were deemed unimportant (unacceptable variables). Acceptable variables were retained for downstream machine learning analyses. Among these variables, hospital stay duration was specifically evaluated as a potential risk factor associated with bacterial co-infection, given its clinical relevance in hospitalized pediatric patients.

Variables selected through the Boruta algorithm were further analyzed using the eXtreme Gradient Boosting (XGBoost) model to predict the risk of RSV-associated bacterial co-infection.

SHapley Additive exPlanations (SHAP) analysis, based on cooperative game theory, was applied to interpret the machine learning model. SHAP quantifies the contribution of each feature to model predictions by assessing its marginal effect across all feature subsets. This approach provides: Global interpretability: The impact of features across the entire dataset. Local interpretability: The effect of features on individual predictions. Visualization tools, including summary plots, force plots, and dependence plots, were used to illustrate feature importance and their directional impact on predictions. SHAP ensures consistent, model-agnostic interpretation for both linear and non-linear models.

All statistical analyses were conducted using the R statistical software package (version 4.2.2) and the free statistical software DCPM_5.38 (version 2.0). Descriptive statistics were calculated for all participants. A two-tailed test was applied, with $P<0.05$ considered statistically significant.

Results

Baseline characteristics

Baseline characteristics of the study population are presented in Table 1. There was no significant difference in age distribution between the two groups ($P=0.311$). Compared with the viral co-infection group, the bacterial co-infection group had a significantly higher rate of abnormal chest X-ray findings (90.3% vs. 83.0%, $P=0.008$) and a longer average length of hospital stay (6.4 days vs. 5.4 days, $P<0.001$). In terms of laboratory parameters, the bacterial co-infection group exhibited significantly higher levels of white blood cell (WBC) count, neutrophil (Neut) count, and platelet (PLT) count than the viral co-infection group ($P<0.05$), indicating a more pronounced systemic inflammatory response.

Feature importance analysis using Boruta algorithm

The distribution of feature importance, ranked by the Boruta algorithm, is visualized in Fig. 2. Features are ranked from low to high importance and visually categorized by color to indicate their relative contributions. The analysis identified several key features as critical predictors strongly associated with the occurrence of RSV-associated bacterial co-infections. These features include: Mean corpuscular hemoglobin (MCH); Mean corpuscular volume (MCV); Neutrophil count (Neut); White blood cell count (WBC); Age (Age.m); Immunoglobulin E (IgE); Creatinine (CREA). These features were classified as "important" by the Boruta algorithm, indicating their substantial contribution to the predictive model.

Univariate logistic regression analysis of factors associated with RSV infection

Univariate logistic regression analyses were performed to identify factors associated with bacterial co-infections in RSV patients. The analysis revealed that length of stay (LOS), immunoglobulin E (IgE) levels, prealbumin (PA), and X-ray findings were key predictors($P<0.05$) (Supplementary Table 1). Other variables included in the model did not show significant associations with bacterial co-infection.

Multivariate logistic regression analysis of factors associated with RSV infection

Multivariate logistic regression analysis revealed that IgE levels, prealbumin (PA) levels, length of hospital stay, and abnormal X-ray findings were significantly associated with bacterial co-infection ($P<0.05$). Other variables included in the model did not show statistically significant associations.(Table 2).

SHAP analysis

SHAP (SHapley Additive exPlanations) analysis identified length of hospital stay (LOS), neutrophil count

