

Construction and Validation of a Clinical Model to Predict Lymph Node Metastasis in Early Hypopharyngeal Squamous Cell Carcinoma

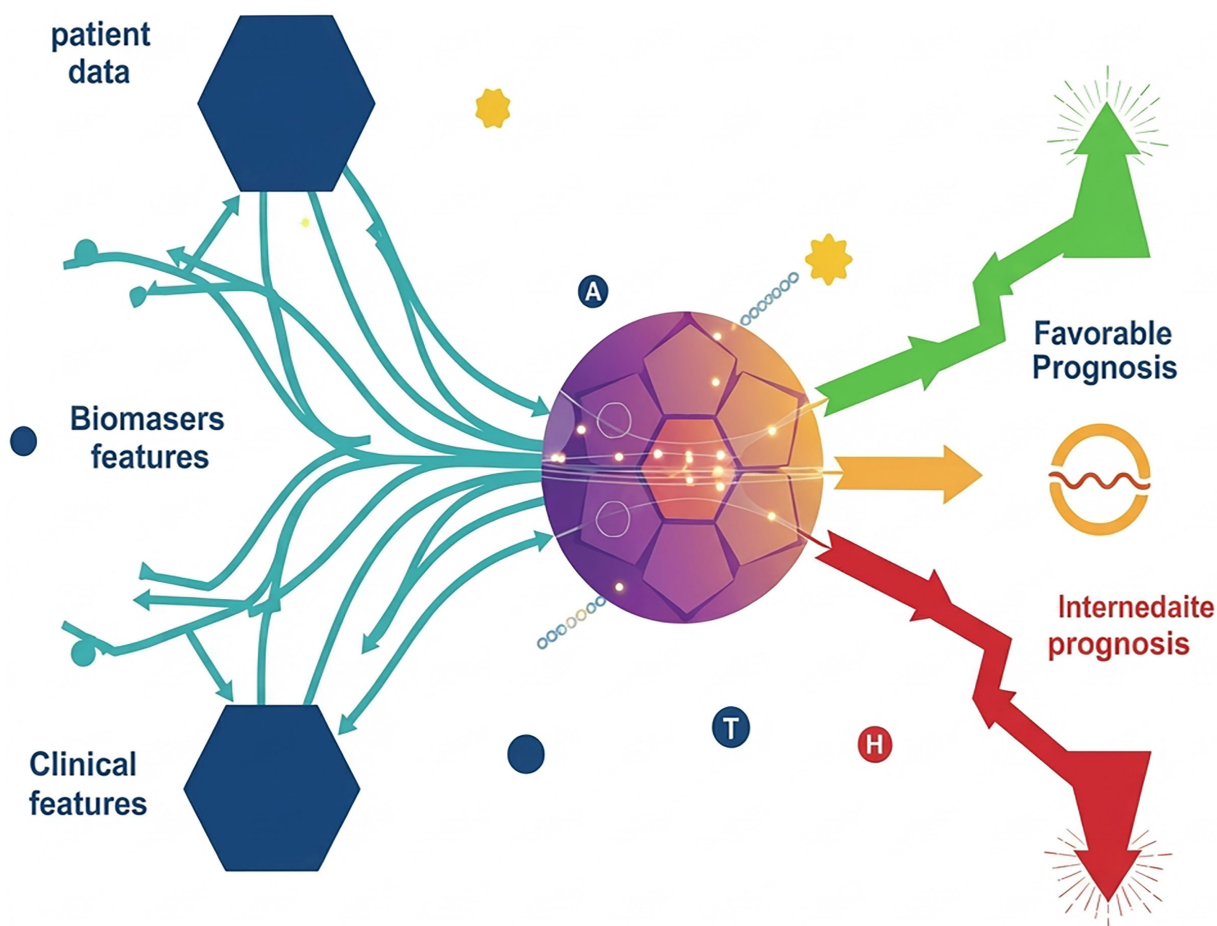
Authors

Keshu Shan, Chunxia Zhu, Xueyou Lv, Jianing Yan, Qier Li

Correspondence

1991473972@qq.com (Q. Li), 419503527@qq.com (J. Yan)

Graphical Abstract



<https://doi.org/10.71321/ggvee953>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Construction and Validation of a Clinical Model to Predict Lymph Node Metastasis in Early Hypopharyngeal Squamous Cell Carcinoma

Keshu Shan¹, Chunxia Zhu¹, Xueyou Lv¹, Jianing Yan^{1*}, Qier Li^{1*}

Received: 2025-12-21 | Accepted: 2026-02-24 | Published online: 2026-03-27

Abstract

Background: Hypopharyngeal squamous cell carcinoma (HSCC) is the most frequent histologic subtype of hypopharyngeal carcinoma and has a poor prognosis. With the development of upper gastrointestinal endoscopy, an increasing number of early HSCC cases can be diagnosed and treated by endoscopic mucosal resection (EMR) and endoscopic submucosal dissection (ESD). However, clinicians lack an effective tool to predict lymph node metastasis (LNM) of early HSCC.

Methods: A total of 410 early HSCC cases were extracted from the Surveillance, Epidemiology, and End Results (SEER) database. Clinical data (50 early HSCC patients) from our hospital were applied for external validation. Univariate and multivariate logistic regression analyses to filter the risk factors were performed. A nomogram model was built to predict LNM in early HSCC patients according to the multivariate logistic regression model.

Results: HSCC patients with lymph node metastasis had worse overall survival rates and a higher risk of long-term mortality ($P = 0.016$). The C-index and AUC of our predictive nomogram model were 0.674 and 0.623-0.726, respectively. The AUC, sensitivity, specificity, PPV, and NPV were 0.680 (95% CI: 0.525-0.835), 80.0%, 46.7%, 50.0%, and 77.8%, respectively, for the validation set.

Conclusion: Our study resulted in a novel nomogram model for predicting LNM that can be a promising tool for assessing the feasibility of EMR/ESD in early HSCC.

Impact: Our study provides a novel method for HSCC preoperative assessment before ESD.

Keywords: Metastasis; Hypopharyngeal; SEER; Nomogram; Prognosis

Introduction

Hypopharyngeal carcinoma (HPC) is a type of head and neck malignant tumour [1]. Hypopharyngeal squamous cell carcinoma (HSCC) is one of the most frequent histologic subtypes of hypopharyngeal carcinoma, accounting for up to 95% of all cases with the worst prognosis and distal survival time, with a 5-year survival rate of only 30–35% [2-3]. Recently, lymph node metastasis (LNM) is confirmed as a crucial factor to the distal prognosis of HSCC [4]. Based on its anatomically heterogenous structure of rich lymphatics, latent HSCC cells can grow reclusively and expeditiously, resulting in distal invasion and metastasis [4]. Up to now, pathologic biopsy still is the gold standard for the diagnosis of LNM. However, several lymph nodes in certain anatomical locations including deep intra-parotid nodes, retropharyngeal nodes or nodes located just posterior to the clavicles cannot be obtained accurately

[5]. Meanwhile, identification of clinical nodal stage always accompanies with lymph node dissection and a large range of surgery. Hence, it is essential to construct a noninvasive method to predict the risk of lymph node metastasis (LNM).

With the advance development of digestive endoscopic technology, the early detection of HPC has become much more possible. Moreover, the main treatment methods, such as endoscopic mucosal resection (EMR) and endoscopic submucosal dissection (ESD), are not limited only to applications in early gastric cancer and oesophageal cancer but are also suitable for HPC [6]. However, the absence of nonlymph node metastasis is an important prerequisite for choosing EMR and ESD [7].

Thus, in our study, we firstly utilized the data obtained from the Surveillance, Epidemiology and End Results (SEER) database to build a nomogram model for predicting lymph node metastasis to provide a promising tool to use before minimally inva-

¹ Department of Gastroenterology, the First Affiliated Hospital of Ningbo University, Ningbo, 315020, China

* Corresponding Author.

sive therapy of early HSCC. Then we used the clinical cases from our hospital for external validation and the sensitivity and specificity were 80.0% and 46.7%, respectively.

Methods

Study design and patients baseline characteristics

The clinical data were extracted from the SEER research plus data with 17 registries and collected via SEER*Stat software (8.4.0.1; www.seer.cancer.gov) with a private ID (account 17443-Nov2021). Informed consent was not required for the use of these registry data, as the SEER database is publicly available and does not include any personal information [8].

The inclusion criteria for patients in our study were as follows: (1) HSCC patients aged over 18 years were diagnosed between 2004 and 2015; (2) patients were definitively confirmed by histopathology as stage Tis, T₀, or T₁ (the 6th American Joint Committee on Cancer stage system); (3) the primary site codes of HSCC were recorded as C12.9, C13.0, C13.1, C13.2, C13.8, or C13.9 in the light of the International Classification of Diseases for Oncology [9]; and (4) patients with whole records of baseline data and long term survival time. The patients meeting the following criteria were eligible for exclusion: (1) patients with no surgical resection; (2) the stage of lymph nodes was Nx or blank; and (3) the tumour size was blank or unknown.

For validation, we retrospectively assessed the clinicopathological data of all 50 consecutive patients with HSCC between 2017 and 2022 at the Affiliated Hospital of Medical School of Ningbo University, China. All of the patients were received surgery and lymph node dissection. All patients provided written informed consent before the surgery and all information of the participants was collected anonymously. All methods were performed in accordance with the relevant guidelines and regulations. This study was approved by the ethics committees of the institutes where the study was conducted (IRB No. KS202210028).

Variables and outcomes

According to the stage of lymph nodes, all of the patients were regrouped into N₀ or N₁₋₃ groups. Patients were classified into ≤ 60 years and > 60 years based on the age of diagnosis. Sex was divided into male or female. The primary site of HSCC was grouped as pyriform sinus (C12.9), postcricoid region (C13.0), aryepiglottic fold (C13.1), posterior wall of hypopharynx (C13.2), overlapping lesion of hypopharynx (C13.8), and hypopharynx (C13.9). Grade was classified into 4 groups: well differentiated (Grade I), moderately differentiated (Grade II), poorly differentiated (Grade III), and undifferentiated (Grade IV). The race of patients was recorded as white, black, Asian or Pacific Islander and others. The tumour size was grouped into ≤ 10 mm and > 10 mm.

Statistical analysis

In our study, the Kaplan–Meier method was used to compare the difference in survival times between the N0 and N1-3 groups by the Gehan-Breslow-Wilcoxon test. Univariate and multivariate logistic regression analyses were performed to filter the risk factors for early HSCC, calculate the P value with a 95% confidence interval (CI) and assess the independent

predictors. A nomogram was built according to the results of multivariate logistic regression, whose performance was evaluated by the C-index and receiver operating characteristic (ROC) curve analysis. The area under the receiver operating characteristic curve (AUC), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were computed as well.

All statistical analyses and figures were carried out via R software (version 4.0.2). *P* < 0.05 was deemed statistically significant.

Results

Patient demographics and clinicopathological characteristics

In total, 410 surgically resected early HSCC patients were included in the study. Among these 410 patients, 193 cases (47.07%) were N₀, and the other 217 cases (52.93%) were N₁₋₃. From a longitudinal perspective, N₁₋₃ patients were more commonly younger, male, and had a worse degree of differentiation grade. The tumour size showed a modest association with lymph node metastasis. Table 1 demonstrates the detailed clinicopathological characteristics noted from the SEER database.

Table 1. Characteristics of early HSCC patients from the SEER database.

Characteristic	N ₀	N ₁₋₃
n (%)	193 (47.07%)	217 (52.93%)
Sex		
Female	56 (13.66%)	45 (10.98%)
Male	137 (33.41%)	172 (41.95%)
Age (years)		
≤60	60 (14.63%)	95 (23.17%)
>60	133 (32.44%)	122 (29.76%)
Race		
White	162 (39.51%)	178 (43.41%)
Black	16 (3.90%)	23 (5.61%)
Asian or Pacific Islander	12 (2.93%)	12 (2.93%)
Others	3 (0.73%)	4 (0.98%)
Differentiation		
Well	19 (4.63%)	13 (3.17%)
Moderate	113 (27.56%)	90 (21.95%)
Poor	61 (14.88%)	106 (25.85%)
Undifferentiated	0 (0%)	8 (1.95%)
Site		
Pyriform sinus	109 (26.59%)	136 (33.17%)
Postcricoid region	6 (1.46%)	5 (1.22%)
Aryepiglottic fold	23 (5.61%)	14 (3.41%)
Posterior wall of hypopharynx	16 (3.90%)	17 (4.15%)
Overlapping lesion of hypopharynx	0 (0%)	2 (0.49%)
Hypopharynx	39 (9.51%)	43 (10.49%)
Tumour size (mm)		
≤10	64 (15.61%)	61 (14.88%)
10-20	129 (31.46%)	156 (38.05%)

N₁₋₃ patients had poor distal outcome

We first compared the distal survival time between the two groups. As shown in Figure 1, HSCC patients with lymph node metastasis had worse overall survival rates and a higher risk of long-term mortality ($P = 0.016$). The median overall survival times of N₀ and N₁₋₃ were 61 and 43 months, respectively.

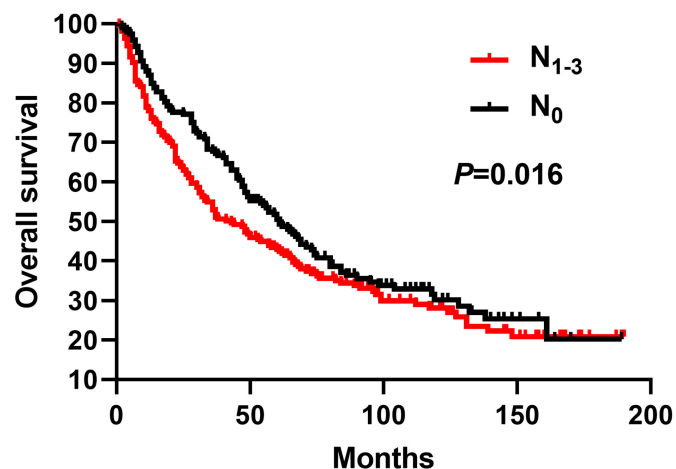


Figure 1. Kaplan-Meier analysis of LNM and overall survival time. The survival analysis displayed that N₁₋₃ patients had poor distal outcome ($P = 0.016$).

Risk factors related to LNM in early HSCC

To systematically predict LNM presence probability, we performed univariate and multivariate logistic regression analyses to determine the remarkable clinical predictors based on the clinical factors mentioned before. In the univariate model, younger age ($P < 0.001$), male sex ($P = 0.059$), and worse degrees of differentiation ($P < 0.001$) were associated with a higher probability of LNM in the 2004-2015 T₁ HSCC cohort (Table 2). Moreover, the results of multivariate logistic regression analysis verified that age, sex, and degrees of differentiation were independent risk factors for assessing the probability of LNM probability in early HSCC patients (Figure 2).

Table 2. Univariate and multivariate logistic regression analyses of factors associated with LNM in early HSCC patients

Variables	Univariate analysis		Multivariate analysis	
	OR (95% CI)	P	OR (95% CI)	P
Age	0.964 (0.945-0.963)	<0.001	0.964 (0.946-0.983)	<0.001
Sex	0.635 (0.396-1.017)	0.059	0.637 (0.398-1.020)	
Grade	2.125 (1.520-2.972)	<0.001	2.117 (1.531-2.929)	<0.001
Site	1.105 (0.648-1.883)	0.715		
Size	1.000 (0.958-1.044)	0.988		
Race	0.999 (0.725-1.375)	0.993		
Constant	0.99	0.998	3.571	0.100

Entry 0.1; Removal: 0.15, $P < 0.001$

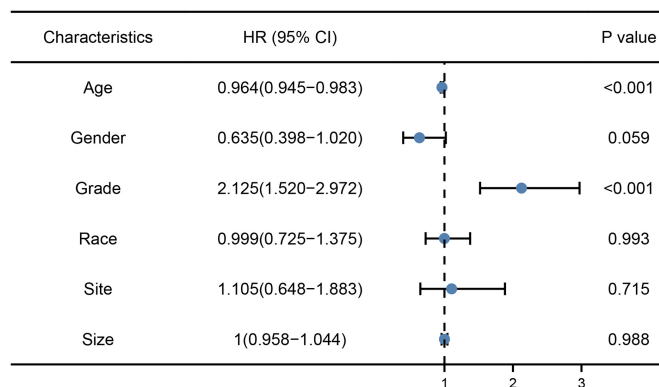


Figure 2. Forest plot showing the independent clinical predictors for LNM in early HSCC patients.

Establishment and validation of the nomogram model for predicting LNM in early HSCC patients

We further constructed a nomogram model to predict the rate of LNM in early HSCC patients according to anterior multivariate logistic regression analysis (Figure 3A). We incorporated sex and degrees of differentiation in the model, and clinicians could easily acquire the probability value by summing scores on each item. The C-index and AUC of our model were 0.674 and 0.623-0.726, respectively, as displayed in Figure 3B.

To further test the actual prediction ability, we collected and used the data of our hospital as the validation sets to examine the model. A detailed description of the baseline is provided in Table 3. We drew the ROC curve, and the AUC for the validation set was 0.680 (95% CI: 0.525-0.835) in Figure 3C. The sensitivity, specificity, PPV, and NPV were 80.0%, 46.7%, 50.0%, and 77.8% for the validation set. All of the results indicated that our model had excellent adaptability to the population with early HSCC.

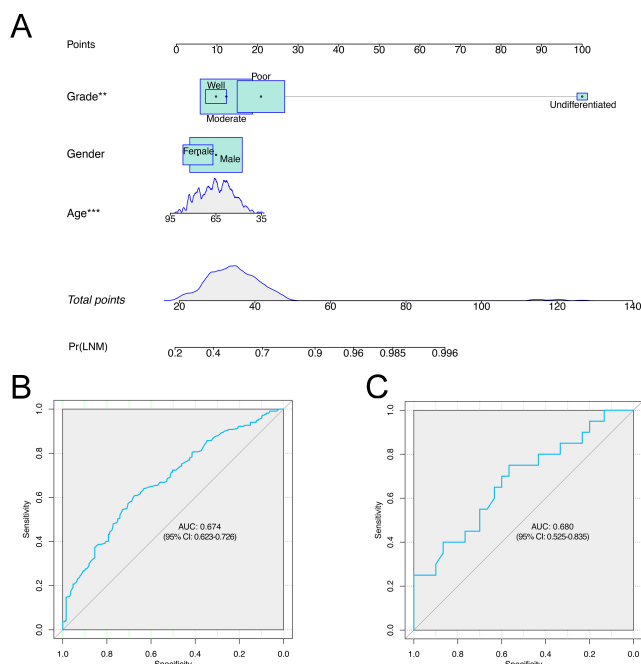


Figure 3. Nomogram model and ROC curves. **A.** A nomogram model for predicting the rates of LNM in early HSCC patients. **B.** ROC curve of our nomogram model. **C.** ROC curve of the validation set.

Table 3. Characteristics of early HSCC patients for external validation.

Characteristic	N ₀	N ₁₋₃
n (%)	27 (54.00%)	23 (46.00%)
Sex		
Female	12 (24.00%)	8 (10.98%)
Male	15 (30.00%)	15 (41.95%)
Age (years)		
≤60	12 (24.00%)	12 (24.00%)
>60	15 (30.00%)	11 (22.00%)
Differentiation		
Well	5 (10.00%)	4 (8.00%)
Moderate	13 (26.00%)	11 (22.00%)
Poor	9 (18.00%)	6 (12.00%)
Undifferentiated	0 (0%)	2 (4.00%)
Site		
Pyriform sinus	10 (20.00%)	7 (14.00%)
Postcricoid region	3 (6.00%)	4 (8.00%)
Aryepiglottic fold	1 (2.00%)	3 (6.00%)
Posterior wall of hypopharynx	10 (20.00%)	6 (12.00%)
Overlapping lesion of hypopharynx	0 (0%)	1 (2.00%)
Hypopharynx	3 (6.00%)	2 (4.00%)
Tumour size (mm)		
≤10	13 (26.00%)	10 (28.00%)
10-20	14 (28.00%)	13 (26.00%)

Discussion

The anatomic location of the hypopharynx is contiguous to the upper gastrointestinal tract, which is important for swallowing, respiration, phonation and even immunomodulation [10]. However, laryngopharyngectomy and pharyngeal reconstruction are still conventional treatment strategies, causing the loss of natural speech and dysphagia [11]. With the development of digestive endoscopy and ESD, early detection, minimally invasive surgery and preservation of anatomical function are becoming increasingly possible. Nevertheless, the absence of lymph node metastasis is the absolute indication for ESD, and

it is essential to construct a satisfactory model to predict LNM [12].

In our study, we first identified that LNM significantly influenced the distal survival time for early HSCC patients. Our clinical samples did not receive this therapy before surgery. Then, a prediction nomogram model for LNM in early HSCC was built according to the SEER database. Moreover, a clinical cohort was used for external validation. The C-index of the model was 0.674 (95% CI: 0.623-0.726), and the AUC, sensitivity, specificity, PPV, and NPV of the external validation were 0.679, 80.0%, 46.7%, 50.0%, and 77.8%, respectively, which showed good discrimination and fitness. In this model, younger age at the time of diagnosis as well as worse degrees of differentiation were potentially related to a higher risk of lymph node metastasis. Younger age was related to much hypermetabolism and rich nutritional support but lower tumour doubling time, causing more aggressive biological behaviours and higher rates of LNM [13]. Degrees of differentiation and low LNM risk are the restricted condition of ESD on the basis of the Japanese Gastric Cancer Association's gastric cancer treatment guidelines [14]. All these results imply that there are many similarities in ESD between HSCC and digestive tumours.

It is essential to define tumour invasion and LNM before EMR/ESD. To date, there is still no consensus among experts or clear recommendations in guidelines about EMR or ESD in early HSCC, but existing studies have clearly demonstrated their feasibility [15-16]. At present, the otolaryngeal endoscope is still less sensitive for the detection of early or subtle mucosal and microvascular architecture changes, so it is difficult to identify the absence of early lesions. Luckily, narrow-band imaging (NBI) and magnifying endoscopy can be helpful in detecting subtle changes in the intraepithelial papillary capillary loop and brownish area in early oesophageal cancer [17]. The Japan Oesophageal Society has developed a mature method to predict the invasion depth based on magnifying endoscopic classification. JES focused on the abnormal microvessels, avascular area and intervacular background colouration, and the overall accuracy of type B microvessels in assessing tumour invasion depth was up to 90.5% [18]. This system has been widely applied in multiple centres and has high accuracy in predicting invasion depth for superficial oesophageal squamous cell carcinoma [19-20]. More interestingly, recent studies

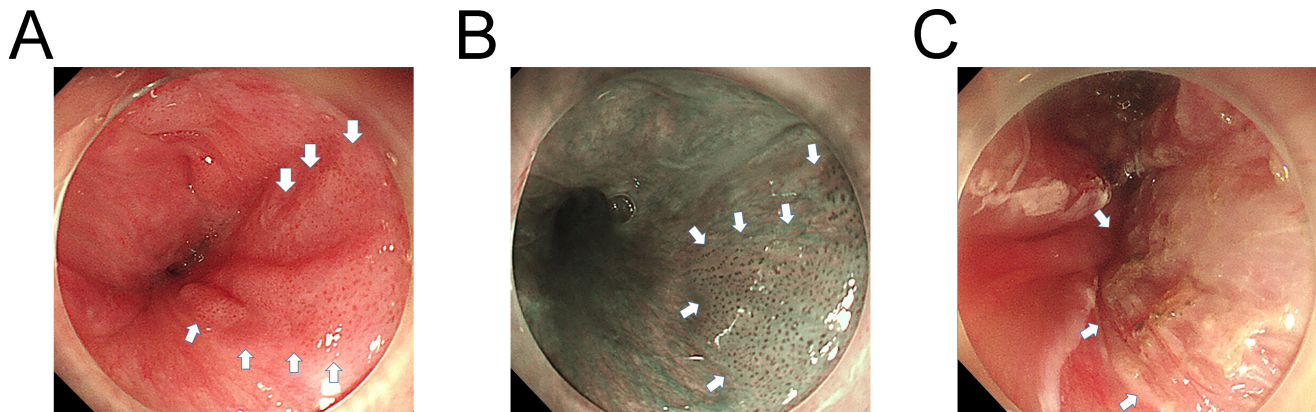


Figure 4. Diagnosis and ESD of early HSCC lesions by gastroscopy. **A.** White light showed a 0-IIa lesion in the right pyriform sinus of the hypopharynx. **B.** NBI showed that the calibre of the IPCL increased notably in the lesion, which was attached to Type B1 vessels. **C.** The lesion was completely resected via ESD using a gastrocope.

have found that NBI can be applied in oral squamous cell carcinoma and laryngopharyngeal malignancies, which contributes to evaluating the differentiated degree and depth of invasion in early HSCC [21–22]. Endoscopists can easily take a biopsy of the lesion in the progress of gastroscopy to obtain detailed pathological evidence, which can lay a solid foundation before EMR/ESD. Figure 4 shows the progress of diagnosing and ESD of early HSCC lesions by NBI and magnifying endoscopy using a gastroscope (GIF-Q260Z; Olympus) in our hospital. The lesion was completely resected, and the final diagnosis was carcinoma in situ (stage $T_{\text{his}}N_0M_0$). This patient is currently under regular follow-up. In a word, we strongly recommend NBI, magnifying endoscopy and biopsy to assess and diagnose the HSCC.

Compared to oesophageal cancer and esophagus cancer as well as gastric cancer, the depth of HPC is unrelated to the T stage, and the T number increases as the tumour size grows in cases of widespread superficial cancer on the basis of the American Joint Committee on Cancer guidelines (2017, 8th ed). The hypopharynx is prone to lymph node metastasis due to the lack of muscularis mucosa and rich lymphatic system when the tumour invades the subepithelial surface [23]. It is well reported that there is no risk of metastasis in intraepithelial cancer, and a larger HSCC correlates with a higher risk of local recurrence [24]. Moreover, on account of the narrow operation space of the hypopharynx and the requirement of ESD treatment to be performed under tracheal intubation anaesthesia, the entire hypopharynx cannot be fully exposed because of the influence of the tracheal cannula according to our clinical experiences. Hence, we selected patients with Tis, T_0 , or T_1 stage from the SEER database for the inclusion criteria in the SEER database. The association between the depth of early HSCC and the risk of LNM should be determined in the future. As such, we propose that clinical HSCC patients are worthy to firstly receive magnifying endoscopy and biopsy to get pathological evidence. Then, our model has the capacity to confirm the indications for endoscopic treatment of early HSCC, which is worthy of validation in a larger, multicentre randomized clinical trial (Figure 5).

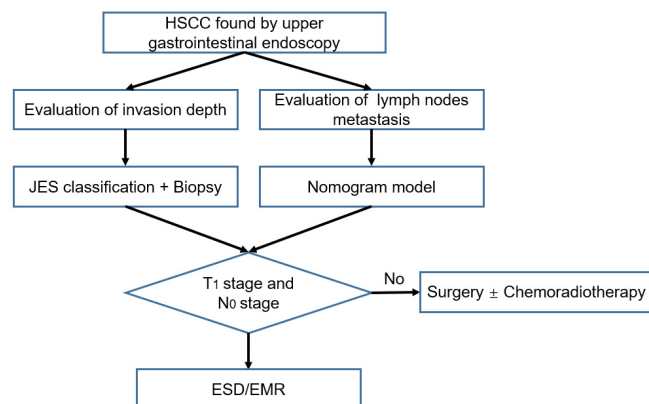


Figure 5. The diagnosis and treatment procedures of HSCC. We designed a novel procedure of HSCC patients received gastrointestinal endoscopy for clinical doctors.

Conclusion

In conclusion, our study resulted in a novel nomogram model for predicting LNM, which can be a promising tool for assessing the feasibility of EMR/ESD in early HSCC.

Abbreviations

AUC: area under the receiver operating characteristic curve; CI: confidence interval; EMR: endoscopic mucosal resection; ESD: endoscopic submucosal dissection; HPC: Hypopharyngeal carcinoma; HSCC: Hypopharyngeal squamous cell carcinoma; LNM: lymph node metastasis; NBI: narrow-band imaging; NPV: negative predictive value; PPV: positive predictive value; ROC: receiver operating characteristic; SEER: Surveillance, Epidemiology, and End Results.

Author Contributions

Jianing Yan designed and performed the research and wrote the paper; Chuxia Zhu and Xueyou Lv provided the ideas about the research and supervised the report; Keshu Shan and Qier Li designed the research and contributed to the analysis of data. KeShu Shan was responsible for the revision of the manuscript for critical intellectual content; all authors issued final approval for the version to be submitted.

Acknowledgements

We sincerely thank all contributors of high-quality data to these accessible public databases.

Funding Information

This study was supported by grants from the Key Scientific and Technological Projects of Ningbo [No. 2021Z133].

Ethics Approval and Consent to Participate

This study was approved by the ethics committees of the First Affiliated Hospital of Ningbo University where the study was conducted (IRB No. KS202210028). The analysis of public data is de-identified, publicly available data from the SEER database. In accordance with US federal regulations (45 CFR Part 46.101(b)(4)), this part of research did not constitute human subjects research requiring full ethical review.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

References

- [1] Jay SC, Kim P, Katherine M, Henry TH, Randal SW, Kian KA, et al. National Cancer Database report on cancer of the head and neck: 10-year update. (2009). *Head Neck*, 31(6), 748-58. <https://doi.org/10.1002/hed.21022>
- [2] Echanique KA, Evans LK, Han AY, Chhetri DK, St John MA. Cancer of the Larynx and Hypopharynx. (2021). *Hematol Oncol Clin North Am*, 35(5), 933-947. <https://doi.org/10.1016/j.hoc.2021.05.005>.
- [3] Meng Z, Zhang L, Huang C, Piao Y, Chen X, & Xian J. Quantitative parameters derived from 18F-fluorodeoxyglucose positron emission tomography/magnetic resonance imaging can accurately estimate the histologic grade of hypopharyngeal squamous cell carcinoma preoperatively. (2022). *Neuroradiology*, 64(11), 2153-2162. <https://doi.org/10.1007/s00234-022-03052-2>.
- [4] Petersen JF, Timmermans AJ, van Dijk BAC, Overbeek LIH, Smit LA, Hilgers FJM, et al. Trends in treatment, incidence and survival of hypopharynx cancer: a 20-year population-based study in the Netherlands. (2018). *Eur Arch Otorhinolaryngol*, 275(1), 181-189. <https://doi.org/10.1007/s00405-017-4766-6>.
- [5] Jung HN, Ryoo I, Suh S, Lee YH, & Kim E. Evaluating the Elasticity of Metastatic Cervical Lymph Nodes in Head and Neck Squamous Cell Carcinoma Patients Using DWI-based Virtual MR Elastography. (2024). *Magn Reson Med Sci*, 23(1), 49-55. <https://doi.org/10.2463/mrms.mp.2022-0082>.
- [6] Shimizu Y, Yamamoto J, Kato M, Yoshida T, Hirota J, Ono Y, et al. Endoscopic submucosal dissection for treatment of early stage hypopharyngeal carcinoma. (2006). *Gastrointest Endosc*, 64(2), 255-259. <https://doi.org/10.1016/j.gie.2006.01.049>.
- [7] Iizuka T, Kikuchi D, Suzuki Y, Tanaka M, & Takeda H. Clinical relevance of endoscopic treatment for superficial pharyngeal cancer: feasibility of techniques corresponding to each location and long-term outcomes. (2021). *Gastrointest Endosc*, 93(4), 818-827. <https://doi.org/10.1016/j.gie.2020.07.039>.
- [8] Cronin K, Ries L, & Edwards B. The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute. (2014). *Cancer*, 120, 3755-3757. <https://doi.org/10.1002/cncr.29049>.
- [9] Yang H, Zeng M, Cao S, & Jin L. Nomograms predicting prognosis for locally advanced hypopharyngeal squamous cell carcinoma. (2022). *Eur Arch Otorhinolaryngol*, 279, 3041-3052. <https://doi.org/10.1007/s00405-021-07109-5>.
- [10] van den Broek MFL, De Boeck I, Kiekens F, Boudewyns A, Vanderveken OM, & Lebeer S. Translating Recent Microbiome Insights in Otitis Media into Probiotic Strategies. (2019). *Clin Microbiol Rev*, 32(4), e00010-18. <https://doi.org/10.1128/CMR.00010-18>.
- [11] Bova R, Goh R, Poulson M, & Coman WB. Total pharyngolaryngectomy for squamous cell carcinoma of the hypopharynx: a review. (2005). *Laryngoscope*, 115, 864-869. <https://doi.org/10.1097/01.MLG.0000158348.38763.5D>.
- [12] Hatta W, Tsuji Y, Yoshio T, Kakushima N, Hoteya S, Doyama H, et al. Prediction model of bleeding after endoscopic submucosal dissection for early gastric cancer: BEST-J score. (2021). *Gut*, 70, 476-484. <https://doi.org/10.1136/gutjnl-2019-319926>.
- [13] Jung J, Kong M, & Hong S. Conventional fractionated helical tomotherapy for patients with small to medium hepatocellular carcinomas without portal vein tumor thrombosis. (2014). *Onco Targets Ther*, 7, 1769-1775. <https://doi.org/10.2147/OTT.S69618>. eCollection 2014.
- [14] Japanese Gastric Cancer A. Japanese gastric cancer treatment guidelines 2014 (ver. 4). (2017). *Gastric Cancer*, 20, 1-19. <https://doi.org/10.1007/s10120-016-0622-4>.
- [15] Iizuka T, Kikuchi D, Hoteya S, Nakamura M, Yamashita S, Mitani T, et al. Clinical advantage of endoscopic submucosal dissection over endoscopic mucosal resection for early mesopharyngeal and hypopharyngeal cancers. (2011). *Endoscopy*, 43(10), 839-843. <https://doi.org/10.1055/s-0031-1271112>.
- [16] Iizuka T, Kikuchi D, Hoteya S, Takeda H, & Kaise M. A new technique for pharyngeal endoscopic submucosal dissection: peroral countertraction (with video). (2012). *Gastrointest Endosc*, 76, 1034-1038. <https://doi.org/10.1016/j.gie.2012.07.013>.
- [17] Hatta W, Koike T, Ogata Y, Kondo Y, Ara N, Uno K, et al. Comparison of Magnifying Endoscopy with Blue Light Imaging and Narrow Band Imaging for Determining the Invasion Depth of Superficial Esophageal Squamous Cell Carcinoma by the Japanese Esophageal Society's Intrapapillary Capillary Loop Classification. (2021). *Diagnostics (Basel, Switzerland)*, 11, 1941. <https://doi.org/10.3390/diagnostics11111941>.
- [18] Oyama T, Inoue H, Arima M, Momma K, Omori T, Ishihara R, et al. Prediction of the invasion depth of superficial squamous cell carcinoma based on microvessel morphology: magnifying endoscopic classification of the Japan Esophageal Society. (2017). *Esophagus*, 14, 105-112. <https://doi.org/10.1007/s10388-016-0527-7>.
- [19] Gotoda T, Hori K, Nakagawa M, Kobayashi S, Toyokawa T, Ishiyama S, et al. A prospective multicenter study of the magnifying endoscopic evaluation of the invasion depth of superficial esophageal cancers. (2022). *Surgical endoscopy*, 36, 3451-3459. <https://doi.org/10.1007/s00464-021-08666-w>.
- [20] Kumagai Y, Takubo K, Kawada K, Ohue M, Higashi M, Ishiguro T, et al. Endocytoscopic Observation of Esophageal Lesions: Our Own Experience and a Review of the Literature. (2022). *Diagnostics (Basel, Switzerland)*, 12(9), 2222. <https://doi.org/10.3390/diagnostics12092222>.
- [21] Ota A, Miyamoto I, Ohashi Y, Chiba T, Takeda Y, & Yamada H. Diagnostic Accuracy of High-Grade Intraepithelial Papillary Capillary Loops by Narrow Band Imaging for Early Detection of Oral Malignancy: A Cross-Sectional Clinicopathological Imaging Study. (2022). *Cancers (Basel)*, 14(10), 2415. <https://doi.org/10.3390/cancers14102415>.
- [22] Ali M, Gupta G, Silu M, Chand D, & Samor V. Narrow band

- imaging in early diagnosis of laryngopharyngeal malignant and premalignant lesions. (2021). *Auris Nasus Larynx*, 49, 676-679. <https://doi.org/10.1016/j.anl.2021.11.008>.
- [23] Muto M, Satake H, Yano T, Minashi K, Hayashi R, Fujii S, et al. Long-term outcome of transoral organ-preserving pharyngeal endoscopic resection for superficial pharyngeal cancer. (2011). *Gastrointest Endosc*, 74, 477-484. <https://doi.org/10.1016/j.gie.2011.04.027>.
- [24] Kinjo Y, Nonaka S, Oda I, Abe S, Suzuki H, Yoshinaga S, et al. The short-term and long-term outcomes of the endoscopic resection for the superficial pharyngeal squamous cell carcinoma. (2015). *Endosc Int Open*, 3, E266-273. <https://doi.org/10.1055/s-0034-1392094>.

