BMC Pulmonary Medicine

RESEARCH

Open Access

Risk prediction and early intervention strategies for persistent SARS-CoV-2 infection in patients with non-Hodgkin lymphoma: a retrospective cohort study

Wailong Zou[1], Jia Zhang[1], Yulin Li[1], Yuwei Cao[1], Jiaxin Li[1], Zhe Zhang[1], Xin Zhang[1], Chuan Song[1], Rui Yang[1], Yaxin Yan[1], Yumin Wang[1], Xinjun Zhang[1], Zhe Xu[2*] and Jichao Chen[1*]

Abstract

Background Patients with non-Hodgkin lymphoma (NHL) face heightened mortality and accelerated disease progression when persistently infected with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). This critical situation underscores the urgent need to identify risk factors and establish early intervention strategies tailored to this vulnerable population. The primary aim of this study was to investigate the risk factors associated with persistent SARS-CoV-2 infection in NHL patients during the COVID-19 pandemic.

Methods A retrospective cohort study was conducted using data from January 2020 to June 2024, obtained from the Aerospace Center Hospital's database, electronic health records, and laboratory archives. Inclusion criteria comprised patients with confirmed NHL and SARS-CoV-2 infection, with persistence defined as positive viral test results beyond 14 days after initial diagnosis. Patients with incomplete medical records or loss of follow-up were excluded. Predictive models were developed and refined using logistic regression and random forest algorithms. The models incorporated data on demographics, comorbidities, laboratory findings, and imaging results. Model performance was evaluated using accuracy, precision, and the area under the receiver operating characteristic curve (AUC-ROC). Validation was conducted on an independent dataset to ensure generalizability, and the best-performing model guided the development of a prediction tool for early risk assessment and intervention.

Results Key risk factors for persistent SARS-CoV-2 infection in NHL patients included advanced age, hypertension, diabetes, immunosuppressed status, low lymphocyte count, elevated C-reactive protein, high body mass index, anemia, reduced CD4 + cell count, and the presence of lung lesions. The random forest model demonstrated superior predictive performance, achieving an AUC of 0.93. The study further highlighted that prompt antiviral therapy, adjustments to immunosuppressive regimens, and enhanced monitoring significantly reduced infection persistence.

Handling Editor: Jun Hyeok Lim

*Correspondence:
Zhe Xu
15301050459@163.com
Jichao Chen
604939512@qq.com

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

© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

Conclusions This study identifies critical risk factors for persistent SARS-CoV-2 infection in NHL patients and underscores the importance of early intervention strategies. These findings may guide clinical decision-making to improve outcomes in this high-risk population.

Keywords Non-Hodgkin's lymphoma, Persistent SARS-CoV-2 infection, Risk factors

Background

Non-Hodgkin's lymphoma (NHL) is a malignancy of the lymphatic system, and its incidence has been increasing globally [1]. Treatment of NHL includes chemotherapy, immunotherapy, and targeted therapy. However, owing to the immunosuppressive nature of the disease itself and its treatment, patients become more susceptible to infections, particularly viral infections [2]. The global outbreak of the novel coronavirus disease (COVID-19) caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has significantly increased the health risks faced by NHL patients [3]. Therefore, understanding and predicting the risk factors for persistent infection in NHL patients during the COVID-19 pandemic and developing effective early intervention strategies are crucial for these patients [4, 5].

Although several studies have explored the impact of COVID-19 on cancer patients, critical gaps persist, particularly in relation to NHL. For instance, Bravaccini et al. examined the clinical outcomes of COVID-19 in cancer patients but did not delve deeply into NHL-specific implications [6]. Similarly, Yang et al. focused on the general characteristics and outcomes of cancer patients with COVID-19, but NHL-specific concerns were not addressed comprehensively [7, 8]. This study aimed to test the hypothesis that patient characteristics, such as impaired immune function and complications at different NHL stages, are important risk factors for persistent SARS-CoV-2 infection in NHL patients.

This retrospective cohort study endeavored to systematically assess the risk factors for persistent infection in NHL patients during the COVID-19 pandemic and propose targeted early intervention strategies. Our specific objectives included identifying the primary risk factors for persistent infection, evaluating the effects of various intervention measures, and furnishing a scientific rationale for clinical decision-making aimed at minimizing COVID-19-related risks in NHL patients. By addressing these research questions, our findings have the potential to influence clinical practice guidelines and public health policies for managing COVID-19 in immunocompromised patients and ultimately enhance the quality of life and survival rates of NHL patients during the ongoing pandemic.

Methods

Study details

We assembled an expert team comprising specialists in respiratory and critical care medicine, infectious diseases, and statistics. Utilizing a retrospective cohort study design, our objectives were to evaluate the clinical characteristics and identify risk factors for persistent SARS-CoV-2 infection among patients with NHL. Furthermore, we aimed to construct and validate predictive models based on these factors. To achieve these goals, we analyzed historical clinical data and patient follow-up records. The data for this study were sourced from the electronic medical record system of the Aerospace Center Hospital, which encompasses over 10 million outpatients and inpatients who visited the hospital between January 2020 and June 2024.