Table 1 Baseline characteristics of RSV patients

Variables	Total (*n* = 1063)	RSV with viral co-detection (*n* = 843)	RSV with bacterial co-detection (*n* = 220)	*P* value
Male, n (%)	661 (62.2)	521 (61.8)	140 (63.6)	0.618
Age				0.311
0 ~ < 1year	265 (30.7)	203 (29.7)	62 (34.6)	
1 ~ < 3year	154 (17.9)	126 (18.4)	28 (15.6)	
3 ~ < 6year	396 (45.9)	313 (45.8)	83 (46.4)	
6 ~ 14year	47 (5.5)	41 (6)	6 (3.4)	
X.rays, n (%)	893 (84.5)	698 (83)	195 (90.3)	0.008
History of allergies, n (%)	98 (9.3)	86 (10.3)	12 (5.5)	0.028
Congenital heart disease, n (%)	24 (2.3)	20 (2.4)	4 (1.8)	0.801
LOS (Days)	5.6 ± 4.3	5.4 ± 3.8	6.4 ± 5.8	< 0.001
WBC ($\times 10^9$/L)	9.4 ± 4.3	9.2 ± 4.2	10.1 ± 4.8	0.003
Neut ($\times 10^9$/L)	4.6 ± 3.5	4.4 ± 3.4	5.0 ± 3.7	0.021
Lymph ($\times 10^9$/L)	3.9 ± 2.3	3.8 ± 2.2	4.0 ± 2.3	0.18
Hb (g/L)	120.7 ± 10.9	120.8 ± 10.9	120.3 ± 11.0	0.544
RBC ($\times 10^{12}$/L)	4.5 ± 0.5	4.5 ± 0.5	4.5 ± 0.5	0.976
HCT	0.4 ± 0.0	0.4 ± 0.0	0.4 ± 0.0	0.763
MCV (fl.)	81.8 ± 5.7	81.8 ± 5.6	81.8 ± 6.2	0.933
MCH (pg)	27.1 ± 2.8	27.1 ± 2.9	27.0 ± 2.4	0.611
PLT ($\times 10^9$/L)	316.2 ± 113.9	311.7 ± 112.7	333.4 ± 117.0	0.012
PA (mg/L)	134.2 ± 40.4	136.1 ± 40.0	127.5 ± 41.1	0.013
AST (U/L)	50.7 ± 36.8	51.8 ± 40.3	46.6 ± 18.5	0.087
TP (g/L)	67.5 ± 5.8	67.6 ± 5.8	67.1 ± 5.8	0.194
ALB (g/L)	44.7 ± 4.0	44.7 ± 4.0	44.8 ± 3.9	0.879
BUN (mmol/L)	3.5 ± 1.3	3.5 ± 1.3	3.4 ± 1.3	0.492
CREA (μmol/L)	25.7 ± 7.0	25.8 ± 7.1	25.3 ± 6.4	0.345
CA (mmol/L)	2.4 ± 0.2	2.4 ± 0.2	2.4 ± 0.1	0.413
MG (mmol/L)	0.9 ± 0.2	0.9 ± 0.2	0.9 ± 0.1	0.499
NA (mmol/L)	137.6 ± 2.5	137.7 ± 2.4	137.4 ± 2.6	0.126
CL (mmol/L)	103.9 ± 2.6	103.9 ± 2.5	103.7 ± 2.8	0.339
CRP (mg/L)	4.5 (1.2, 13.7)	4.3 (1.1, 13.3)	5.3 (1.3, 16.7)	0.075
PCT (ng/ml)	0.1 (0.1, 0.2)	0.1 (0.1, 0.2)	0.1 (0.1, 0.3)	0.076
ALT (U/L)	20.0 (15.0, 27.0)	20.0 (16.0, 27.0)	19.0 (14.0, 27.0)	0.236
TBIL (μmol/L)	5.4 (4.0, 7.6)	5.3 (4.0, 7.4)	5.4 (4.3, 8.0)	0.178
IgE (IU/ML)	64.6 (19.5, 175.5)	69.6 (20.4, 199.0)	41.2 (14.1, 109.5)	< 0.001

Note: Documented allergy history (including milk and protein allergies) was present in both infants and older children, based on medical records and caregiver reports

LOS (Length of Stay), WBC (White Blood Cell count), Neut (Neutrophil count), Lymph (Lymphocyte count), Hb (Hemoglobin), RBC (Red Blood Cell count), HCT (Hematocrit), MCV (Mean Corpuscular Volume), MCH (Mean Corpuscular Hemoglobin), PLT (Platelet count), PA (Prealbumin), AST (Aspartate Aminotransferase), TP (Total Protein), ALB (Albumin), BUN (Blood Urea Nitrogen), CREA (Creatinine), CA (Calcium), MG (Magnesium), NA (Sodium), CL (Chloride), CRP (C-Reactive Protein), PCT (Procalcitonin), ALT (Alanine Aminotransferase), TBIL (Total Bilirubin), IgE (Immunoglobulin E)

(Neut), white blood cell count (WBC), and immunoglobulin E (IgE) as the most influential predictors of bacterial co-infections in RSV patients. LOS had the highest SHAP value, indicating a strong positive association with infection risk, followed by Neut and WBC. In contrast, IgE showed a significant negative SHAP value, suggesting a potential protective effect. Other variables, such as lymphocyte count (Lymph), mean corpuscular volume (MCV), mean corpuscular hemoglobin (MCH), and creatinine (CREA), contributed marginally to the prediction model (Fig. 3).

Comparison of RSV co-infection detection results across different age groups

The detection rates of various respiratory pathogens co-infected with RSV across different age groups are summarized in Table 3. RSV co-infection with **influenza A viruses (INFA, including H1N1)** was significantly more common in the 3 < 6 years age group ($P = 0.016$), suggesting that this age group may have higher susceptibility to these viral pathogens. No statistically significant differences were observed in the distribution of other co-infections, including human rhinovirus (HRV), human bocavirus (HBoV), Mycoplasma pneumoniae