Gut Microbiota Affects Cardiovascular Diseases via Inflammatory Cytokines: a Bidirectional Two-sample Mendelian Randomization Study and Mediation Analysis

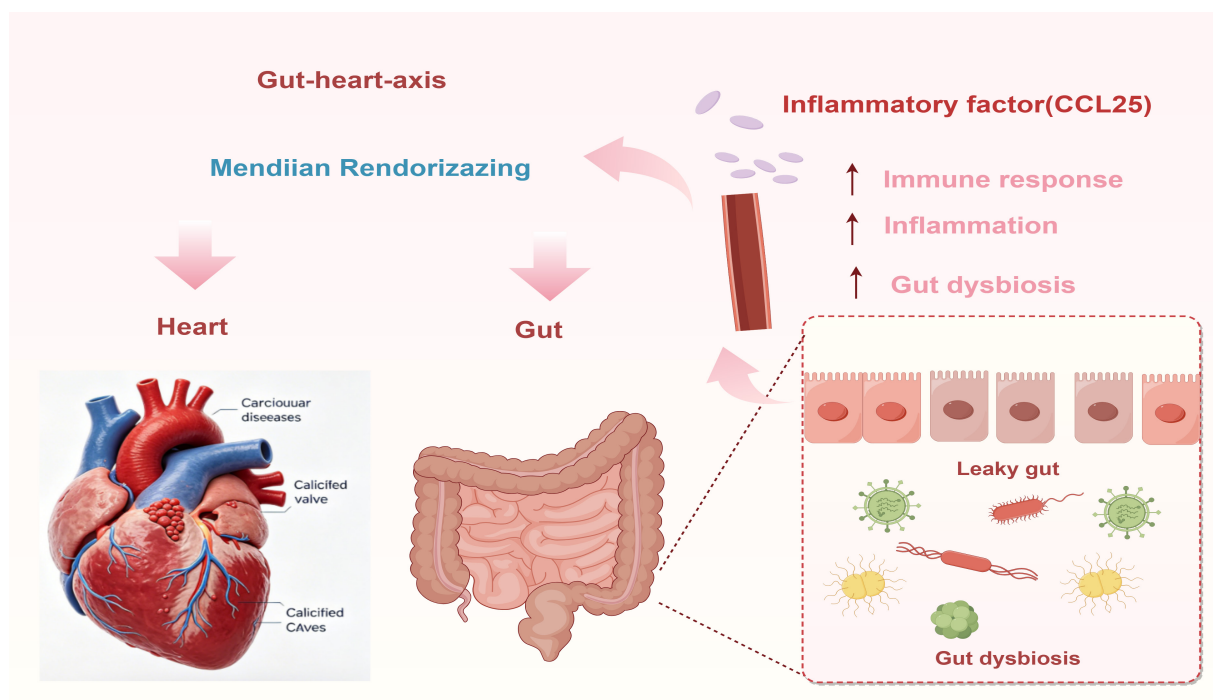
Authors

Xuekun Hou, Yunxiao Meng, Yuxuan Zhang, Xianghan Zheng, Hongqi Guo, Yichen Zhang, Hongmei Yuan, Zefei Chu

Correspondence

chuzefei@hebcm.edu.cn (Z. Chu)

Graphical Abstract



<https://doi.org/10.71321/c18kt054>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Gut Microbiota Affects Cardiovascular Diseases via Inflammatory Cytokines: a Bidirectional Two-sample Mendelian Randomization Study and Mediation Analysis

Xuekun Hou¹, Yunxiao Meng¹, Yuxuan Zhang¹, Xianghan Zheng¹, Hongqi Guo¹, Yichen Zhang¹, Hongmei Yuan¹, Zefei Chu^{1*}

Received: 2026-03-11 | Accepted: 2026-04-10 | Published online: 2026-04-21

Abstract

Background: Cardiovascular diseases (CVDs) arise from complex interactions between genetic predisposition and environmental exposures, manifesting as persistent inflammatory disorders. Although accumulating evidence has implicated gut microbiota in the pathogenesis of CVDs, whether this association reflects a causal relationship remains to be firmly established. Therefore, the present study sought to examine the potential causal links between gut microbiota composition, circulating inflammatory cytokines, and susceptibility to CVDs.

Methods: Summary statistics for 196 gut microbiota taxa, 91 cytokines, and 8 CVD subtypes were extracted from the largest published genome-wide association studies (GWAS) to conduct bidirectional two-sample Mendelian randomization (MR) analysis. The inverse variance weighted (IVW) method was used as the primary statistical approach. We investigated the causal associations between gut microbiota and 8 CVD subtypes, including coronary artery disease (CAD), coronary atherosclerosis (CAS), heart failure (HF), stroke, atrial fibrillation (AF), angina pectoris (AP), calcific aortic valve stenosis (CAVS), and aortic aneurysm. Multivariable MR (MVMR) analysis was further performed to explore the potential mediating role of cytokines in the causal pathways from gut microbiota to CVDs.

Results: MR analysis identified causal associations between 58 gut microbiota taxa, 46 inflammatory cytokines, and CVDs. Specifically, 34 positive and 24 negative causal effects were observed between gut microbiota and CVDs, along with 16 positive and 30 negative causal effects between cytokines and CVDs. Additionally, mediation analysis identified C-C motif chemokine 25 (CCL25) level as a key mediator in the causal relationship from gut microbiota to CVDs.

Conclusions: Our findings support a causal link between gut microbiota, inflammatory cytokines, and CVDs, with CCL25 mediating the protective effect of specific gut taxa against CAVS. These identified biomarkers offer novel insights into the pathophysiological mechanisms of CVDs and may inform the development of novel strategies for CVD prevention, diagnosis, and treatment.

Keywords: Gut microbiota; Inflammatory Cytokines; Cardiovascular diseases; Mendelian randomization

Introduction

Cardiovascular diseases (CVDs) represent a significant global health burden and are a leading cause of mortality, prompting widespread public health concern [1]. Genomics and high-throughput sequencing analysis represent contemporary methodologies for investigating the role of gut microbiota in the development of CVDs [2]. The human intestine hosts a symbiotic community of 10 to 100 trillion microbial cells that coexist with the host [3]. The gut microbiota functions as a complex ecosystem, acting as virtual endocrine organs and playing a regulatory role in the occurrence and development of CVDs [4]. One such microbiota-dependent metabolite, trimethylamine N-oxide (TMAO), has been implicated in the pathologic processes of heart failure and may serve as an ear-

ly warning marker for identifying individuals at risk of disease progression [5]. The relationship between gut microbiota and heart disease is elucidated through the gut-heart axis [6]. Dysbiosis of the gut microbiome has been associated with dysregulated immune activity, particularly the enhanced secretion of inflammatory cytokines implicated in the pathogenesis of cardiovascular diseases [7]. Despite the acknowledged role of inflammation in CVD progression, studies examining the interplay between gut microbiota and cytokine profiles among CVD patients remain scarce.

Changes in gut microbiota abundance are implicated in the onset of inflammation, a key factor in the pathophysiology of CVDs [8]. Epidemiological research has identified key interleukins as playing a role in the pathogenesis of CVDs [9]. It appears that both gut microbiota and cytokines can influence

¹ Faculty of Integrated Traditional Chinese and Western Medicine, Hebei University of Chinese Medicine, Shijiazhuang, 050200, China

* Corresponding Author.

the progression of CVDs. It is hypothesized that cytokines may serve as mediators in the pathway linking gut microbiota to CVDs.

Randomized controlled trials (RCTs) have traditionally been utilized to establish causal relationships between gut microbiota or cytokines and CVDs. Nevertheless, ethical considerations and financial constraints restrict the widespread implementation of RCTs in clinical settings. Mendelian randomization (MR) analysis has emerged as a valuable tool for addressing issues of causality in epidemiological data by utilizing genetic variants as instrumental variables [10]. The advancement of Genome-wide Association Studies (GWASs) and increased accessibility to GWAS data have significantly enhanced the methodology of MR analysis in recent years. While numerous studies have identified protective and pathogenic relationships between gut microbiota and cardiovascular diseases (CVDs), the potential role of inflammatory cytokines as mediators in these relationships remains unexplored [11].

This study aimed to systematically evaluate the causal links between gut microbiota composition, inflammatory cytokines, and 8 CVD subtypes via bidirectional two-sample MR and MVMR analysis, with a secondary objective to explore the mediating role of inflammatory cytokines in the gut microbiota-CVD pathways.

Methods

Study Design

This study employed multiple single-nucleotide polymorphisms (SNPs) as instrumental variables under three core assumptions to evaluate the causal effects of gut microbiota and inflammatory cytokines on eight cardiovascular diseases: (i) the relevance assumption, which dictates that IVs must be highly associated with the exposure; (ii) the independence assumption, which stipulates that IVs are not influenced by confounding variables; and (iii) the exclusion restriction assumption, which posits that IVs impact the outcome solely through the exposure and not through alternative pathways [12]. In our study, we employed bidirectional two-sample MR analysis and MVMR analysis, following a structured approach illustrated in Figure 1. Specifically, Step 1 involved evaluating the bidirectional causal relationships between 196 gut microbiota and 8 cardiovascular diseases (CVDs); Step 2 entailed identifying bidirectional causality between 91 inflammatory cytokines and 8 CVDs; Step 3 focused on screening the 91 inflammatory cytokines as potential mediators in the pathways linking gut microbiota to CVDs. To estimate and quantify mediator effects, we utilized two-step MR and MVMR methodologies.

Of note, bidirectional MR analysis used exactly the same IV selection criteria, statistical methods (IVW as the primary approach), and significance threshold ($P < 0.05$) as the forward causal analysis to ensure the comparability of forward and reverse results.

Data Sources

This study utilized summary statistics obtained from publicly available data from Genome-Wide Association Studies (GWAS) conducted on populations of European ancestry. Detailed information regarding ethical approval and informed consent can be provided in the relevant publications.

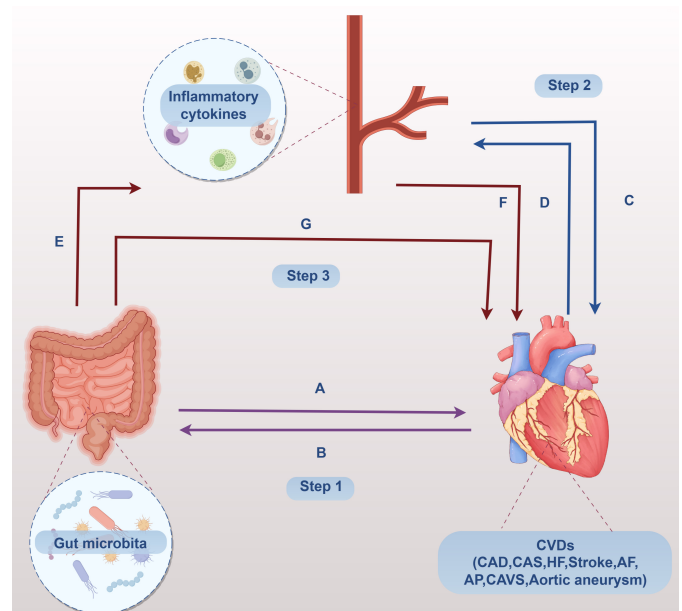


Figure 1. Overview of the study design for the bidirectional two-sample Mendelian randomization (MR) and mediation analysis. The study included three core steps: Step 1, bidirectional two-sample MR analysis to assess the causal relationship between 196 gut microbiota taxa and 8 cardiovascular disease (CVD) subtypes; Step 2, bidirectional two-sample MR analysis to assess the causal relationship between 91 circulating inflammatory cytokines and 8 CVD subtypes; Step 3, two-step mediation MR and multivariate MR (MVMR) analysis to identify the mediating role of inflammatory cytokines in the causal pathway from gut microbiota to CVDs.

Genetic data for gut microbiota were obtained from previously published GWAS summary statistics curated by the MiBioGen consortium, which analyzed genome-wide genotypes and 16S fecal microbiome data from 18,340 individuals across 24 cohorts to assess host genetic influences on microbial composition [13]. The dataset encompassed 196 known bacterial taxa, spanning 9 phyla, 16 classes, 20 orders, 32 families, and 119 genera. Summary-level genetic data for inflammatory cytokines were derived from a separate study involving 8,293 participants, which profiled 91 cytokines and growth factors [14]. The data sources for CAD, CAS, HF, stroke, AF, AP, CAVS and aortic aneurysm used in this study are shown in Table 1.

Instrumental Variables Selection

Instrumental variable selection proceeded in two steps. First, SNPs significantly associated with gut microbiota were identified using a genome-wide significance threshold of $P < 1 \times 10^{-5}$. For inflammatory cytokines, a threshold of $P < 5 \times 10^{-6}$ was applied to retain a sufficient number of valid IVs, given the smaller sample size of the cytokine GWAS; this stricter threshold also mitigates the risk of weak instrument bias [15]. Second, to ensure SNP independence, linkage disequilibrium clumping was performed with an r^2 threshold of 0.001 within a 1000 kb window, using the 1000 Genomes Project European population as the reference panel [16]. To ensure that the effects of a SNPs on exposure and outcome were influenced by the same allele, we performed data harmonization; palindromic SNPs with ambiguous allele frequencies were excluded from the analysis to avoid bias from mismatched effect directions [17].

Third, the genetic instrument's effectiveness was assessed through the F-statistic, and SNPs with an F-statistic greater than 10 were selected to prevent bias from a weak instrument [18]. Formula for calculating F statistic for each SNP as

$$F = \frac{R^2}{1 - R^2} (N - 2)$$

In the formula, R^2 refers to the variance explained by a model's fixed effects part, N indicates sample size. R^2 can be calculated as [19]

$$R^2 = \frac{\beta^2}{(\beta^2 + se^2 \times N)}$$

In this formula, β means effect size for the genetic variant of interest; se means a standard error for β .

MR analysis

Bidirectional Causality Analysis

We conducted two-sample MR analysis to estimate the bidirectional causal effects between gut microbiota abundance, inflammatory cytokines and CVDs, respectively step 1 and step 2 in Figure 1. We used the inverse variance weighted (IVW) method as a main method for MR analysis because it allows for robust causal estimation despite heterogeneity [20]. A pleiotropy-robust MR analysis was conducted using four MR methods (weighted median, weighted mode, MR-Egger, and MR-PRESSO) to investigate whether pleiotropy could contribute to MR estimation bias [21]. Mendelian randomization estimates are presented as odds ratios with 95% confidence intervals. A causal association was considered significant when the inverse-variance weighted method yielded $P < 0.05$ and the direction of effect was consistent with that obtained from MR-Egger regression.

Mediation Analysis

A two-step MR design was used to test the causal mediation effect of inflammatory cytokines (mediator) on the relationship between the gut microbiota (exposure) and CVDs (outcome), as step 3 shown in Figure 1. A two-step approach has fewer biases than a multivariable approach [22]. In the first step

of two-step MR, we perform two-sample MR to estimate the causal effect between exposure and mediator (Figure 1E), mediator and outcome (Figure 1F), exposure and outcome (Figure 1G) separately. In the second step, we conducted a MVMR between exposure and mediator to outcome, in order to explore whether inflammatory cytokines are confounding factors affecting the causal relationship between gut microbiota and CVDs. We evaluated mediation effects using the product of coefficients method ($\beta_E \times \beta_F$) with 95% CIs estimated via the delta method, which directly assessed the significance of the mediation effect that gut microbiota on CVDs through inflammatory cytokines. The total effect represents $\beta_G + \beta_E \beta_F$ [23]. Mediation effect percentage is measured by the mediated proportion = $(\beta_E \times \beta_F) / (\beta_G)$ [24].

Sensitivity Analysis

Sensitivity analysis was conducted to assess the robustness of the causal associations. In the results, we used the cochrane's Q test [25] to assess heterogeneity between SNPs. Scatter plots were used to visualize the SNP-to-exposure and SNP-to-outcome effects, at the same time, leave-one-out analyses was performed to determine whether the association was driven by a single SNP [26]. To assess horizontal pleiotropy, we employed the MR-Egger regression intercept and the MR-PRESSO global test [27]; outlier SNPs identified via MR-PRESSO were removed to correct for horizontal pleiotropy when necessary. Sensitivity analyses were also performed for all mediation pathways to validate the robustness of the mediation effects, with no significant heterogeneity or horizontal pleiotropy detected for the significant pathways [28]. All analyses were conducted using the TwoSampleMR (version 0.5.8) package and MRPRESSO (version 1.0) packages in R software (Version 4.3.2).

Results

Instrumental Variable Selection

First, we screened 2559 SNPs (Supplementary Table 1) from 196 bacterial taxa as IVs with a significance threshold of $P < 1 \times 10^{-5}$. To obtain more comprehensive information, we performed LD-clumping ($r^2 < 0.001$, distance = 10,000kb) [29], harmonizing and testing pleiotropy for all the IVs to reduce correlations between the SNPs. All retained IVs had an F-statistic > 10 , indicating no risk of weak instrument bias. Second, we

Table 1. Summary of the GWAS data used in the MR analysis.

Data Source	GWAS ID	Trait	Sample size	Year	Population
IEU Open GWAS	ebi-a-GCST005195	Coronary artery disease (CAD)	122,733	2017	European
	ebi-a-GCST009541	Heart failure (HF)	977,323	2020	European
	ebi-a-GCST90018793	Angina pectoris (AP)	470,931	2021	European
	ebi-a-GCST90018783	Aortic aneurysm	479,194	2021	European
FINNGEN	finn-b-I9_CORATHER	Coronary atherosclerosis (CAS)	211,203	2021	European
	finn-b-C_STROKE	Stroke	180,862	2021	European
	finn-b-I9_AF	Atrial fibrillation (AF)	138,994	2021	European
	finngen_R9_I9_CAVS_OPERATED	Calcific aortic valve stenosis (CAVS)	377,277	2019	European

identified 1203 SNPs associated with 91 inflammatory cytokines by the threshold value of $P < 5 \times 10^{-6}$ (Supplementary Table 2).

Causal Effects of Gut Microbiota and Cytokines on the Risk of 8 Types of CVD

CAD

The results for the primary analyses show 77 SNPs belong to six gut microbial taxa (including 3 families, 2 genus, and 1 order) were associated with CAD (Supplementary Table 3, Figure 2). The detailed information of the 77 SNPs is shown in Supplementary Table 4.

As presented in Figure 2, the results of the MR analyses suggested that 4 gut microbial taxa were causally associated with the risk of CAD. The family XI (OR=1.055, 95% CI=1.004 - 1.109, $P=0.034$), family Lachnospiraceae (OR=1.097, 95% CI=1.019 - 1.181, $P=0.014$), family Veillonellaceae (OR=1.065, 95% CI=1.011 - 1.122, $P=0.019$), genus Oxalobacter (OR=1.062, 95% CI=1.019 - 1.106, $P=0.004$) increased risk for developing CAD. At the same time, Genetic prediction of two gut microbial taxa (genus Parabacteroides, order Lactobacillales) was associated with a significantly lower risk of CAD. The genus Parabacteroides (OR=0.866, 95% CI=0.784 - 0.956, $P=0.005$), order Lactobacillales (OR=0.866, 95% CI=0.784 - 0.956, $P=0.005$), decreased the risk of CAD.

As shown in Figure 3, Supplementary Table 4, the MR analysis results indicated that 9 cytokines may increase the incidence of CAD. The T-cell surface glycoprotein CD5 levels (OR=1.161, 95% CI=1.032 - 1.308, $P=0.013$), Hepatocyte growth factor levels (OR=1.121, 95% CI=1.007 - 1.247, $P=0.036$) significantly increased the risk of CAD. The Interleukin-5 levels (OR=0.889, 95% CI=0.813 - 0.973, $P=0.011$), Interleukin-8 levels (OR=0.942, 95% CI=0.888-1.000, $P=0.049$), Neurturin levels (OR=0.907, 95% CI 0.835-0.985, $P=0.021$) decreased the risk of CAD.

CAS

A total of 7 gut microbiota (including 1 class, 1 family, 1 genus, 2 orders and 2 phyla) were causally related to CAS (Supplementary Table 3, Figure 2). Details of related 76 SNPs are shown in Supplementary Table 4.

Figure 2 shows the genus Turicibacter (OR=1.906, 95% CI=1.014 - 1.184, $P=0.022$) suggested to increase the risk of CAS by the IVW method. Six gut microbial taxa (class Gammaproteobacteria, family Bifidobacteriaceae, order Bifidobacteriales, order Burkholderiales, phylum Actinobacteria, phylum Firmicutes) has an decreased risk of CAS. The class Gammaproteobacteria (OR=0.829, 95% CI=0.727 - 0.945, $P=0.005$) and phylum Firmicutes (OR=0.876, 95% CI=0.786 - 0.976, $P=0.017$) significantly decreased the risk of CAS.

Figure 3 shows that C-C motif chemokine 4 levels, Fibroblast growth factor 5 levels and Interleukin-22 receptor subunit alpha-1 levels seemed to be risk factors to CAS. The Interleukin-22 receptor subunit alpha-1 levels significantly increased the incidence of CAS. The Leukemia inhibitory factor receptor levels was a protective factor for CAS.

HF

A total of 9 gut microbiota (including 2 class, 4 genus and 3 orders 2 phylums) were causally associated with HF (Supplementary Table 3, Figure 2). Details of related 91 SNPs are shown in Supplementary Table 4.

As Figure 2 shows that five gut microbial taxa (class Negativicutes, genus Eubacterium eligens group, genus Eubacterium oxidoreducens group, genus Flavonifractor, order Selenomonadales) seemed to be risk factors to HF. The genus Eubacterium eligens group (OR=1.126, 95% CI=1.017 - 1.247, $P=0.023$) and genus Flavonifractor (OR=1.144, 95% CI=1.031 - 1.270, $P=0.012$) considerably elevated risk for HF. Four gut microbial taxa (class Bacteroidia, genus Anaerostipes, order Bacillales and order Bacteroidales) seemed to be protective factors to HF. The genus Anaerostipes (OR=0.899, 95% CI=0.825 - 0.978, $P=0.014$) significantly reduced the risk for HF.

In Figure 3, four cytokines (T-cell surface glycoprotein CD5 levels, Macrophage colony-stimulating factor 1 levels, Interleukin-2 levels, Tumor necrosis factor levels) were associated with a increased risk of HF. The T-cell surface glycoprotein CD5 levels (OR=1.100, 95% CI=1.016 - 1.191, $P=0.019$) and Tumor necrosis factor levels (OR=1.110, 95% CI=1.029 - 1.199, $P=0.007$) significantly increased the risk of HF. The Interleukin-17C levels (OR=0.907, 95% CI=0.832 - 0.989, $P=0.027$), Leukemia inhibitory factor receptor levels (OR=0.913, 95% CI=0.849 - 0.981, $P=0.014$) and Sulfotransferase 1A1 levels (OR=0.950, 95% CI=0.905 - 0.998, $P=0.041$) decreased the risk of HF.

Stroke

A total of 10 gut microbiota (including 4 family, 5 genus, 1 order) had a causal relationship with stroke (Supplementary Table 3, Figure 2). Details of related 97 SNPs are shown in Supplementary Table 4.

As shown in Figure 2, the family Bacteroidales S24-7 group (OR=1.109, 95% CI=1.025 - 1.200, $P=0.010$), genus Holdemanella (OR=1.080, 95% CI=1.014 - 1.151, $P=0.017$) and genus Marvinbryantia (OR=1.111, 95% CI=1.002 - 1.231, $P=0.046$) suggested to be risk factors for stroke. Seven gut microbial taxa (family Alcaligenaceae, family Defluviitaleaceae, family Desulfovibrionaceae, genus Eggerthella, genus Faecalibacterium, genus Ruminococcus torques group) may be protective factors for stroke. The family Desulfovibrionaceae (OR=0.875, 95% CI=0.783 - 0.979, $P=0.020$) and genus Ruminococcus torques group (OR=0.769, 95% CI=0.631 - 0.936, $P=0.009$) were significantly lower risk for stroke.

Figure 3 shows that C-X-C motif chemokine 11 levels (OR=1.096, 95% CI=1.012 - 1.187, $P=0.025$), FGF5 (OR=1.053, 95% CI=1.022 - 1.085, $P=0.001$), Interleukin-2 receptor subunit beta levels (OR=1.123, 95% CI=1.026 - 1.230, $P=0.012$) seemed to be risk factors to stroke. The Cystatin D levels (OR=0.963, 95% CI=0.935 - 0.992, $P=0.014$) and Stem cell factor levels (OR=0.954, 95% CI=0.910 - 1.000, $P=0.048$) were protective factors to stroke.

AF

The results suggest a causal relationship between 9 gut microbiota (including 1 family, 5 genus, 1 order and 2 phylum) and AF (Supplementary Table 3, Figure 2). Details of related 106 SNPs are shown in Supplementary Table 4.

As shown in Figure 2, the family Pasteurellaceae (OR=1.107, 95% CI=1.021 - 1.199, $P=0.013$), genus Ruminiclostridium5 (OR=1.142, 95% CI=1.017 - 1.283, $P=0.025$) and order Pasteurellales (OR=1.107, 95% CI=1.021 - 1.199, $P=0.013$) suggest to be risk factors to AF. Six gut microbial taxa (genus Bifidobacterium, genus Eggerthella, genus Howardella, genus

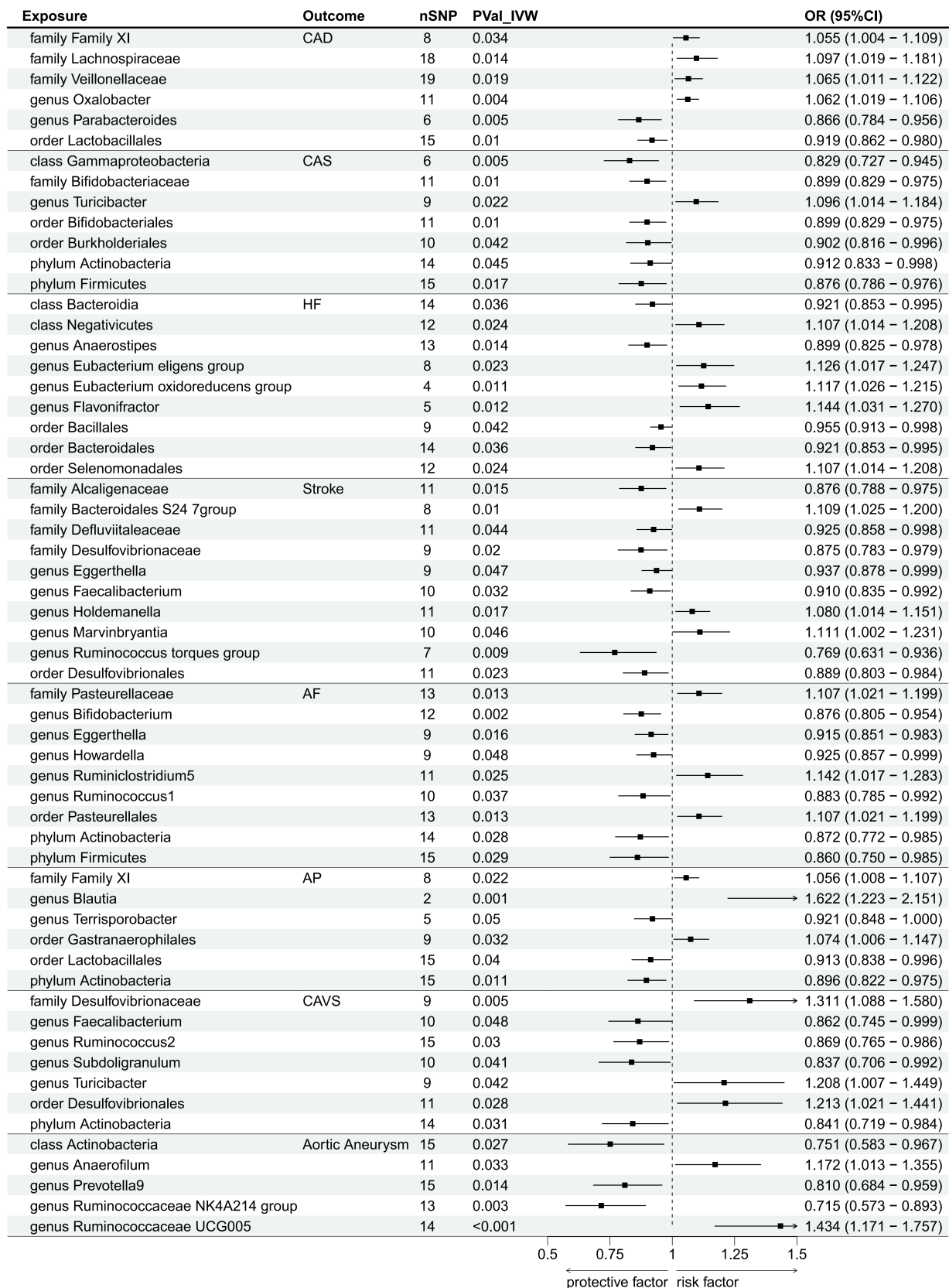


Figure 2. Mendelian randomization results of the causal effects between gut microbiota taxa and 8 CVD subtypes, estimated using the inverse variance weighted (IVW) method. Results are presented as odds ratios (ORs) with 95% confidence intervals (CIs). Red squares indicate a significant positive causal association (increased CVD risk, $P < 0.05$), and blue squares indicate a significant negative causal association (decreased CVD risk, $P < 0.05$).

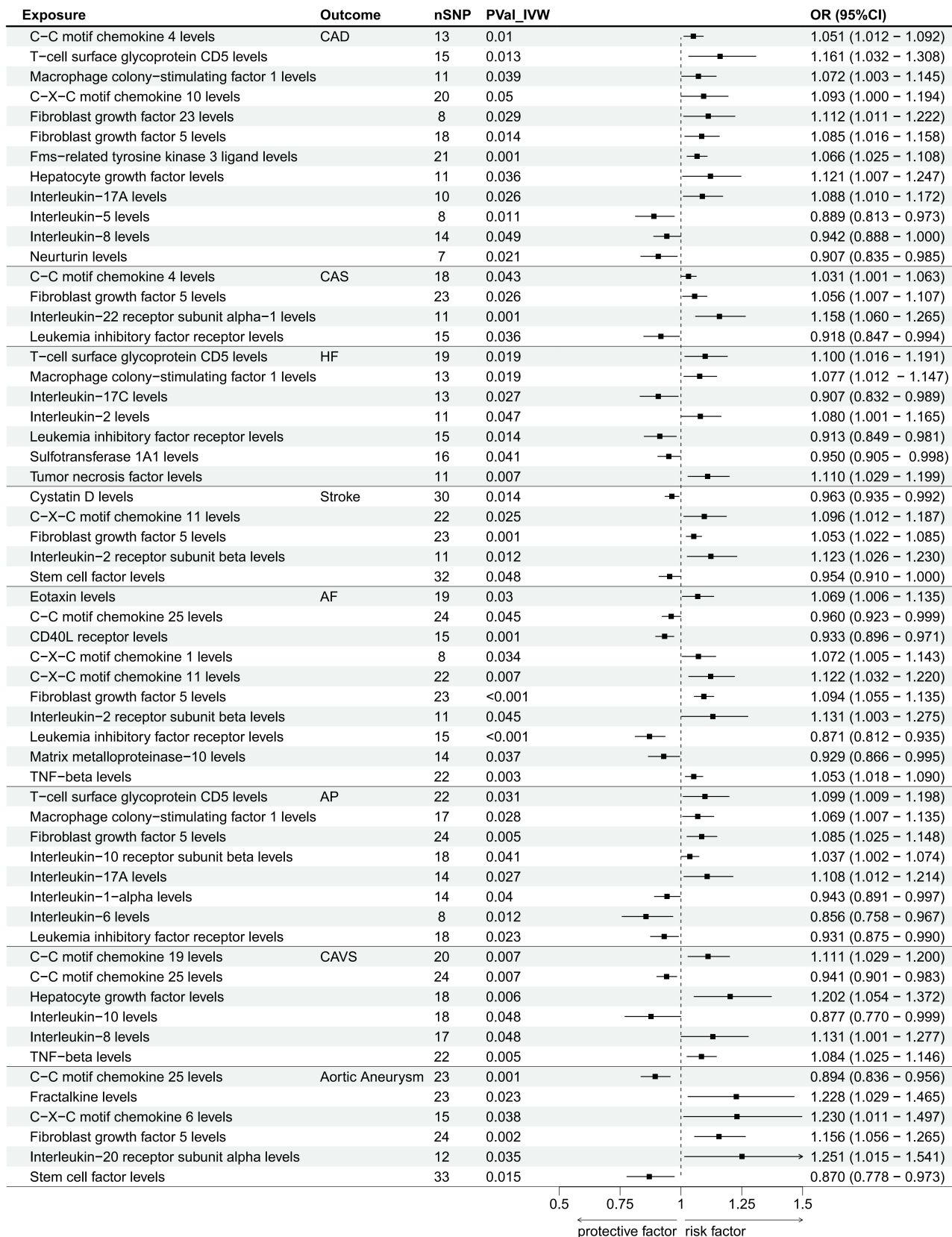


Figure 3. Mendelian randomization results of the causal effects between circulating inflammatory cytokines and 8 CVD subtypes, estimated using the inverse variance weighted (IVW) method. Results are presented as odds ratios (ORs) with 95% confidence intervals (CIs). Red squares indicate a significant positive causal association (increased CVD risk, $P < 0.05$), and blue squares indicate a significant negative causal association (decreased CVD risk, $P < 0.05$).