Sample selection criteria

The inclusion criteria were as follows: Patients who met the NHL diagnostic criteria and were confirmed to have SARS-CoV-2 infection between January 2020 and June 2024, with persistent PCR positivity that was not attributed to Paxlovid rebound (specifically, excluding those who showed symptom recurrence or PCR positivity between 2 and 8 days post-Paxlovid treatment completion). The exclusion criteria were as follows: Patients with other severe illnesses that could affect their expected survival.

Sample grouping

- Persistent infection group: NHL patients confirmed with persistent SARS-CoV-2 infection, defined as virus RNA detected ≥ 14 days after symptom onset or initial positive test result. During the survey, samples were collected weekly for the first 2 weeks of enrollment.
- Non-persistent infection group: NHL patients were confirmed to have SARS-CoV-2 infection but did not develop a persistent infection.

Data collection

Basic patient information, clinical characteristics, laboratory test results, imaging findings, treatments, and follow-up records were extracted. Based on recent literature related to persistent SARS-CoV-2 infection in oncology patients and expert panel recommendations, variables specifically associated with SARS-CoV-2 infection were

prioritized, including age, sex, underlying diseases (e.g., hypertension and diabetes), immunosuppressive status, lymphocyte count, C-reactive protein (CRP) level, body mass index (BMI), anemia status, CD4 + cell count, and lung lesions. CD4 + cell count was determined using flow cytometry, IgG levels were measured using ELISA, and viral load was quantified using real-time quantitative PCR (RT-qPCR).

Data preprocessing

- Missing data and outliers were addressed and corrected.
- Continuous variables were standardized.
- Categorical variables were encoded for analysis.

Research design

Clinical characteristics and risk factor analysis

- Data were collected from eligible NHL patients between January 2020 and June 2024.
- Random forest feature importance analysis was employed to identify the primary risk factors contributing to persistent SARS-CoV-2 infection.

Prediction model development and validation

- After expert panel discussions and literature review, approximately 600 patients were deemed necessary for the study based on the prevalence of persistent SARS-CoV-2 infection in NHL patients and sample size calculations. Consequently, 660 eligible subjects from January 2020 to June 2023 were included, with 110 and 550 cases of persistent and non-persistent infections, respectively, for model training and initial validation.
- From July 2023 to June 2024, an additional 450 NHL patients infected with SARS-CoV-2 were recruited for model validation.
- Multiple machine learning algorithms (e.g., logistic regression and random forest) were utilized to model and train the dataset.
- Model performance was evaluated using sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve (AUC-ROC) in the training set. The best-performing model was further assessed in the validation set to ensure its robustness to the new data.

Intervention strategy development

Based on the feature importance analysis and clinical expert input, personalized intervention strategies for high-risk NHL patients were formulated, encompassing intensified monitoring, preventive treatments, lifestyle modifications, AI-assisted health management, and multidisciplinary collaboration.

Statistical analyses

SPSS 25.0 (IBM Corp., Armonk, NY, USA), or R software version 3.2.2 (version 3.2.2; R Foundation for Statistical Computing, Vienna, Austria), was used for statistical analyses. Continuous variables were analyzed using t-tests or Mann–Whitney U tests, and categorical variables were analyzed using chi-square or Fisher's exact tests. Statistical significance was set at $p < 0.05$. After conducting a thorough power analysis, we have determined that, to achieve the desired statistical power (set at 0.8) and given significance level (0.05), and taking into account the expected effect size (i.e., the difference in the risk of persistent SARS-CoV-2 infection among NHL patients) as well as the heterogeneity of the patient population, approximately 50 to 300 subjects are needed per group. This sample size is intended to ensure that we can accurately assess and predict the risk of persistent SARS-CoV-2 infection in NHL patients, while also proposing effective early intervention strategies. This sample size is based on currently available information and assumptions and may need to be adjusted appropriately according to specific circumstances in actual research.

Ethical considerations

This study was approved by the Ethics Committee of Aerospace Central Hospital, the Ethical Mention Approval Number: Jinghang Medical Ethics Review 2022 No. (075). Given that it is a retrospective study, after comprehensively considering the nature, purpose, potential risks and benefits of this research, as well as the rights and privacy of patients, informed consent is exempted. Strict confidentiality measures were implemented to protect patient privacy, and all data were used solely for this study.

Expert consultation and discussion

Experts in relevant fields were consulted and engaged in discussions to validate and assess the study findings and offer suggestions for optimizing the diagnostic and treatment strategies.

Results

Basic characteristics and treatment regimens

Table 1 outlines the basic patient characteristics and treatment plans for persistent and non-persistent SARS-CoV-2 infections in the experimental cohort.