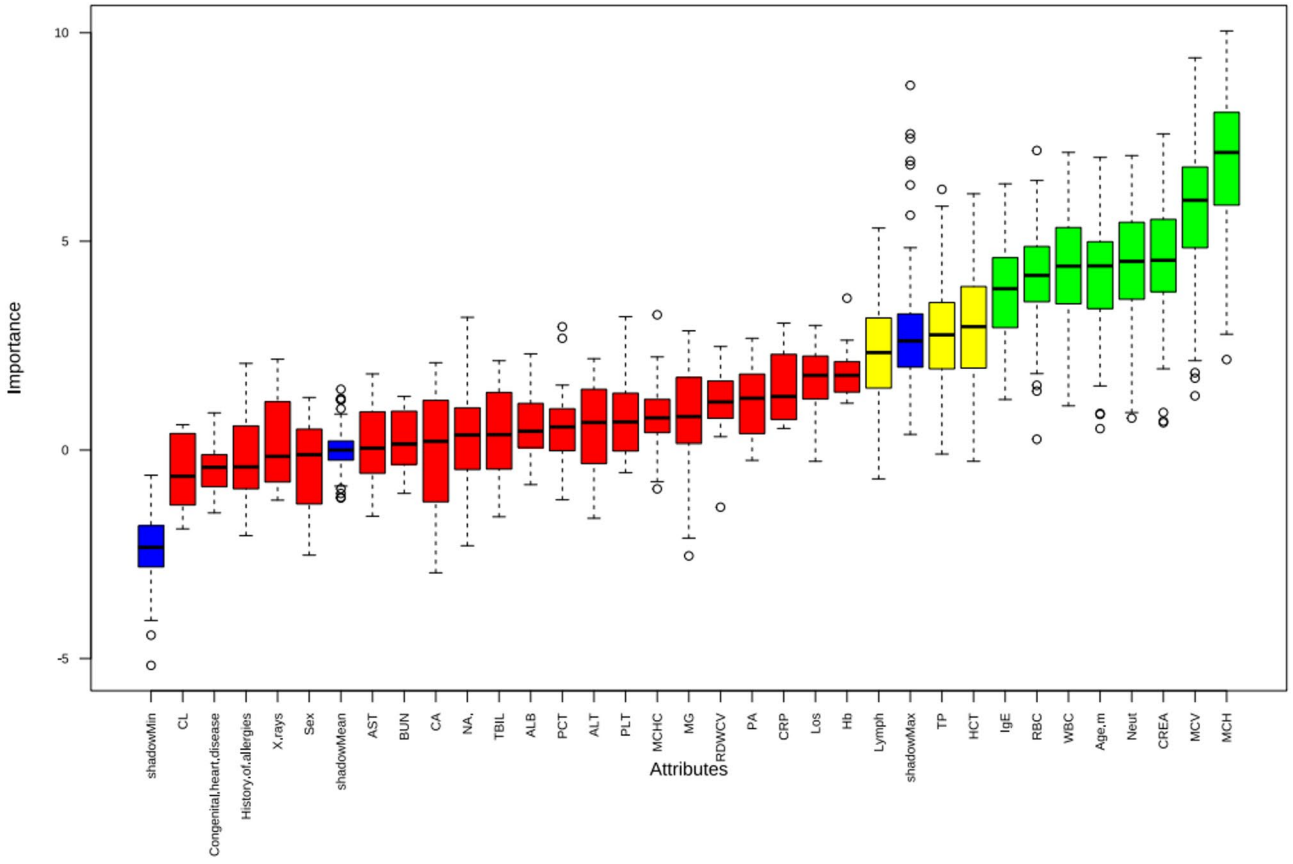

Fig. 2 Feature importance analysis of predictors for RSV-associated bacterial co-infections in RSV patients using the Boruta algorithm. Box plots represent the importance distribution of each feature, ranked from low to high. Features are categorized by color: blue for shadow features (used as controls), red for unimportant variables, yellow for tentative variables, and green for important variables

Table 2 Multivariate logistic regression analysis of factors associated with RSV infection

Variables	Multivariate Logistic Regression Analysis		
	Hazard ratio	95%CI	*P*
Age.m	0.992	0.991 (0.978–1.003)	0.210
IgE	0.999	0.999 (0.998–0.999)	0.019
CREA	1.001	1.001(0.966–1.034)	0.873
TP	0.983	0.983 (0.950–1.017)	0.618
PA	0.994	0.994(0.990 ~ 0.998)	0.002
MCH	0.979	0.979 (0.810–1.076)	0.801
MCV	1.089	1.089 (0.887–1.315)	0.468
Lymph	0.639	0.639 (0.369–1.085)	0.073
Neut	0.691	0.691 (0.409–1.146)	0.111
WBC	1.481	1.481 (0.923–2.412)	0.08
LOS	1.052	1.051(1.010–1.098)	0.015
X rays	1.936	1.936(1.193 ~ 3.142)	0.008

Immunoglobulin E (IgE), Creatinine (CREA), Total Protein (TP), Prealbumin (PA), Mean Corpuscular Hemoglobin (MCH), Mean Corpuscular Volume (MCV), Lymphocyte count (Lymph), Neutrophil count (Neut), White Blood Cell count (WBC), Length of Stay (LOS)

(MP), human parainfluenza virus (HPIV), and human coronavirus (HCoV). However, RSV co-infection with human metapneumovirus (HMPV) was most frequently detected in the 6 to 14 years group, though the difference was marginally non-significant ($P = 0.059$).