Ruminococcus 1, phylum Actinobacteria, phylum Firmicutes) may be protective factors for AF. The phylum Firmicutes (OR=0.860, 95% CI=0.750 - 0.985, $P=0.029$) significantly decreased the incidence of AF.

Figure 3 shows that six inflammatory cytokines (Eotaxin levels, C-X-C motif chemokine 1 levels, C-X-C motif chemokine 11 levels, FGF5 levels, Interleukin-2 receptor subunit beta levels, TNF-beta levels) seemed to be risk factors for AF. The Interleukin-2 receptor subunit beta levels (OR=1.131, 95% CI=1.003 - 1.275, $P=0.045$) significantly increased the risk of AF. Four inflammatory cytokines (C-C motif chemokine 25 levels, CD40L receptor levels, Leukemia inhibitory factor receptor levels, Matrix metalloproteinase-10 levels) suggested to be protective factors to AF. Among them, the Leukemia inhibitory factor receptor levels (OR=0.871, 95% CI=0.812 - 0.935, $P<0.001$) significantly decreased the risk of AF.

AP

The IVW method revealed 52 causal associations of gut microbiota on AP, including 1 family, 1 genus, 2 orders and 1 phylum (Figure 2, Supplementary Table 3). Details of related 52 SNPs are shown in Supplementary Table 4. The risk of AP may be increased by three gut microbiota traits, family XI (OR=1.056, 95% CI=1.008 - 1.107, $P=0.022$), genus Blautia (OR=1.622, 95% CI=1.223 - 2.151, $P=0.010$), order Gastranaerophilales (OR=1.074, 95% CI=1.006 - 1.147, $P=0.032$). The genus Terrisporobacter (OR=0.921, 95% CI=0.848 - 1.000, $P=0.050$), order Lactobacillales (OR=0.913, 95% CI=0.838 - 0.996, $P=0.040$), phylum Actinobacteria (OR=0.896, 95% CI=0.822 - 0.975, $P=0.011$) may reduce the risk of AP.

Described in Figure 3, five inflammatory cytokines (T-cell surface glycoprotein CD5 levels, Macrophage colony-stimulating factor 1 levels, FGF5 levels, Interleukin-10 receptor subunit beta levels, Interleukin-17A levels) may increase the risk of AP (Supplementary Table 3). Among them, the Interleukin-17A levels (OR=1.108, 95% CI=1.012 - 1.214, $P=0.027$) significantly increased the risk of AP. At the same time, the Interleukin-1-alpha levels (OR=0.943, 95% CI=0.891 - 0.997, $P=0.040$), Interleukin-6 levels (OR=0.856, 95% CI=0.758 - 0.967, $P=0.012$), Leukemia inhibitory factor receptor levels (OR=0.931, 95% CI=0.875 - 0.990, $P=0.023$) may decrease the risk of AP.

CAVS

Based on the IVW method, the results suggested 78 causal relationships between gut microbiota and CAVS, including 1 family, 4 genus, 1 order and 1 phylum (Figure 2, Supplementary Table 3). Details of related 78 SNPs are shown in Supplementary Table 4. The family Desulfovibrionaceae (OR=1.311, 95% CI=1.088 - 1.580, $P=0.005$), genus Turicibacter (OR=1.208, 95% CI=1.007 - 1.449, $P=0.042$), order Desulfovibrionales (OR=1.213, 95% CI=1.021 - 1.441, $P=0.028$) seemed to be risk factors to CAVS. There were four gut microbiota (genus Faecalibacterium, genus Ruminococcus 2, genus Subdoligranulum, phylum Actinobacteria) was associated with an decreased risk of CAVS. The phylum Actinobacteria (OR=0.841, 95% CI=0.719 - 0.984, $P=0.031$) significantly decreased risk of CAVS.

As shown in Figure 3, four inflammatory cytokines (C-C motif chemokine 19 levels, Hepatocyte growth factor levels, Interleukin-8 levels, TNF-beta levels) seemed to increase the risk of CAVS. The Hepatocyte growth factor levels (OR=1.202, 95% CI=1.054 - 1.372, $P=0.006$) significantly increased risk of

CAVS. The C-C motif chemokine 25 levels (OR=0.941, 95% CI=0.901 - 0.983, $P=0.007$) and Interleukin-10 levels (OR=0.877, 95% CI=0.770 - 0.999, $P=0.048$) decreased the risk of CAVS.

Aortic Aneurysm

Genetic prediction of five gut microbiota (including 1 class and 4 genus) was associated with a causal effect of Aortic aneurysm (Figure 2, Supplementary Table 3). Details of related 68 SNPs are shown in Supplementary Table 4. The genus Anaerofilum (OR=1.172, 95% CI=1.013 - 1.355, $P=0.033$) and genus Ruminococcaceae UCG005 (OR=1.434, 95% CI=1.171 - 1.757, $P<0.001$) may increase the risk of Aortic aneurysm. The class Actinobacteria (OR=0.751, 95% CI=0.583 - 0.967, $P=0.027$), genus Prevotella 9 (OR=0.810, 95% CI=0.684 - 0.959, $P=0.014$), genus Ruminococcaceae NK4A214 group (OR=0.715, 95% CI=0.573 - 0.893, $P=0.003$) may be protective factors to Aortic aneurysm.

A total of six inflammatory cytokines were associated with Aortic aneurysm (Figure 3, Supplementary Table 4). Four inflammatory cytokines (Fractalkine levels, C-X-C motif chemokine 6 levels, FGF5 levels, Interleukin-20 receptor subunit alpha levels) were risk factors for aortic aneurysm, the Interleukin-20 receptor subunit alpha levels (OR=1.251, 95% CI=1.015 - 1.541, $P=0.035$) significantly increased risk of aortic aneurysm. The C-C motif chemokine 25 levels (OR=0.894, 95% CI=0.836 - 0.956, $P=0.001$) and Stem cell factor levels (OR=0.870, 95% CI=0.778 - 0.973, $P=0.015$) had a protective causal effect on Aortic aneurysm.

Bi-Directional Causal Effects of CVDs on Gut Microbiota and Cytokines

CAD

When assessing the bi-directional causal effect of CAD on gut microbiota and cytokines, we found evidence suggesting a genetic predisposition to CAD was linked to specific gut microbiota, with lower abundance in one family, five genus and one order, and higher abundance in one family, four genus and 1 order (Supplementary Table 7). The most significant result in lower abundance was the genus Actinomyces (OR=0.926, 95% CI=0.872 - 0.983, $P=0.011$), and in higher abundance was the genus Gordonibacter (OR=1.098, 95% CI=1.010 - 1.193, $P=0.027$). CAD was associated with a lower abundance in 8 cytokines (Supplementary Table 8), the most significant result was the TNF-beta levels (OR=0.943, 95% CI=0.894 - 0.996, $P=0.034$).

CAS

CAS has a causal relationship on nine gut microbiota (Supplementary Table 7), with lower abundance in two genus (genus Dialister (OR=0.956, 95% CI=0.915 - 1.000, $P=0.050$), genus Family XIII UCG001 (OR=0.927, 95% CI=0.887 - 0.968, $P=0.001$)), and with higher abundance in one class, one family, four genus and one order, the most significant result was the genus Veillonella (OR=1.083, 95% CI=1.030 - 1.140, $P=0.002$). CAS was associated with a lower abundance in 4 cytokines (Supplementary Table 8), the most significant result was the T-cell surface glycoprotein CD5 levels (OR=0.953, 95% CI=0.910 - 0.999, $P=0.046$).

HF

HF was associated with lower abundance in one class, two

family, three genus and one order (Supplementary Table 7), the most significant result was the genus *Victivallis* (OR=0.787, 95% CI=0.635 - 0.974, $P=0.028$). HF has a causal relationship with higher abundance in one family, six genus and one order, the most significant result was the genus *Lachnospiraceae* UCG001 (OR=1.157, 95% CI=1.027 - 1.303, $P=0.016$). HF has a causal association with Interleukin-24 levels (OR=1.094, 95% CI=1.005 - 1.191, $P=0.038$).

Stroke

Stroke has a causal relationship with lower abundance in two genus (genus *Marvinbryantia* (OR=0.902, 95% CI=0.816 - 0.998, $P=0.045$), genus *Prevotella* 9 (OR=0.888, 95% CI=0.805 - 0.979, $P=0.017$)), and with higher abundance in two class (Supplementary Table 7), three genus, the most significant result was the genus *Lactococcus* (OR=1.230, 95% CI=1.050 - 1.440, $P=0.01$). As shown in Supplementary Table 8, there was no reverse effect between stroke and cytokines.

AF

AF was associated with lower abundance in 1 family and 7 genus (Supplementary Table 7), the most significant result was the genus *Prevotella*7 (OR=0.943, 95% CI=0.890-0.998, $P=0.044$), and with high abundance in genus *Butyrivimonas* (OR=1.035, 95% CI=1.001-1.069, $P=0.041$) and genus *Intestini-bacter* (OR=1.134, 95% CI=1.003-1.066, $P=0.034$). As shown in Supplementary Table 8, there was no reverse effect between AF and cytokines.

AP

AP has a causal relationship with lower abundance in 7 gut microbiota (including 1 class, 1 family, 4 genus and 1 order), the most significant result was the genus *Allisonella* (OR=0.891, 95% CI=0.803-0.989, $P=0.030$). AP was associated with higher abundance in 1 class, 1 family, 3 genus and 1 order, the most significant result was the genus *Veillonella* (OR=1.097, 95% CI=1.038-1.159, $P=0.001$). AP was associated with a lower abundance in 8 cytokines (Supplementary Table 8), the most significant result was the T-cell surface glycoprotein CD5 levels (OR=0.913, 95% CI=0.879-0.948, $P=0.001$).

CAVS

CAVS was associated with lower abundance in 7 gut microbiota (including 1 class, 1 family, 3 genus, 1 order and 1 phylum, Supplementary Table 7), the most significant result was the genus *Ruminococcus gnavus* group (OR=0.947, 95% CI=0.896-1, $P=0.005$), and with high abundance in family XI (OR=1.140, 95% CI=1.027-1.264, $P=0.014$). As shown in Supplementary Table 8, there was no reverse effect between CAVS and cytokines.

Aortic Aneurysm

Aortic aneurysm was associated with lower abundance in 13 gut microbiota (including 2 class, 4 family, 4 genus, 2 order and 1 phylum, Supplementary Table 7), the most significant result was the genus *Eubacterium nodatum* group (OR=0.906, 95% CI=0.827-0.993, $P=0.035$), and with high abundance in the genus *Catenibacterium* (OR=1.119, 95% CI=1.004-1.246, $P=0.041$). Aortic aneurysm was associated with a lower abundance in 4 cytokines (Supplementary Table 8), the most significant result was the Leukemia inhibitory factor levels

(OR=0.958, 95% CI=0.932-0.986, $P=0.003$), and with high abundance in Interleukin-18 levels (OR=1.031, 95% CI=1.004-1.058, $P=0.024$).

Sensitivity Analyses

For the robustness of the estimates, we further conducted a sensitivity and pleiotropy analysis to avoid horizontal pleiotropy, with all intercept P values >0.05 and global test P values >0.05 , indicating no significant horizontal pleiotropy for all identified causal associations. Cochran's Q tests demonstrated no significant heterogeneity ($P > 0.05$, Supplementary Table 9).

Mediation Analysis of Gut Microbiota, Inflammatory Cytokines, and CVDs

To explore the underlying potential mechanisms involving in the occurrence and development of CVDs, we performed mediation analysis to identify the causal pathway from gut microbiota to CVDs mediated by cytokines. Initially, the results of two-sample MR indicate that gut microbiota has causal effects on cytokines. We identified 22 associations of gut microbiota to cytokines (CAD,3; CAS,3; HF,4; Stroke,1; AF,5; AP,1; CAVS,3; Aortic Aneurysm,2) (Supplementary Table 10). Secondly, MVMR was applied to explore the significant causal association between cytokines with CVDs after correcting for the gut microbiota. The reliability of results was verified by sensitivity analysis, indicating that no heterogeneity and horizontal pleiotropy existed (Supplementary Table 9).

Notably, the majority of tested pathways showed only a non-significant trend of mediation effect without statistical support, and thus were not considered as valid mediating pathways. In summary, we identified 2 mediating relationships, including 2 gut microbiota-CAVS causal pathways mediated by 1 cytokine (Table 2). The mediation analysis reveals that C-C motif chemokine 25 levels exhibit significant negative mediation effects to CAVS, respectively on genus *Subdoligranulum* ($\beta = -0.017$, 95% CI [-0.035, -0.002], $P=0.034$) with 9.55% proportion, and phylum *Actinobacteria* ($\beta = -0.018$, 95% CI [-0.035, -0.005], $P=0.024$) with 10.41% proportion.

Discussion

Emerging evidence has established strong associations between gut microbiota-derived metabolites [30]—particularly trimethylamine-N-oxide and short-chain fatty acids—and the pathogenesis of cardiovascular diseases [31], including ischemic heart disease [32], heart failure, and hypertension [33]. Beyond metabolic pathways, the gut microbiota also exerts profound effects on host immune responses, modulating key cellular processes such as immunoregulation, apoptosis, and inflammatory signaling [34-36]. Chronic inflammation and autoimmune mechanisms have likewise been implicated in cardiovascular pathology [37-38]. Of particular interest, recent studies have identified autoantibodies directed against C-X-C motif chemokine receptor 3, a G-protein-coupled receptor known to participate in atherosclerotic progression [39]. Nevertheless, the intricate nature of cardiovascular diseases and the diverse composition of gut microbiota present challenges in comprehensively elucidating the impact of gut microbiota on cardiovascular conditions through observational investiga-

tions. Furthermore, the potential mediating role of inflammatory factors in signaling pathways remains uncertain.

This study utilized MR analysis to assess the causal relationship between gut microbiota, inflammatory cytokines and CVDs. The investigation examined potential causal associations between 196 gut microbiota abundance variables and 8 types of CVDs (CAD, HF, AP, Aortic aneurysm, CAS, AF, CAVS), identifying a potential causal relationship involving 58 gut microbiota, 46 inflammatory cytokines, and 8 types of CVDs. Furthermore, mediation analysis results provided support for the involvement of altered inflammatory cytokines in the pathogenesis of CVDs.

Our research findings indicate a potential causal relationship between specific gut microbiota and various cardiovascular diseases, with certain gut microbiota exhibiting a consistent causal role across different types of cardiovascular diseases. Specifically, our results suggest that a high abundance of the phylum Actinobacteria may serve as a protective factor against four cardiovascular diseases (coronary artery stenosis, atrial fibrillation, aortic aneurysm, and calcific aortic valve stenosis). A clinical trial conducted in northwestern China further supported this notion, demonstrating elevated levels of Actinobacteria in the fecal samples of healthy individuals compared to those with hypertension [40]. Collinsella, a dominant taxon of the family Coriobacteriaceae within the phylum Actinobacteria [41], has been found to have elevated abundance in patients with type 2 diabetes and symptomatic carotid atherosclerotic stenosis [42-43].

Additionally, our research indicates that FGF5 levels are a key risk factor for stroke, atrial fibrillation, angina pectoris, and aortic aneurysm. Previous studies from the 1990s have demon-

strated the active role of fibroblast growth factors (FGFs) in regulating endothelial cells (EC) and smooth muscle cells (SMC) [44]. Moreover, numerous studies have demonstrated the significance of inflammatory responses in the vascular system as key contributors to atherosclerosis, with basic Fibroblast Growth Factor (bFGF) playing a role in vascular inflammation [45]. FGF5, a member of the FGF family, has been linked to vascular remodeling and hypertension in previous clinical studies, which aligns with our finding that FGF5 increases the risk of multiple CVD subtypes. Furthermore, our findings suggest that elevated levels of Macrophage colony-stimulating factor 1 may be associated with an increased risk of coronary artery disease, heart failure, and angina pectoris. Conversely, decreased levels of C-C motif chemokine 25 may be linked to a reduced risk of atrial fibrillation, calcific aortic valve stenosis, and aortic aneurysm. Prior research has demonstrated that a deficiency in M-CSF can significantly decrease the progression of atherosclerosis in mice with hyperlipidemia. Additionally, a study has demonstrated that macrophage colony-stimulating factor plays a crucial role in neointimal thickening formation following arterial injury [46]. Chemokines, in turn, contribute to disease pathogenesis by directing leukocyte migration to inflammatory sites [47-48]. Studies utilizing serum proteomic profiling in heart failure cohorts have demonstrated an age-related increase in circulating C-C motif chemokine ligand 17 (CCL17) levels, which correlate with cardiac insufficiency [49]. Nevertheless, the impact of C-C motif chemokine 25 levels on the development of cardiovascular diseases remains poorly understood.

Furthermore, our study investigated the reverse causal effects of 8 CVD subtypes on gut microbiota and cytokine levels

Table 2. Mediation analysis of the causal pathways from gut microbiota to cardiovascular diseases via inflammatory cytokines.

Exposure	Mediator	Outcome	Total effect	Direct effect	Mediation effect (95% CI)	P value	Mediation Proportion
family Veillonellaceae	Macrophage colony-stimulating factor 1 levels		0.063	0.057	0.006 (0, 0.016)	0.134	9.55%
genus Parabacteroides	C-X-C motif chemokine 10 levels	CAD	-0.144	-0.123	-0.021 (-0.058, 0.002)	0.189	14.56%
order Lactobacillales	Interleukin-5 levels		-0.084	-0.076	-0.008 (-0.024, 0.004)	0.253	9.51%
phylum Actinobacteria	Leukemia inhibitory factor receptor levels	CAS	-0.092	-0.079	-0.013 (-0.033, 0)	0.104	14.06%
genus Bifidobacterium	C-C motif chemokine 25 levels		-0.132	-0.125	-0.007 (-0.017, 0)	0.162	5.29%
phylum Actinobacteria	C-C motif chemokine 25 levels	AF	-0.699	-0.690	-0.009 (-0.023, 0)	0.134	1.29%
phylum Actinobacteria	Leukemia inhibitory factor receptor levels		-0.699	-0.681	-0.018 (-0.043, 0)	0.102	2.57%
phylum Actinobacteria	Leukemia inhibitory factor receptor levels	AP	-0.110	-0.099	-0.011 (-0.028, 0)	0.116	9.98%
genus Subdoligranulum	C-C motif chemokine 25 levels	CAVS	-0.178	-0.161	-0.017 (-0.035, -0.002)	0.034	9.55%
phylum Actinobacteria	C-C motif chemokine 25 levels		-0.173	-0.155	-0.018 (-0.035, -0.005)	0.024	10.41%
genus Ruminococcaceae NK4A214 group	Fractalkine levels	Aortic Aneurysm	-0.390	-0.361	-0.029 (-0.071, -0.001)	0.107	7.44%

through bidirectional MR analysis. The findings indicated that 4 CVD subtypes exhibited bidirectional causal effects on the abundance of class Gammaproteobacteria, family Pasteurellaceae, and order Pasteurellales, while 3 CVD subtypes had bidirectional causal effects on levels of Interleukin-12 receptor subunit beta and T-cell surface glycoprotein CD5. These reverse causal findings suggest that the occurrence and progression of CVDs may further exacerbate gut microbiota dysbiosis and systemic inflammatory dysregulation, forming a vicious cycle of "microbiota dysregulation - inflammatory activation - CVD progression", which provides new insights for the long-term management of CVD patients. Additionally, we conducted further analysis to identify potential mediators of inflammatory cytokines in the pathway linking gut microbiota to cardiovascular diseases. Our research revealed that C-C motif chemokine 25 levels act as a mediator in the pathways linking the genera *Subdoligranulum* and phylum *Actinobacteria* to CAVS, with mediation effects of 9.55% and 10.41% respectively.

This study represents the initial exploration of the causal connections between gut microbiota, inflammatory cytokines, and eight CVDs utilizing a thorough and integrated MR Framework. Moreover, we established a pathway from gut microbiota to CVDs through the mediation analysis of inflammatory factors. In order to assess the validity of our findings in light of potential violations of Mendelian randomization assumptions, we conducted various sensitivity analyses using multiple Mendelian randomization approaches. Nonetheless, this study is subject to certain limitations. Firstly, the sample size for cardiovascular disease subtypes was inadequate. Secondly, the study only included populations of European descent due to limited availability of genome-wide association study data from other populations. Lastly, additional experiments and in-depth analysis of clinical patient data will be necessary to confirm the results.

Despite these advancements, our study has several limitations that warrant consideration. First, the sample sizes for CVD subtypes, particularly rare conditions such as CAVS and aortic aneurysm, were relatively small, which may reduce the statistical power to detect causal associations and increase the risk of type II errors. Second, the exclusive use of GWAS data from European populations limits the generalizability of our findings to other ethnic groups. Given the known heterogeneity in gut microbiota composition and genetic susceptibility to CVDs across populations, future studies should prioritize multi-ethnic cohorts to validate these associations. Third, while MR analysis minimizes confounding and reverse causation, it cannot fully replace experimental validation. The mechanisms linking specific bacterial taxa (e.g., *Actinobacteria*) to CVDs protection remain speculative, necessitating functional studies using animal models or in vitro systems to elucidate microbial metabolite production, host immune interactions, and vascular remodeling pathways.

To address these gaps, we propose the following research priorities: (1) Large-scale longitudinal studies integrating multi-omics data (metagenomics, metabolomics, and proteomics) to track dynamic microbiota-cytokine-CVD interactions; (2) Randomized controlled trials testing dietary, probiotic, or anti-cytokine interventions targeting identified mediators (e.g., CCL25 or FGF5); (3) Development of ethnic-specific microbiota biomarkers for CVD risk stratification; and (4) Mech-

anistic investigations into how bidirectional CVD-microbiota effects influence disease progression (e.g., through gut-barrier integrity or systemic inflammation).

Additionally, we emphasize the need for standardization in microbiota nomenclature and cytokine measurement across studies to enhance comparability. Future MR analyses should also explore sex-specific and age-stratified causal effects, given the demographic variability in CVD presentations.

Conclusion

In this study, a comprehensive Mendelian randomization analysis was conducted to investigate the relationships between gut microbiota, inflammatory cytokines, and eight types of CVDs. The analysis revealed 34 positive causal effects and 24 negative causal effects between genetic predisposition in gut microbiota and CVDs, as well as 16 protective factors and 30 risk factors between cytokines and CVDs. We further identified two significant mediating pathways, in which CCL25 mediated the protective effect of *Subdoligranulum* and *Actinobacteria* against CAVS. The gut microbiota taxa and inflammatory cytokines identified in this investigation may serve as novel biomarkers for CVD risk stratification and management, as well as enhance our understanding of the pathophysiological mechanisms underlying CVDs. In conclusion, our findings provide novel genetic evidence for the gut-heart axis, while the outlined limitations highlight the necessity for translational research to bridge these genetic associations with clinical applications.

Abbreviations

Angina Pectoris: AP; Atrial Fibrillation: AF; Calcific Aortic Valve Stenosis: CAVS; Cardiovascular Diseases: CVDs; C-C Motif Chemokine 25: CCL25; C-C Motif Chemokine Ligand 17: CCL17; Confidence Intervals: CI; Coronary Atherosclerosis: CAS; Coronary Artery Disease: CAD; Endothelial Cells: EC; Fibroblast Growth Factor 5: FGF5; Genome-Wide Association Studies: GWAS; Heart Failure: HF; Instrumental Variables: IVs; Inverse Variance Weighted: IVW; Linkage Disequilibrium: LD; Macrophage Colony-Stimulating Factor 1: M-CSF; Mendelian Randomization: MR; Multivariable Mendelian Randomization: MVMR; Odds Ratios: OR; Randomized Controlled Trials: RCTs; Single-Nucleotide Polymorphisms: SNPs; Smooth Muscle Cells: SMC; Trimethylamine N-Oxide: TMAO.

Author Contributions

Xuekun Hou: Conceptualization, Methodology, Formal Analysis, Writing – Original Draft Preparation. Yunxiao Meng: Methodology, Formal Analysis, Writing – Original Draft Preparation. Yuxuan Zhang: Formal Analysis, Data Curation. Xianghan Zheng: Formal Analysis, Data Curation. Hongqi Guo: Validation, Visualization. Yichen Zhang: Validation, Visualization. Hongmei Yuan: Project Administration, Writing – Review & Editing. Zefei Chu: Conceptualization, Writing – Review & Editing, Supervision, Funding Acquisition. All authors read and approved the final manuscript.

Acknowledgements

Not Applicable.

Funding Information

The project was financially supported by Medical Science Research Project of Hebei Provincial Health Commission (20251061/20251060); Scientific Research Project for Institutions of Higher Education, Department of Education of Hebei Province (QN2025430/QN2025264); HBUCLM College Students' Innovation Training Program(S202514432008/S202514432025/S202514432080).

Ethics Approval and Consent to Participate

Not Applicable.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

References

- [1] Hemmati M, Kashanipoor S, Mazaheri P, Alibabaei F, Babaeizad A, Asli S, et al. (2023). Importance of gut microbiota metabolites in the development of cardiovascular diseases (CVD). *Life Sci*, 329, 121947. <https://doi.org/10.1016/j.lfs.2023.121947>
- [2] Chen X, Zhang H, Ren S, Ding Y, Remex NS, Bhuiyan MS, et al. (2023). Gut microbiota and microbiota-derived metabolites in cardiovascular diseases. *Chin Med J (Engl)*, 136(19), 2269–2284. <https://doi.org/10.1097/cm9.0000000000002206>
- [3] Jia Q, Li H, Zhou H, Zhang X, Zhang A, Xie Y, et al. (2019). Role and Effective Therapeutic Target of Gut Microbiota in Heart Failure. *Cardiovasc Ther*, 2019, 5164298. <https://doi.org/10.1155/2019/5164298>
- [4] Nesci A, Carnuccio C, Ruggieri V, D'Alessandro A, Di Giorgio A, Santoro L, et al. (2023). Gut Microbiota and Cardiovascular Disease: Evidence on the Metabolic and Inflammatory Background of a Complex Relationship. *Int J Mol Sci*, 24(10). <https://doi.org/10.3390/ijms24109087>
- [5] Zhang Y, Wang Y, Ke B, & Du J. (2021). TMAO: how gut microbiota contributes to heart failure. *Transl Res*, 228, 109–125. <https://doi.org/10.1016/j.trsl.2020.08.007>
- [6] Blöbaum L, Witkowski M, Wegner M, Lammel S, Schencke PA, Jakobs K, et al. (2023). Intestinal Barrier Dysfunction and Microbial Translocation in Patients with First-Diagnosed Atrial Fibrillation. *Biomedicines*, 11(1). <https://doi.org/10.3390/biomedicines11010176>
- [7] Lin CH, Chen CC, Chiang HL, Liou JM, Chang CM, Lu TP, et al. (2019). Altered gut microbiota and inflammatory cytokine responses in patients with Parkinson's disease. *J Neuroinflammation*, 16(1), 129. <https://doi.org/10.1186/s12974-019-1528-y>
- [8] Soysal P, Arik F, Smith L, Jackson SE, & Isik AT. (2020). Inflammation, Frailty and Cardiovascular Disease. *Adv Exp Med Biol*, 1216, 55–64. https://doi.org/10.1007/978-3-030-33330-0_7
- [9] Haybar H, Bandar B, Torfi E, Mohebbi A, & Saki N. (2023). Cytokines and their role in cardiovascular diseases. *Cytokine*, 169, 156261. <https://doi.org/10.1016/j.cytoto.2023.156261>
- [10] Gagnon E, Mitchell PL, Manikpurage HD, Abner E, Taba N, Esko T, et al. (2023). Impact of the gut microbiota and associated metabolites on cardiometabolic traits, chronic diseases and human longevity: a Mendelian randomization study. *J Transl Med*, 21(1), 60. <https://doi.org/10.1186/s12967-022-03799-5>
- [11] Dai XC, Yu Y, Zhou SY, Yu S, Xiang MX, & Ma H. (2024). Assessment of the causal relationship between gut microbiota and cardiovascular diseases: a bidirectional Mendelian randomization analysis. *BioData Min*, 17(1), 6. <https://doi.org/10.1186/s13040-024-00356-2>
- [12] Lawlor DA, Harbord RM, Sterne JA, Timpson N, & Davey Smith G. (2008). Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat Med*, 27(8), 1133–1163. <https://doi.org/10.1002/sim.3034>
- [13] Kurilshikov A, Medina-Gomez C, Bacigalupe R, Radjabzadeh D, Wang J, Demirkan A, et al. (2021). Large-scale association analyses identify host factors influencing human gut microbiome composition. *Nat Genet*, 53(2), 156–165. <https://doi.org/10.1038/s41588-020-00763-1>
- [14] Ahola-Olli AV, Würtz P, Havulinna AS, Aalto K, Pitkänen N, Lehtimäki T, et al. (2017). Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet*, 100(1), 40–50. <https://doi.org/10.1016/j.ajhg.2016.11.007>
- [15] Ji D, Chen WZ, Zhang L, Zhang ZH, & Chen LJ. (2024). Gut microbiota, circulating cytokines and dementia: a Mendelian randomization study. *J Neuroinflammation*, 21(1), 2. <https://doi.org/10.1186/s12974-023-02999-0>
- [16] Jiang S, Yu C, Lv B, He S, Zheng Y, Yang W, et al. (2023). Two-sample Mendelian randomization to study the causal association between gut microbiota and atherosclerosis. *Front Immunol*, 14, 1282072. <https://doi.org/10.3389/fimmu.2023.1282072>
- [17] Shi Q, Wang Q, Wang Z, Lu J, & Wang R. (2023). Systemic inflammatory regulators and proliferative diabetic retinopathy: A bidirectional Mendelian randomization study. *Front Immunol*, 14, 1088778. <https://doi.org/10.3389/fimmu.2023.1088778>
- [18] Burgess S, & Thompson SG. (2011). Avoiding bias from weak instruments in Mendelian randomization studies. *Int*

- J Epidemiol, 40(3), 755–764. <https://doi.org/10.1093/ije/dyr036>
- [19] Luo S, Li W, Li Q, Zhang M, Wang X, Wu S, et al. (2023). Causal effects of gut microbiota on the risk of periodontitis: a two-sample Mendelian randomization study. *Front Cell Infect Microbiol*, 13, 1160993. <https://doi.org/10.3389/fcimb.2023.1160993>
 - [20] Cai D, Chen J, Wu Y, & Jiang C. (2022). No causal association between tea consumption and 7 cardiovascular disorders: A two-sample Mendelian randomization study. *Front Genet*, 13, 989772. <https://doi.org/10.3389/fgene.2022.989772>
 - [21] De La Barrera B, & Manousaki D. (2023). Serum 25-Hydroxyvitamin D Levels and Youth-Onset Type 2 Diabetes: A Two-Sample Mendelian Randomization Study. *Nutrients*, 15(4). <https://doi.org/10.3390/nu15041016>
 - [22] Richmond RC, Hemani G, Tilling K, Davey Smith G, & Relton CL. (2016). Challenges and novel approaches for investigating molecular mediation. *Hum Mol Genet*, 25(R2), R149–r156. <https://doi.org/10.1093/hmg/ddw197>
 - [23] Zheng X, Liu M, Wu Z, & Jia Z. (2023). Identifying the causal relationship between sedentary behavior and heart failure: Insights from a Mendelian randomization study and mediation analysis. *Clin Cardiol*, 46(9), 1082–1089. <https://doi.org/10.1002/clc.24101>
 - [24] Yang J, Nie D, Chen Y, Liu Z, Li M, Gong C, et al. (2022). The role of smoking and alcohol in mediating the effect of gastroesophageal reflux disease on lung cancer: A Mendelian randomization study. *Front Genet*, 13, 1054132. <https://doi.org/10.3389/fgene.2022.1054132>
 - [25] Cohen JF, Chalumeau M, Cohen R, Korevaar DA, Khoshnood B, & Bossuyt PM. (2015). Cochran's Q test was useful to assess heterogeneity in likelihood ratios in studies of diagnostic accuracy. *J Clin Epidemiol*, 68(3), 299–306. <https://doi.org/10.1016/j.jclinepi.2014.09.005>
 - [26] Burgess S, & Thompson SG. (2017). Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*, 32(5), 377–389. <https://doi.org/10.1007/s10654-017-0255-x>
 - [27] Zhang Z, Xu W, Zheng Y, Chen C, Kang X, Chen D, et al. (2024). Causal relationship between psoriasis vulgaris and dementia: Insights from Mendelian randomization analysis. *Exp Dermatol*, 33(1), e14984. <https://doi.org/10.1111/exd.14984>
 - [28] Verbanck M, Chen CY, Neale B, & Do R. (2018). Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*, 50(5), 693–698. <https://doi.org/10.1038/s41588-018-0099-7>
 - [29] Lu W, Cen J, Dai Q, Tao H, & Peng L. (2023). Gut microbiota does not play a mediating role in the causal association between inflammatory bowel disease and several its associated extraintestinal manifestations: a Mendelian randomization study. *Front Immunol*, 14, 1296889. <https://doi.org/10.3389/fimmu.2023.1296889>
 - [30] Zhang Y, Zhang S, Li B, Luo Y, Gong Y, Jin X, et al. (2022). Gut microbiota dysbiosis promotes age-related atrial fibrillation by lipopolysaccharide and glucose-induced activation of NLRP3-inflammasome. *Cardiovasc Res*, 118(3), 785–797. <https://doi.org/10.1093/cvr/cvab114>
 - [31] Zhao J, Zhang Q, Cheng W, Dai Q, Wei Z, Guo M, et al. (2023). Heart-gut microbiota communication determines the severity of cardiac injury after myocardial ischaemia/reperfusion. *Cardiovasc Res*, 119(6), 1390–1402. <https://doi.org/10.1093/cvr/cvad023>
 - [32] Chen W, Zhao Y, Zhao Q, Zhou Y, Ma C, Dong L, et al. (2026). Lachnospiraceae-Derived Extracellular Vesicles Mediate the Cardioprotective Effects of Barley Leaf in Myocardial Infarction by Improving Intestinal Stem Cell Function. *J Extracell Vesicles*, 15(3), e70250. <https://doi.org/10.1002/jev2.70250>
 - [33] Zou Y, Song X, Liu N, Sun W, & Liu B. (2022). Intestinal Flora: A Potential New Regulator of Cardiovascular Disease. *Aging Dis*, 13(3), 753–772. <https://doi.org/10.14336/ad.2021.1022>
 - [34] Liu Q, Li Y, Song X, Wang J, He Z, Zhu J, et al. (2020). Both gut microbiota and cytokines act to atherosclerosis in ApoE^{-/-} mice. *Microb Pathog*, 138, 103827. <https://doi.org/10.1016/j.micpath.2019.103827>
 - [35] Georgakis MK, Malik R, Burgess S, & Dichgans M. (2022). Additive Effects of Genetic Interleukin-6 Signaling Downregulation and Low-Density Lipoprotein Cholesterol Lowering on Cardiovascular Disease: A 2x2 Factorial Mendelian Randomization Analysis. *J Am Heart Assoc*, 11(1), e023277. <https://doi.org/10.1161/jaha.121.023277>
 - [36] Al Bander Z, Nitert MD, Mousa A, & Naderpoor N. (2020). The Gut Microbiota and Inflammation: An Overview. *Int J Environ Res Public Health*, 17(20). <https://doi.org/10.3390/ijerph17207618>
 - [37] Chen Y, & Zeng L. (2023). Peripheral Inflammatory Factors and Acute Myocardial Infarction Risk: A Mendelian Randomization Study. *Glob Heart*, 18(1), 55. <https://doi.org/10.5334/gh.1269>
 - [38] Ridker PM, Tuttle KR, Perkovic V, Libby P, & MacFadyen JG. (2022). Inflammation drives residual risk in chronic kidney disease: a CANTOS substudy. *Eur Heart J*, 43(46), 4832–4844. <https://doi.org/10.1093/eurheartj/ehac444>
 - [39] Müller FS, Aherrahrou Z, Grasshoff H, Heidorn MW, Humrich JY, Johanson L, et al. (2023). Autoantibodies against the chemokine receptor 3 predict cardiovascular risk. *Eur Heart J*, 44(47), 4935–4949. <https://doi.org/10.1093/eurheartj/ehad666>
 - [40] Lv J, Wang J, Yu Y, Zhao M, Yang W, Liu J, et al. (2023). Alterations of gut microbiota are associated with blood pressure: a cross-sectional clinical trial in Northwestern China. *J Transl Med*, 21(1), 429. <https://doi.org/10.1186/s12967-023-04176-6>
 - [41] Valassi E, Manichanh C, Amodru V, Fernández PG, Gaztambide S, Yañez F, et al. (2023). Gut microbial dysbiosis in patients with Cushing's disease in long-term remission. Relationship with cardiometabolic risk. *Front Endocrinol (Lausanne)*, 14, 1074757. <https://doi.org/10.3389/fendo.2023.1074757>
 - [42] Lambeth SM, Carson T, Lowe J, Ramaraj T, Leff JW, Luo L, et al. (2015). Composition, Diversity and Abundance of Gut Microbiome in Prediabetes and Type 2 Diabetes. *J Diabetes Obes*, 2(3), 1–7. <https://doi.org/10.15436/2376-0949.15.031>
 - [43] Karlsson FH, Fåk F, Nookaew I, Tremaroli V, Fagerberg B, Petranovic D, et al. (2012). Symptomatic atherosclerosis is associated with an altered gut metagenome. *Nat Commun*, 3, 1245. <https://doi.org/10.1038/ncomms2266>