Table 1 Basic characteristics and treatment regimens in patients with NHL and persistent vs. non-persistent SARS-CoV-2 infections

Characteristics	Persistent SARS-CoV-2 Infection Group ($n=110$)	Non-persistent SARS-CoV-2 Infection Group ($n=550$)	*p*-value
Age (years)	55.2 ± 12.8	45.1 ± 10.5	< 0.001
Sex (M/F)	60/50	277/273	0.32
Hypertension	71 (64.5%)	192 (34.9%)	< 0.001
Diabetes	33 (30.0%)	108 (19.6%)	0.002
Immunosuppressed State	76 (69.1%)	76 (13.8%)	< 0.001
Lymphocyte Count ($\times 10^9$/L)	0.7 ± 0.4	1.2 ± 0.6	< 0.001
C-Reactive Protein (mg/L)	53.5 ± 14.1	20.3 ± 10.5	< 0.001
Body Mass Index (BMI)	27.5 ± 4.1	24.3 ± 3.8	< 0.001
Average Disease Duration (d)	28.0 ± 12.1	10.0 ± 4.1	< 0.001
Renal Insufficiency	49 (44.5%)	53 (9.6%)	< 0.001
Anemia	35 (31.8%)	78 (14.2%)	< 0.001
COVID-19 Vaccination	89 (80.9%)	431 (78.4%)	0.34
Number of Vaccinations	2.5 ± 0.8	2.7 ± 0.6	0.08
Comparison of Treatment Regimens			
Antiviral Treatment	91 (82.7%)	407 (74.0%)	0.55
Immunomodulators	46 (41.8%)	207 (37.6%)	0.32
Respiratory Support	18 (16.4%)	82 (14.9%)	0.68
Steroid Treatment	25 (22.7%)	106 (19.3%)	0.38
Anticoagulation Therapy	48 (43.6%)	211 (38.4%)	0.15
Comparison of Specific Antiviral Drugs			
Nirmatrelvir/Ritonavir	51 (46.4%)	247 (44.9%)	0.98
Xenutide/Ritonavir	29 (26.4%)	148 (26.9%)	0.45
Molnupiravir	26 (23.6%)	99 (18.0%)	0.40
Azvudine	14 (12.7%)	77 (14.0%)	0.75

The results are shown as mean ± SD or number (%)

COVID-19, coronavirus disease; NHL, NHL, non-Hodgkin's lymphoma; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

Comparison of clinical symptoms, disease course, and laboratory and imaging findings between persistent and non-persistent NHL patients with SARS-CoV-2 infections

Table 2 compares the differences in clinical symptoms, disease course, and laboratory and imaging findings between NHL patients with persistent and non-persistent SARS-CoV-2 infections. The results indicated no significant differences in clinical symptoms between the groups ($p > 0.05$). The listed symptoms reflect the main symptoms observed throughout the patients' entire illness, rather than just those prior to hospital admission. Additionally, the table presents detailed information on the disease course and complications, as well as significant differences in CD4 + T cell counts and incidence of lung lesions based on laboratory tests.

Univariate analysis

Univariate analysis identified age, hypertension, diabetes, immunosuppression, lymphocyte count, CRP level, BMI, anemia, CD4 + cell count, and lung lesions as significant risk factors for persistent SARS-CoV-2 infection.

Multivariate regression analysis

Multivariate logistic regression analysis further confirmed hypertension, immunosuppression, lymphocyte count, and lung lesions as independent risk factors (Table 3).

Regression equation and risk-scoring table

A regression equation and risk-scoring table (Table 4) were constructed to predict persistent SARS-CoV-2 infection based on significant risk factors:

$$\text{Log}\left(\frac{p}{1-p}\right) = -0.7838 + 1.053 \quad \text{Hypertension} + 1.288$$

Immunosuppression+ (– 0.å755) Lymphocyte count + 1. å391 Lung lesions

Nomogram for predicting persistent SARS-CoV-2 infection in patients with NHL

As shown in Fig. 1, we created a forest plot based on the results of the multivariate logistic regression analysis, illustrating the associations between individual independent risk factors and persistent SARS-CoV-2 infection.

AUC curve

We assessed the predictive accuracy of the multivariate logistic regression model by plotting the ROC curve and calculating the AUC. The results are shown in Fig. 2.

Table 2 Clinical characteristics, disease course, complications, and laboratory/imaging findings in NHL patients with persistent vs. non-persistent SARS-CoV-2 infections