Comparison of bacterial detection in RSV-positive patients with bacterial co-infections across different age groups

Significant age-related differences were observed in the distribution of Gram-positive bacterial pathogens among RSV-infected children ($\chi^2 = 10.01$, $P = 0.02$), with the highest detection rate in the 3 to < 6 years group (14.38%). In contrast, the distribution of Gram-negative bacteria did not differ significantly across age groups ($\chi^2 = 2.63$, $P = 0.452$). Among individual pathogens, *Staphylococcus aureus* was most frequently detected in the < 1 year age group (5.34%, $P < 0.001$), while *Streptococcus pneumoniae* showed a significant predominance in the 3 to < 6 years group (13.71%, $P < 0.001$). *Escherichia coli* was exclusively detected in the < 1 year group (2.21%, $P = 0.01$), A detailed breakdown of pathogen distribution across age groups is provided in Table 4.

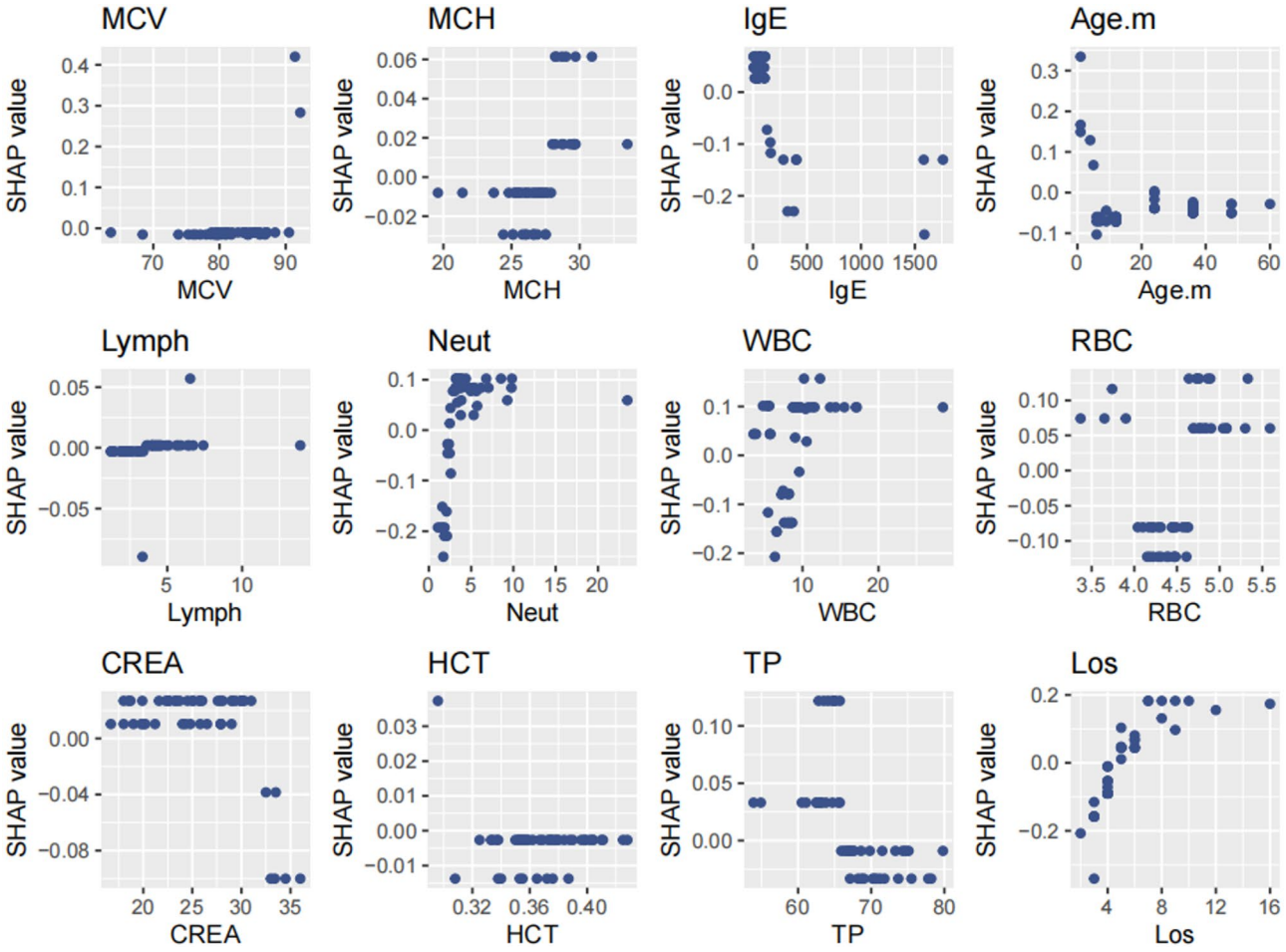

Fig. 3 SHAP summary plot of predictive features for bacterial co-infections in RSV patients. SHAP scatter plots show the impact of individual feature values on model predictions. Each point represents a patient. Features are ordered by importance (top to bottom). Positive SHAP values indicate an increased predicted risk of bacterial infection, while negative values indicate a reduced risk