- [44] Liu Y, Liu Y, Deng J, Li W, & Nie X. (2021). Fibroblast Growth Factor in Diabetic Foot Ulcer: Progress and Therapeutic Prospects. *Front Endocrinol (Lausanne)*, 12, 744868. <https://doi.org/10.3389/fendo.2021.744868>
- [45] Sinha SK, Miikeda A, Fouladian Z, Mehrabian M, Edillor C, Shih D, et al. (2021). Local M-CSF (Macrophage Colony-Stimulating Factor) Expression Regulates Macrophage Proliferation and Apoptosis in Atherosclerosis. *Arterioscler Thromb Vasc Biol*, 41(1), 220–233. <https://doi.org/10.1161/atvbaha.120.315255>
- [46] Mishra V, Sinha SK, & Rajavashisth TB. (2016). Role of macrophage colony-stimulating factor in the development of neointimal thickening following arterial injury. *Cardiovasc Pathol*, 25(4), 284–292. <https://doi.org/10.1016/j.carpath.2016.04.003>
- [47] Rabe T, Lazar K, Cambroner C, Goelz R, & Hamprecht K. (2020). Human Cytomegalovirus (HCMV) Reactivation in the Mammary Gland Induces a Proinflammatory Cytokine Shift in Breast Milk. *Microorganisms*, 8(2). <https://doi.org/10.3390/microorganisms8020289>
- [48] Chang TT, & Chen JW. (2016). Emerging role of chemokine CC motif ligand 4 related mechanisms in diabetes mellitus and cardiovascular disease: friends or foes? *Cardiovasc Diabetol*, 15(1), 117. <https://doi.org/10.1186/s12933-016-0439-9>
- [49] Zhang Y, Ye Y, Tang X, Wang H, Tanaka T, Tian R, et al. (2022). CCL17 acts as a novel therapeutic target in pathological cardiac hypertrophy and heart failure. *J Exp Med*, 219(8). <https://doi.org/10.1084/jem.20200418>

The Mechanism of Platelet-Rich Plasma Added to Skin Flap Postoperative for the Influence of Skin Flap Survival

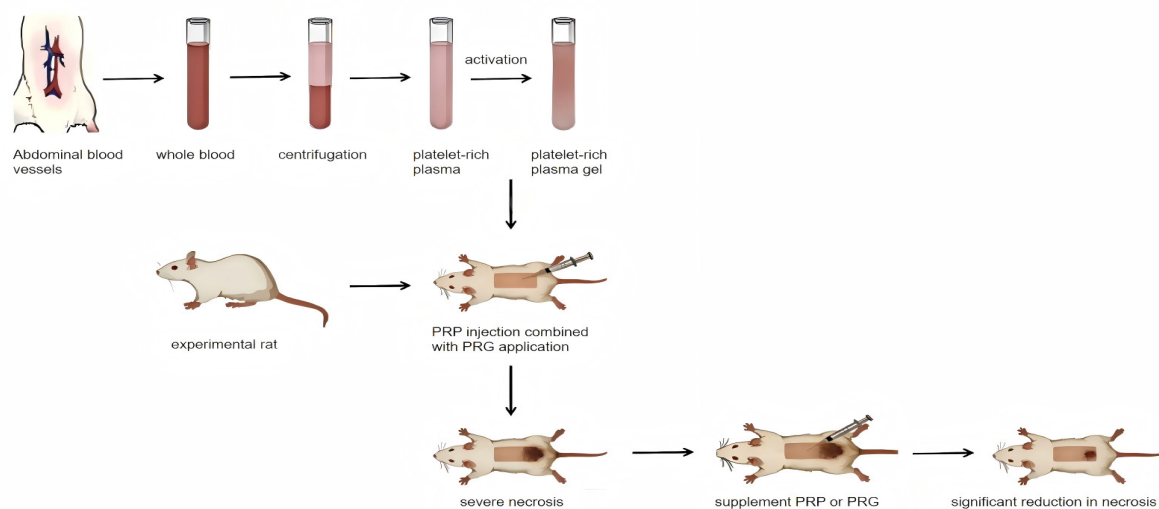
Authors

Chengyang Dong, Zean Lin, Linsen Fang

Correspondence

shaoshangke@126.com (L. Fang)

Graphical Abstract



<https://doi.org/10.71321/0x94xp56>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

The Mechanism of Platelet-Rich Plasma Added to Skin Flap Postoperative for the Influence of Skin Flap Survival

Chengyang Dong¹, Zean Lin¹, Linsen Fang^{2*}

Received: 2025-12-17 | Accepted: 2026-04-14 | Published online: 2026-05-16

Abstract

Background: This study investigated the effects of postoperative platelet-rich plasma (PRP) supplementation on skin flap survival in a rat model. **Methods:** A random-pattern skin flap model was established in rats. During surgery, PRP injection combined with platelet-rich gel (PRG) application was performed. On postoperative day 3, PRP or PRG was supplemented into the flap via uniform point injections. Necrotic area was recorded at postoperative day 7. Flap defect length and terminal edema severity were assessed by contrast-enhanced ultrasonography. Tissue specimens were collected for histopathological and immunohistochemical analysis. Relative expression levels of vascular endothelial growth factor (VEGF), platelet-derived growth factor (PDGF), and cluster of differentiation 34 (CD34) were measured. **Results:** The group receiving intraoperative PRP injection combined with PRG application, followed by postoperative PRG supplementation, showed significantly shorter tissue defect lengths and less terminal tissue edema compared to other groups ($P < 0.01$). Relative expression levels of VEGF, PDGF, and CD34 were significantly higher than those in other groups ($P < 0.01$). **Conclusion:** Postoperative PRP and PRG supplementation reduced flap defect length and edema severity, increased growth factor expression, and improved flap survival. PRG supplementation produced the most pronounced benefits. **Keywords:** free flap; Platelet-Rich Plasma Gel; Platelet-Rich Plasma; ultrasound contrast; VEGF; CD34

Introduction

Skin flap transplantation is commonly used for reconstructing extensive burn wounds and deep injuries secondary to trauma or congenital defects. However, persistent issues such as delayed wound healing and incomplete recovery after flap surgery have been observed [1-2]. Research indicates that poor wound healing is associated with growth factor deficiency, as these factors play a crucial role in wound repair. Platelet-rich plasma (PRP) and platelet-rich gel (PRG), both platelet-rich blood products, contain activated platelets that release multiple growth factors including vascular endothelial growth factor (VEGF). These factors synergistically promote cell proliferation, migration, and angiogenesis, providing multiple positive effects in wound healing and serving as an effective adjunctive therapy to improve outcomes for difficult-to-heal wounds post-skin flap surgery [3-4]. However, partial necrosis may still occur after PRP-injected skin flaps. To investigate this, the research team developed a rat model of free flap. During surgery, 50 μ l/cm²PRP was injected and 50 μ l/cm²PRG

was applied. Postoperatively, an additional 100 μ l/cm²PRP or PRG was administered. The study aimed to determine whether postoperative PRP supplementation could further increase growth factor levels, and enhance flap survival probability, thereby providing new insights for clinical PRP therapy in treating non-healing wounds.

Experimental Animals and Related Materials

Experimental Animals

86 healthy SPF-grade SD female rats (7-8 weeks old, 200-230g) were used in this study. 36 rats were allocated for experimental modeling, while 50 rats were used for PRP preparation. The rats were procured from the Experimental Animal Center of Anhui Medical University (License: SYXK (Wan) 2022-001) and maintained in a controlled environment with 20-28°C temperature and 45-75% humidity. Each cage contained four rats, with free access to food and water. The experiment was approved by the Anhui Medical University Laboratory Animal Ethics

¹ Department of Burn, First Affiliated Hospital of Anhui Medical University, Hefei 230022, China

² The First Affiliated Hospital of Anhui Medical University / Anhui Provincial Public Health Clinical Center Wound Repair and Plastic Surgery, Hefei 230012, China

* Corresponding Author.

Committee (Approval No.: LLSC20252051).

Main Reagents and Instruments

80°C ultra-low temperature freezer, DW-HL5 purchased from Zhongke Meiling Low-Temperature Technology; Excellent Low-Speed Benchtop Centrifuge, L3-5K purchased from Hunan Kecheng Instrument Equipment Co., Ltd.; Ultrasound Diagnostic Instrument, ACUSON Sequoia purchased from Siemens Medical Systems Co., Ltd.; Ruiwode Enhanced Small Animal Anesthesia Machine, R540 purchased from Shenzhen Ruiwode Life Technology Co., Ltd.; Fully Automatic Blood Cell Analyzer, BC-5390 CRP purchased from Mindray Medical International Co., Ltd.; High-Resolution Biological Microscope, DP74 purchased from Olympus. PDGF Antibody Reagents purchased from Hefei Shanben Biotechnology. VEGF antibodies were procured from Affinity; CD34 antibodies from Protein Tech; calcium gluconate injection from Sichuan Meidakanghua Pharmaceutical Co., Ltd.; and sulfur hexafluoride microbubbles for injections from Bracco Suisse SA.

PRP and PRG Preparation

Two female SD rats were randomly selected from 50 SD rats. The anesthetized rats were placed on the experimental table in supine position, and their limbs were fixed. The abdominal area was fully exposed, and the skin was cut along the white line of the abdomen to expose the abdominal vein. The blood was collected by a blood needle and 5ml of whole blood was taken. The whole blood was placed in 10% sodium citrate anti-coagulation tube.

The whole blood was centrifuged at 200g for 20 minutes, and the whole blood was divided into red blood cell layer, white blood cell layer and plasma layer. Plasma and leukocyte layer were collected using a 1ml syringe, followed by a second 200g centrifugation for 10 minutes. The supernatant (top 3/4) was discarded, and the remaining 1/4 was homogenized to obtain plasma-rich plasma (PRP). Mix 1000 U of thrombin lyophilized powder with 1mL of 10% calcium gluconate injection to prepare the activator. Then, mix PRP with the activator in a 10:1 ratio to obtain PRG. The gel-like PRG is left undisturbed for 10-20 minutes until liquefaction, after which filtration yields the liquefied PRG.

Collect 0.1ml of platelet-rich plasma in a sterile EP tube and perform platelet count using the BC-5390 CRP blood cell analyzer. Ensure strict aseptic techniques throughout the procedure to prevent contamination.

Experimental Grouping

The experiment was divided into two parts: one group received 0.5ml saline injection in the flap during operation, and the other group received 0.25mlPRP injection in the flap combined with 0.25mlPRG application. After operation, the two parts received 0.5ml saline, 0.5mlPRP and 0.5mlPRG respectively. Rats were randomly divided into six groups: N+N, N+P, N+G, PG+N, PG+P, and PG+G.

Model Design Concept

The dorsal skin of rats exhibits abundant blood supply, primarily supplied by three arterial regions: the intercostal artery and lumbar artery in the central zone, the deep ilio-rotator artery in the lateral zone, and the thoracodorsal artery in the scapular zone [5]. Our preliminary experiments demonstrated that, given

rats' rich subcutaneous vascular supply and strong repair capacity, even retaining a single thoracodorsal artery as the primary donor would result in minimal necrotic area at the flap's distal end postoperatively. Therefore, this study proposes a free flap model where major vessels are severed intraoperatively, leaving only the dermis, subdermis, and microvascular network in the subcutaneous tissue for blood supply. This design increases the distal necrotic area of the flap, facilitating postoperative evaluation of PRP's therapeutic effects.

Skin Preparation

Place the rats in a small animal anesthesia machine. After inducing anesthesia with 3.0% isoflurane for 3 minutes, adjust the isoflurane concentration to 2.0%. Position the anesthetized rats prone on the experimental table. Use a small razor to remove back hair, then apply an experimental animal depilatory agent to eliminate residual hair roots. After thorough depilation, wipe off the depilatory agent with povidone-iodine to avoid skin burns.

Flap Design and Procedure

A 1×5cm rectangular skin flap was designed on the back of the rats, with the scapula as the pedicle. After routine disinfection and draping, the rat's skin and superficial fascia were incised along the marked line. The superficial fascia was dissected from the deep fascia using a tissue dissection scissors, and the flap was lifted in its entirety. The three major arteries supplying the flap were ligated to ensure the free flap was obtained. Following complete hemostasis, 50μl/cm² of PRP injection and 50μl/cm² of PRG application were performed at the flap base. The flap was then sutured in situ with 5-0 sutures, re-disinfected, and kept warm. Upon recovery, the rats were housed individually in cages to prevent mutual tearing.

Postoperative Addition of PRP or PRG

On the third postoperative day, anesthetized rats were placed prone on the experimental platform. Photographs were taken to record the necrotic area of the skin flap, which was then analyzed using Image Pro Plus software to calculate the necrotic area. After routine disinfection and draping, 0.5 ml of PRP was drawn with a 1ml syringe and uniformly injected as a spot at 1-2 mm from the flap edge. The other treatment group received the same dose of liquid PRG, while the control group received the same dose of saline. After anesthesia recovery, the rats were returned to their cages and continued under standard feeding conditions.

Ultrasound Contrast

Imaging and Flap Sampling On the seventh postoperative day, rats were anesthetized again and photographed to document necrotic flap area. The images were analyzed using Image Pro Plus software to calculate necrotic area. Rats were positioned supine on the experimental table, routinely disinfected, and gauze was laid. One side of the jugular vein was isolated and punctured with a cannula, which was then sutured (see figure below). A 0.5ml dose of sulfur hexafluoride microbubble contrast agent was injected uniformly via the cannula, followed by ultrasound contrast imaging guided by 10MHz ultrasound. Under contrast imaging, the length of the visual defect and the thickness of edema at the flap margin were recorded. A small amount of whole blood was taken for platelet count before eu-

thanizing the rats for sampling. The samples were processed separately for pathological immunohistochemistry.

Immunohistochemistry

The mouse skin samples were embedded in paraffin to prepare white slides for HE staining and immunohistochemical detection. HE staining was used to observe the gross morphology of tissues and infiltration extent of inflammatory markers in sections. Immunohistochemical staining was employed to detect the relative expression levels of VEGF and PDGF in tissues. Five randomly selected 400x field observations from each well-processed slide were scored, with the final score calculated as the average of five evaluations. Positive results were defined as brown-yellow or brown-brown staining particles localized to cytoplasm or cell membranes, with the percentage of positive cells counted as a percentage of skin tissue cells. Scoring criteria were as follows: 1) Staining: No staining = 0 points; light yellow = 1 point; light brown = 2 points; brown-brown = 3 points. 2) Positive cell percentage: 0% = 0 points; 1%–24% = 1 point; 25%–49% = 2 points; 50%–74% = 3 points; 75%–100% = 4 points. 3) The final score was determined by multiplying staining intensity by the percentage of positive cells. Statistical analysis was performed on the final scores of each group.

Microvessel Counting

Under the CD34 immunostaining field magnified 400 times, 5 non-adjacent areas were randomly selected for scoring. The number of microvessels was counted and averaged to determine the vascular density of each sample. Statistical analysis was performed on the microvessel counts of rats in each group.

Statistical Methods

Statistical analysis was performed using SPSS software. Quantitative data conforming to normal distribution were presented as mean $\pm$ standard deviation ($\bar{X} \pm s$). One-way ANOVA was employed for inter-group comparisons. For pairwise comparisons with homogeneous variances, the Bonferroni method was used. A P-value < 0.05 was considered statistically significant.

Results

Platelet Count Analysis

Through statistical analysis of platelet counts in whole blood, intraoperative PRP, and postoperative PRP from all groups: The platelet count in whole blood of all groups was $(622.53 \pm 52.70) \times 10^9 / L$, with no significant intergroup differences ($F=1.128$; $P=0.367$). The platelet count in intraoperative PRP was $(2469.50 \pm 225.75) \times 10^9 / L$, which was 3.966 times that of whole blood, with no significant intergroup difference ($F=2.479$; $P=0.139$). The platelet count in postoperative PRP was $(2287.81 \pm 168.65) \times 10^9 / L$, which was 3.674 times that of whole blood, with no significant intergroup difference ($F=2.409$; $P=0.118$). (Table 1)

Contrast-enhanced Ultrasound

On the seventh postoperative day, contrast-enhanced ultrasound revealed no significant differences in normal tissue thickness among all rat groups in this study ($P>0.05$). Measurements of defect length and terminal edema in skin flaps showed that the PG+G group exhibited the smallest defect length and lightest terminal edema compared to other groups ($P<0.05$) (Figure 1). Compared with the N+N group, the N+P and N+G groups demonstrated significantly reduced terminal edema and defect severity ($P<0.001$). In contrast, the PG+N group exhibited more pronounced terminal edema and defect severity than the PG+P group ($P<0.001$), while the PG+G group showed marked improvement ($P<0.05$).

Flap Necrosis Status

On the third postoperative day, all rat groups exhibited varying degrees of necrosis in the distal flap areas (Table 2). The PRP-treated group showed significantly reduced necrotic area compared to the saline-treated group ($P<0.05$). By the seventh day, the PG+P and PG+G groups demonstrated statistically significant reduction in necrotic area ($P<0.001$) compared to other groups. Notably, the N+P and N+G groups showed marked necrosis reduction ($P<0.001$) relative to the N+N group, while the PG+P and PG+G groups also exhibited decreased necrotic area ($P<0.001$) compared to the PG+N group. However, no significant difference was observed between PG+P and PG+G groups ($P>0.05$), nor between N+P and N+G groups ($P>0.05$). These results indicate that the saline-treated group showed further expansion of necrotic area by the seventh day, whereas

Table 1. Platelet count in whole blood and PRP of rats in each group ($\bar{X} \pm s$, $10^9 / L$)

Group	Number	whole blood	Intraoperative PRP	Postoperative PRP
N+N	6	—	—	—
N+P	6	613.33 ± 26.80	—	2443.75 ± 187.45
N+G	6	613.33 ± 38.28	—	2307.25 ± 157.86
PG+N	6	655.33 ± 51.86	2425.50 ± 113.66	—
PG+P	6	637.50 ± 59.00	2644.50 ± 30.23	2197.25 ± 160.10
PG+G	6	622.50 ± 49.68	2338.50 ± 254.98	2203.81 ± 55.43
F		1.128	2.479	2.409
P		0.367	0.139	0.118

Note: Data are expressed as mean $\pm$ standard deviation ($\bar{X} \pm s$), and one-way analysis of variance was performed. Bonferroni method was used for inter-group comparisons in this dataset. Statistical analysis showed no significant differences in platelet counts among the groups in whole blood, intraoperative PRP, and postoperative additional PRP. And PRP was not used at the horizontal line.

the PRP or PRG-treated groups maintained stable necrotic extent compared to the third-day baseline.

HE Staining

From HE staining of rat skin flaps postoperatively in all groups showed chronic inflammatory cell infiltration and epidermal shedding in all flap tissues. The N+N group exhibited the most pronounced inflammatory cell infiltration and epidermal shedding compared to other treatment groups (Figure 2A). In contrast, PG+G and PG+P groups demonstrated significantly fewer inflammatory cells and reduced epidermal shedding, with well-preserved epithelial cells (Figure 2E-F).

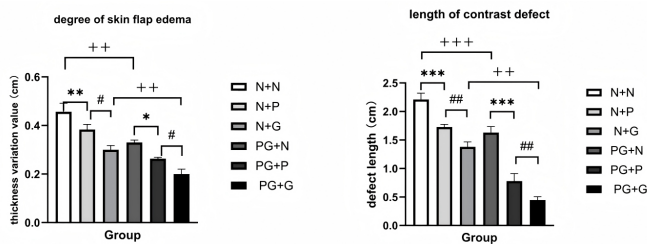


Figure 1. Comparison of flap edema degree and defect length on POD 7.

Comparison of flap edema severity and defect length on postoperative day 7 in six groups of rats. Thirty-six rats were randomly divided into a saline group (n=18, treated with normal saline) and an experimental group (n=18, treated with PRP+PRG). On postoperative day 3, the rats were further subdivided into three groups: an additional saline group (n=6), an additional PRP group (n=6), and an additional PRG group (n=6). On postoperative day 7, flap edema severity and defect length were measured using contrast-enhanced ultrasound. Data were presented as mean $\pm$ standard deviation (SD). Statistical analysis was performed using One-way ANOVA.

Note: * $P<0.05$ when comparing the additional PRP group; ** $P<0.01$; *** $P<0.001$; # $P<0.05$ when comparing with the additional PRG group; ## $P<0.01$; ### $P<0.001$; + $P<0.05$ when comparing with the group receiving PRP combined with PRG during surgery, ++ $P<0.01$; +++ $P<0.001$. X-axis: Group N+N; Group N+P; Group N+G; Group PG+N; Group PG+P; Group PG+G. Y-axis: edema thickness of skin flap (cm); length of contrast defect (cm).

Table 2. The necrotic area of skin flaps ($\bar{X} \pm s$, cm^2)

Group	Number	D3	D7
N+N	6	3.37 $\pm$ 0.27	4.01 $\pm$ 0.19
N+P	6	3.31 $\pm$ 0.19	3.34 $\pm$ 0.17 ²
N+G	6	3.39 $\pm$ 0.29	3.28 $\pm$ 0.16 ^{2,3}
PG+N	6	1.14 $\pm$ 0.13 ⁵	2.68 $\pm$ 0.26 ²
PG+P	6	1.57 $\pm$ 0.19 ⁵	1.60 $\pm$ 0.17 ^{1,2}
PG+G	6	1.31 $\pm$ 0.24 ⁵	1.34 $\pm$ 0.25 ^{1,2,4}
F		131.479	166.698
P		<0.001	<0.001

Note: Data are expressed as mean $\pm$ standard deviation ($\bar{X} \pm s$), and one-way analysis of variance was performed. Bonferroni method was used for inter-group comparisons in this dataset. 1 indicates $P<0.001$ compared to PG+N group; 2 indicates $P<0.001$ compared to N+N group; 3 indicates $P>0.05$ compared to N+P group; 4 indicates $P>0.05$ compared to PG+P group; 5 indicates $P<0.05$ compared to N+N, N+P, and N+G groups.

Immunohistochemical Analysis

On the seventh postoperative day, immunohistochemical analysis was performed on the rat skin flap specimens to detect the levels of growth factors including VEGF and PDGF. Results showed that: Compared with the N+N group, all other treatment groups exhibited significantly increased expression levels of PDGF and VEGF ($P<0.05$). Among the treatment groups with significant differences from the N+N group, the PG+G group demonstrated notably higher expression levels of PDGF and VEGF compared to other groups ($P<0.05$) (Figure 3, 4, and 5). Compared with the group receiving PRP postoperatively, all PRG groups showed marked increases in growth factor levels. Specifically, the N+G group exhibited significantly higher expression levels than the N+P group ($P<0.05$). The PG+G group showed significantly higher growth factor expression levels compared to the PG+P group, while the PG+N group showed significantly lower levels than the PG+P group ($P<0.05$) (Table 3).

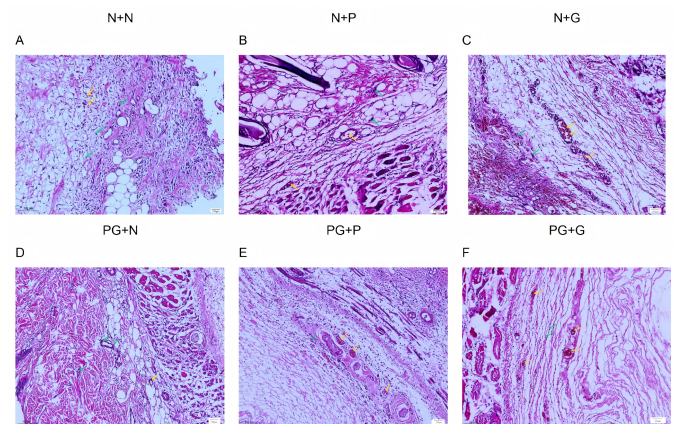


Figure 2. HE staining of rat skin flaps.

HE staining was performed to observe the integrity of tissue sections, the extent of inflammatory cell infiltration, and the number of blood vessels. HE staining revealed randomly selected sections. In group N+N, extensive inflammatory cell infiltration, vascular congestion, and tissue necrosis were observed. Group N+P with added PRP and Group N+G with added PRG showed better preservation of cell morphology and an increased number of blood vessels. Compared to group PG+N and group PG+P, group PG+G exhibited better tissue morphology, milder inflammatory infiltration, and a higher number of blood vessels. Yellow arrows indicate blood vessels; Green arrows indicate inflammatory cell infiltration. Scale =100 μm . HE staining: Hematoxylin-eosin staining.

Microvessel Count Analysis

Microvessel counts were determined through CD34 detection. Compared with other groups, the PG+G group showed a significantly higher number of microvessels ($P<0.001$). All other treatment groups exhibited statistically significant increases in microvessel count compared to the N+N group ($P<0.05$) (Figure 6-7). The N+G group demonstrated a significantly higher microvessel count than the N+P group ($P<0.05$), while the PG+P group showed a markedly higher count than the PG+N group ($P<0.01$) (Table 4).

Discussion

A random flap is a localized skin flap primarily supplied by the dermal

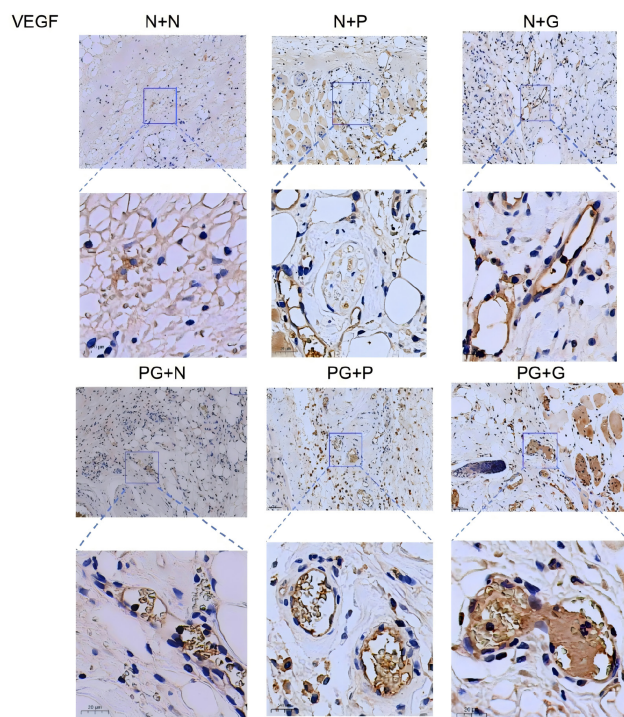


Figure 3. VEGF staining results

The relative expression levels of VEGF in tissues were detected by immunohistochemical staining. Five randomly selected sections with satisfactory staining were scored, and the final score was calculated as the average of the five scores. Positive results were defined as brown-yellow or brown-brown staining granules localized on the cytoplasm or cell membrane. The group receiving additional PRP or PRG showed a higher proportion of brownish-yellow areas and a higher final score. The group receiving saline supplementation exhibited a lower proportion of light yellow area scores. Scale = 20 μ m.

layer, subdermal layer, and microvascular network within subcutaneous tissue, excluding dominant axial cardiovascular structures or large-caliber arteries with accompanying veins [6]. Due to unstable blood supply and limited perfusion range, the distal portion of such flaps is prone to ischemic necrosis [7]. Historically, it was believed that maintaining fixed length-to-width ratios was essential for achieving high flap survival rates. However, with advancements in anatomy, Milton et al. designed flaps with varying length-to-width ratios on pig backs—though all sharing the same blood supply base—revealing that flap survival depends not on specific ratios but on the underlying vascular supply [8]. This groundbreaking conclusion challenged the rigid constraints of fixed length-to-width ratios in random flaps. Rat dorsal skin is primarily supplied by three parallel vascular networks: the central region, fed by the posterior perforating branches of the intercostal and lumbar arteries, serves as the primary blood supply source; the lateral region is supplied by the deep ilio-rotator artery; and the scapular region by the thoracolumbar artery [5]. This three-zone vascular "map" provides an anatomical basis for experimental design of random flaps. Additionally, the necrotic areas in rat flaps remain stable and can be precisely quantified using Image Pro Plus to calculate necrosis percentages, ensuring objective and reliable data. Based on these principles, the research team developed a rat model for random flap experiments, surgically isolating the axial cardiovascular structures within the region while preserving only the dermal layer and subcutaneous microvascular network to induce distal necrosis, with

necrosis areas analyzed using Image Pro Plus software.

The essence of skin flap healing lies in establishing new blood supply and completing tissue repair under ischemic conditions. In the early postoperative period, the flap's blood supply is entirely dependent on the pedicle. Flaps distant from the pedicle experience ischemia and hypoxia due to inadequate blood perfusion, leading to lactic acid accumulation, cellular acidosis, reduced ATP production, and insufficient cellular energy supply [9]. As platelets aggregate and activate, temporary stromal formation releases inflammatory factors, initiating the inflammatory response. Excessive neutrophil activation, oxygen free radical bursts, tissue edema, and microthrombus formation cause ischemia-reperfusion injury, further exacerbating flap necrosis [10]. Around the third postoperative day, under the regulation of hypoxia-inducible factor-1 α (HIF-1 α), endothelial cells of the original blood vessels in the flap are activated by growth factors including vascular endothelial growth factor (VEGF), platelet-derived growth factor (PDGF). These cells proliferate and migrate to form new capillary buds, interconnecting to create a functional microvascular network that generates fresh blood supply to promote flap healing [10-11].

Platelet-rich plasma (PRP), a platelet-concentrated biological agent isolated through gradient centrifugation from autologous whole blood, contains platelets 3-5 times more abundant than peripheral blood. Platelet-rich gel (PRG), a gelatinous substance produced by adding activators to PRP, shares the ability to enhance growth factor concentrations and reduce inflammatory cytokine responses [12-13].

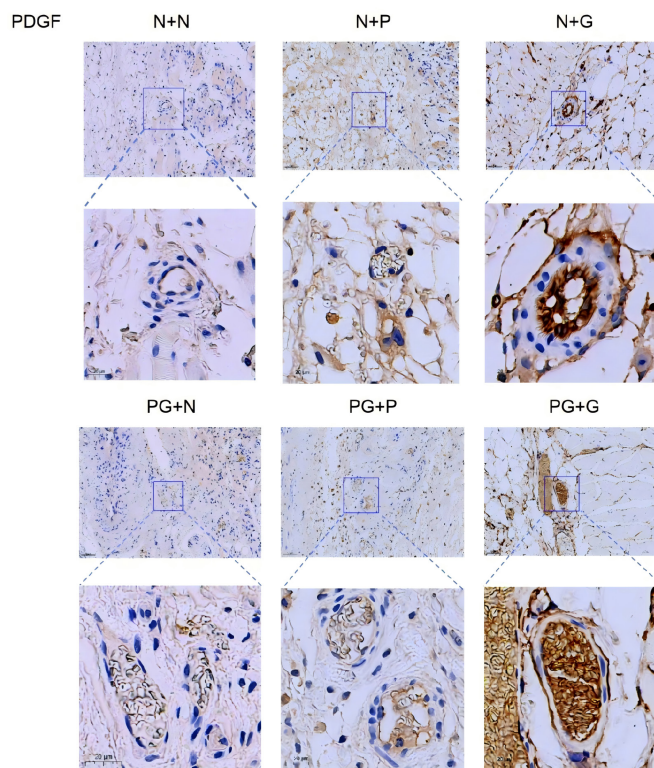


Figure 4. PDGF staining results.

The relative expression levels of PDGF in tissues were detected by immunohistochemical staining. Five randomly selected sections with satisfactory staining were scored, and the final score was calculated as the average of the five scores. Positive results were defined as brown-yellow or brown-brown staining granules localized on the cytoplasm or cell membrane. The group receiving additional PRP or PRG showed a higher proportion of brownish-yellow areas and a higher final score. Scale = 20 μ m.

13]. However, due to the weight limitations of rats, autologous blood supply could not be achieved. This study therefore used allogeneic rat blood for PRP and PRG preparation. No significant immune rejection was observed during the experiment, consistent with Akbarzadeh S et al.'s findings [14]. Beyond its high platelet concentration, PRP contains multiple key growth factors, particularly platelet-derived growth factor (PDGF), vascular endothelial growth factor (VEGF). ELISA confirmed that these factors in PRP are 3-4 times higher than in ordinary plasma while maintaining high activity.