Indicator/Symptom	Persistent SARS-CoV-2 Infection Group ($n=110$)	Non-persistent SARS-CoV-2 Infection Group ($n=550$)	*p*-value
Clinical Symptoms			
Fever	60 (54.5%)	297 (54.0%)	0.28
Average Temperature (°C)	38.2 ± 0.5	38.1 ± 0.4	0.18
Cough	66 (60.0%)	325 (59.1%)	0.21
Dyspnea	13 (11.8%)	57 (10.4%)	0.18
Oxygen Therapy Required	7 (6.4%)	36 (6.5%)	0.43
Fatigue	34 (30.9%)	161 (29.3%)	0.34
Loss of Smell/Taste	2 (1.8%)	11 (2.0%)	0.09
Disease Course			
Average Disease Duration (d)	28.0 ± 12.1	10.0 ± 4.1	< 0.001
Median Disease Duration (d)	32	5	
Range of Disease Duration (d)	15–90	3–14	
Complications			
Myocarditis (n, %)	4 (3.6%)	19 (3.5%)	0.28
Thromboembolism (n, %)	2 (1.8%)	10 (1.8%)	0.13
Acute Kidney Injury (n, %)	2 (1.8%)	9 (1.6%)	0.37
Blood Parameters			
White Blood Cell Count ($\times 10^9$/L)	6.2 ± 1.8	5.9 ± 1.6	0.06
D-Dimer (μg/mL)	0.9 ± 0.4	0.8 ± 0.4	0.08
Infection Characteristics			
Duration of Persistent Infection (d)	30.2 ± 15.1	5.1 ± 4.1	< 0.01
Viral Load (Mean Ct Value)	34.3 ± 3.5	35.0 ± 3.2	0.41
Duration of Symptoms (d)	11.1 ± 7.1	10.2 ± 6.1	0.32
Recurrence	15 (13.6%)	69 (12.5%)	0.67
Immune Response Indicators			
CD4 + T Cell Count (cells/μL)	420.8 ± 200.7	800.4 ± 150.9	< 0.001
IgG Level (g/L)	11.7 ± 2.1	12.5 ± 2.5	0.15
Imaging Findings			
Lung Lesions (n, %)	95 (86.4%)	160 (29.1%)	< 0.001
Bilateral Lung Lesions (n, %)	86 (78.2%)	146 (26.5%)	< 0.001
Unilateral Lung Lesion (n, %)	9 (8.2%)	14 (2.5%)	< 0.01

NHL, Non-Hodgkin's lymphoma; SARS-CoV-2: Severe acute respiratory syndrome coronavirus 2

Table 3 Independent risk factors for persistent SARS-CoV-2 infection in NHL patients

Risk Factor	Persistent SARS-CoV-2 Group ($n=110$)	Non-persistent SARS-CoV-2 Group ($n=550$)	Adjusted OR (95% CI)	Adjusted *p*-value
Hypertension (n, %)	71 (64.5%)	192 (34.9%)	2.21 (1.61–3.03)	< 0.001
Immunosuppressive Status (n, %)	76 (69.1%)	76 (13.8%)	5.72 (3.89–8.41)	< 0.001
Lymphocyte Count ($\times 10^9$/L, mean ± SD)	0.7 ± 0.4	1.2 ± 0.6	0.45 (0.32–0.63)	< 0.001
Lung Lesions (n, %)	95 (86.4%)	160 (29.1%)	5.29 (3.86–7.25)	< 0.001

CI, confidence interval; NHL, NHL, non-Hodgkin's lymphoma; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

Model's performance in predicting persistent SARS-CoV-2 infections in patients with NHL

We evaluated the model's performance using a validation set of 450 patients. The model had 85% accuracy, 78% sensitivity, 88% specificity, and AUC value of 0.93.

Discussion

Recent studies hint at a higher susceptibility and severity of COVID-19 among NHL patients but do not provide definitive answers to several pertinent questions specific to NHL patients [8]. Key unanswered questions include the following. How does NHL impact the susceptibility and severity of COVID-19? Does COVID-19 accelerate NHL progression or affect treatment effectiveness? What is the mortality rate among NHL patients with COVID-19? What are the long-term health consequences? What are best practices for managing NHL in patients with COVID-19? Understanding and accurately

Table 4 Risk scoring table

Risk Factor	Rounded Coefficient	Score
Hypertension	1.053	1
Immunosuppression	1.288	1
Lymphocyte count ($\times 10^9$/L, note negative direction)	−0.755	−1
Lung lesions	1.391	1

Notes

1 Lymphocyte Count: The coefficient is negative, indicating that a higher lymphocyte count is associated with lower risk. Therefore, when calculating the total score, if the lymphocyte count is above a certain threshold (which needs to be determined based on actual data), a negative score is given (i.e., subtract 1 point). If it is below the threshold, no score is given or a positive score is given based on specific circumstances (but in this table, for simplification, we assume that all patients' lymphocyte counts are below this threshold, so they all lose 1 point). However, in practical applications, a more refined approach may be to assign scores based on specific lymphocyte count values, such as dividing into several ranges, each corresponding to a different score

2 Other Risk Factors: For hypertension, immunosuppression, and lung lesions, if the patient has the risk factor, the corresponding score is given

Risk Score Calculation:

Add the scores for each risk factor the patient has, to obtain the total risk score

Note that since the score for lymphocyte count is negative, when calculating the total score, if the lymphocyte count is above the threshold, the negative score needs to be subtracted from the total score (i.e., add 1 point because negative times negative equals positive). But in this simplified table, we assume that all patients' lymphocyte counts result in them losing 1 point

Risk Level Classification

The higher the total score, the higher the risk of the patient developing persistent SARS-CoV-2 infection

Different thresholds can be set based on the total score to classify risk levels, such as:

Total score ≤ 1: low risk

2 ≤ Total score ≤ 3: moderate risk

Total score ≥ 4: high risk

Please note that the above risk level classification is based on a simplified risk scoring table and may need to be adjusted in practical applications based on more detailed data and clinical judgment

Important Note:

This risk scoring table is only an example, and it needs to be validated and optimized based on specific research data and clinical backgrounds when used in practice

When using this table, please be sure to consider other factors that may affect the patient's risk and assess the patient's overall health status and medical history comprehensively

predicting the risk factors for persistent SARS-CoV-2 infection in patients with NHL will facilitate the formulation of effective early intervention strategies, which can significantly mitigate adverse clinical outcomes. Our study not only provides a scientific foundation for clinical treatment decisions but also serves as pivotal support for the development of public health prevention and control strategies.