Table 3 Comparison of RSV co-infection rates across different age groups [n (%)]

Viral pathogen	0 ~ < 1year (*n* = 543)	1 ~ < 3year (*n* = 188)	3 ~ < 6year (*n* = 299)	6 ~ 14year (*n* = 33)	χ^2	*P* value
HRV	64(11.8%)	19(10.11%)	42(14.05%)	4 (12.12%)	1.81	0.612
HBoV	10(1.84%)	5 (2.66%)	7(2.34%)	0 (0.00%)	1.26	0.737
MP	20(3.68%)	11 (5.85%)	19(6.35%)	4 (12.12%)	6.82	0.078
HPIV	14(2.58%)	4 (2.13%)	4(1.34%)	0 (0.00%)	2.18	0.535
HCOV	9(1.66%)	4 (2.13%)	0(0.00%)	0 (0.00%)	6.23	0.101
INFA	8(1.47%)	6 (3.19%)	24 (8.03%)	5(15.15%)	10.37	0.016
HMPV	13 (2.39%)	7 (3.72%)	5 (1.67%)	3(9.09%)	7.44	0.059

Although *Mycoplasma pneumoniae* is an atypical bacterium, it was grouped with viral pathogens in this study due to the shared PCR-based detection methodology. HRV (Human Rhinovirus), HBov (Human Bocavirus), MP(Mycoplasma Pneumonia), HPIV (Human Parainfluenza Virus), HCOV (Human Coronavirus), INFA (Influenza A), HMPV (Human Metapneumovirus)

Imaging detection in children with positive respiratory syncytial virus

Age-stratified analysis of radiographic findings in RSV-positive children with viral co-infections revealed that pulmonary infection (19.66%, $P = 0.004$) and bronchiolitis (5.13%, $P = 0.134$) were more commonly observed in the 3 ~ < 6 years age group. Increased lung markings (7.41%, $P = 0.038$) and normal imaging findings (25.93%, $P = 0.004$) were most frequently seen in the 6 ~ 14 years group. Bronchitis was the most common radiographic abnormality across all age groups (66.67%, $P = 0.027$) (Table 5). In children with RSV and bacterial co-infections, bronchitis was also the predominant radiographic

Table 4 Age-stratified bacterial detection in children with confirmed RSV infection and positive bacterial cultures [n (%)]

Pathogen	0 ~ < 1year (*n* = 543)	1 ~ < 3year (*n* = 188)	3 ~ < 6year (*n* = 299)	6 ~ 14year (*n* = 33)	χ^2	*P* value
G-bacteria	55 (10.13)	18(9.57)	21(6.70)	2(6.06)	2.63	0.452
Klebsiella pneumoniae	2(0.37)	0(0.0)	0(0.0)	0(0.0)	1.92	0.59
Escherichia coli	12(2.21)	0(0.0)	0(0.0)	0(0.0)	11.62	0.01
Acinetobacter baumannii	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81
Moraxella catarrhalis	16(2.95)	13(6.91)	12(4.01)	1(3.03)	5.87	0.12
Haemophilus influenzae	18(3.31)	5(2.66)	8(2.68)	1(3.03)	0.37	0.95
Enterobacter aerogenes	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81
Burkholderia cepacia	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81
Enterobacter cloacae	2(0.37)	0(0.0)	0(0.0)	0(0.0)	1.91	0.59
Serratia marcescens	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81
Pseudomonas aetuginosa	1(0.18)	0(0.0)	1(0.33)	0(0.0)	0.76	0.86
G + bacteria	62(11.42)	11(5.85)	43(14.38)	2(6.06)	10.01	0.02
Staphylococcus aureus	29(5.34)	2(1.06)	2(0.67)	1(3.03)	17.00	< 0.001
Streptococcus pneumoniae	32(5.89)	9(4.79)	41(13.71)	0(0.0)	22.66	< 0.001
Bacillus licheniformis	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81
Streptococcus pyogenes	0(0.0)	0(0.0)	0(0.0)	1(3.03)	31.24	< 0.001
Fungi	1	1	1	0	0.72	0.88
Candida albicans	1(0.18)	1(0.53)	1(0.33)	0(0.0)	0.72	0.87
Bacterial co-infection	2	1	0	0	1.50	0.68
Streptococcus pneumoniae / Moraxella catarrhalis	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81
Streptococcus pneumoniae / Haemophilus influenzae	0(0.0)	1(0.53)	0(0.0)	0(0.0)	4.66	0.20
Haemophilus influenzae / Moraxella catarrhalis	1(0.18)	0(0.0)	0(0.0)	0(0.0)	0.96	0.81