VEGF, a 45-kD heparin-binding dimeric glycoprotein, is recognized as the most potent regulator in angiogenesis induction [15]. By binding to its receptor VEGFR2, VEGF activates the P13K/AKT signaling pathway, which phosphorylates and activates endothelial nitric oxide synthase (eNOS) in endothelial cells. This increases NO production, thereby inducing vasodilation, promoting angiogenesis, and improving vascular permeability [16]. Kryger Z, Vourtsis SA et al. created rat skin flaps and administered VEGF suspension to the distal 1/3 flap through various delivery methods [17-18]. They found that under conditions of hypoxia and endothelial cell damage, regardless of delivery method, VEGF could enhance blood flow and flap survival in rat skin flaps. In our study, Point injections of PRP or PRG were administered at the flap margins, with photographic documentation followed by Image Pro Plus analysis of necrotic areas. The analysis of the data revealed that the saline group showed significantly increased necrotic area by day 7 compared to day 3. In contrast, the PRP or PRG groups demonstrated no significant difference in necrotic extent between days 3 and 7. These findings suggest that while PRP supplementation cannot reverse necrotic outcomes, it effectively delays the progression of necrosis. Immunohistochemical analysis further demonstrated markedly higher growth factor levels in PRP/PRG groups compared to saline control, with PRG supplementation showing more substantial improvements. The results indicate that PRP/PRG supplementation enhances tissue growth factor concentrations, activates vascular endothelial cells to promote angiogenesis, and delays necrotic progression, thereby improving necrosis rates. However, these effects were limited in flaps already showing significant necrosis.

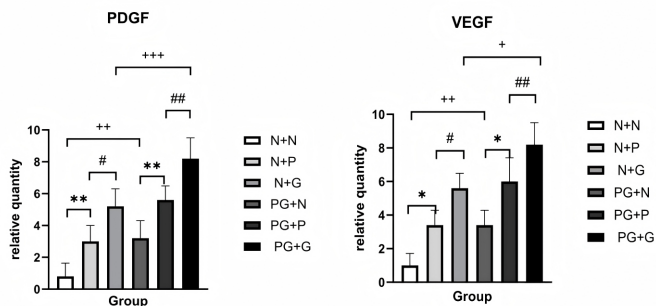


Figure 5. Relative expression levels of VEGF and PDGF

The relative expression levels of VEGF and PDGF in tissues were detected by immunohistochemical staining. Five randomly selected sections with satisfactory staining were scored, and the final score was calculated as the average of the five scores. Positive results were defined as brown-yellow or brown-brown staining granules localized on the cytoplasm or cell membrane. Data were presented as mean $\pm$ standard deviation (SD). Statistical analysis was performed using One-way ANOVA. * $P < 0.05$ when comparing the additional PRP group; ** $P < 0.01$; *** $P < 0.001$; # $P < 0.05$ when comparing with the additional PRG group; # $P < 0.01$; # $P < 0.001$; + $P < 0.05$ when comparing with the group receiving PRP combined with PRG during surgery, ++ $P < 0.01$; +++ $P < 0.001$. X-axis: Group N+N; Group N+P; Group N+G; Group PG+N; Group PG+P; Group PG+G. Y-axis: relative quantity.

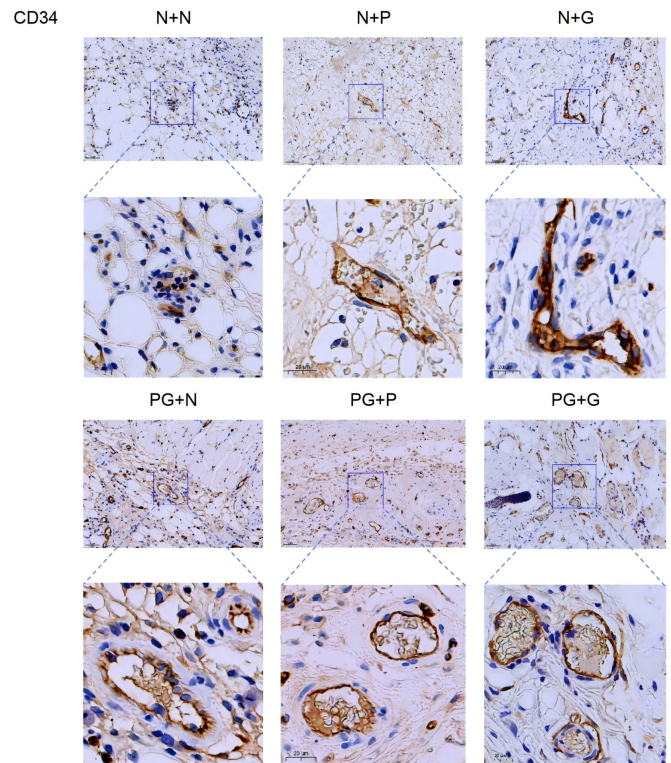


Figure 6. CD34 staining results.

Under the background of CD34 staining, 5 non-adjacent areas were randomly selected for scoring. The number of microvessels was counted and averaged to determine the vascular density of each sample. The stained sections revealed that the group supplemented with PRP and PRG exhibited a higher number of microvessels.

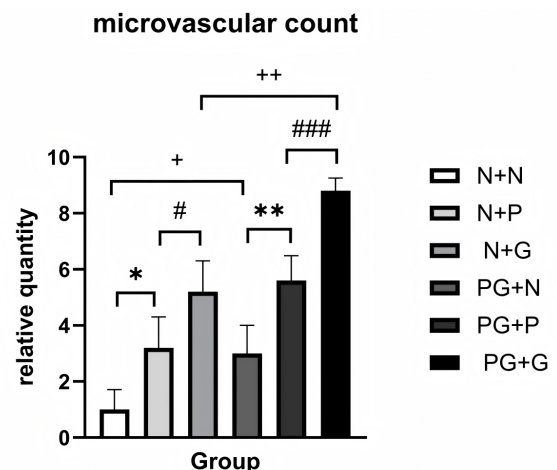


Figure 7. Microvessel Counting

Microvessel density (MVD) was expressed as the number of CD34-positive vessels per high-power field under light microscopy, with 5 randomly selected fields counted per slide. Data were presented as mean $\pm$ standard deviation (SD). Statistical analysis was performed using One-way ANOVA. * $P < 0.05$ when comparing the additional PRP group; ** $P < 0.01$; *** $P < 0.001$; # $P < 0.05$ when comparing with the additional PRG group; # $P < 0.01$; # $P < 0.001$; + $P < 0.05$ when comparing with the group receiving PRP combined with PRG during surgery, ++ $P < 0.01$; +++ $P < 0.001$. X-axis: Group N+N; Group N+P; Group N+G; Group PG+N; Group PG+P; Group PG+G. Y-axis: relative quantity.

Table 3. Statistical table of immunohistochemical results ($\bar{X} \pm s$).

Group	Number	PDGF			VEGF		
		Intensity	Percentage	Total	Intensity	Percentage	Total
N+N	6	0.60±0.89	0.40±0.55	0.80±0.84	0.80±0.45	1.00±0.71	1.00±0.71
N+P	6	1.60±0.55	2.00±0.45	3.00±1.00 ²	1.60±0.55	2.20±0.45	3.40±0.89 ²
N+G	6	2.40±0.55	2.20±0.45	5.20±1.10 ^{3,5}	2.20±0.45	2.60±0.55	5.60±0.89 ^{3,5}
PG+N	6	1.80±0.45	1.80±0.45	3.2±1.10 ²	1.60±0.55	2.20±0.45	3.40±0.89 ²
PG+P	6	2.20±0.45	2.60±0.55	5.60±0.89 ^{4,5}	2.20±0.45	2.80±0.84	6.00±1.41 ^{4,5}
PG+G	6	2.80±0.45	3.00±0.71	8.20±1.30 ^{1,4,5}	2.80±0.45	3.00±0.71	8.20±1.30 ^{1,4,5}
F				29.867			28.873
P				<0.001			<0.001

Note: Data are expressed as mean $\pm$ standard deviation ($\bar{x} \pm s$), and one-way analysis of variance was performed. Bonferroni method was used for inter-group comparisons in this dataset. 1 indicates $P < 0.05$ compared to PG+P group; 2 indicates $P < 0.05$ compared to N+N group; 3 indicates $P < 0.05$ compared to N+P group; 4 indicates $P < 0.05$ compared to PG+N group; 5 indicates $P < 0.001$ compared to N+N group.

Table 4. Statistical table of microvascular count results ($\bar{X} \pm s$).

Group	Number	Microvascular Count
N+N	6	1.00±0.71
N+P	6	3.20±1.10 ²
N+PG	6	5.20±1.10 ^{1,3}
PG+N	6	3.00±1.00 ²
PG+P	6	5.60±0.89 ^{1,4}
PG+PG	6	8.80±0.45 ^{1,5}
F		44.539
P		<0.001

Note: Data are expressed as mean $\pm$ standard deviation ($\bar{x} \pm s$), and one-way analysis of variance was performed. Bonferroni method was used for inter-group comparisons. 1 indicates $P < 0.001$ compared to N+N group; 2 indicates $P < 0.05$ compared to N+N group; 3 indicates $P < 0.05$ compared to N+P group; 4 indicates $P < 0.01$ compared to PG+N group; 5 indicates $P < 0.001$ compared to PG+P group.

Ischemic necrosis of rat skin flaps typically manifests as characteristic blackening, crusting, hardened texture, and absence of blood flow upon needle puncture at the distal flap. While these macroscopic signs may initially suggest vascular impairment, they fail to accurately reflect the pathological and physiological state of subcutaneous tissues. Epidermal appearance often fails to correlate with the extent of subcutaneous edema or the defect range in microcirculatory perfusion imaging. Contrast-enhanced ultrasonography (CEUS) utilizes contrast agents such as sulfur hexafluoride microbubbles, with their diameter strictly controlled within 2-5 μ m. These microbubbles are completely metabolized through respiratory pathways within 15 minutes, eliminating dependence on hepatic and renal metabolic systems. Data from a large-scale multicenter clinical study involving 460,000 patients demonstrated an adverse event rate of merely 0.034%, with severe allergic reactions occurring at an even lower rate of 0.001% highlighting CEUS' advantages of high stability, excellent imaging quality, low diffusion, high biocompatibility, and safety [19-21]. In this study, CEUS revealed that compared to the group receiving only saline postoperatively, the group receiving PRP showed significantly smaller imaging defects and milder edema. Notably, the PG+G group demonstrated the best results, with both the length of imaging defects and flap edema being significantly lower than other treatment groups.

Conclusion

The findings demonstrate that postoperative application of platelet-rich plasma (PRP) or platelet-rich gel (PRG) significantly improves subcutaneous tissue defect coverage and reduces flap edema. By elevating growth factor concentrations (eg: VEGF), these treatments enhance vascular regeneration in recipient sites, thereby improving flap survival rates and delaying necrosis. Notably, PRG exhibits superior efficacy. This study provides novel therapeutic strategies for reducing postoperative necrosis and accelerating wound healing in clinical flap transplantation.

Abbreviations

Colony Differentiation Antigen 34: CD34; Vascular Endothelial Growth Factor: VEGF; Platelet-Derived Growth Factor: PDGF; Platelet-Rich Plasma Gel: PRG; Platelet-Rich Plasma: PRP; Nitric Oxide: NO; Vascular Endothelial Growth Factor Receptor 2: VEGFR2; Contrast-Enhanced Ultrasound: CEUS; Hypoxia-Inducible Factor-1 α : HIF-1 α .

Author Contributions

Linsen Fang: Conceptualization, Methodology, Resources, Supervision, Project administration, Funding acquisition. Chengyang Dong: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - Original Draft, Writing - Review & Editing, Visualization. Zean Lin: Investigation. Writing - Original Draft. All authors read and approved the final manuscript.

Acknowledgements

Not Applicable.

Funding Information

Clinical Medicine Peak Discipline Construction Project of Anhui Medical University (Grant No. 9301001815).

Ethics Approval and Consent to Participate

Not Applicable.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

References

- [1] Ershadifar, S., Colback, A., Basmaci, U. N., Wilson, M., Birkeland, A. C., and Silverman, D. A. Predictors of Donor-Site Wound Complications Following Fibula Free Flap Reconstruction. *OTO Open*. 2025;9(2):e70126. <https://doi.org/10.1002/oto2.70126>
- [2] Yang, J., Qin, X., Hou, L., and Liu, Y. Risk prediction models for complications after flap repair surgery: a systematic review and meta-analysis. *BMC Surgery*. 2025;25(1):398. <https://doi.org/10.1186/s12893-025-03072-8>
- [3] Marx, R. E. Platelet-rich plasma: evidence to support its use. *Journal of Oral and Maxillofacial Surgery*. 2004;62(4):489–496. <https://doi.org/10.1016/j.joms.2003.12.003>
- [4] Everts, P., Onishi, K., Jayaram, P., Lana, J. F., and Mautner, K. Platelet-Rich Plasma: New Performance Understandings and Therapeutic Considerations in 2020. *International Journal of Molecular Sciences*. 2020;21(20):7794. <https://doi.org/10.3390/ijms21207794>
- [5] Yang, D., and Morris, S. F. Comparison of two different delay procedures in a rat skin flap model. *Plastic and Reconstructive Surgery*. 1998;102(5):1591–1597. <https://doi.org/10.1097/00006534-199810000-00039>
- [6] Zhang, D., Jin, C., Han, T., Chen, J., Ali Raza, M., Li, B., et al. Sinomenine promotes flap survival by upregulating eNOS and eNOS-mediated autophagy via PI3K/AKT pathway. *International Immunopharmacology*. 2023;116:109752. <https://doi.org/10.1016/j.intimp.2023.109752>
- [7] Harder, Y., Amon, M., Erni, D., and Menger, M. D. Evolution of ischemic tissue injury in a random pattern flap: a new mouse model using intravital microscopy. *Journal of Surgical Research*. 2004;121(2):197–205. <https://doi.org/10.1016/j.jss.2004.03.026>
- [8] Milton, S. H. Pedicled skin-flaps: the fallacy of the length: width ratio. *British Journal of Surgery*. 1970;57(7):502–508. <https://doi.org/10.1002/bjs.1800570705>
- [9] Lee, J. H., You, H. J., Lee, T. Y., and Kang, H. J. Current Status of Experimental Animal Skin Flap Models: Ischemic Preconditioning and Molecular Factors. *International Journal of Molecular Sciences*. 2022;23(9):5234. <https://doi.org/10.3390/ijms23095234>
- [10] Landén, N. X., Li, D., and Ståhle, M. Transition from inflammation to proliferation: a critical step during wound healing. *Cellular and Molecular Life Sciences*. 2016;73(20):3861–3885. <https://doi.org/10.1007/s00018-016-2268-0>
- [11] Eming, S. A., Martin, P., and Tomic-Canic, M. Wound repair and regeneration: mechanisms, signaling, and translation. *Science Translational Medicine*. 2014;6(265):265sr6. <https://doi.org/10.1126/scitranslmed.3009337>
- [12] Collins, T., Alexander, D., and Barkatali, B. Platelet-rich plasma: a narrative review. *EFORT Open Reviews*. 2021;6(4):225–235. <https://doi.org/10.1302/2058-5241.6.200017>
- [13] Boswell, S. G., Cole, B. J., Sundman, E. A., Karas, V., and Fortier, L. A. Platelet-rich plasma: a milieu of bioactive factors. *Arthroscopy*. 2012;28(3):429–439. <https://doi.org/10.1016/j.arthro.2011.10.018>
- [14] Akbarzadeh, S., McKenzie, M. B., Rahman, M. M., and Cleland, H. Allogeneic Platelet-Rich Plasma: Is It Safe and Effective for Wound Repair? *European Surgical Research*. 2021;62(1):1–9. <https://doi.org/10.1159/000514223>
- [15] Ferrara, N., and Kerbel, R. S. Angiogenesis as a therapeutic target. *Nature*. 2005;438(7070):967–974. <https://doi.org/10.1038/nature04483>
- [16] Carmeliet, P., and Jain, R. K. Molecular mechanisms and clinical applications of angiogenesis. *Nature*. 2011;473(7347):298–307. <https://doi.org/10.1038/nature10144>
- [17] Vourtsis, S. A., Spyriounis, P. K., Agogiannis, G. D., Ionac, M., and Papalois, A. E. VEGF application on rat skin flap survival. *Journal of Investigative Surgery*. 2012;25(1):14–19. <https://doi.org/10.3109/08941939.2011.593693>
- [18] Kryger, Z., Zhang, F., Dogan, T., Cheng, C., Lineaweaver, W. C., and Buncke, H. J. The effects of VEGF on survival of a random flap in the rat: examination of various routes of administration. *British Journal of Plastic Surgery*. 2000;53(3):234–239. <https://doi.org/10.1054/bjps.1999.3315>
- [19] Shang, Y., Xie, X., Luo, Y., Nie, F., Luo, Y., Jing, X., et al. Safety findings after intravenous administration of sulfur hexafluoride microbubbles to 463,434 examinations at 24 centers. *European Radiology*. 2023;33(2):988–995. <https://doi.org/10.1007/s00330-022-09108-4>
- [20] Newsome, I. G., and Dayton, P. A. Acoustic Angiography: Superharmonic Contrast-Enhanced Ultrasound Imaging for Noninvasive Visualization of Microvasculature. *Methods in Molecular Biology*. 2022;2393:641–655. https://doi.org/10.1007/978-1-0716-1803-5_34
- [21] Lv, K., Zhai, H., Jiang, Y., Liang, P., Xu, H. X., Du, L., et al. Prospective assessment of diagnostic efficacy and safety of Sonazoid™ and SonoVue® ultrasound contrast agents in patients with focal liver lesions. *Abdominal Radiology*. 2021;46(10):4647–4659. <https://doi.org/10.1007/s00261-021-03010-1>

Association of Peritoneal Dialysis Effluent Interleukin-17 with Peritoneal Functional Decline and Clinical Significance: A Prospective Cohort Study

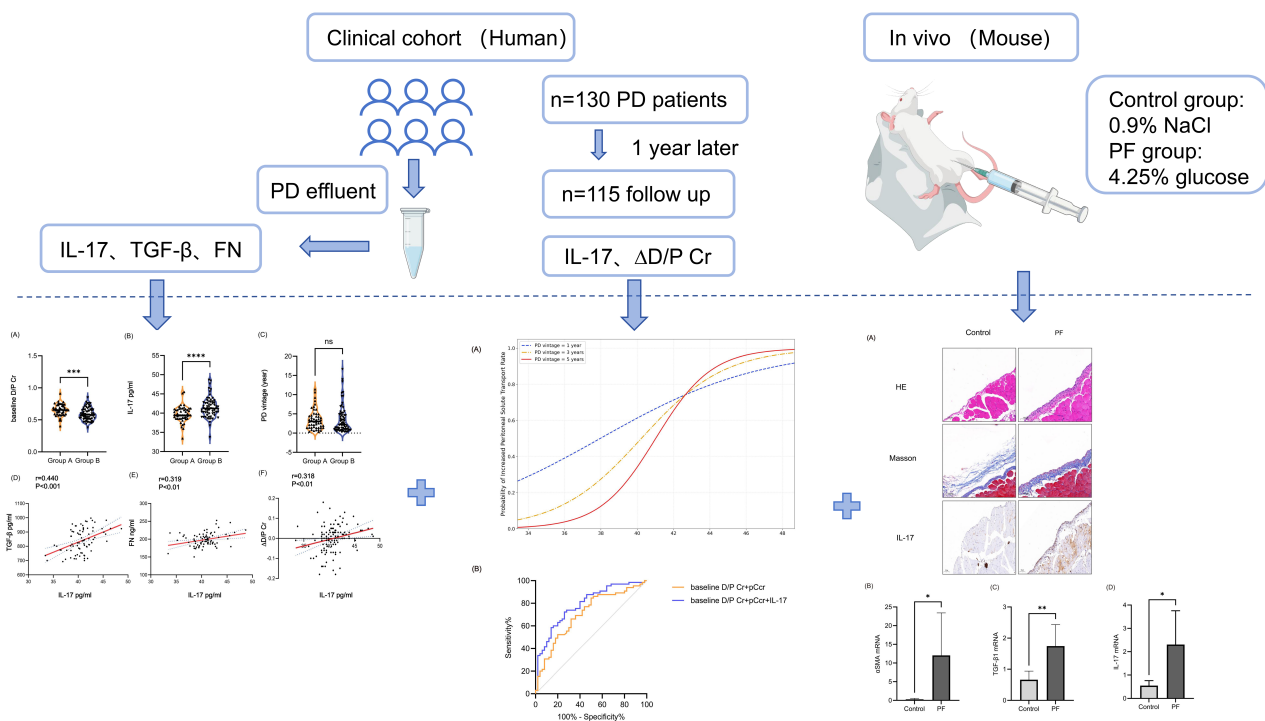
Authors

Lejia Song, Lu Li, Guang Chen, Qiufeng Wang, Qingqing Rao, Huaina Dou, Li Zhang, Pei Zhang

Correspondence

zhangpei@ahmu.edu.cn (P. Zhang)

Graphical Abstract



<https://doi.org/10.71321/g94y8w65>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Association of Peritoneal Dialysis Effluent Interleukin-17 with Peritoneal Functional Decline and Clinical Significance: A Prospective Cohort Study

Lejia Song^{1,2†}, Lu Li^{1,2†}, Guang Chen^{1,2}, Qiufeng Wang¹, Qingqing Rao^{1,2}, Huaina Dou², Li Zhang², Pei Zhang^{1,2*}

Received: 2026-02-19 | Accepted: 2026-04-15 | Published online: 2026-05-26

Abstract

Background: Long-term peritoneal dialysis (PD) frequently induces chronic peritoneal inflammation, fibrosis, and increased peritoneal solute transport rate (PSTR). This study assessed whether effluent interleukin-17 (IL-17) predicts increased PSTR and relates to peritoneal fibrosis.

Methods: This prospective cohort enrolled 130 PD patients, of whom 115 completed 1-year follow-up. Effluent IL-17, transforming growth factor- β (TGF- β), and fibronectin (FN) were measured by enzyme-linked immunosorbent assay (ELISA). Peritoneal function was assessed by dialysate-to-plasma creatinine ratio (D/P Cr) at baseline and 1 year. We used correlation analysis, multivariable regression, interaction analysis, and Receiver operating characteristic (ROC) analysis to evaluate the predictive value of effluent IL-17 for increased PSTR. A mouse peritoneal fibrosis model was used to evaluate local IL-17 expression by immunohistochemistry and real-time quantitative PCR (RT-qPCR).

Results: Effluent IL-17 was higher in patients with increased PSTR and correlated positively with TGF- β , FN, and Δ D/P Cr. Multivariable analysis showed that IL-17 independently predicted 1-year D/P Cr after adjustment for baseline D/P Cr and clinical covariates. PD vintage significantly modified the association between IL-17 and increased PSTR. Adding IL-17 to baseline D/P Cr and peritoneal creatinine clearance improved prediction of increased PSTR, with the AUC increasing from 0.705 to 0.789. Furthermore, immunohistochemical and qPCR results showed that IL-17 expression was elevated in fibrotic peritoneal tissues of PF mice compared with control mice.

Conclusion: Effluent IL-17 is associated with peritoneal fibrosis and longitudinal PSTR increase, supporting its potential as a noninvasive biomarker for early risk stratification in PD patients.

Keywords: Peritoneal dialysis; Peritoneal fibrosis; Peritoneal function; Biomarker; IL-17

Introduction

Peritoneal dialysis (PD) is a vital treatment option for end-stage renal disease (ESRD) worldwide, as evidenced by its simplicity, cost-effectiveness, and ability to preserve residual renal function [1]. The efficacy of this therapeutic modality relies heavily upon maintaining the architectural and physiological integrity of the peritoneal membrane. However, chronic inflammation and peritoneal injury are induced by chronic exposure to bioincompatible dialysate (characterized by hypertonicity, high glucose, and acidity). These pathological changes include mesothelial cell denudation, neovascularization, and excessive extracellular matrix deposition, eventually progressing to sub-mesothelial fibrosis [2]. Such structural remodelling compromises vascular permeability and disrupts transport kinetics, driving ultrafiltration failure (UFF) and, ultimately, technique failure [3]. The peritoneal equilibration test (PET) remains the clinical standard for evaluating peritoneal function through the

quantification of solute transport and ultrafiltration capacity. This assessment relies on measuring the dialysate-to-plasma creatinine ratio (D/P Cr), the 4-hour to 0-hour dialysate glucose ratio (D/D0 glucose), and the 4-hour ultrafiltration volume (4 h UF) [4]. A faster peritoneal solute transport rate (PSTR) is predictive of adverse clinical outcomes. Specifically, an increase of 0.1 units in D/P Cr is associated with a significantly elevated risk of technique failure, hospitalization, and all-cause mortality [5-6]. However, while PET effectively reflects current transport characteristics, it lacks the sensitivity required for the early detection of peritoneal function decline.

While specific effluent biomarkers have been identified, their clinical applications remain limited. Interleukin-6 (IL-6) is correlated with solute transport, inflammation, and fibrosis. However, its diagnostic specificity is compromised by interference from systemic inflammation and interindividual variability [7]. Similarly, cancer antigen 125 (CA125) has been recognized as a biomarker for mesothelial cell mass. Nevertheless, CA125

1 Department of Nephropathy, The First Affiliated Hospital of Anhui Medical University, Hefei, China.

2 Anhui Medical University, Hefei, China.

† These authors have contributed equally to this work.

* Corresponding Author.

shows limited prognostic value for adverse outcomes, such as peritonitis or ultrafiltration failure, when used as a single marker [8]. Therefore, the identification of highly specific biomarkers to accurately characterize peritoneal pathological mechanisms remains a critical research priority.

Interleukin-17A (IL-17) is the signature cytokine of the T helper 17 (Th17) lineage. It plays a critical role in regulating inflammation and angiogenesis in autoimmune pathologies, such as multiple sclerosis and rheumatoid arthritis [9]. Mechanistically, IL-17 orchestrates a continuous inflammatory response by enhancing Th17 cell differentiation through the IL-6/TGF- β /ROR γ t (Interleukin-6/ transforming growth factor β / retinoic acid-related orphan receptor gamma t) axis and recruiting immune cells via CCL20-dependent chemotaxis [10–11]. While IL-17 levels are negligible in the peritoneum of healthy individuals, this cytokine is significantly upregulated in the peritoneum of patients receiving PD. These elevated levels originate primarily from infiltrating Th17 and $\gamma\delta$ T cells [12]. Moreover, accumulating evidence suggests that IL-17 promotes fibrosis across various organ systems. For example, it promotes myocardial fibrosis by activating cardiac fibroblasts, induces pulmonary fibrosis through mitochondrial dysfunction-mediated epithelial apoptosis, and exacerbates liver fibrosis by enhancing TGF- β receptor signalling [13–15]. In hypertensive nephropathy and IgA nephropathy, studies have shown that IL-17 activates downstream signalling cascades, notably NF- κ B, to induce the release of inflammatory mediators and aberrant extracellular matrix deposition, consequently driving the progression of renal fibrosis [16–17]. However, research concerning the expression of IL-17 in peritoneal tissue and its correlation with impaired peritoneal function remains limited.

Therefore, in this study, the expression of IL-17 in PD effluent was prospectively evaluated, and its ability to predict longitudinal peritoneal function decline was evaluated, with the goal of identifying a novel noninvasive biomarker for early risk stratification.

Methods

Study Design and Population

This prospective cohort study included 130 patients who received regular PD follow-up at the Department of Nephrology, the First Affiliated Hospital of Anhui Medical University, between September 2023 and September 2025. The follow-up duration was 1 year. This research was approved by the Ethics Committee of the First Affiliated Hospital of Anhui Medical University (approval number: PJ 2025-08-97).

Eligible patients were aged between 18 and 80 years, on stable peritoneal dialysis replacement therapy (PD vintage $\geq$ 3 months) using only commercially available peritoneal dialysis solutions with 1.5% or 2.5% glucose, and had a stable PD regimen (no adjustments for at least 1 month). Exclusion criteria included peritonitis within the past month, autoimmune diseases (e.g., lupus nephritis, rheumatoid arthritis, Sjögren's syndrome, and chronic lymphocytic thyroiditis), infections (e.g., hepatitis B virus, hepatitis C virus, and syphilis), malignant or benign tumors, and use of antibiotics or hormones at the time of sample collection.

All animal procedures were approved by the Animal Ethics Committee of Anhui Medical University (Approval Number:

LLSC20241406).

Male C57BL/6 mice (20–24 g, $n = 12$) were randomly allocated into two groups (control group $n = 6$, PF group $n = 6$). To establish a peritoneal fibrosis (PF) model, mice in the experimental group received daily intraperitoneal injections of 2 mL of 4.25% glucose-based dialysis solution for 28 consecutive days. The control group received an equal volume of isotonic saline (0.9% NaCl) following the same schedule.

Data Collection and Variables

General clinical data included the following: (1) demographic characteristics, such as age, sex, height, weight, and body mass index (BMI); (2) PD vintage, defined as the time from PD initiation to study enrolment (years); (3) comorbidities such as diabetes mellitus, hypertension, and cardiovascular disease; and (4) laboratory examinations, including routine biochemical parameters (renal function, blood urea nitrogen (BUN), serum creatinine (Scr), and parathyroid hormone (PTH)), electrolytes (serum calcium and serum phosphate (P)), nutritional indicators (serum albumin [Alb] and haemoglobin [Hb]), lipid metabolism parameters (total cholesterol [TC] and triglyceride [TG]), inflammatory markers (neutrophil count [NE], lymphocyte count [LYM], monocyte count [MONO], and C-reactive protein [CRP]), and corrected serum calcium concentration calculated as: corrected serum calcium (mmol/L) = serum calcium + 0.02 $\times$ (40 – Alb [g/L]).

Cardiac indices included left ventricular ejection fraction (EF), left atrial diameter (LA), left ventricular diameter (LVD), and left ventricular posterior wall thickness (LVPWD).

Peritoneal Function Measurement

On the night before the PET, patients were required to retain peritoneal dialysis fluid for 8–12 hours. On the test day morning, after emptying the bladder and bowels, patients assumed a standing position, and the existing PD fluid was completely drained from the peritoneal cavity within 20 minutes. Subsequently, 2 L of 2.5% glucose PD solution was rapidly infused into the peritoneal cavity within 10 minutes, with patients turning from side to side after each 400 mL administration. Upon completion of infusion, the dialysis tubing was clamped, and the PD fluid was retained in the peritoneal cavity for 4 hours. First, original PD fluid was collected immediately (0 minutes) after complete infusion to obtain a 0-hour PD fluid sample. At exactly 2 hours of dwell time, venous blood and PD effluent samples were collected. After maintaining the dialysis position for precisely 4 hours, patients switched to a seated or standing posture, and the PD fluid in the peritoneal cavity was completely drained into a collection bag within 20 minutes. Following drainage, the total volume of effluent was measured and recorded. Subsequently, additional PD effluent samples were collected from the drainage bag [4].

The following parameters were calculated:

Dialysate-to-plasma creatinine ratio (D/P Cr) = dialysate creatinine (4 h) / plasma creatinine (2 h)

Dialysate glucose ratio (D/D0 glucose) = dialysate glucose (4 h) / dialysate glucose (0 h)

At 12 months (± 1 month) after baseline, peritoneal function parameters D/P Cr and D/D0 glucose were evaluated again.

Definitions:

Δ D/P Cr = D/P Cr 1 year later – baseline D/P Cr

Δ D/D0 glucose = D/D0 glucose 1 year later – baseline D/D0

standard deviations (means $\pm$ SDs), whereas parameters violating normality assumptions are presented as median values accompanied by interquartile ranges (IQRs). When variable relationships between indicators followed a normal distribution and satisfied the assumptions of homogeneity of variance and independence, Pearson correlation analysis was applied; otherwise, Spearman correlation analysis was used. We implemented univariate and multivariate linear regression models to assess the predictive variables affecting Δ D/P Cr. In addition, we used binary logistic regression to analyse the interaction effect between IL-17 concentration and PD vintage on the binary outcome of increased PSTR (Δ D/P Cr > 0). We utilized Receiver operating characteristic (ROC) analysis to evaluate the prognostic capability of IL-17 for Δ D/P Cr > 0 . Two-tailed tests were performed with $\alpha = 0.05$, and the results were significant at the 0.05 level. All statistical analyses were performed with SPSS, GraphPad Prism, and R software.

Results

Baseline Characteristics

The study population comprised 130 patients receiving PD. All participants were included in the baseline characteristic analysis. As detailed in Table 1, 42.31% of the initial cohort were male, with a median age of 54 years (IQR 44.25–60.00). The mean BMI was 23.10 ± 3.14 kg/m². At baseline, the median peritoneal effluent IL-17 concentration was 40.49 pg/mL (IQR 39.00–41.79 pg/mL). For peritoneal transport parameters, the mean baseline D/P Cr was 0.62 ± 0.10 , and the median baseline D/D0 glucose was 0.42 (IQR 0.36–0.47). After 1 year of follow-up, 15 patients were censored (due to death, transplantation, or loss to follow-up), leaving 115 patients (88.5%) for the final outcome analysis. The baseline characteristics of the analyzed cohort (N = 115) are also presented in Table 1 (second column) and were comparable to the initial cohort (all $P > 0.05$), indicating no significant selection bias. Detailed comparisons are provided in Supplementary Table 1. Participants were dichotomized based on PSTR into a non-increased PSTR group (Group A: Δ D/P Cr ≤ 0 , n = 50) and an increased PSTR group (Group B: Δ D/P Cr > 0 , n = 65). Statistical analysis revealed that baseline D/P Cr (Figure 2A, $P = 0.001$), D/D0 glucose ($P = 0.002$), peritoneal creatinine clearance (pCcr) ($P = 0.013$), and IL-17 concentration (Figure 2B, $P < 0.0001$) differed significantly between the two groups, whereas PD vintage (Figure 2C) and other parameters showed no significant differences.

Correlation between Effluent IL-17 and Changes in PSTR

To quantitatively assess the correlation between cytokines and profibrotic factors, the associations between effluent IL-17 and both TGF- β and FN were calculated. Correlation analysis revealed a marked positive correlation between IL-17 and both TGF- β ($r = 0.440$, $P < 0.001$) and FN ($r = 0.319$, $P < 0.01$). The scatter plot distributions illustrating these linear associations are depicted in Figure 2D and 2E. Furthermore, the positive correlation between IL-17 and Δ D/P Cr ($r = 0.318$, $P < 0.01$) suggests the clinical application value of IL-17 (Figure 2F).

Predictive Value of IL-17 for Peritoneal Transport Status

To identify the vital predictors of PSTR, we constructed univar-

iate and multivariate regression models (Table 2). Univariate analysis revealed that only baseline D/P Cr was independent predictor of D/P Cr one year later. However, in the multivariate regression analysis incorporating PD vintage, age, CRP, baseline D/P Cr, pCcr and IL-17 as covariates, baseline D/P Cr ($\beta = 0.791$, $P < 0.001$) and IL-17 ($\beta = 0.006$, $P = 0.022$) were identified as independent predictors of D/P Cr one year later. In the multivariate model, after adjustment for other variables, each 1 pg/mL increase in effluent IL-17 concentration predicted an average increase of 0.006 units in 1-year D/P Cr. Similarly, each 1-unit increase in baseline D/P Cr was associated with an average increase of 0.791 units in 1-year D/P Cr.