In this innovative study, we employed a multifaceted approach, integrating various research methods to comprehensively analyze the clinical characteristics and risk factors associated with persistent SARS-CoV-2 infection in NHL patients. This comprehensive analysis enabled us to construct a robust and reliable prediction model that facilitates early identification and intervention in high-risk NHL patients.

The novelty of our study lies in its comprehensive and integrated approach, which distinguishes it from previous studies. By leveraging a diverse array of research methods and analyzing a wide range of clinical data, we have developed a prediction model that holds promise for significantly improving patient outcomes and advancing our understanding of persistent SARS-CoV-2 infection in the NHL patient population.

Since its first report in Wuhan, China, at the end of 2019, SARS-CoV-2 has spread rapidly, leading to a global COVID-19 pandemic. While the virus is typically cleared from the respiratory tract within 9 days of symptom onset, viral RNA can be detected for an average of 17 days [9]. However, some case reports have indicated that viral RNA can persist in the upper respiratory tract and feces for several months [9]. This abnormally prolonged viral presence, known as "persistent infection," is generally defined as the detection of viral RNA ≥ 14 days after symptom onset or the first positive test [9]. SARS-CoV-2 not only directly attacks the respiratory mucosa but can also increase the severity and duration of infection in patients with NHL by disrupting their immune system. As recently reported, immunosuppressive therapy, severity of underlying diseases, and viral load may be associated with persistent SARS-CoV-2 infection [10]. Furthermore, early intervention measures, such as intensified monitoring, antiviral therapy, and personalized management strategies, have been shown to reduce persistent infections and improve patient outcomes [11].

In patients with NHL, persistent SARS-CoV-2 infection can lead to more severe consequences, including higher hospitalization rates, intensive care requirements, and mortality [12]. NHL is a highly heterogeneous hematological malignancy, and its incidence increases with age. Using univariate analysis, we identified age, hypertension, diabetes, immunosuppressive status, lymphocyte count, CRP level, BMI, anemia, CD4 + cell count, and lung lesions to be the main risk factors for persistent SARS-CoV-2 infection, consistent with the findings of previous studies [13]. The immune function of patients with NHL is often suppressed by the disease itself or the treatment, making them more susceptible to SARS-CoV-2 infection and more likely to develop a persistent infection [14]. As reported, 70% of patients with NHL receiving immunosuppressive therapy experience immune suppression, which increases the risk of persistent infection [15]. Additionally, underlying conditions such as hypertension and diabetes also increase the risk of severe disease progression after infection. In patients with NHL, persistent infections have been reported in 65% of those with hypertension and in 30% with diabetes [16]. Chronic inflammatory states, anemia, low lymphocyte counts, and high CRP levels are all closely associated with the occurrence of persistent infection [17].

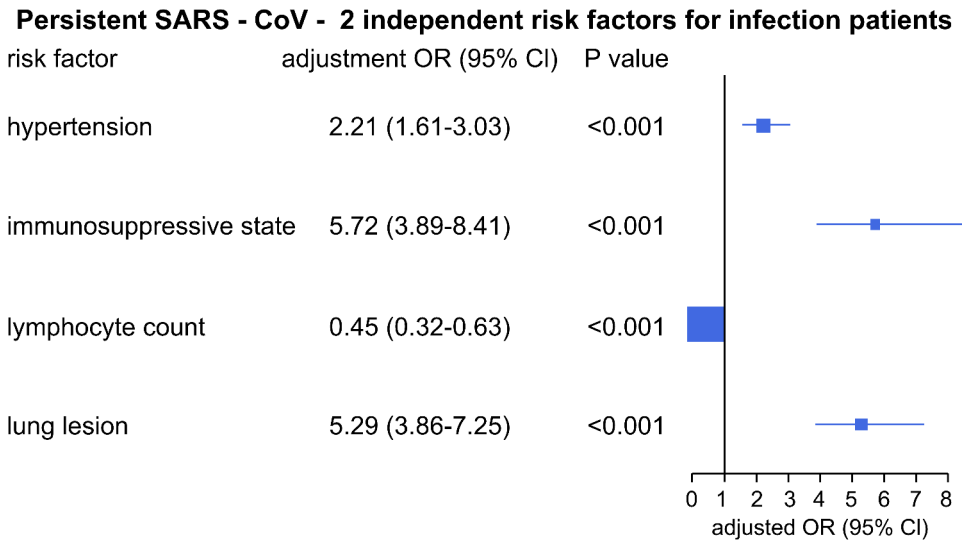

Fig. 1 Independent risk factors for persistent SARS-CoV-2 infection in patients with NHL. NHL, non-Hodgkin's lymphoma; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2 In the forest plot, the boxes for each risk factor are all positioned to the right of 1, and their horizontal lines do not cross 1, indicating that these factors significantly increase the risk of persistent SARS-CoV-2 infection. The box for lymphocyte count is positioned to the left of 1, indicating a negative correlation with persistent SARS-CoV-2 infection, and its confidence interval also does not cross 1, demonstrating statistical significance