Table 5 Radiographic findings in RSV-positive children with viral co-infections by age group [n (%)]

Indicators	0 ~ < 1 year (*n* = 440)	1 ~ < 3 year (*n* = 137)	3 ~ < 6 year (*n* = 239)	6 ~ 14 year (*n* = 27)	χ^2	*P* Value
Lung infection	68(15.74%)	26(19.55%)	46(19.66%)	0(0.00%)	13.36	0.004
Texture thickening	12(2.78%)	9(6.77%)	6(2.56%)	2(7.41%)	8.42	0.038
Capillary bronchitis	16(3.70%)	4(3.01%)	12(5.13%)	0(0.00%)	5.58	0.134
Bronchitis	261(60.42%)	63(47.37%)	132(56.41%)	18(66.67%)	9.16	0.027
Nothing out of the ordinary	75(17.36)	31(23.31%)	38(16.24%)	7(25.93%)	13.26	0.004
Unchecked	8	4	5	0	-	-

Distribution of chest radiographic findings in RSV-positive pediatric patients with viral co-infections, stratified by age group. Data are presented as counts and percentages (%)

finding, although the differences among age groups were not statistically significant ($P > 0.05$) (Table 6).

Discussion

This study conducted a retrospective descriptive analysis of 1,063 hospitalized children with RSV infection to investigate the epidemiological characteristics, clinical features, and laboratory predictors of bacterial co-infection. In addition, an interpretable machine learning model was developed to enable early risk identification. First, we found that 20.7% of RSV-positive patients had bacterial co-infections, with Streptococcus pneumoniae being the most common pathogen. Second, multivariate logistic regression analysis identified IgE, PA, and LOS as independent factors associated with bacterial co-infection. Third, we established a predictive framework incorporating Boruta-based feature selection, XGBoost modeling, and SHAP interpretability analysis, which improved both the predictive performance and transparency of the model. This approach further confirmed LOS, neutrophil count, WBC count, and IgE level as the most important predictive features. Finally, age-stratified imaging analysis revealed that bronchitis was the most common radiographic manifestation across all age groups. All predictive variables used in this study were derived from routine laboratory testing, enhancing the clinical applicability of our model. These findings offer a feasible and practical tool for the early identification of bacterial co-infection in RSV-positive children and may support more precise risk stratification and individualized clinical interventions.

The association between RSV infection and bacterial co-infection has been inconsistently reported in previous studies. On the one hand, some research has highlighted

Table 6 Radiographic findings in RSV-positive children with bacterial co-infections by age group [n (%)]

Indicators	0 ~ < 1 year (*n* = 103)	1 ~ < 3 year (*n* = 51)	3 ~ < 6 year (*n* = 60)	6 ~ 14 year (*n* = 6)	χ^2	*P* Value
Lung infection	12(11.88%)	7(13.73%)	10(16.95%)	1(20.00%)	0.97	0.807
Texture thickening	3(2.97%)	4(7.84%)	3(5.08%)	0(0.00%)	2.05	0.561
Capillary bronchitis	1(0.99%)	7(13.73%)	3(5.08%)	0(0.00%)	6.47	0.091
Bronchitis	76(75.25%)	27(52.94%)	34(57.63%)	4(80.00%)	4.84	0.184
Nothing out of the ordinary	9(8.91%)	6(11.76%)	9(15.25%)	1(20.00%)	1.53	0.675
Unchecked	2	0	1	1	-	-

Distribution of chest radiographic findings in RSV-positive pediatric patients with bacterial co-infections, stratified by age group. Data are presented as counts and percentages (%)

potential synergistic interactions between RSV and Streptococcus pneumoniae, which may aggravate disease severity [16, 18, 19]. However, most studies have not explored the clinical predictors of such co-infections [20, 21]. On the other hand, some investigations have assessed inflammatory or immunologic markers in RSV-infected children [22, 23], but often relied on univariate analysis or traditional regression models, lacking the ability to assess complex feature interactions. For example, Abhijeet R Sonawane et al. reported a diverse respiratory microbiome in RSV-infected children but did not quantify key risk factors [24]. Acacia Ozturk et al. observed that longer hospital stays may reflect disease severity, but the link to specific laboratory markers remained unclear [25]. You Li et al. emphasized the clinical importance of RSV-bacterial co-infection but did not propose any actionable predictive framework [26]. These studies often relied on statistical associations without offering predictive or practical tools. In contrast, our study employed machine learning methods to identify and interpret multivariate predictors. First, we applied interpretable algorithms to establish a predictive framework with quantified feature importance. Second, we visualized global and individual-level risk contributions using SHAP analysis. Third, our model was constructed entirely from clinically accessible variables, making it more suitable for bedside application. These features differentiate our study from previous literature and represent a step forward in early risk assessment for pediatric RSV infections.