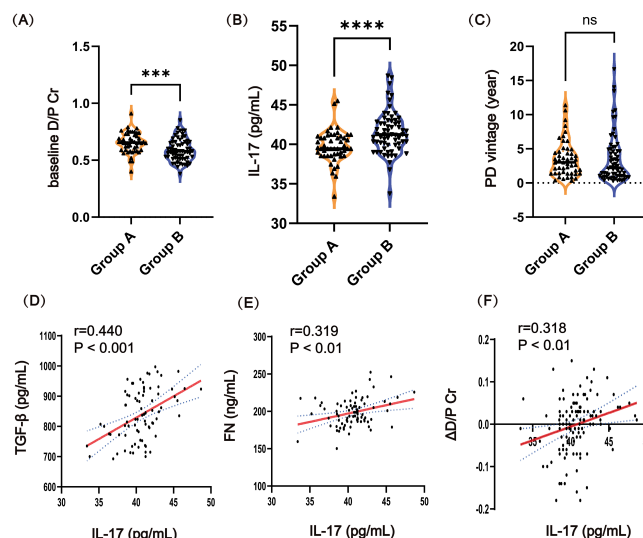


Figure 2. Comparisons of baseline indicators and correlation analysis of IL-17.

(A) Comparison of baseline dialysate-to-plasma creatinine ratio (D/P Cr) between non-increased PSTR group (Group A) and increased PSTR group (Group B). (B) Comparison of peritoneal effluent IL-17 levels between two groups. (C) Comparison of PD vintage between two groups. (D) Correlation between effluent IL-17 and TGF- β . (E) Correlation between effluent IL-17 and fibronectin (FN). (F) Correlation between effluent IL-17 and Δ D/P Cr.

Note: *** $P < 0.001$, **** $P < 0.0001$

To further investigate the predictive value of IL-17 for changes in peritoneal function, we evaluated its association with 1-year D/D0 glucose (Table 3). Univariate analysis identified baseline D/P Cr, baseline D/D0 glucose, and pCcr as significant correlates of 1-year D/D0 glucose. However, after adjustment for baseline peritoneal function parameters, only baseline D/D0 glucose ($\beta = 0.474$, $P < 0.001$) and IL-17 ($\beta = -0.005$, $P = 0.040$) remained independent predictors of 1-year D/D0 glucose.

We constructed a binary logistic regression model to assess the joint effect of IL-17 levels and PD vintage on an increased PSTR. To explore this interaction, an interaction term (IL-17 $\times$ PD vintage) was included in the model. Notably, PD vintage significantly modified the association between IL-17 and peritoneal transport status (P for interaction = 0.028). As shown in Figure 3A, the predictive effect of IL-17 on faster PSTR was weaker in the short-term dialysis group (1 year). However, this association became significantly stronger in patients with longer dialysis durations (3 and 5 years). The increasingly steep

Table 1. Baseline Characteristics of the Overall Study Population and Patients Who Completed Follow-up.

Index	Total (n=130)	Follow-up (n=115)		t/Z/X ²	P value
		Group A (n=50)	Group B (n=65)		
Gender				0.416	0.519
Male (%)	55 (42.31%)	23 (46.00%)	26 (40.00%)		
Female (%)	75 (57.69%)	27 (54.00%)	39 (60.00%)		
Hypertension				1.410	0.235
No	29 (22.31%)	13 (26.00%)	11 (16.92%)		
Yes	101 (77.69%)	37 (74.00%)	54 (83.08%)		
Diabetes				1.441	0.230
No	114 (87.69%)	46 (92.00%)	55 (84.62%)		
Yes	16 (12.31%)	4 (8.00%)	10 (15.38%)		
age (year)	54.00 (44.25, 60.00)	52 (40.5,59)	54 (49,60)	-1.115	0.265
BMI (kg/m ²)	23.10 ± 3.14	23.17 (20.88,25.03)	22.86 (20.44,25.75)	-0.166	0.868
PD vintage (year)	2.60 (1.21, 4.71)	3.01 (1.4,4.66)	2.23 (0.97,4.94)	-0.852	0.394
baseline D/P Cr	0.62 ± 0.10	0.65±0.09	0.59±0.1	3.579	<0.001***
baseline D/D0 Glucose	0.42 (0.36, 0.47)	0.39±0.08	0.44±0.08	-3.184	0.002**
Urea kinetics					
Peritoneal Kt/V	1.54 ± 0.47	1.6±0.51	1.52±0.43	0.846	0.399
Renal Kt/V	0.14 (0, 0.6)	0.08 (0,0.46)	0.21 (0,0.66)	-0.953	0.341
Total Kt/V	1.87 ± 0.52	1.88±0.59	1.9±0.48	-0.269	0.789
Creatinine clearance (L/wk/1.73 m ²)					
pCcr	38.79±8.95	41.41±9.04	37.11±9	2.534	0.013*
rCcr	4.38 (0,26.12)	2.18 (0,14.23)	9.1 (0,28.69)	-1.268	0.205
Total Ccr	50.13 (42.88,61.76)	50.52 (43.09,57.51)	50.24 (41.77,62.89)	-0.195	0.846
SCr (umol/L)	959.66±271.67	990.06±274.29	949.87±279.93	0.770	0.443
BUN (mmol/L)	20.16±5.54	20.3±5.43	20.02±5.51	0.275	0.784
PTH (pg/mL)	277 (120.75, 397)	289.5 (182.25,412.25)	261 (125.5,379.5)	-1.162	0.245
Hb (g/L)	109.71 ± 20.31	106.36±17.93	109.88±20.78	-0.954	0.342
ALB (g/L)	37.85 ± 3.87	37.48±4.12	38.36±3.49	-1.237	0.219
TG (mmol/L)	1.53 (1.07, 2.48)	1.42 (1.03,2.42)	1.9 (1.19,2.9)	-1.704	0.088
TC (mmol/L)	4.76 ± 1.32	4.61±1.08	4.83±1.43	-0.902	0.369
Glucose (mmol/L)	4.87 (4.3, 5.72)	4.84 (4.2,5.49)	4.97 (4.42,6.09)	-0.886	0.376
Corrected calcium (mmol/L)	2.35 (2.21, 2.45)	2.36 (2.21,2.47)	2.34 (2.18,2.44)	-0.753	0.451
P (mmol/L)	1.80 ± 0.47	1.78±0.47	1.85±0.47	-0.889	0.376
CaxP product	4.10 (3.28, 4.84)	3.97 (3.43,4.53)	4.31 (3.22,5.08)	-0.739	0.460
NE (x10 ⁹ /L)	4.19 (3.34, 5.4)	4.05 (3.47,5.17)	4.27 (3.3,5.54)	-0.553	0.580
LYM (x10 ⁹ /L)	1.32 (1.1, 1.65)	1.26 (1.03,1.6)	1.41 (1.14,1.73)	-1.786	0.074
MONO (x10 ⁹ /L)	0.41 (0.31, 0.54)	0.43 (0.29,0.55)	0.41 (0.32,0.55)	-0.186	0.852
PLT (x10 ⁹ /L)	197.62 ± 66.19	195 (144.25,228.5)	190 (167.5,249)	-0.920	0.358
CRP (mg/L)	1.63 (0.6, 3.59)	1.24 (0.39,2.63)	1.79 (0.73,3.86)	-1.334	0.182
EF (%)	62 (60, 65.25)	62 (60,66)	63 (60,66)	-0.368	0.713
LA (cm)	3.70 ± 0.49	3.76 (3.32,4.22)	3.55 (3.35,4.05)	-0.866	0.386
LVD (cm)	4.77 (4.33, 5.06)	4.72 (4.38,5.06)	4.8 (4.33,5.07)	-0.316	0.752
LVPWD (cm)	0.95 (0.89, 1.04)	0.98 (0.9,1.07)	0.94 (0.88,1.03)	-1.672	0.094
D/P Cr 1 year later	-	0.59 (0.54,0.65)	0.61 (0.57,0.71)	-2.047	0.041*
D/D0 Glucose 1 year later	-	0.44 (0.41,0.50)	0.44 (0.36,0.47)	-1.534	0.125
IL-17 (pg/mL)	40.49 (39, 41.79)	39.45 (38.55,41.04)	41.25 (39.9,42.9)	-3.858	<0.0001****

Note: Group A: non-increased PSTR group; Group B: increased PSTR group. Normally distributed continuous variables are expressed as mean ± standard deviation (mean ± SD); non-normally distributed continuous variables are expressed as median (interquartile range) [M (P25, P75)]; categorical variables are expressed as number (%). *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

slopes indicate that the deleterious effect of elevated IL-17 on faster PSTR is exacerbated by prolonged PD exposure.

Incremental Predictive Value of IL-17 over Standard Clinical Metrics

While the standard clinical model (including baseline D/P Cr and pCcr) showed acceptable accuracy in predicting faster PSTR ($\Delta D/P\ Cr > 0$) (the area under the curve [AUC] = 0.705, 95% CI: [0.609–0.801]), the addition of effluent IL-17 significantly enhanced the accuracy. Specifically, the comprehensive model (IL-17 + baseline D/P Cr + pCcr) achieved the optimal predictive accuracy, with an AUC of 0.789 (95% CI: 0.706–0.872), a sensitivity of 72.3%, and a specificity of 74.0%. DeLong test confirmed that this comprehensive model was statistically superior to the base model ($P < 0.05$; Figure 3B and Table 4). These findings highlight the incremental predictive value of IL-17, indicating that it captures changes in peritoneal function that cannot be reflected by baseline transport status.

Validation of IL-17 Expression in A Mouse Peritoneal Fibrosis Model

To investigate the specific role of IL-17 in the pathogenesis of peritoneal fibrosis *in vivo*, we established a murine PF model by daily intraperitoneal injection of 4.25% glucose-based dialysis fluid. As shown in Figure 4A, haematoxylin and eosin (HE) staining revealed notable thickening of the peritoneum and increased inflammatory cell infiltration in the fibrosis group compared with the control group. Masson staining showed that collagen deposition was significantly increased in fibrotic peritoneal tissue. Moreover, immunohistochemistry (IHC) demonstrated marked infiltration of IL-17-positive cells within the fibrotic tissues. These results verified the successful induction of the PF model and indicated that IL-17 expression was upregulated in the fibrotic peritoneum.

qPCR Assessment of Fibrotic Markers and IL-17 mRNA Expression in Murine Peritoneal Tissue

To further quantify the extent of peritoneal injury and inflammatory status at the molecular level, real-time quantitative PCR (qPCR) was employed to determine the mRNA expression

Table 2. Univariable and Multivariable Linear Regression of Influencing Factors on D/P Cr.

Variables	Univariate Analysis		Multivariate Analysis	
	β (95%CI)	P-value	β (95%CI)	P-value
PD vintage (years)	0.003 (-0.003, 0.009)	0.284	0.001 (-0.004, 0.005)	0.820
Age(years)	0.001 (-0.0002, 0.003)	0.084	0.001 (0.000, 0.002)	0.205
CRP(mg/L)	0.002 (-0.004, 0.007)	0.499	0.0003 (-0.003, 0.004)	0.853
ALB (g/L)	-0.008 (-0.012, -0.003)	0.001**	0.0002 (-0.003, 0.004)	0.927
Baseline D/P Cr	0.742 (0.623, 0.861)	<0.001***	0.791 (0.648, 0.934)	< 0.001***
pCcr	0.003 (0.001, 0.005)	0.002**	-0.001 (-0.003, 0.001)	0.248
IL-17(pg/mL)	0.003 (-0.004, 0.010)	0.411	0.006 (0.001, 0.011)	0.022*

Note: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table 3. Univariable and Multivariable Linear Regression of Influencing Factors on D/D 0 glucose.

Variables	Univariate Analysis		Multivariate Analysis	
	β (95%CI)	P-value	β (95%CI)	P-value
PD vintage (years)	-0.004 (-0.008, 0.001)	0.129	-0.002 (-0.006, 0.002)	0.307
ALB (g/L)	0.005 (0.001, 0.008)	0.017*	0.0001 (-0.003, 0.003)	0.946
Baseline D/P Cr	-0.473 (-0.594, -0.352)	<0.001***	-0.166 (-0.370, 0.039)	0.111
Baseline D/D0 Glucose	0.626 (0.484, 0.768)	<0.001***	0.474 (0.238, 0.711)	<0.001***
peritoneal Kt/V	0.020 (-0.011, 0.052)	0.207	0.031 (0.005, 0.057)	0.022*
pCcr	-0.002 (-0.004, -0.001)	0.002**	0.000 (-0.002, 0.002)	0.922
IL-17(pg/mL)	-0.003 (-0.009, 0.003)	0.282	-0.005 (-0.010, -0.0002)	0.040*

Note: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

Table 4. Comparison of ROC Curve Parameters for Prediction Models of faster PSTR.

Groups	AUC	95%CI	Optimal Cutoff	Sensitivity (%)	Specificity (%)
Model 1 IL-17	0.710	0.617-0.804	39.86	76.9	58.0
Model 2 baseline D/P Cr	0.694	0.597-0.790	0.605	63.1	72.0
Model 3 pCcr	0.633	0.529-0.737	37.88	53.8	72.0
Model 4 baseline D/P Cr + pCcr	0.705	0.609-0.801	-	84.6	50.0
Model 5 baseline D/P Cr + pCcr+ IL-17	0.789	0.706-0.872	-	72.3	74.0

levels of fibrotic markers (α -SMA and TGF- β) and IL-17 in the experimental groups. As shown in Figure 4B-D, mRNA expression of α -SMA in peritoneal tissues was significantly higher in the PF group than in the control group ($P < 0.05$). Similarly, the mRNA levels of TGF- β in the PF group were significantly higher than those in the control group ($P < 0.01$). Furthermore, IL-17 mRNA expression in the PF group showed a marked up-regulation compared with the control group ($P < 0.05$). These transcriptional findings were consistent with the previously observed trend in IL-17 protein expression as determined by immunohistochemical staining.

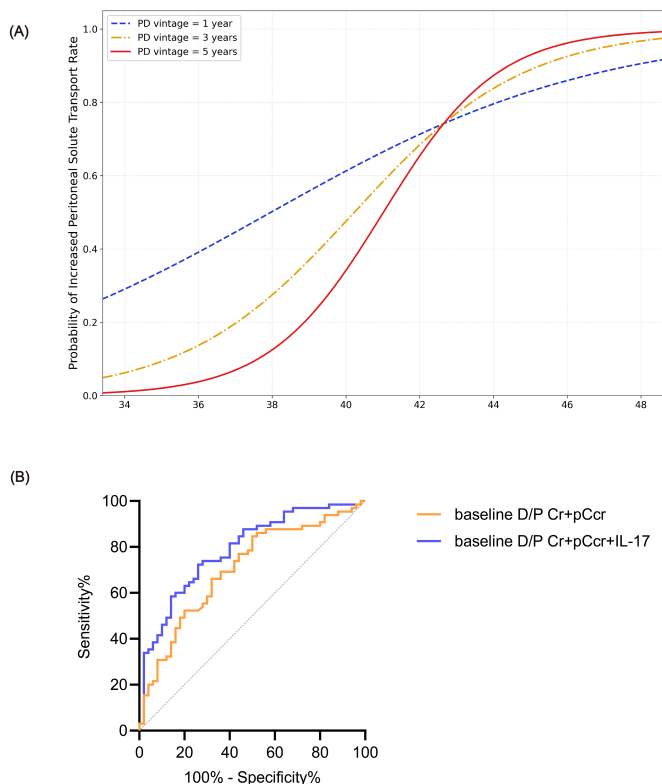


Figure 3. Interaction effect and receiver operating characteristic (ROC) curve analysis.

(A) Predicted probability plot showing the interaction between IL-17 concentration and PD vintage on faster PSTR. (B) ROC curves of baseline model (D/P Cr + pCcr) and combined model (D/P Cr + pCcr + IL-17) for predicting faster PSTR.

Discussion

This study represents the first prospective cohort analysis to establish that effluent IL-17 serves as a reliable predictor of faster PSTR. There is a statistically significant correlation between elevated PSTR and increased mortality rate, as well as a trend toward a higher technical failure rate [18-19]. However, traditional clinical criteria still lack the sensitivity required for early monitoring of PSTR increase. Our study bridges this gap and reveals that compared with the standard clinical model, the model incorporating IL-17 has significant predictive value. Additionally, the histological evidence from the mouse peritoneal fibrosis model further supports the feasibility of using IL-17 as an early biomarker for peritoneal fibrosis. Univariate regression analysis demonstrated that baseline D/

P Cr was the most robust predictor of peritoneal transport status at the 1-year follow-up ($\beta = 0.742$, $P < 0.001$). This finding underscores the inherent stability of peritoneal membrane characteristics, where initial solute transport rates—likely reflecting baseline microvascular density and anatomical integrity—largely dictate subsequent functional trajectories. Consequently, adjusting for initial functional status is imperative when isolating the contributions of auxiliary pathogenic factors. Notably, after adjusting for baseline D/P Cr and systemic inflammatory markers (CRP and albumin), local peritoneal IL-17 levels remained independently associated with the progression of D/P Cr ($\beta = 0.006$, $P = 0.022$). These data suggest that IL-17-mediated local immune-inflammatory responses may serve as an independent driver of peritoneal transport acceleration. Even in the absence of systemic inflammation, elevated intraperitoneal IL-17 may herald a shift toward a high-transporter phenotype within the following year.

In an experimental fibrosis model induced by glucose-rich dialysis fluid, significant peritoneal thickening, abnormal collagen deposition, and increased IL-17 expression were observed. These morphological changes were consistent with the findings of González-Mateo et al. and Ferrantelli et al., who explained that IL-17 blockade attenuated peritoneal fibrosis [20-21], further confirming the role of IL-17 in peritoneal structural deterioration. This profibrotic potential was further supported by in vivo evidence. These pathological alterations impair peritoneal function, especially by affecting solute transport and

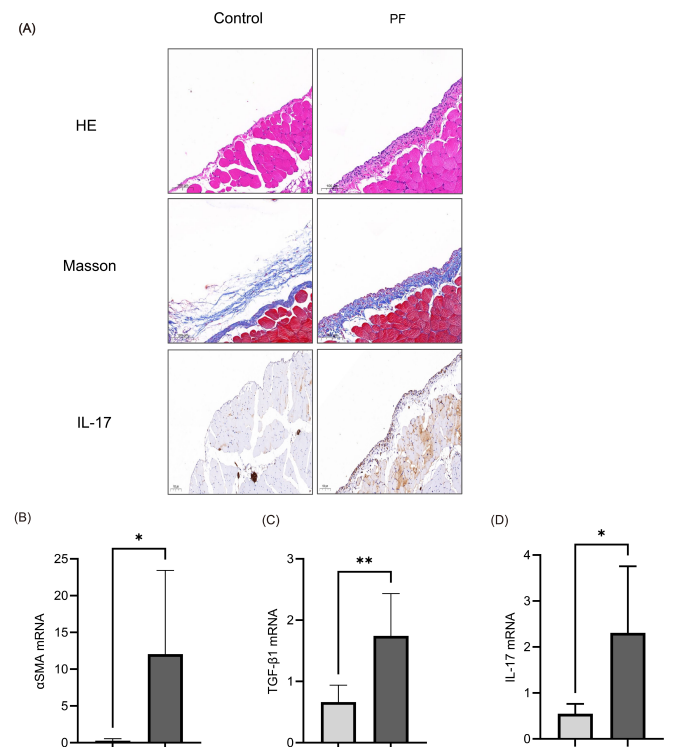


Figure 4. Histological staining and mRNA expression in mouse peritoneal tissues.

(A) Representative images of HE staining, Masson staining, and IL-17 immunohistochemical staining in peritoneal tissues from control and PF mice (scale bar = 100 μ m, 100 $\times$ magnification). Relative mRNA expression of (B) α -SMA mRNA, (C) TGF- β 1 mRNA, and (D) IL-17 mRNA in peritoneal tissues. * $P < 0.05$, ** $P < 0.01$.

ultrafiltration, ultimately resulting in dialysis technique failure [22]. Previous studies have indicated that bioincompatible dialysis fluids induce Th17 polarization and IL-17 secretion [23]. This process activates mesothelial cells, prompting the release of proinflammatory mediators that modulate the local inflammatory microenvironment. This cascade initiates and amplifies peritoneal damage, laying the pathological foundation for the progression of fibrosis. Local peritoneal microinflammation, which intensifies with increasing PD vintage, serves as a pivotal factor driving structural remodelling, including fibrosis, vascular lesions, and abnormal lymphatic vessel formation [24].

Our data also revealed robust positive correlations between effluent IL-17 and both TGF- β ($r = 0.440$) and FN ($r = 0.319$). TGF- β is a well-recognized promoter of mesothelial-to-mesenchymal transition (MMT), and FN drives extracellular matrix (ECM) remodelling [25-26]. Mechanistically, we hypothesize that these associations stem from a pathogenic positive feedback loop in which IL-17 upregulates TGF- β expression [26]. In accordance with previous findings, IL-17 activates the ERK and TGF- β /Smad pathways to drive ECM production [16, 27]. Simultaneously, TGF- β enhances the inflammatory response by inducing Th17 polarization. This interaction accelerates the transition from chronic inflammation to irreversible fibrosis. This irreversible pathological damage results in a significant reduction in peritoneal capacity, ultimately compromises the clinical success of dialysis. Such pathological change in peritoneal structure can only be diagnosed via peritoneal biopsy. However, its clinical application is limited due to the invasive nature of the procedure. Clinically, the results of PET are used to assess the function of the peritoneum.

Moreover, our research demonstrates that the comprehensive model incorporating IL-17 has greater predictive value than the basic model that includes only baseline D/P Cr and pCcr. ROC analysis confirmed this enhanced efficacy, and the results were validated by the DeLong test ($P < 0.05$). These findings suggest that the effluent IL-17 has independent clinical prognostic value. Furthermore, our logistic regression model revealed a statistically significant interaction between IL-17 levels and PD duration ($P = 0.028$). As shown in Figure 3A, higher effluent IL-17 levels predicted an increased risk of faster PSTR, and this association was more pronounced in patients with long-term PD. Conversely, no such trend was observed in the short-term treatment group. These results suggest that the prognostic discriminatory ability of IL-17 strengthens with increasing dialysis vintage.

These observations align with those of Poniewierska-Baran et al. [15] and Shao et al. [28], supporting the hypothesis that prolonged exposure to bioincompatible dialysis fluids continuously activates Th17 cells. As a well-recognized inflammatory factor, IL-17 was found to be significantly positively correlated with serum CRP levels and PD vintage. This positive correlation is in accordance with the hypothesis that a chronic inflammatory response occurs in the peritoneum of PD patients. Additionally, the positive relationship between IL-17 levels and corrected serum calcium levels suggests that IL-17 may be involved in calcium and phosphorus metabolism disorders. Consequently, IL-17 acts as an important time-varying factor within the peritoneal microenvironment. For patients with a long dialysis duration, close monitoring of IL-17 levels is essential for timely clinical intervention.

This study has the following limitations: the single-centre study design may limit its generalizability. Although the 12-month follow-up confirmed the predictive value of IL-17, it may not fully capture long-term structural outcomes.

Conclusion

In conclusion, this study demonstrated that IL-17 in the effluent serves as a potent biomarker for predicting faster PSTR. By elucidating the synergistic role of IL-17 in peritoneal fibrosis, our findings reveal a candidate biomarker for early detection and clinical intervention. Identifying high-risk populations enables the timely adjustment of treatment regimens, ultimately offering a strategy to optimize survival prospects within the PD population.

Abbreviations

4 h UF - 4-hour ultrafiltration volume; Alb - Albumin; AUC - Area Under the Curve; BMI - Body mass index; BUN - Blood urea nitrogen; CA125 - Cancer antigen 125; CRP - C-reactive protein; D/D0 glucose - Dialysate-to-initial dialysate glucose ratio; D/P Cr - Dialysate-to-plasma creatinine ratio; EF - Ejection fraction; ELISA - Enzyme-linked immunosorbent assay; ESRD - End-stage renal disease; FN - Fibronectin; Hb - Haemoglobin; IHC - Immunohistochemistry; IL-17 - Interleukin-17; IL-6 - Interleukin-6; LA - Left atrial diameter; LVD - Left ventricular diameter; LVPWD - Left ventricular posterior wall thickness; LYM - Lymphocyte count; MMT - Mesothelial-to-mesenchymal transition; MONO - Monocyte count; NE - Neutrophil count; NF- κ B - Nuclear factor kappa-B; pCcr - Peritoneal creatinine clearance; PD - Peritoneal Dialysis; PET - Peritoneal equilibration test; PF - Peritoneal Fibrosis; PSTR - Peritoneal solute transport rate; PTH - Parathyroid hormone; rCcr - Renal creatinine clearance; ROC - Receiver Operating Characteristic; RORyt - Retinoic acid-related orphan receptor gamma t; RT-qPCR - Real-time Quantitative PCR; Scr - Serum creatinine; TC - Total cholesterol; TG - Triglyceride; TGF- β - Transforming growth factor β ; Th17 - T helper 17; UFF - Ultrafiltration failure.

Author Contributions

Lejia Song: Conceptualization, Methodology, Investigation, Formal analysis, Writing-original draft. Lu Li: Methodology, Supervision, Resources. Guang Chen: Validation, Software. Qiufeng Wang: Resources, Data curation. Qingqing Rao: Supervision, Investigation. Huaina Dou: Formal Analysis, Visualization. Li Zhang: Investigation, Software. Pei Zhang (Corresponding Author): Conceptualization, Funding Acquisition, Supervision, Validation, Writing-Review & Editing; All authors read and approved the final manuscript.

Acknowledgements

We would like to thank all the patients who participated in this study and the medical staff at the Department of Nephrology, the First Affiliated Hospital of Anhui Medical University, for

their assistance in data collection and patient follow-up.

Funding Information

This work was supported by the Anhui Provincial Natural Science Foundation (Grant No. 2408085MH208).

Ethics Approval and Consent to Participate

The studies involving human participants were reviewed and approved by the Ethics Committee of the First Affiliated Hospital of Anhui Medical University (Approval No. PJ 2025-08-97). The animal study protocol was approved by the Animal Ethics Committee of Anhui Medical University (Approval Number: LLSC20241406). The human study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy or ethical restrictions.

References

- [1] Cho Y, Bello AK, Levin A, Lunney M, Osman MA, Ye F, et al. (2021). Peritoneal Dialysis Use and Practice Patterns: An International Survey Study. *Am J Kidney Dis*, 77(3), 315–325. <https://doi.org/10.1053/j.ajkd.2020.05.032>
- [2] Suryantoro SD, Thaha M, Sutanto H, & Firdaus S. (2023). Current Insights into Cellular Determinants of Peritoneal Fibrosis in Peritoneal Dialysis: A Narrative Review. *J Clin Med*, 12(13). <https://doi.org/10.3390/jcm12134401>
- [3] Teitelbaum I. (2021). Peritoneal Dialysis. *N Engl J Med*, 385(19), 1786–1795. <https://doi.org/10.1056/NEJMr2100152>
- [4] Gu J, Bai E, Ge C, Winograd J, & Shah AD. (2023). Peritoneal equilibration testing: Your questions answered. *Perit Dial Int*, 43(5), 361–373. <https://doi.org/10.1177/08968608221133629>
- [5] Brimble KS, Walker M, Margetts PJ, Kundhal KK, & Rabbat CG. (2006). Meta-analysis: peritoneal membrane transport, mortality, and technique failure in peritoneal dialysis. *J Am Soc Nephrol*, 17(9), 2591–2598. <https://doi.org/10.1681/asn.2006030194>
- [6] Mehrotra R, Ravel V, Streja E, Kuttykrishnan S, Adams SV, Katz R, et al. (2015). Peritoneal Equilibration Test and Patient Outcomes. *Clin J Am Soc Nephrol*, 10(11), 1990–2001. <https://doi.org/10.2215/cjn.03470315>
- [7] Graterol Torres F, Molina M, Soler-Majoral J, Romero-González G, Rodríguez Chitiva N, Troya-Saborido M, et al. (2022). Evolving concepts on inflammatory biomarkers and malnutrition in chronic kidney disease. *Nutrients*, 14, 4297. <https://doi.org/10.3390/nu14204297>
- [8] Prasad N, Chaturvedi S, Singh H, Udimula MP, Rawat A, Jeyakumar M, et al. (2025). Peritoneal Dialysis -Associated Fibrosis: Emerging Mechanisms and Therapeutic Opportunities. *Front Pharmacol*, 16, 1635624. <https://doi.org/10.3389/fphar.2025.1635624>
- [9] Iwakura Y, Ishigame H, Saijo S, & Nakae S. (2011). Functional specialization of interleukin-17 family members. *Immunity*, 34(2), 149–162. <https://doi.org/10.1016/j.immuni.2011.02.012>
- [10] Marchant V, Tejera-Muñoz A, Marquez-Expósito L, Rayego-Mateos S, Rodríguez-Díez RR, Tejedor L, et al. (2020). IL-17A as a Potential Therapeutic Target for Patients on Peritoneal Dialysis. *Biomolecules*, 10(10). <https://doi.org/10.3390/biom10101361>
- [11] Witowski J, Kamhieh-Milz J, Kawka E, Catar R, & Jörres A. (2018). IL-17 in Peritoneal Dialysis-Associated Inflammation and Angiogenesis: Conclusions and Perspectives. *Front Physiol*, 9, 1694. <https://doi.org/10.3389/fphys.2018.01694>
- [12] Rodríguez-Díez R, Aroeira LS, Orejudo M, Bajo MA, Hefferman JJ, Rodríguez-Díez RR, et al. (2014). IL-17A is a novel player in dialysis-induced peritoneal damage. *Kidney Int*, 86(2), 303–315. <https://doi.org/10.1038/ki.2014.33>
- [13] Huang L. (2024). The role of IL-17 family cytokines in cardiac fibrosis. *Front Cardiovasc Med*, 11, 1470362. <https://doi.org/10.3389/fcvm.2024.1470362>
- [14] Xiao H, Peng L, Jiang D, Liu Y, Zhu L, Li Z, et al. (2022). IL-17A promotes lung fibrosis through impairing mitochondrial homeostasis in type II alveolar epithelial cells. *J Cell Mol Med*, 26(22), 5728–5741. <https://doi.org/10.1111/jcmm.17600>
- [15] Poniewierska-Baran A, Tabarkiewicz J, Radzikowska U, Tynecka M, & Eljaszewicz A. (2024). The pivotal role of IL-17A in hepatic stellate cell activation. *Cent Eur J Immunol*, 49(4), 331. <https://doi.org/10.5114/ceji.2024.146900>
- [16] Liu F, Wang J, Sun Z, & Yu X. (2025). Rehmannioside A alleviates renal inflammation and fibrosis in hypertensive nephropathy via AT1R/MAPK14/IL-17 signaling pathway. *Biochem Biophys Res Commun*, 776, 152237. <https://doi.org/10.1016/j.bbrc.2025.152237>
- [17] Zhou YN, Xia JK, Shi CR, He Y, & Shang SL. (2025). Cross-talk Between Th17 Cells and Renal Tubular Epithelial Cells Promotes Fibrotic Progression in IgA Nephropathy. *Curr Med Sci*, 45(3), 626–639. <https://doi.org/10.1007/s11596-025-00068-6>
- [18] Chung SH, Heimbürger O, & Lindholm B. (2008). Poor outcomes for fast transporters on PD: the rise and fall of a clinical concern. *Semin Dial*, 21(1), 7–10. <https://doi.org/10.1111/j.1525-139X.2007.00327.x>
- [19] Huang G, Wang Y, Shi Y, Ma X, Tao M, Zang X, et al. (2021). The prognosis and risk factors of baseline high peritoneal transporters on patients with peritoneal dialysis. *J Cell Mol Med*, 25(18), 8628–8644. <https://doi.org/10.1111/jcmm.16819>
- [20] González-Mateo GT, Fernández-Míllara V, Bellón T, Liappas

- G, Ruiz-Ortega M, López-Cabrera M, et al. (2014). Paricalcitol reduces peritoneal fibrosis in mice through the activation of regulatory T cells and reduction in IL-17 production. *PLoS One*, 9(10), e108477. <https://doi.org/10.1371/journal.pone.0108477>
- [21] Ferrantelli E, Liappas G, Vila Cuenca M, Keuning ED, Foster TL, Vervloet MG, et al. (2016). The dipeptide alanyl-glutamine ameliorates peritoneal fibrosis and attenuates IL-17 dependent pathways during peritoneal dialysis. *Kidney Int*, 89(3), 625–635. <https://doi.org/10.1016/j.kint.2015.12.005>
- [22] Ito Y, Sun T, Tawada M, Kinashi H, Yamaguchi M, Katsuno T, et al. (2024). Pathophysiological Mechanisms of Peritoneal Fibrosis and Peritoneal Membrane Dysfunction in Peritoneal Dialysis. *Int J Mol Sci*, 25(16). <https://doi.org/10.3390/ijms25168607>
- [23] Helmke A, Hüsing AM, Gaedcke S, Brauns N, Balzer MS, Reinhardt M, et al. (2021). Peritoneal dialysate-range hypertonic glucose promotes T-cell IL-17 production that induces mesothelial inflammation. *Eur J Immunol*, 51(2), 354–367. <https://doi.org/10.1002/eji.202048733>
- [24] Li XR, Yang SK, Zeng BY, Tian J, Liu W, & Liao XC. (2023). Relationship between peritoneal solute transport and dialysate inflammatory markers in peritoneal dialysis patients: A cross-sectional study. *Nefrologia (Engl Ed)*, 43(3), 335–343. <https://doi.org/10.1016/j.nefro.2022.12.001>
- [25] Huang Q, Xiao R, Lu J, Zhang Y, Xu L, Gao J, et al. (2022). Endoglin aggravates peritoneal fibrosis by regulating the activation of TGF- β /ALK/Smads signaling. *Front Pharmacol*, 13, 973182. <https://doi.org/10.3389/fphar.2022.973182>
- [26] Lurje I, Gaisa NT, Weiskirchen R, & Tacke F. (2023). Mechanisms of organ fibrosis: Emerging concepts and implications for novel treatment strategies. *Mol Aspects Med*, 92, 101191. <https://doi.org/10.1016/j.mam.2023.101191>
- [27] Zhang Y, Feng W, Peng X, Zhu L, Wang Z, Shen H, et al. (2022). Parthenolide alleviates peritoneal fibrosis by inhibiting inflammation via the NF- κ B/ TGF- β /Smad signaling axis. *Lab Invest*, 102(12), 1346–1354. <https://doi.org/10.1038/s41374-022-00834-3>
- [28] Shao X, Yao L, Fu J, He M, & Zhang P. (2024). Differential expression and clinical significance of IGF2BP3 in peritoneal dialysate of patients with varying duration of peritoneal dialysis. *Clin Transl Sci*, 17(4), e13774. <https://doi.org/10.1111/cts.13774>

The Relationship between Blood Pressure Response Index and Acute Kidney Injury in Patients with Septic Shock: A Retrospective Study Based on the MIMIC-IV Database

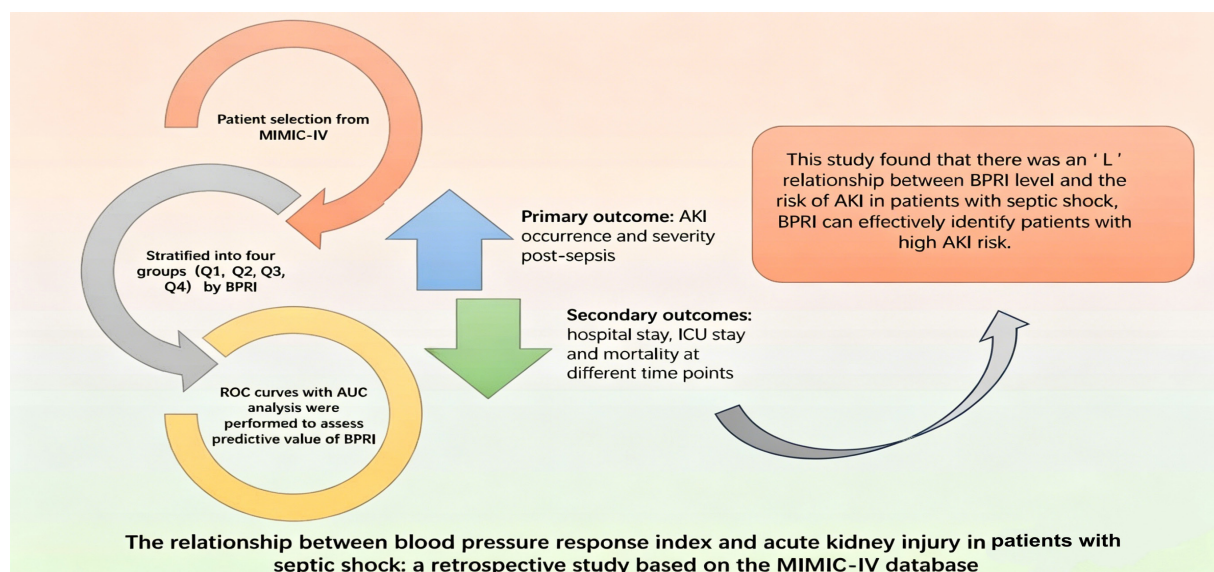
Authors

Jiahui Liu, Xinghua Chen, Jinglan Wu, Li Yao

Correspondence

yl2018ay@163.com (L. Yao)

Graphical Abstract



<https://doi.org/10.71321/61pvj246>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

The Relationship between Blood Pressure Response Index and Acute Kidney Injury in Patients with Septic Shock: A Retrospective Study Based on the MIMIC-IV Database

Jiahui Liu^{1,2}, Xinghua Chen^{1,2}, Jinglan Wu^{1,2}, Li Yao^{1,2*}

Received: 2026-01-19 | Accepted: 2026-04-15 | Published online: 2026-05-26

Abstract

Background: The blood pressure response index (BPRI) is the ratio of mean arterial pressure (MAP) to the vasoactive-inotropic score (VIS). It accurately reflects the response to vasoactive drugs in patients with septic shock and assesses the patient's vascular status. However, the relationship between BPRI and the occurrence of acute kidney injury (AKI) in patients with septic shock has never been studied.