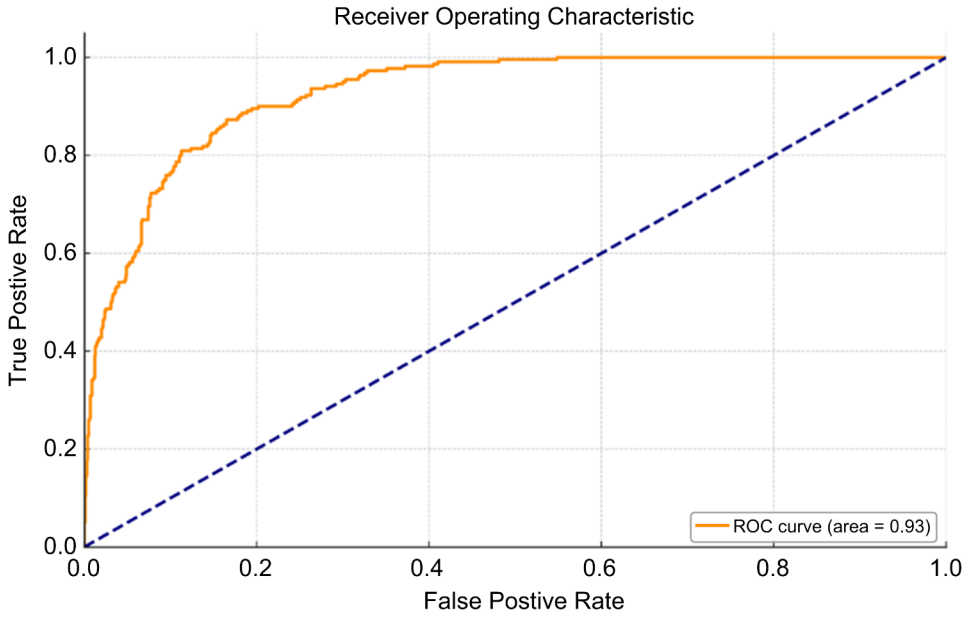

Fig. 2 ROC curve for predicting the risk of persistent SARS-CoV-2 infection in NHL patients NHL, non-Hodgkin's lymphoma; ROC, Receiver Operating Characteristic; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2

Patients with persistent SARS-CoV-2 infection have significantly lower lymphocyte counts (averaging $0.8 \pm 0.5 \times 10^9$/L) [18], significantly higher CRP levels (averaging 50.4 ± 15.2 mg/L) [19], and significantly reduced CD4 + cell counts (averaging 450 ± 110 cells/μL) [20] when compared to those in patients without persistent infection.

Among patients with persistent infections, 85.5% presented with lung lesions [21]. Based on the multivariate logistic regression analysis, we identified hypertension, immunosuppressive status, low lymphocyte count, and lung lesions as independent risk factors for persistent SARS-CoV-2 infection in patients with NHL. Our findings are consistent with those of recent studies. For example, a 2021 systematic review indicated that hypertension is associated with increased severity and mortality in COVID-19 [22]. Immunosuppressive status, as seen in patients receiving chemotherapy and biological

agents, has also been confirmed as an important risk factor for persistent COVID-19 and poor prognosis [23]. A low lymphocyte count, reflecting impaired immune function, was also reported to be a poor prognostic factor in COVID-19 patients in 2022 [24]. Lung lesions, as markers of COVID-19 severity, have been proven by multiple studies to correlate with long-term viral load [25, 26].

We constructed a prediction model for persistent SARS-CoV-2 infection in patients with NHL using a random forest algorithm, based on a comprehensive dataset encompassing clinical information, laboratory test results, and treatment records. The model's efficacy was evaluated using a validation set of 450 patients, yielding impressive results.

The validation process showed that the model demonstrated an accuracy of 85%, correctly predicting persistent infection in 383 out of 450 patients, sensitivity of 78%, indicating its ability to accurately identify patients with actual persistent infection; and specificity of 88%, indicating its ability to correctly distinguish patients without persistent infection. The area under the ROC curve was 0.93, further confirming the model's robust performance in differentiating high- and low-risk patients.