The associations observed between low IgE levels, elevated WBC and neutrophil counts, and bacterial co-infection in RSV-infected children may reflect important underlying immunological mechanisms. IgE is involved in mucosal defense and type 2 helper T cell (Th2) immune responses; its reduction may indicate impaired immune regulation, which could compromise the host's ability to prevent bacterial co-infection. Elevated white blood cells and neutrophils are classic markers of bacterial infection, indicating an activated systemic inflammatory response. RSV infection has also been shown to disrupt epithelial barriers in the airways, facilitating bacterial adherence and co-infection. In addition, prolonged hospitalization may reflect more severe illness or delayed recovery and may increase the risk of nosocomial exposure. These mechanisms likely interact to increase the risk of bacterial co-infection in children with RSV.

A major strength of this study lies in the development of an interpretable risk prediction model using routinely available clinical variables. SHAP analysis allowed us to understand the contribution of individual features to both overall model output and specific patient predictions. For example, in children presenting with markedly decreased IgE and elevated WBC levels on admission, SHAP values indicated a higher predicted risk of bacterial co-infection, suggesting that clinicians may consider early microbiological testing or empiric antibiotic therapy in such cases. Furthermore, the use of machine learning models with interpretable outputs bridges the gap between complex analytics and clinical practice. This approach has the potential to support individualized risk stratification, triage prioritization, and early intervention in pediatric patients with RSV.

This study has several limitations. First, it is a retrospective single-center study, and the findings may not be generalizable to other populations. To mitigate this, we employed multiple imputation to address missing data and used cross-validation to improve model robustness. Second, the dataset did not include treatment variables, such as prior antibiotic use, which may influence co-infection risk. To control for this, we restricted model input to laboratory indicators obtained within 24 h of admission to reflect the early disease state. Third, while SHAP enhances interpretability, the observational nature of this study limits causal inference. Accordingly, we interpreted associations conservatively. In addition, this study did not include an RSV-negative control group, which limits direct comparisons of clinical or laboratory indicators between infected and uninfected patients. Future studies may incorporate such controls to enhance comparative analysis. Future prospective multi-center studies are needed to validate our findings and explore the real-world utility of the proposed model in diverse healthcare settings.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12879-025-11086-z.

Supplementary Material 1

Acknowledgements
We gratefully thank Jie Liu, PhD (Department of Vascular and Endovascular Surgery, Chinese PLA General Hospital), for his helpful review and comments regarding the manuscript.

Author contributions
Q.X. and X.Q. designed the study. Y.X., J.Y., and Z.F did the bioinformatics analysis. T.Y. and Z.F. collected patient information and serum. Y.F. performed the ELISA assay. Q.X and T.Y. written the article. All authors read and approved the final version of the manuscript.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability
Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
The study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the Medical Ethics Committee of The Second Affiliated Hospital of Wenzhou Medical University (Ethical approval number: [2024-K-045-01]). Due to the retrospective nature of the study and the use of anonymized clinical data, the requirement for informed consent was waived by the ethics committee.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Clinical trial
Not applicable.

Author details
[1]Department of Clinical Laboratory, The Second Affiliated Hospital & Yuying Children's Hospital of Wenzhou Medical University, Wenzhou 325000, Zhejiang, China