Methods: This study retrospectively analyzed patients diagnosed with sepsis and using vasoactive drugs in the MIMIC-IV database. The patients were divided into four groups according to the BPRI level. The primary outcome is whether AKI occurs after the diagnosis of sepsis. The association between BPRI and outcome was assessed by constructing logistic regression models and plotting restricted cubic spline (RCS) curves. The predictive value of the BPRI for outcome was evaluated by plotting receiver operating characteristic (ROC) curves and comparing areas under the curves (AUC).

Results: 5524 patients with septic shock were finally included. The median age of the patients was 68.8 years, and 84.3% developed AKI after the diagnosis of sepsis. In a fully adjusted model, BPRI and the development of AKI were significantly correlated (OR [95%CI], 0.988 [0.979, 0.996], $P=0.006$). The results of the RCS curves suggested an "L"-shaped association and the inflection point was located at the level of 3.98. On the left side of the inflection point, there was a decreasing trend in the risk of AKI in patients as BPRI levels increased (OR [95%CI], 0.82 [0.747, 0.9], $P<0.001$). On the right side of the inflection point, there was no statistically significant association (OR [95%CI], 0.995 [0.985, 1.01], $P=0.212$). In addition, the BPRI showed favorable predictive value for outcome (AUC [95% CI], 0.624 [0.601, 0.639]).

Conclusion: We identified a distinct non-linear association between BPRI and AKI risk in septic shock, characterized by a critical inflection point at 3.98. Below this threshold, the likelihood of AKI rises sharply as BPRI decreases. Integrating perfusion pressure and vasopressor load, the BPRI serves as a readily available metric to stratify renal risk and recognize vascular hyporesponsiveness.

Keywords: septic shock; acute kidney injury; blood pressure response index; intensive care unit; MIMIC-IV database

Introduction

Sepsis is a life-threatening organ dysfunction caused by dysregulation of the body's response to infection and is one of the leading causes of death in intensive care unit (ICU) patients [1]. The incidence of acute kidney injury (AKI) in patients with sepsis is up to 70% or more, with prolonged hospitalization cycles and approximately one-fold higher morbidity and mortality rates in patients with concurrent AKI compared to sepsis patients without concurrent AKI [2-3]. In patients with septic shock, persistent renal insufficiency of perfusion is a major cause of AKI [4]. The Surviving Sepsis Campaign guidelines recommend controlling patients' mean arterial pressure (MAP) above 65 mmHg [5].

However, in distributed shock due to infection, although suf-

ficient cardiac output and large circulation pressure can be maintained, persistent organ hypoperfusion may still occur due to the loss of automatic regulation of vasoconstriction and decreased arterial tension [6-7]. Moreover, due to the widespread use of vasoactive drugs, MAP-based monitoring may not reflect the actual organ perfusion of the patient [8-9]. The blood pressure response index (BPRI) is the ratio between MAP and the vasoactive-inotropic score (VIS) [10-11]. Recently, it has been found that the BPRI accurately responds to vasoactive medications in patients with septic shock, allowing for a convenient assessment of the patient's vascular status [11]. In multicenter cohort studies, BPRI has been shown to be strongly associated with poor patient prognosis [11]. However, the relationship between BPRI and the development of AKI in patients with septic shock is unclear.

¹ Anhui Medical University, Anhui, Hefei, 230032.

² The Second People's Hospital of Hefei, Intensive Care Unit, Anhui, Hefei, 230011.

* Corresponding Author.

In this study, we hypothesize that BPRI and the occurrence of AKI in septic shock patients are closely related and can be used as an accurate predictor. We will test this hypothesis by retrospectively analyzing a large sample of septic shock patients in the Medical Information Mart for Intensive Care-IV database.

Materials and Methods

Data Sources

All data for this study were obtained from the Beth Israel Deaconess Medical Center (BIDMC) intensive care unit database (MIMIC-IV 2.2) [12]. The database covers the period from 2008 to 2019 and includes detailed medical information on nearly 200,000 hospitalizations as well as over 50,000 patients admitted to the ICU. In addition, MIMIC-IV data are rigorously de-identified to ensure that patient privacy is effectively protected. It is worth mentioning that all patients included in this database were adults (age >18 years). The institutional review board of BIDMC waived the declaration of ethical review and informed consent and approved the sharing of research resources. One of the authors has applied for access to this database (Certificate number: 58951192).

Study Objects

All patients were admitted to the ICU for the first time and stayed for more than 24 hours. These patients were diagnosed with sepsis on the first day of ICU admission and were treated with vasoactive drugs. Sepsis was defined using the Third International Consensus on Sepsis and Septic Shock (Sepsis-3) as an increase of at least 2 points in the sequential organ failure assessment (SOFA) score from baseline on the basis of infection or suspected infection [13].

Formulas and Outcomes

VIS formula: dopamine ($\mu\text{g/kg/min}$) + dobutamine ($\mu\text{g/kg/min}$) + epinephrine ($\mu\text{g/kg/min}$) $\times 100$ + norepinephrine ($\mu\text{g/kg/min}$) $\times 100$ + vasopressin (U/kg/min) $\times 10,000$ + milrinone ($\mu\text{g/kg/min}$) $\times 10$ [14]. We chose the MAP with maximal support of vasoactive drugs and calculated the BPRI based on the ratio of MAP to VIS. Patients were categorized into four groups based on quartiles of BPRI: Q1 group (BPRI<1.90), Q2 group (BPRI: 1.90-4.03), Q3 group (BPRI: 4.02-8.84), and Q4 group (BPRI>8.84). The primary outcome of the study was whether AKI occurred after the diagnosis of septic shock and during ICU hospitalization. Secondary outcomes included: AKI staging, renal replacement therapy, in-hospital all-cause mortality, 30-day, 180-day, and 365-day all-cause mortality. The diagnosis and staging of AKI were determined according to the guidelines of Kidney Disease: Improving Global Outcomes (KDIGO-2012) [15]. The study flow chart is shown in Figure 1.

Data Collection

The following information was collected from the database by Navicat Premium 16.0 software. 1) General information: age, gender, weight, race, smoking history, type of ICU admission. 2) Comorbidities: hypertension, diabetes mellitus (DM), coronary heart disease (CHD), heart failure (HF), chronic obstructive pulmonary disease (COPD), chronic liver disease (CLD), chronic kidney disease (CKD). 3) Disease severity scores: acute

physiology score-III (APS-III), SOFA score, charles comorbidity index (CCI), glasgow coma scale (GCS) score. 4) Various inspections on the first day in the ICU: white blood cell (WBC) count, red blood cell (RBC) count, platelet (PLT) count, hemoglobin (Hb), serum creatinine (SCr), blood urea nitrogen (BUN), lactic acid, bicarbonate, pH, prothrombin time (PT), activated partial thromboplastin time (APTT), international normalized ratio (INR), sodium, potassium, chlorine, urine volume. 5) Treatments received on the first day of ICU admission: diuretics, Sedative drugs, and invasive mechanical ventilation (IMV).

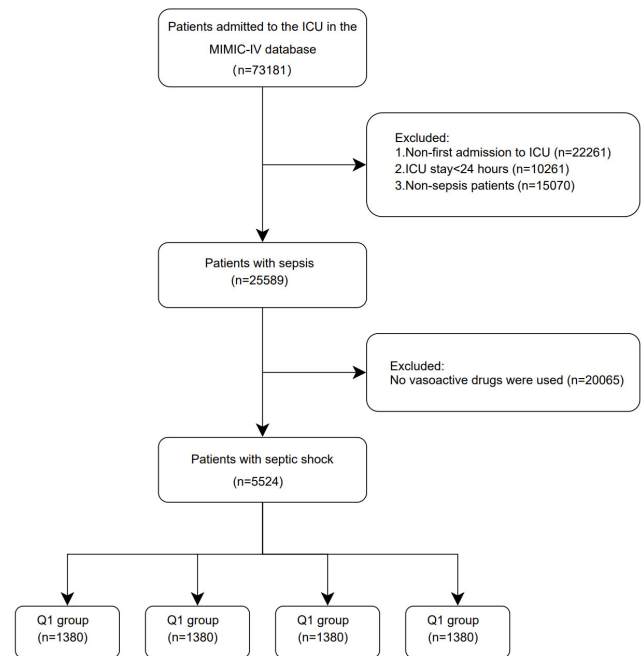


Figure 1. Flow chart.

Statistical Analysis

Statistical analyses were done through R software (version 4.3.3). Categorical variables were expressed as frequencies (proportions) and continuous variables as medians (interquartile range). Categorical variables were compared by chi-square test. Continuous variables were skewed and compared by the Mann-Whitney U test.

Missing data were observed in less than 15% of the variables. To address this, we employed multiple imputation by chained equations (MICE) to generate five imputed datasets, assuming that data were missing at random (MAR).

The correlation between BPRI and the occurrence of AKI was assessed by univariate and multivariate logistic regression models. We adjusted for potential confounders through three models based on factors from previous studies that may have influenced study variables or outcomes. Model 1 adjusted for general information (age, gender, weight, race, smoking history). Model 2 added comorbidities and various scores (hypertension, DM, CHD, HF, CKD, APS-III, SOFA score, and GCS score) to Model 1. Model 3 adds various tests and treatments (WBC count, PLT count, Hb, lactic acid, pH, PT, INR, potassium, diuretics, sedative drugs, IMV) to Model 2. Restricted cubic (RCS) curves were plotted within the framework of the unadjusted model and Model 3 to further assess associations between study variables and outcomes.

Analyze the predictive value of the BPRI for endpoints by plotting the receiver operating characteristic (ROC) curve and comparing the area under the curve (AUC). Incorporate study metrics into existing disease severity scores and compare whether the BRI improves the accuracy of the scores in predicting study outcomes. Predictive performance was compared using the Delong test. In addition, to further assess differences in the impact of BPRI in different populations, we also performed subgroup analyses in adjusted model 3. Two-sided *P* values below 0.05 were considered statistically significant. To verify the robustness of our findings, we performed sensitivity analyses. Specifically, considering the potential use of phenylephrine in septic shock, we recalculated the VIS by incorporating phenylephrine (weighted $\times 10$) and repeated the regression analysis to assess whether the association between BPRI and AKI remained consistent.

Result

Baseline Data Analysis

A final total of 5524 patients with septic shock were included. In the study cohort, the median age of the patients was 68.8 years, 57.6% were female, and 64.1% were white. Patients were divided into 4 groups based on BPRI levels, and the baseline information for each group is shown in Table 1. The patients were also grouped according to the occurrence of AKI, and the baseline information of the two groups is shown in Supplementary Table 1. The profile of study variables and outcomes for each group of patients is shown in Table 2. In the study cohort, norepinephrine was used most frequently (85.5%), followed by vasopressin (27.9%), and dobutamine (5.1%) least frequently. The incidence of AKI in the total population was 84.3%. The odds of developing AKI were significantly higher in all groups of patients as the level of BPRI decreased ($P < 0.001$). In addition, patients in the Q1 group had significantly higher final AKI stage, odds of receiving RRT, hospitalization period, and in-hospital all-cause mortality than patients in the other three groups ($P < 0.001$). Kaplan-Meier survival analysis showed a stepwise distribution of 30-, 180-, and 365-day survival rates in each group of patients as BPRI levels decreased (Figure 2).

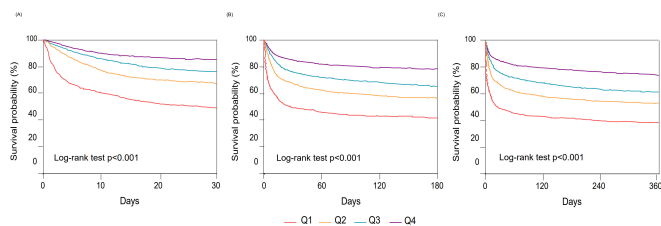


Figure 2. Kaplan-Meier survival analysis at 30 (A), 180 (B), and 365 days (C) for four groups of patients.

Correlation Analysis

Through constructing a multifactorial logistic regression model, we found that BPRI and the incidence of AKI in septic shock patients during ICU hospitalization were strongly associated. The results were still significant in the unadjusted model (OR [95%CI], 0.982 [0.975, 0.988], $P < 0.001$) and the fully adjusted model (OR [95%CI], 0.988 [0.979, 0.996], $P = 0.006$). The Q4

group with the lowest incidence of AKI was set as the reference group. The risk of AKI in most groups of patients in each model was significantly higher than that in the Q4 group (Table 3). In both the unadjusted and fully adjusted models, the plotted RCS curves suggested an “L” shaped relationship between the level of BPRI and the risk of occurrence of AKI (Both P for no-linear < 0.001 , Figure 3). Threshold effects analysis suggests that in the adjusted model, the inflection point of the BPRI is located at the 3.98 level. On the left side of the inflection point, there was a decreasing trend in the risk of AKI in patients as BPRI levels increased (OR [95%CI], 0.82 [0.747, 0.9], $P < 0.001$). On the right side of the inflection point, there was no statistically significant change in the risk of AKI in patients as BPRI levels increased (OR [95%CI], 0.995 [0.985, 1.01], $P = 0.212$) (Supplementary Table 2).

To evaluate the potential influence of phenylephrine, we also recalculated the BPRI by incorporating this agent into the VIS formula. The modified index retained a strong independent association with AKI in multivariable models. The RCS analysis (Supplementary Figure 1) reproduced the “L-shaped” association seen in the primary cohort, identifying a nearly identical inflection point at 4.03. This consistency supports the validity of BPRI as a marker for renal risk, independent of vasopressor choice.

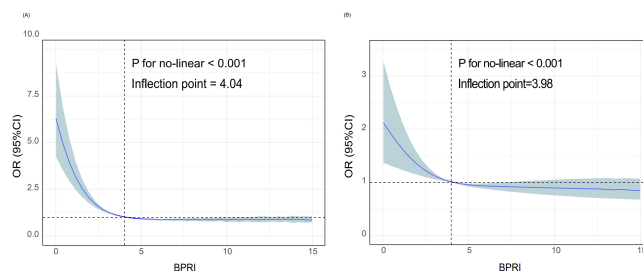


Figure 3. Correlation between BPRI and AKI incidence odds ratio in unadjusted model (A) and adjusted model 3 (B).

Predictive Analysis

ROC analysis indicated that the BPRI demonstrated moderate discriminative ability for predicting AKI (AUC 0.624 [95% CI 0.601–0.639]). Notably, the composite index showed superior predictive performance compared to its individual components, MAP (AUC 0.542 [0.521–0.563]) and VIS (AUC 0.618 [0.599–0.637]), suggesting that integrating vasopressor load with pressure targets offers better risk stratification than either metric alone. Regarding in-hospital mortality, the BPRI exhibited relatively higher prognostic accuracy (AUC 0.697 [0.681–0.712]). Furthermore, the inclusion of BPRI yielded incremental value when combined with established disease severity scores, improving the predictive accuracy of both APS-III and SOFA models (Table 4, Figure 4).

Subgroup analysis

Subgroups of patients were analyzed using age, gender, hypertension, SOFA score, and sedative drugs as stratification factors. The results showed that in different populations, the risk of AKI was higher in groups Q1, Q2, and Q3 compared to group Q4. Interaction tests showed that there was no significant interaction between study variables and stratification factors (Figure 5).

Table 1. Baseline characteristics of patients in each group.

Variables	Total (n=5524)	Q1 group (n=1380)	Q2 group (n=1382)	Q3 group (n=1381)	Q4 group (n=1381)	P-value
General information						
Age, year	68.8 (57.8, 79.3)	67.5 (57.0, 78.7)	69.6 (57.3, 80.2)	69.7 (58.2, 79.4)	68.4 (59.0, 78.7)	.021
Gender, female	3181 (57.6%)	782 (56.7%)	775 (56.1%)	790 (57.2%)	834 (60.4%)	.098
Weight, kg	79.1 (66.6, 95.0)	77.2 (65.0, 93.4)	78.2 (65.8, 94.0)	79.0 (66.7, 93.1)	81.2 (68.5, 98.9)	<.001
Race						<.001
White	3542 (64.1%)	798 (57.8%)	881 (63.7%)	901 (65.2%)	962 (69.7%)	
Black	389 (7%)	102 (7.4%)	94 (6.8%)	102 (7.4%)	91 (6.6%)	
Other	1593 (28.8%)	480 (34.8%)	407 (29.5%)	378 (27.4%)	328 (23.8%)	
Smoking history	1567 (28.4%)	353 (25.6%)	400 (28.9%)	428 (31%)	386 (28%)	.016
Type of ICU						<.001
MICU	2668 (48.3%)	730 (52.9%)	717 (51.9%)	708 (51.3%)	513 (37.1%)	
SICU	1041 (18.8%)	332 (24.1%)	292 (21.1%)	258 (18.7%)	159 (11.5%)	
Other	1815 (32.9%)	318 (23%)	373 (27%)	415 (30.1%)	709 (51.3%)	
Complications						
Hypertension	3720 (67.3%)	875 (63.4%)	911 (65.9%)	946 (68.5%)	988 (71.5%)	<.001
DM	1705 (30.9%)	400 (29%)	394 (28.5%)	452 (32.7%)	459 (33.2%)	.008
CHD	1887 (34.2%)	357 (25.9%)	448 (32.4%)	481 (34.8%)	601 (43.5%)	<.001
HF	1512 (27.4%)	340 (24.6%)	381 (27.6%)	394 (28.5%)	397 (28.7%)	.059
COPD	407 (7.4%)	102 (7.4%)	104 (7.5%)	115 (8.3%)	86 (6.2%)	.208
Stroke	404 (7.3%)	100 (7.2%)	123 (8.9%)	80 (5.8%)	101 (7.3%)	.020
CID	611 (11.1%)	202 (14.6%)	162 (11.7%)	155 (11.2%)	92 (6.7%)	<.001
CKD	271 (4.9%)	69 (5%)	82 (5.9%)	62 (4.5%)	58 (4.2%)	.162
Various scores						
APS, III	64.0 (45.0, 90.0)	86.0 (65.0, 110.0)	68.0 (49.0, 92.0)	56.0 (42.0, 77.0)	49.0 (35.0, 71.0)	<.001
SOFA score	10.0 (7.0, 12.0)	12.0 (10.0, 15.0)	10.0 (8.0, 13.0)	9.0 (7.0, 11.0)	7.0 (5.0, 10.0)	<.001
CCI	6.0 (4.0, 8.0)	6.0 (4.0, 8.0)	6.0 (4.0, 8.0)	6.0 (4.0, 8.0)	6.0 (4.0, 7.0)	<.001
GCS score	13.0 (8.0, 15.0)	11.0 (6.0, 14.0)	12.0 (7.0, 14.0)	13.0 (9.0, 15.0)	14.0 (10.0, 15.0)	<.001
VIS	18.1 (8.0, 38.4)	55.8 (48.1, 80.0)	25.4 (20.0, 30.1)	12.0 (10.0, 15.0)	5.0 (3.0, 6.0)	<.001
BPRI	4.0 (1.9, 8.8)	1.3 (0.9, 1.6)	2.8 (2.3, 3.4)	6.1 (4.8, 7.3)	15.3 (11.9, 24.3)	<.001
Various inspections						
WBC count, $\times 10^9$ /L	12.0 (8.7, 16.5)	14.0 (9.7, 20.1)	12.3 (9.0, 17.1)	11.1 (8.0, 15.0)	11.3 (8.4, 14.9)	<.001
RBC count, $\times 10^{12}$ /L	3.2 (2.8, 3.6)	3.2 (2.8, 3.6)	3.2 (2.8, 3.6)	3.2 (2.8, 3.6)	3.3 (2.9, 3.7)	<.001
PLT, $\times 10^9$ /L	144.0 (101.0, 204.0)	127.0 (76.0, 190.0)	147.0 (101.0, 211.0)	153.0 (109.0, 213.0)	145.0 (108.0, 199.0)	<.001
Hb, mg/dl	9.6 (8.5, 10.8)	9.6 (8.4, 10.8)	9.6 (8.4, 10.8)	9.6 (8.5, 10.8)	9.8 (8.7, 10.9)	.005
SCr, mg/dl	1.2 (0.8, 2.1)	1.7 (1.0, 2.8)	1.2 (0.8, 2.1)	1.1 (0.8, 1.9)	1.1 (0.8, 1.5)	<.001
BUN, mg/dl	25.0 (15.0, 41.0)	31.0 (20.0, 47.0)	27.0 (16.0, 41.0)	23.0 (14.0, 39.0)	21.0 (14.0, 32.0)	<.001
Lactic acid, mmol/L	1.6 (1.2, 2.1)	1.9 (1.4, 2.6)	1.5 (1.2, 2.0)	1.4 (1.1, 1.8)	1.5 (1.1, 2.0)	<.001
Bicarbonate, mEq/L	1.5 (1.2, 2.0)	1.7 (1.3, 2.4)	1.5 (1.2, 2.0)	1.4 (1.1, 1.8)	1.5 (1.1, 2.0)	<.001
pH, units	7.4 (7.3, 7.4)	7.4 (7.3, 7.4)	7.4 (7.3, 7.4)	7.4 (7.3, 7.4)	7.4 (7.3, 7.4)	<.001
PT, s	14.8 (13.2, 18.0)	16.6 (13.9, 21.9)	14.9 (13.3, 18.1)	14.4 (12.9, 16.8)	14.1 (12.9, 15.9)	<.001
APTT, s	33.0 (28.8, 42.5)	37.7 (30.7, 50.1)	33.8 (29.2, 42.9)	31.9 (28.2, 39.8)	31.0 (27.8, 37.3)	<.001
INR	1.3 (1.2, 1.6)	1.5 (1.3, 2.0)	1.4 (1.2, 1.6)	1.3 (1.2, 1.5)	1.3 (1.2, 1.5)	<.001
Sodium, mEq/L	138.0 (135.0, 141.0)	137.0 (134.0, 141.0)	138.0 (135.0, 142.0)	139.0 (136.0, 142.0)	138.0 (135.0, 141.0)	<.001
Potassium, mEq/L	4.1 (3.8, 4.5)	4.2 (3.8, 4.8)	4.1 (3.7, 4.5)	4.1 (3.7, 4.4)	4.1 (3.8, 4.4)	<.001
Chlorine, mEq/L	104.0 (100.0, 108.0)	104.0 (98.0, 108.0)	105.0 (101.0, 109.0)	105.0 (101.0, 109.0)	104.0 (100.0, 108.0)	<.001
Urine volume, ml	1335.5 (704.5, 2225.0)	896.0 (365.5, 1765.0)	1245.0 (655.0, 2140.0)	1470.0 (860.0, 2325.0)	1665.0 (1050.0, 2494.0)	<.001
MAP, mmHg	72.7 (68.5, 77.5)	71.6 (66.9, 76.0)	72.5 (68.5, 76.7)	72.7 (68.7, 77.8)	74.3 (70.1, 79.3)	<.001
Received treatment						
Diuretics	1445 (26.2%)	296 (21.4%)	315 (22.8%)	350 (25.3%)	484 (35%)	<.001
Sedative drugs	4221 (76.4%)	1221 (88.5%)	1062 (76.8%)	918 (66.5%)	1020 (73.9%)	<.001
IMV	3938 (71.3%)	1142 (82.8%)	1002 (72.5%)	837 (60.6%)	957 (69.3%)	<.001

Note: Values are the median (interquartile range), or n (%).

Discussion

In this large retrospective cohort study of 5,524 patients with septic shock, we identified a distinct non-linear association between the Blood Pressure-to-Resistivity Index (BPRI) and the risk of developing AKI. A critical inflection point was observed at approximately 3.98. Below this threshold, every 1-unit decrease in BPRI was associated with an 18% increase in the risk of AKI during the ICU stay. Lower BPRI levels were not only independent predictors of severe AKI and the need for renal replacement therapy (RRT) but also correlated significantly with prolonged hospital stays and increased short- and long-term mortality. Notably, our sensitivity analysis incorporating phenylephrine confirmed the robustness of these findings, yielding a nearly identical threshold (4.03). To the best of our knowledge, this is the first study to elucidate the relationship between BPRI—a novel metric integrating perfusion pressure and vasopressor load—and renal outcomes in septic shock.

Renal perfusion insufficiency due to septic shock is the most important cause of AKI in these patients. The kidney is extremely sensitive to hypoxia, and the tubular epithelial cells in the thick segment of the ascending branch of the renal medullary collaterals have a very high oxygen extraction rate, and

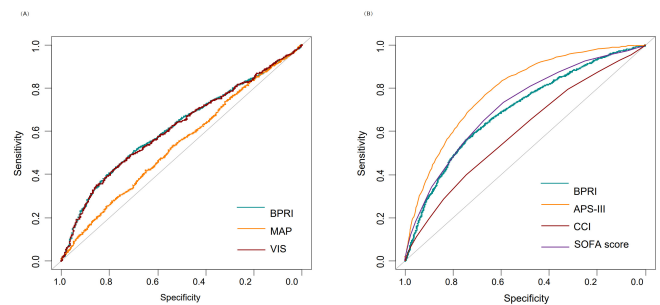


Figure 4. ROC curves for each variable in predicting the occurrence of AKI (A) and in-hospital all-cause mortality (B) in patients with septic shock.

Table 2. Hemodynamic drugs and study outcomes.

	Total (n=5524)	Q1 (n=1380)	Q2 (n=1382)	Q3 (n=1381)	Q4 (n=1381)	P-value
Hemodynamic drugs						
Dopamine	579 (10.5%)	196 (14.2%)	123 (8.9%)	139 (10.1%)	121 (8.8%)	<.001
Dobutamine	282 (5.1%)	132 (9.6%)	62 (4.5%)	32 (2.3%)	56 (4.1%)	<.001
Epinephrine	1021 (18.5%)	331 (24%)	171 (12.4%)	126 (9.1%)	393 (28.5%)	<.001
Norepinephrine	4722 (85.5%)	1373 (99.5%)	1341 (97%)	1251 (90.6%)	757 (54.8%)	<.001
Vasopressin	1542 (27.9%)	887 (64.3%)	419 (30.3%)	133 (9.6%)	103 (7.5%)	<.001
Milrinone	302 (5.5%)	58 (4.2%)	84 (6.1%)	68 (4.9%)	92 (6.7%)	<.001
Outcomes						
AKI	4658 (84.3%)	1287 (93.3%)	1186 (85.8%)	1089 (78.9%)	1096 (79.4%)	<.001
AKI-stages						
0	866 (15.7%)	93 (6.7%)	196 (14.2%)	292 (21.1%)	285 (20.6%)	<.001
1	700 (12.7%)	132 (9.6%)	161 (11.6%)	184 (13.3%)	223 (16.1%)	<.001
2	1845 (33.4%)	378 (27.4%)	456 (33%)	452 (32.7%)	559 (40.5%)	<.001
3	2113 (38.3%)	777 (56.3%)	569 (41.2%)	453 (32.8%)	314 (22.7%)	<.001
RRT	610 (11%)	304 (22%)	162 (11.7%)	83 (6%)	61 (4.4%)	<.001
Hospitalization time, day	9.7 (5.5, 16.8)	10.2 (3.5, 19.5)	10.8 (6.3, 18.1)	9.6 (5.9, 16.2)	8.6 (5.6, 13.7)	<.001
Length of stay in ICU, day	4.1 (2.3, 8.1)	5.0 (2.6, 11.0)	5.0 (2.9, 9.1)	3.8 (2.3, 7.1)	3.1 (1.9, 5.9)	<.001
In-hospital mortality	1536 (27.8%)	676 (49%)	406 (29.4%)	275 (19.9%)	179 (13%)	<.001
30 days mortality	1688 (30.6%)	702 (50.9%)	450 (32.6%)	329 (23.8%)	207 (15%)	<.001
180 days mortality	2191 (39.7%)	802 (58.1%)	605 (43.8%)	479 (34.7%)	305 (22.1%)	<.001
365 days mortality	2391 (43.3%)	851 (61.7%)	653 (47.3%)	535 (38.7%)	352 (25.5%)	<.001

Note: Values are the median (interquartile range), or n (%).

Table 3. Multifactorial logistic regression modeling to evaluate the association between BPRI and the incidence of AKI.

Variables	Unadjusted model	Model 1	Model 2	Model 3
BPRI	0.982 (0.975, 0.988) ***	0.976 (0.969, 0.983) ***	0.996 (0.989, 1.001)	0.988 (0.979, 0.996) **
Q1	3.599 (2.820, 4.630)***	4.493 (3.492, 5.830) ***	2.218 (1.669, 2.968) ***	1.957 (1.453, 2.651) ***
Q2	1.573 (1.290, 1.923) ***	1.780 (1.449, 2.191) ***	1.138 (0.905, 1.432)	1.322 (1.036, 1.690) **
Q3	0.970 (0.807, 1.165)	1.050 (0.868, 1.270)	0.843 (0.686, 1.036)	1.113 (0.893, 1.388)
Q4	Ref.	Ref.	Ref.	Ref.
P for trend	<0.001	<0.001	0.004	<0.001

Note: P-value: * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Model 1: After adjusting for age, gender, weight, race, and smoking history; Model 2: adjusted for model 1, additionally adjusted for hypertension, DM, CHD, HF, CKD, APS-III, SOFA score, and GCS score; Model 3: adjusted for model 2, additionally adjusted for WBC count, PLT count, Hb, Lactic acid, pH, PT, INR, potassium, diuretics, sedative drugs, and IMV.

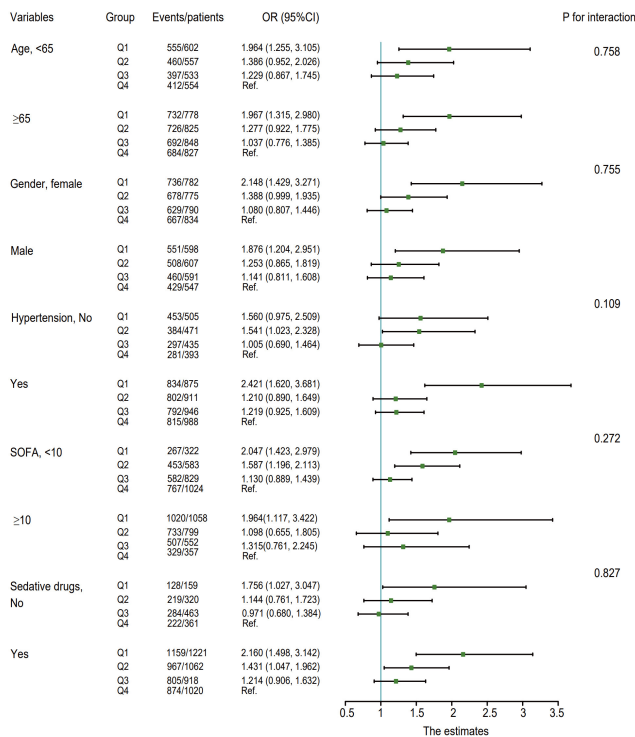


Figure 5. Subgroup analysis with AKI incidence as outcome event.

about 80% of the oxygen delivery is utilized by them [16]. Reduced oxygen supply to them due to decreased renal perfusion pressure makes them more susceptible to renal dysfunction [17]. In our study cohort, all patients with sepsis required treatment with vasoactive drugs and therefore exhibited a higher incidence of AKI relative to other studies [18-19]. A prospective, exploratory cohort study found that in the early stages of diagnosis of infectious shock, patients in the group that developed AKI had significantly lower MAP levels than those in the group that did not develop AKI (MAP: 67 ± 15 mmHg VS. 68 ± 17 mmHg, $P=0.043$) [20]. Meiping Wang et al. A prospective multicenter cohort study found that among patients with a diagnosis of sepsis and AKI, patients who developed early persistent AKI generally had a lower MAP than patients with late or early transient AKI [21].

As recommended by sepsis management guidelines, norepinephrine is commonly used as a first-line pressor-boosting agent in adult patients with septic shock [13]. In addition combination vasopressin therapy is recommended for patients whose MAP remains difficult to maintain, and positive

inotropic agents such as dobutamine are often required for patients with concomitant cardiac insufficiency. The use of vasoactive drugs in patients with infectious shock and the patient's response to the drugs have some individual differences [22-23]. The VIS score can be used to quantitatively assess the strength of hemodynamic support. A retrospective study of 1935 adult patients undergoing cardiovascular surgery found that VIS was strongly associated with the development of postoperative AKI in patients (OR [95% CI], 1.191 [1.11, 1.34], $P<0.001$) [24]. The BPRI combines hemodynamic status and support strength and can be used to accurately respond to a patient's response to vasoactive medications. Crucially, in our subgroup analyses, the protective effect of higher BPRI was more pronounced in patients with hypertension. This likely reflects the rightward shift of the renal autoregulatory curve induced by chronic hypertension. These patients may require higher perfusion pressures to overcome increased varying vascular resistance and maintain glomerular filtration, making them more sensitive to drops in BPRI. In our study, we found that when the BPRI was below the level of 3.98, the risk of AKI in patients increased sharply with the decrease in BPRI. More attention needs to be paid to such patients even in the presence of circulatory stabilization.