These findings highlight the potential of our model as a valuable decision-support tool for clinicians. By precisely identifying NHL patients more prone to developing persistent SARS-CoV-2 infection, the model could enable timely implementation of aggressive treatments or preventive measures, ultimately improving patient outcomes. The seamless integration of this model into electronic medical record systems or clinical decision-support systems could facilitate real-time automated risk assessments. Additionally, its alignment with existing clinical guidelines and treatment protocols offers a scientific foundation for personalized treatment plans.

Despite the model's strong performance, several limitations merit attention. These include the representativeness of the dataset, comprehensiveness of feature selection, and risk of model overfitting. Addressing these limitations requires further research and refinement. Future studies should aim to include larger sample sizes, explore additional biomarkers as predictive features, and develop dynamic prediction models to further enhance the model's predictive accuracy and clinical applicability.

This study has some limitations. First, it had a relatively small sample size, which may have affected the external validity of the results. Second, the retrospective study design may have resulted in incomplete information, reporting bias, data bias, and omission of confounding factors, potentially affecting the universality and accuracy of the findings. Third, data derived from single-center or region-specific patients may limit the generalizability of the results. Finally, the inability to control for all potential confounding factors may have affected the interpretation of the results.

Conclusions

This study aimed to explore the key risk factors for persistent SARS-CoV-2 infection among patients with NHL. Through in-depth analysis, we successfully identified these risk factors, providing important insights into the mechanisms underlying persistent SARS-CoV-2 infection in NHL patients. The results of this study provide a reference for clinical practice and will help clinicians more accurately assess the infection risk of NHL patients and take corresponding preventive measures to reduce the incidence of infection. The risk factors identified in this study serve as a foundation for future research to further explore SARS-CoV-2 infection in NHL patients. Although we do not directly propose early intervention strategies, the identified risk factors provide a scientific basis for formulating targeted prevention and control measures. For NHL patients, especially those with high-risk factors, enhanced monitoring, personalized treatment recommendations, and health education should be provided to improve their self-management abilities and reduce the risk of infection.

Abbreviations

AUC-ROC	Area Under the Receiver Operating Characteristic Curve
BMI	Body Mass Index
COVID-19	Coronavirus Disease 2019
CRP	C-Reactive Protein
Ct	Cycle threshold
ELISA	Enzyme-Linked Immunosorbent Assay
IgG	Immunoglobulin G
NHL	Non-Hodgkin's Lymphoma
PCR	Polymerase Chain Reaction
RT-qPCR	Real-Time quantitative Polymerase Chain Reaction
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SPSS	Statistical Package for the Social Sciences
WBC	White Blood Cell

Acknowledgements

The authors would like to thank the patients; their families; and all the researchers, including physicians, nursing staff, and laboratory technicians, who participated in this study.

Author contributions

WLZ, JZ, YLL, ZX, and JCC conceived the study and wrote the manuscript. YWC and JXL collected data and critically revised the manuscript. ZZ, XZ, and CS performed statistical analyses. YXY, YMW, XJZ, and RY collected data. All the authors have read and approved the final version of the manuscript. JZ, YLL, and CS visited and verified the data, and WLZ, ZX, JZ, and JCC were responsible for the decision to submit the manuscript. All authors read and approved the final manuscript.

Funding

This study was supported by the National Natural Science Foundation of China (grant no. 82300089). This funding source played a crucial role in facilitating the research process for this study. Specifically, they provided essential resources for conducting the study, including financial support for data collection, analysis, and interpretation. However, the funding source had no direct involvement in the study design, writing of the report, or decision to submit the paper for publication. The authors maintained complete independence and autonomy in all aspects of the research and reporting process. Therefore, any conclusions or recommendations presented in this

study are solely those of the authors and do not necessarily reflect the views or opinions of the funding source.

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
The study protocol was approved by the Institutional Review Board of the Aerospace Center Hospital (no. 2022-075) and was conducted according to the tenets of the Declaration of Helsinki. The need for informed consent was waived because of the retrospective nature of the study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Pulmonary and Critical Care Medicine, Aerospace Center Hospital, Beijing 100049, China
[2]Department of Infectious Diseases, The Fifth Medical Center of Chinese PLA General Hospital, Beijing 100039, China

Received: 24 November 2024 / Accepted: 24 January 2025
Published online: 15 March 2025