Received: 19 February 2025 / Accepted: 6 May 2025
Published online: 14 May 2025

References

1. Bem RA, van Woensel JBM. The changing landscape of respiratory syncytial virus infections. Pathogens (Basel Switzerland). 2023;12(10).
2. Duan Y, Liu Z, Zang N, Cong B, Shi Y, Xu L, Jiang M, Wang P, Zou J, Zhang H, et al. Landscape of respiratory syncytial virus. Chin Med J. 2024;137(24):2953–78.
3. Kopera E, Czajka H, Zapolnik P, Mazur A. New insights on respiratory syncytial virus prevention. Vaccines. 2023;11(12).
4. Kuhdari P, Brosio F, Malaventura C, Stefanati A, Orsi A, Icardi G, Gabutti G. Human respiratory syncytial virus and hospitalization in young children in Italy. Ital J Pediatr. 2018;44(1):50.
5. Young M, Smitherman L. Socioeconomic impact of RSV hospitalization. Infect Dis Therapy. 2021;10(Suppl 1):35–45.
6. Zar HJ. New advances in RSV: is prevention attainable? Pediatr Pulmonol. 2025;60(Supppl 1):S120–S122
7. Yang T, Lian H, Liao J, Zeng Y, Li J, Lin C, Lin M. Epidemiological characteristics and meteorological factors of acute respiratory infections (ARIs) in hospitalized children in Eastern Guangdong, China. Sci Rep. 2024;14(1):25518.
8. Anderson J, Do LAH, van Kasteren PB, Licciardi PV. The role of respiratory syncytial virus G protein in immune cell infection and pathogenesis. EBioMedicine. 2024;107:105318.
9. Hu M, Bogoyevitch MA, Jans DA. Impact of respiratory syncytial virus infection on host functions: implications for antiviral strategies. Physiol Rev. 2020;100(4):1527–94.
10. Klink T, Rankin DA, Piya B, Spieker AJ, Faouri S, Shehabi A, Williams JV, Khuri-Bulos N, Halasa NB. Evaluating the diagnostic accuracy of the WHO severe acute respiratory infection (SARI) criteria in middle Eastern children under two years over three respiratory seasons. PLoS ONE. 2020;15(4):e0232188.
11. Rankin DA, Peetluk LS, Deppen S, Slaughter JC, Katz S, Halasa NB, Khankari NK. Diagnostic models predicting paediatric viral acute respiratory infections: a systematic review. BMJ Open. 2023;13(4):e067878.
12. Wang N, Li H, Song JJ, Song JH, Zhang Y. Research progress on detection methods for human respiratory syncytial virus. Chin J Prev Med. 2024;58(8):1143–9.
13. Soni A, Kabra SK, Lodha R. Respiratory syncytial virus infection: an update. Indian J Pediatr. 2023;90(12):1245–53.
14. Pacheco GA, Galvez NMS, Soto JA, Andrade CA, Kalergis AM. Bacterial and viral coinfections with the human respiratory syncytial virus. Microorganisms. 2021;9(6).
15. Zhang XL, Zhang X, Hua W, Xie ZD, Liu HM, Zhang HL, Chen BQ, Chen Y, Sun X, Xu Y, et al. Expert consensus on the diagnosis, treatment, and prevention of respiratory syncytial virus infections in children. World J Pediatrics: WJP. 2024;20(1):11–25.
16. Torres AR, Gaio V, Melo A, Lança M, Barreto M, Gomes L, Azevedo I, Bandeira T, Lito D, Guiomar R, et al. RSV-bacterial co-infection is associated with increased illness severity in hospitalized children - results from a prospective Sentinel surveillance study. J Med Virol. 2025;97(2):e70209.
17. Kassim A, Omuse G, Premji Z, Revathi G. Comparison of clinical laboratory standards Institute and European committee on antimicrobial susceptibility testing guidelines for the interpretation of antibiotic susceptibility at a university teaching hospital in Nairobi, Kenya: a cross-sectional study. Ann Clin Microbiol Antimicrob. 2016;15:21.
18. Manna S, McAuley J, Jacobson J, Nguyen CD, Ullah MA, Sebina I, Williamson V, Mulholland EK, Wijburg O, Phipps S et al. Synergism and antagonism of bacterial-viral coinfection in the upper respiratory tract. mSphere. 2022;7(1):e0098421.
19. Besteman SB, Bogaert D, Bont L, Mejias A, Ramilo O, Weinberger DM, Dagan R. Interactions between respiratory syncytial virus and Streptococcus pneumoniae in the pathogenesis of childhood respiratory infections: a systematic review. Lancet Respiratory Med. 2024;12(11):915–32.
20. Vartiainen P, Jukarainen S, Rhedin SA, Prinz A, Hartonen T, Vabalas A, Viippola E, Rodosthenous RS, Koskelainen S, Liu A, et al. Risk factors for severe respiratory syncytial virus infection during the first year of life: development and validation of a clinical prediction model. Lancet Digit Health. 2023;5(11):e821–30.
21. The Lancet Digital. Equitable prediction of respiratory syncytial virus risk in infants. Lancet Digit Health. 2023;5(11):e749.
22. Okuyan O, Elgormus Y, Dumur S, Sayili U, Uzun H. New generation of systemic inflammatory markers for respiratory syncytial virus infection in children. Viruses. 2023;15(6).
23. Sawatzky J, Soo J, Conroy AL, Bhargava R, Namasopo S, Opoka RO, Hawkes MT. Biomarkers of systemic inflammation in Ugandan infants and children hospitalized with respiratory syncytial virus infection. Pediatr Infect Dis J. 2019;38(8):854–9.
24. Sonawane AR, Tian L, Chu CY, Qiu X, Wang L, Holden-Wiltse J, Grier A, Gill SR, Caserta MT, Falsey AR, et al. Microbiome-transcriptome interactions related to severity of respiratory syncytial virus infection. Sci Rep. 2019;9(1):13824.
25. Ozturk A, Chan M, Khan JR, Hu N, McMullan B, Britton PN, Bartlett A, Kandasamy R, Saravanos GL, Prentice B et al. Clinical characteristics and in-hospital outcomes associated with RSV-ALRI compared to other viral ALRI in hospitalised children under two years old. J Infect Dis. 2024;jiae543. https://doi.org/10.1093/infdis/jiae543

26. Li Y, Wang X, Blau DM, Caballero MT, Feikin DR, Gill CJ, Madhi SA, Omer SB, Simões EAF, Campbell H, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in children younger than 5 years in 2019: a systematic analysis. Lancet (London England). 2022;399(10340):2047–64.

Publisher's note