References

1. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-hodgkin lymphoma: the Lugano classification. J Clin Oncol. 2014;32:3059–68.
2. Kuderer NM, Choueiri TK, Shah DP, Shyr Y, Rubinstein SM, Rivera DR, et al. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. Lancet. 2020;395:1907–18.
3. He W, Chen L, Chen L, Yuan G, Fang Y, Chen W, et al. COVID-19 in persons with haematological cancers. Leukemia. 2020;34:1637–45.
4. Saito M, Mori A, Ishio T, Kobayashi M, Tsukamoto S, Kajikawa S, et al. Initial efficacy of the COVID-19 mRNA vaccine booster and subsequent breakthrough omicron variant infection in patients with B-cell non-hodgkin's lymphoma: a single-center cohort study. Viruses. 2024;16:328.
5. Roschewski M, Lionakis MS, Sharman JP, Roswarski J, Goy A, Monticelli MA, et al. Inhibition of Bruton tyrosine kinase in patients with severe COVID-19. Sci Immunol. 2020;5:eabd0110.
6. Bravaccini S, Nicolini F, Zanoni M, Gaimari A, Cerchione C, Maltoni R, et al. Why the complications of COVID-19 patients differ in elderly and young cancer patients. Transl Oncol. 2022;26:101541.
7. Yang F, Shi S, Zhu J, Shi J, Dai K, Chen X. Clinical characteristics and outcomes of cancer patients with COVID-19. J Med Virol. 2020;92:2067–73.
8. Teixeira Ferreira R, Cardoso Ferreira I, Carmona S, Montalvão A, Santos AI. Spontaneous remission of high-grade non-hodgkin lymphoma after SARS-CoV-2 infection: a case report. Clin Nucl Med. 2024;49:e77–9.
9. Machkovech HM, Hahn AM, Garonzik Wang J, Grubaugh ND, Halfmann PJ, Johnson MC, et al. Persistent SARS-CoV-2 infection: significance and implications. Lancet Infect Dis. 2024;24:e453–62.
10. Avanzato VA, Matson MJ, Seifert SN, Pryce R, Williamson BN, Anzick SL, et al. Case study: prolonged infectious SARS-CoV-2 shedding from an asymptomatic immunocompromised individual with cancer. Cell. 2020;183:1901–12..e9.
11. Aydillo T, Gonzalez-Reiche AS, Aslam S, van de Guchte A, Khan Z, Obla A, et al. Shedding of viable SARS-CoV-2 after immunosuppressive therapy for cancer. N Engl J Med. 2020;383:2586–8.
12. D'Abramo A, Vita S, Maffongelli G, Beccacece A, Agrati C, Cimini E, et al. Clinical management of patients with B-cell depletion agents to treat or prevent prolonged and severe SARS-COV-2 infection: defining a treatment pathway. Front Immunol. 2022;13:911339.
13. Corcione S, Lupia T, Raviolo S, Montrucchio G, Trentalange A, Curtoni A, et al. Putative invasive pulmonary aspergillosis within medical wards and intensive care units: a 4-year retrospective, observational, single-centre study. Intern Emerg Med. 2021;16:1619–27.
14. Tripathy AS, Trimbake D, Suryawanshi PV, Tripathy SP, Gurav YK, Potdar VA, et al. Peripheral lymphocyte subset alteration in patients with COVID-19 having differential clinical manifestations. Indian J Med Res. 2022;155:136–47.
15. Paces J, Strizova Z, Smrz D, Cerny J. COVID-19 and the immune system. Physiol Res. 2020;69:379–88.
16. Brunetti NS, Davanzo GG, de Moraes D, Ferrari AJR, Souza GF, Muraro SP, et al. SARS-CoV-2 uses CD4 to infect T helper lymphocytes. eLife. 2023;12:e84790.
17. Vojdani A, Vojdani E, Saidara E, Maes M, Persistent. SARS-CoV-2 infection, EBV, HHV-6 and other factors may contribute to inflammation and autoimmunity in long COVID. Viruses. 2023;15:400.
18. Nasserie T, Hittle M, Goodman SN. Assessment of the frequency and variety of persistent symptoms among patients with COVID-19: a systematic review. JAMA Netw Open. 2021;4:e2111417.
19. Li Y, Li H, Song C, Lu R, Zhao Y, Lin F, et al. Early prediction of disease progression in patients with severe COVID-19 using C-reactive protein to albumin ratio. Dis Markers. 2021;2021:6304189.
20. Meckiff BJ, Ramírez-Suástegui C, Fajardo V, Chee SJ, Kusnadi A, Simon H, et al. Imbalance of regulatory and cytotoxic SARS-CoV-2-Reactive $CD4^+$ T cells in COVID-19. Cell. 2020;183:1340–e135316.
21. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid-mechanisms, risk factors, and management. BMJ. 2021;374:n1648.
22. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med. 2020;8:e21.
23. Lee LYW, Cazier JB, Angelis V, Arnold R, Bisht V, Campton NA, et al. COVID-19 mortality in patients with cancer on chemotherapy or other anticancer treatments: a prospective cohort study. Lancet. 2020;395:1919–26.
24. Diao B, Wang C, Tan Y, Chen X, Liu Y, Ning L, et al. Reduction and functional exhaustion of T cells in patients with coronavirus disease 2019 (COVID-19). Front Immunol. 2020;11:827.
25. Xie J, Wu W, Li S, Hu Y, Hu M, Li J, et al. Clinical characteristics and outcomes of critically ill patients with novel coronavirus infectious disease (COVID-19) in China: a retrospective multicenter study. Intensive Care Med. 2020;46:1863–72.
26. Guo T, Fan Y, Chen M, Wu X, Zhang L, He T, et al. Cardiovascular implications of fatal outcomes of patients with coronavirus disease 2019 (COVID-19). JAMA Cardiol. 2020;5:811–8.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.