The current diagnosis of AKI is based primarily on increased SCr or decreased urine output. However, sepsis also reduces peripheral perfusion, causing a decrease in muscle blood supply, which results in a decrease in creatine production, leading to a lag in the increase in serum creatinine compared to the actual impairment of renal function [25-26]. And decreased urine output may also be diluted by aggressive fluid resuscitation as well as diuretic use, resulting in failure to diagnose AKI in a timely manner [27]. It is difficult to avoid delays in developing a diagnosis and implementing treatment that are based on the fact that renal failure has already occurred. In patients with septic shock, accurate assessment of renal perfusion status is critical for effective prevention of AKI [28]. Currently, ultrasonic Doppler is most commonly used to approximate renal perfusion by measuring the resistivity index of the interlobar arteries [29]. However, ultrasound Doppler results are more dependent on the skill and experience of the operator. Different pathologies (e.g., renal artery stenosis) may result in similar ultrasound presentations, which adds to the complexity of clinical judgment [30]. Furthermore, abdominal ultrasound can be affected by conditions such as obesity or the presence of abdominal and intestinal gas, which can affect the accuracy of the measurements [31]. In contrast, the BPRI is derived from readily available bedside parameters and requires no special-

Table 4. Predictive performance of each model for different outcomes

Models	AUC (95%CI)	Models	AUC (95%CI)	P for comparison
Occurrence of AKI				
BPRI	0.624 (0.601, 0.639)			
MAP	0.542 (0.521, 0.563)			
VIS	0.618 (0.599, 0.637)			
In-hospital mortality				
BPRI	0.697 (0.681, 0.712)			
APS- III	0.789 (0.777, 0.802)	+BPRI	0.796 (0.783, 0.808)	<0.001
SOFA score	0.713 (0.699, 0.729)	+BPRI	0.731 (0.717, 0.746)	<0.001
CCI	0.601 (0.584, 0.618)	+BPRI	0.697 (0.682, 0.712)	0.063

ized equipment. Our findings suggest that BPRI may serve as a surrogate marker for renal perfusion status, reflecting the systemic vascular tone. Further prospective validation is warranted to confirm its utility in guiding hemodynamic therapy. This study is subject to several limitations. Given its retrospective, single-center design, inherent biases preclude definitive causal inference, necessitating validation in larger, multi-center cohorts. A key methodological constraint is the calculation of BPRI at a single time point—the maximum vasopressor dosage. While this cross-sectional approach may introduce lead-time bias due to variable intervals from admission to peak support, we prioritized this specific juncture as it captures the nadir of vascular responsiveness. Theoretically, this represents the phase of maximal hemodynamic instability and potential renal hypoxic insult. Although longitudinal tracking would offer a more granular temporal profile, the peak-dose BPRI serves as a pragmatic indicator of the worst-case physiological status. Furthermore, the lack of baseline MAP data for many patients complicates the assessment of relative hypotension. Despite statistical adjustments for hypertension and heart failure, unmeasured variability in pre-morbid blood pressure remains a potential confounder. Regarding predictive performance, while the AUC of 0.624 indicates moderate discrimination, the BPRI consistently outperformed its individual components (MAP and VIS). Future research should explore whether integrating this index into existing prognostic models enhances their clinical utility.

Conclusions

Our findings establish the BPRI as a significant independent predictor of renal outcomes and survival in septic shock. Specifically, a BPRI below 4.0 signals a sharp increase in AKI risk. As a composite metric, it captures the interplay between perfusion pressure and vasopressor load more effectively than either parameter alone. Although its standalone discrimination is moderate, the BPRI adds important prognostic value, helping clinicians stratify patients based on vascular responsiveness early in their ICU course.

Abbreviations

Acute Kidney Injury: AKI; Acute Physiology Score-III: APS-III; Activated Partial Thromboplastin Time: APTT; Area Under the Curve: AUC; Blood Pressure Response Index: BPRI; Blood Urea Nitrogen: BUN; Charlson Comorbidity Index: CCI; Coronary Heart Disease: CHD; Confidence Interval: CI; Chronic Kidney Disease: CKD; Chronic Liver Disease: CLD; Chronic Obstructive Pulmonary Disease: COPD; Diabetes Mellitus: DM; Glasgow Coma Scale: GCS; Hemoglobin: Hb; Heart Failure: HF; Hazard Ratio: HR; Intensive Care Unit: ICU; Length of Intensive Care Unit Stay: ICU stay; Invasive Mechanical Ventilation: IMV; International Normalized Ratio: INR; Mean Arterial Pressure: MAP; Medical Intensive Care Unit: MICU; Medical Information Mart for Intensive Care IV: MIMIC-IV; Odds Ratio: OR; Platelet: PLT; Prothrombin Time: PT; Quartile: Q; Red Blood Cell: RBC; Restricted Cubic Spline: RCS; Reference: Ref.; Receiver Operating Characteristic: ROC; Renal Replacement Therapy: RRT; Serum Creatinine: SCr; Surgical Intensive Care Unit: SICU; Sequential

Organ Failure Assessment: SOFA; Vasoactive Inotropic Score: VIS; White Blood Cell: WBC; Length of Hospital Stay: hospital stay.

Author Contributions

J.H.L., X.H.C., and L.Y. conceived and designed the research, and also participated in the review and revision of the manuscript. J.H.L. and X.H.C. drafted the initial manuscript and collected the data. Z.H.H. and J.L.W. conducted the data analysis and interpretation. All authors read and approved the final draft.

Acknowledgments

We are grateful to all the developers and researchers who created and maintained the MIMIC-IV database.

Funding Information

Not Applicable.

Ethics Approval and Consent to Participate

This study for the MIMIC-IV database was approved by the review boards of the Massachusetts Institute of Technology and Beth Israel Deaconess Medical Center, and the patient information was de-identified so that there were no ethical concerns.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

References

- [1] Zampieri FG, Bagshaw SM, & Semler MW. (2023). Fluid Therapy for Critically Ill Adults With Sepsis: A Review. *Jama*, 329(22), 1967–1980. <https://doi.org/10.1001/jama.2023.7560>
- [2] Oppert M, Engel C, Brunkhorst FM, Bogatsch H, Reinhart K, Frei U, et al. (2008). Acute renal failure in patients with severe sepsis and septic shock—a significant independent risk factor for mortality: results from the German Prevalence Study. *Nephrol Dial Transplant*, 23(3), 904–909. <https://doi.org/10.1093/ndt/gfm610>

- [3] Zarbock A, Nadim MK, Pickkers P, Gomez H, Bell S, Joannidis M, et al. (2023). Sepsis-associated acute kidney injury: consensus report of the 28th Acute Disease Quality Initiative workgroup. *Nat Rev Nephrol*, 19(6), 401–417. <https://doi.org/10.1038/s41581-023-00683-3>
- [4] Peerapornratana S, Manrique-Caballero CL, Gómez H, & Kellum JA. (2019). Acute kidney injury from sepsis: current concepts, epidemiology, pathophysiology, prevention and treatment. *Kidney Int*, 96(5), 1083–1099. <https://doi.org/10.1016/j.kint.2019.05.026>
- [5] Evans L, Rhodes A, Alhazzani W, Antonelli M, Coopersmith CM, French C, et al. (2021). Surviving sepsis campaign: international guidelines for management of sepsis and septic shock 2021. *Intensive Care Med*, 47(11), 1181–1247. <https://doi.org/10.1007/s00134-021-06506-y>
- [6] Font MD, Thyagarajan B, & Khanna AK. (2020). Sepsis and Septic Shock - Basics of diagnosis, pathophysiology and clinical decision making. *Med Clin North Am*, 104(4), 573–585. <https://doi.org/10.1016/j.mcna.2020.02.011>
- [7] Monnet X, Lai C, & Teboul JL. (2023). How I personalize fluid therapy in septic shock? *Crit Care*, 27(1), 123. <https://doi.org/10.1186/s13054-023-04363-3>
- [8] Mok G, Hendin A, Reardon P, Hickey M, Gray S, & Yadav K. (2021). Macrocirculatory and Microcirculatory Endpoints in Sepsis Resuscitation. *J Intensive Care Med*, 36(12), 1385–1391. <https://doi.org/10.1177/0885066620982585>
- [9] Frenkel A, Abuhasira R, Fisher L, Bichovsky Y, Zlotnik A, Novack V, et al. (2023). Vasopressors and Mean Arterial Pressure in Septic Shock: Do We Bend the Rules for Young Adults? *Isr Med Assoc J*, 25(11), 741–746.
- [10] Belletti A, Leroise CC, Zangrillo A, & Landoni G. (2021). Vasoactive-Inotropic Score: Evolution, Clinical Utility, and Pitfalls. *J Cardiothorac Vasc Anesth*, 35(10), 3067–3077. <https://doi.org/10.1053/j.jvca.2020.09.117>
- [11] Chen Y, Jiang H, Wei Y, Qiu Y, Su L, Chen J, et al. (2024). Blood pressure response index and clinical outcomes in patients with septic shock: a multicenter cohort study. *EBioMedicine*, 106, 105257. <https://doi.org/10.1016/j.ebiom.2024.105257>
- [12] Johnson AEW, Bulgarelli L, Shen L, Gayles A, Shammout A, Horng S, et al. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Sci Data*, 10(1), 1. <https://doi.org/10.1038/s41597-022-01899-x>
- [13] Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *Jama*, 315(8), 801–810. <https://doi.org/10.1001/jama.2016.0287>
- [14] Han J, Pinsino A, Sanchez J, Takayama H, Garan AR, Topkara VK, et al. (2019). Prognostic value of vasoactive-inotropic score following continuous flow left ventricular assist device implantation. *J Heart Lung Transplant*, 38(9), 930–938. <https://doi.org/10.1016/j.healun.2019.05.007>
- [15] KDIGO 2021 Clinical Practice Guideline for the Management of Glomerular Diseases. (2021). *Kidney Int*, 100(4s), S1–s276. <https://doi.org/10.1016/j.kint.2021.05.021>
- [16] Eckardt KU, Bernhardt WM, Weidemann A, Warnecke C, Rosenberger C, Wiesener MS, et al. (2005). Role of hypoxia in the pathogenesis of renal disease. *Kidney Int Suppl*, 10.1111/j.1523-1755.2005.09909.x(99), S46–51. <https://doi.org/10.1111/j.1523-1755.2005.09909.x>
- [17] Nath KA, & Norby SM. (2000). Reactive oxygen species and acute renal failure. *Am J Med*, 109(8), 665–678. [https://doi.org/10.1016/s0002-9343\(00\)00612-4](https://doi.org/10.1016/s0002-9343(00)00612-4)
- [18] Liu J, Xie H, Ye Z, Li F, & Wang L. (2020). Rates, predictors, and mortality of sepsis-associated acute kidney injury: a systematic review and meta-analysis. *BMC Nephrol*, 21(1), 318. <https://doi.org/10.1186/s12882-020-01974-8>
- [19] White KC, Serpa-Neto A, Hurford R, Clement P, Laupland KB, See E, et al. (2023). Sepsis-associated acute kidney injury in the intensive care unit: incidence, patient characteristics, timing, trajectory, treatment, and associated outcomes. A multicenter, observational study. *Intensive Care Med*, 49(9), 1079–1089. <https://doi.org/10.1007/s00134-023-07138-0>
- [20] Badin J, Boulain T, Ehrmann S, Skarzynski M, Bretagnol A, Buret J, et al. (2011). Relation between mean arterial pressure and renal function in the early phase of shock: a prospective, explorative cohort study. *Crit Care*, 15(3), R135. <https://doi.org/10.1186/cc10253>
- [21] Wang M, Wang X, Zhu B, Li W, Jiang Q, Zuo Y, et al. (2023). The effects of timing onset and progression of AKI on the clinical outcomes in AKI patients with sepsis: a prospective multicenter cohort study. *Ren Fail*, 45(1), 1–10. <https://doi.org/10.1080/0886022x.2022.2138433>
- [22] Annane D, Ouanes-Besbes L, de Backer D, Du B, Gordon AC, Hernández G, et al. (2018). A global perspective on vasoactive agents in shock. *Intensive Care Med*, 44(6), 833–846. <https://doi.org/10.1007/s00134-018-5242-5>
- [23] Priyanka P, Chang CH, Chawla LS, Kellum JA, Clermont G, & Murugan R. (2022). VASOPRESSOR-RESISTANT HYPOTENSION, COMBINATION VASOPRESSOR THERAPY, AND SHOCK PHENOTYPES IN CRITICALLY ILL ADULTS WITH VASODILATORY SHOCK. *Shock*, 58(4), 260–268. <https://doi.org/10.1097/shk.0000000000001980>
- [24] Hou K, Chen Q, Zhu X, Shen X, Zou L, Mu X, et al. (2021). Correlation Between Vasoactive-Inotropic Score and Postoperative Acute Kidney Injury after Cardiovascular Surgery. *Heart Surg Forum*, 24(2), E282–e292. <https://doi.org/10.1532/hcf.3537>
- [25] Hoste EA, Damen J, Vanholder RC, Lameire NH, Delanghe JR, Van den Hauwe K, et al. (2005). Assessment of renal function in recently admitted critically ill patients with normal serum creatinine. *Nephrol Dial Transplant*, 20(4), 747–753. <https://doi.org/10.1093/ndt/gfh707>
- [26] Makris K, & Spanou L. (2016). Acute Kidney Injury: Diagnostic Approaches and Controversies. *Clin Biochem Rev*, 37(4), 153–175.
- [27] Prowle JR, Echeverri JE, Ligabo EV, Ronco C, & Bellomo RJNRN. (2010). Fluid balance and acute kidney injury. 6(2), 107–115.
- [28] Harrois A, Grillot N, Figueiredo S, & Duranteau J. (2018). Acute kidney injury is associated with a decrease in cortical renal perfusion during septic shock. *Crit Care*, 22(1), 161. <https://doi.org/10.1186/s13054-018-2067-0>
- [29] Le Dorze M, Bouglé A, Deruddre S, & Duranteau JJS. (2012). Renal Doppler ultrasound: a new tool to assess renal perfusion in critical illness. 37(4), 360–365.
- [30] Malik J, Lomonte C, Meola M, de Bont C, Shahverdyan R, Rotmans JI, et al. (2021). The role of Doppler ultrasonography in vascular access surveillance-controversies continue. *J Vasc Access*, 22(1_suppl), 63–70. <https://doi.org/10.1177/0885066620982585>

- org/10.1177/1129729820928174
- [31] Koratala A, & Reisinger N. (2022). Venous Excess Doppler Ultrasound for the Nephrologist: Pearls and Pitfalls. *Kidney Med*, 4(7), 100482. <https://doi.org/10.1016/j.xkme.2022.100482>

A Unique Case of Migraine Without Aura Triggered by Palpation of V2-innervated Chronic Facial Folliculitis

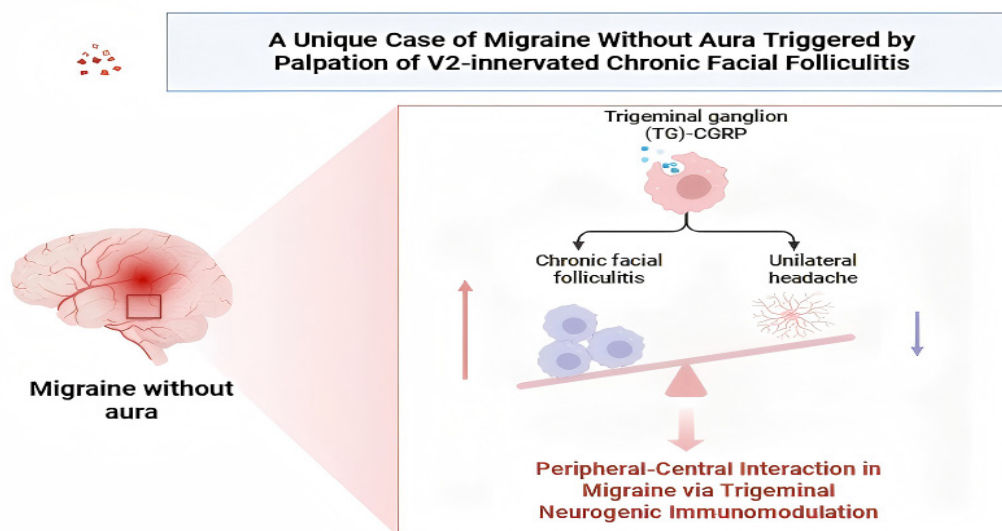
Authors

Dan Xiao, Yu Wang

Correspondence

yfy126330@fy.ahmu.edu.cn (Y. Wang)

Graphical Abstract



<https://doi.org/10.71321/jynd6h07>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Unique Case of Migraine Without Aura Triggered by Palpation of V2-innervated Chronic Facial Folliculitis

Dan Xiao¹, Yu Wang^{1*}

Received: 2026-03-09 | Accepted: 2026-04-19 | Published online: 2026-05-28

Abstract

Accumulating research suggests that peripheral pathological changes participate in the pathogenesis of migraine, and supporting evidence has been reported in many clinical cases. However, few cases have directly demonstrated the bidirectional interaction between migraine and peripheral pathological lesions. Here, we present a case of severe migraine without aura (MwoA) whose attacks were elicited by palpation of chronic folliculitis lesions around the unilateral nasal ala. This unique case reveals a distinct peripheral-central interaction in migraine, which may be mediated by neurogenic immunomodulation of the trigeminal system.

Keywords: Facial folliculitis; Migraine without aura; Trigeminal nerve; Neurogenic immunomodulation; Peripheral trigger; CGRP

Introduction

It has long been established that migraine patients frequently exhibit tenderness in multiple pericranial regions, including the shoulder, neck, sternocleidomastoid muscle, anterior temporal area and occipital insertion sites. A large proportion of these tender regions act as classic trigger points that can induce migraine attacks, which strongly suggests that peripheral pathological mechanisms play a non-negligible role in the onset and recurrence of migraine [1]. As a common otogenic inflammatory disease, petrous apicitis is often accompanied by headache as a prominent clinical manifestation, and several clinical case reports have pointed out that the headache phenotype induced by petrous apicitis is highly consistent with the characteristics of migraine in some patients [2-3]. Recently, Mancini et al. reported a rare case where a patient with petrous apicitis developed de novo migraine without aura (MwoA), and the migraine attacks achieved sustained remission after the patient received mastoidectomy combined with a long-course antibiotic therapy targeting the primary petrous apicitis [4]. This finding further confirms that inflammatory lesions of the petrous apex can cause pathological changes in the trigeminal nerve fibers innervating the area, which in turn initiates and maintains the recurrent attacks of migraine.

On the basis of previous research, we report a clinical case of

severe MwoA triggered by palpation of chronic facial folliculitis around the unilateral nasal ala, a peripheral lesion that was refractory to conventional antibiotic therapy but completely resolved after effective migraine prophylactic treatment. This case not only provides new clinical evidence for the peripheral trigger mechanism of migraine, but also for the first time reveals that the central neural modulation of migraine can in turn regulate the progression of peripheral inflammatory lesions. It further enriches the understanding of the peripheral-central interaction mechanism in migraine and provides a novel perspective for exploring the cross-regulation between the trigeminal nervous system and peripheral immune inflammation.

Case Presentation

A 45-year-old male presented to our department with an 8-month history of severe headaches, which were solely induced by facial contact activities such as face washing and shaving that involved the facial folliculitis lesions around his left nasal ala. Each episode of contact or palpation of this affected area was capable of precipitating an acute severe headache attack. Preceding the onset of the severe headache, an abrupt, transient sharp stabbing pain would radiate instantaneously from the left nasal ala to the surrounding regions of

¹ Department of Neurology, Epilepsy and Headache Group, the First Affiliated Hospital of Anhui Medical University, Jixi Road 218, Hefei 230022, China

* Corresponding Author.

the left face. The headache was strictly unilateral on the left side, localizing to the left frontal, temporal and parietal lobes, and occasionally extending to the entire left cerebral hemisphere. Characteristically, the headache manifested as throbbing and pulsating pain, with occasional stabbing sensations, and had an average intensity of 8 to 10 on a 10-point numerical rating scale. Associated symptoms included nausea (with or without vomiting), photophobia, phonophobia and irritability. The headache was exacerbated by head movement, ambulation and physical exertion, while mild relief was achieved with recumbency and rest. Each headache attack typically persisted for half a day or resolved spontaneously following an overnight sleep.

On admission to our department, neurological and physical examinations yielded unremarkable findings, and brain magnetic resonance imaging (MRI) scans were normal. Routine blood tests, as well as additional investigations including erythrocyte sedimentation rate and screening for sexually transmitted diseases, all fell within normal reference ranges. Psychological assessments for anxiety and depression were performed during a headache-free interval, revealing a Hamilton Anxiety Rating Scale (HAMA) score of 10 (a score ≥ 14 indicates mild anxiety) and a Hamilton Depression Rating Scale (HAM-D) score of 8 (a score > 7 indicates mild depression). Dermatological examination of the skin around the left nasal ala identified small pustular lesions and several keloids with varying degrees of pigmentation, which the patient reported were sequelae of previous pustular lesions (Figure 1). Based on the clinical features of the headache and the normal results of all auxiliary examinations except for the local cutaneous findings, a diagnosis of migraine was made. The headache attacks conformed to the diagnostic criteria for migraine without aura (MwoA) as defined by the International Classification of Headache Disorders, 3rd edition (ICHD-3), with the only exception being the specific peripheral trigger factor.



Figure 1. Clinical manifestations of chronic facial folliculitis adjacent to the left nosewing. The skin over the left nosewing region exhibited small purulent papules (white arrows) and multiple keloidal lesions with varying degrees of pigmentation (black arrows), which the patient described as sequelae of prior purulent folliculitis. These lesions were restricted to the cutaneous area innervated by the second division (V2) of the trigeminal nerve and served as the sole trigger for migraine without aura (MwoA) attacks in this patient.

The patient was initiated on prophylactic treatment with sodium valproate (500 mg twice daily) and discontinued all topical antibiotic ointments for 2 months. During the first week of treatment, palpation of the pustular lesions still triggered migraine attacks, but pain intensity and duration were markedly reduced. One week after starting sodium valproate, the pustular lesions around the left nasal ala resolved completely, and palpation or contact stimuli such as face washing no longer induced headaches. During the subsequent 2 months of treatment and 6-month follow-up, the patient remained free of

migraine attacks and recurrent pustular folliculitis.

Ten months after initial sodium valproate initiation (and 8 months after drug discontinuation), pustular lesions relapsed around the left nasal ala, and their palpation again triggered migraine attacks identical to the original presentation. Each headache typically lasted 12 hours or resolved overnight after sleep. On one occasion, topical lidocaine hydrochloride gel applied to the pustules reduced headache intensity after attack onset; repeated application prevented palpation-triggered headaches, but therapy was stopped due to its short duration of action (only 2-6 hours). The patient was restarted on sodium valproate at 500 mg twice daily. Five days after resumption, pustular lesions resolved fully, and no headaches were triggered by skin palpation or face washing. The patient has remained on continuous sodium valproate with sustained remission up to manuscript submission.

Written informed consent was obtained from the patient for the publication of this case report, and all identifying information of the patient has been anonymized to protect privacy.

Discussion

Migraine attacks can be triggered by multiple factors via trigeminal nerve (TN) or non-trigeminal pathways [5-7]. Although trigeminal nerve block is effective for migraine, direct TN stimulation as an independent trigger of migraine attacks has rarely been reported. In this case, severe migraine without aura was exclusively elicited by palpation of pustular folliculitis in the V2 distribution in a patient with no prior headache history, providing direct clinical evidence that peripheral TN stimulation can initiate migraine under sensitized conditions. Psychiatric factors were excluded by normal anxiety and depression scores, and contralateral symmetrical skin palpation did not induce headaches. Topical lidocaine hydrochloride gel transiently blocked palpation-induced headaches, supporting the key role of peripheral trigeminal terminal activation in migraine attack initiation [8].

Recent studies have demonstrated microstructural changes and myelin sheath disruption in the TN of migraine patients [9], suggesting that migraine may induce functional and structural abnormalities in the TN [10]. Notably, our case is the first to show that antibiotic-refractory chronic facial folliculitis in the V2 territory completely resolved after migraine prophylaxis, indicating that central migraine-related neural activity can regulate peripheral cutaneous immune responses through the trigeminal nerve. The trigeminal nerve acts as a critical bridge for neuro-immune crosstalk, and its neuropeptides mediate peripheral immune regulation and inflammation [11-12]. Folliculitis is mainly caused by *Staphylococcus aureus* and *Streptococcus pyogenes* [13]. Bacterial toxins activate nociceptors, followed by calcitonin gene-related peptide (CGRP) release, which inhibits neutrophil phagocytosis and prolongs infection [14]. Meanwhile, CGRP from the trigeminal ganglion (TG) maintains peripheral and central sensitization, increasing migraine susceptibility [15]. TG-derived CGRP also modulates macrophage polarization and exerts local immunomodulatory effects [12]. We therefore propose a vicious cycle: nasal ala folliculitis activates TG neurons and increases CGRP release, leading to trigeminal and central sensitization that enables palpation to trigger migraine. Concurrently, CGRP suppresses

local antibacterial immunity, worsening chronic infection. This bidirectional interaction between peripheral inflammation and central migraine pathogenesis highlights a novel peripheral-central mechanism mediated by CGRP-related neurogenic immunomodulation.

Conclusion

This case observation demonstrates a bidirectional regulatory relationship between migraine attacks and chronic cutaneous infection in the trigeminal nerve (TN) innervated facial region. Moreover, this peripheral-central interaction in migraine pathogenesis is likely mediated by calcitonin gene-related peptide (CGRP) secreted from trigeminal ganglion (TG) neurons.

Abbreviations

Migraine Without Aura: MwoA; Second branch of trigeminal nerve: V2; Calcitonin Gene-Related Peptide: CGRP; Magnetic Resonance Imaging: MRI; Hamilton Anxiety Rating Scale: HAMA; Hamilton Depression Rating Scale: HAM-D; International Classification of Headache Disorders, 3rd edition: ICHD-3; Trigeminal Nerve: TN; Trigeminal Ganglion: TG.

Author Contributions

Conception YW, DX. Acquisition of data YW, DX. Data interpretation YW, DX. Manuscript draft DX. Revision for intellectual content YW. All Authors read and approved the manuscript.

Acknowledgements

We acknowledge the patient for his patience in the following up.

Funding Information

This work was supported by Natural Science grants to Yu Wang (Grant No. 81671290, 82071460) from the National Natural Science Foundation of China.

Ethics Approval and Consent to Participate

Not Applicable.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

All data needed to evaluate the conclusions in the paper are present in the paper or the Supplementary Materials. Additional data related to this paper may be requested from the authors.

References

- [1] DO T P, HELDARSKARD G F, KOLDING L T, HVEDSTRUP J, SCHYTZ H W. Myofascial trigger points in migraine and tension-type headache [J]. *The journal of headache and pain*, 2018, 19(1): 84.<http://doi.org/10.1186/s10194-018-0913-8>.
- [2] PLODPAI Y, HIRUNPAT S, KIDDEE W. Gradenigo's syndrome secondary to chronic otitis media on a background of previous radical mastoidectomy: a case report [J]. *Journal of medical case reports*, 2014, 8: 217.<http://doi.org/10.1186/1752-1947-8-217>.
- [3] ISERSON K V. The ethics of emergency medicine [J]. *The Journal of emergency medicine*, 1985, 3(2): 161-2.[http://doi.org/10.1016/0736-4679\(85\)90050-2](http://doi.org/10.1016/0736-4679(85)90050-2).
- [4] MANCINI A J, GLASSMAN R D, CHANG Y M, BURSTEIN R, ASHINA S. Headache in Petrous Apicitis: A Case Report of Chronic Migraine-like Headache Due to Peripheral Pathology [J]. *Headache*, 2019, 59(10): 1821-6.<http://doi.org/10.1111/head.13643>.
- [5] TAI M S, YET S X E, LIM T C, POW Z Y, GOH C B. Geographical Differences in Trigger Factors of Tension-Type Headaches and Migraines [J]. *Current pain and headache reports*, 2019, 23(2): 12.<http://doi.org/10.1007/s11916-019-0760-6>.
- [6] SILVA-NÉTO R P, PERES M F, VALENÇA M M. Odorant substances that trigger headaches in migraine patients [J]. *Cephalalgia : an international journal of headache*, 2014, 34(1): 14-21.<http://doi.org/10.1177/0333102413495969>.
- [7] SILVA-NÉTO R P, RODRIGUES Â B, CAVALCANTE D C, FERREIRA P H, NASI E P, SOUSA K M, et al. May headache triggered by odors be regarded as a differentiating factor between migraine and other primary headaches? [J]. *Cephalalgia : an international journal of headache*, 2017, 37(1): 20-8.<http://doi.org/10.1177/0333102416636098>.
- [8] RAM S, KUMAR S K, CLARK G T. Using oral medications, infusions and injections for differential diagnosis of orofacial pain [J]. *Journal of the California Dental Association*, 2006, 34(8): 645-54.
- [9] GUYURON B, YOHANNES E, MILLER R, CHIM H, REED D, CHANCE M R. Electron microscopic and proteomic comparison of terminal branches of the trigeminal nerve in patients with and without migraine headaches [J]. *Plastic and reconstructive surgery*, 2014, 134(5): 796e-805e.<http://doi.org/10.1097/prs.0000000000000696>.
- [10] ZIEGLER C, MAY A. Facial presentations of migraine, TACs, and other paroxysmal facial pain syndromes [J]. *Neurology*, 2019, 93(12): e1138-e47.<http://doi.org/10.1212/wnl.00000000000008124>.
- [11] BARON R, BINDER A, WASNER G. Neuropathic pain: diagnosis, pathophysiological mechanisms, and treatment [J]. *The Lancet Neurology*, 2010, 9(8): 807-19.[http://doi.org/10.1016/S1473-3099\(10\)70111-1](http://doi.org/10.1016/S1473-3099(10)70111-1).

- org/10.1016/s1474-4422(10)70143-5.
- [12] TAO J, WANG X, XU J. Expression of CGRP in the Trigeminal Ganglion and Its Effect on the Polarization of Macrophages in Rats with Temporomandibular Arthritis [J]. Cellular and molecular neurobiology, 2024, 44(1): 22.<http://doi.org/10.1007/s10571-024-01456-7>.
- [13] JACOBS M R, JONES R N, GIORDANO P A. Oral beta-lactams applied to uncomplicated infections of skin and skin structures [J]. Diagnostic microbiology and infectious disease, 2007, 57(3 Suppl): 55s-65s.<http://doi.org/10.1016/j.diagmicrobio.2006.11.020>.
- [14] PINHO-RIBEIRO F A, BADDAL B, HAARSMA R, O'SEAGHDHA M, YANG N J, BLAKE K J, et al. Blocking Neuronal Signaling to Immune Cells Treats Streptococcal Invasive Infection [J]. Cell, 2018, 173(5): 1083-97.e22.<http://doi.org/10.1016/j.cell.2018.04.006>.
- [15] IYENGAR S, JOHNSON K W, OSSIPOV M H, AURORA S K. CGRP and the Trigeminal System in Migraine [J]. Headache, 2019, 59(5): 659-81.<http://doi.org/10.1111/head.13529>.

Enhancing Early Sepsis Prediction with Temporal bias EHRs Data

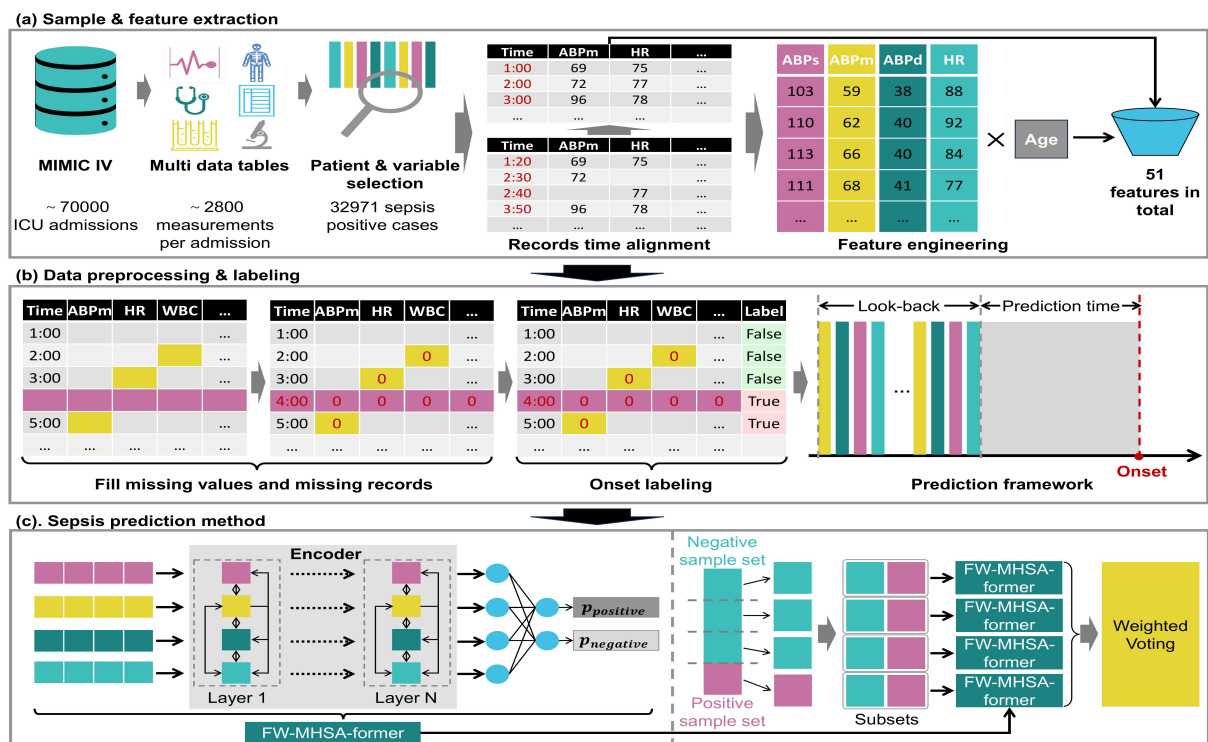
Authors

Mireaye Abudurexiti, Jianshu Wang, Pengfei Zhang, Fangfang Liu, Zhiqiang Jia

Correspondence

12664494@qq.com (Z. Jia)

Graphical Abstract



<https://doi.org/10.71321/4t9gsy05>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Enhancing Early Sepsis Prediction with Temporal bias EHRs Data

Mireaye Abudurexiti^{1†}, Jianshu Wang^{2†}, Pengfei Zhang¹, Fangfang Liu¹, Zhiqiang Jia^{1*}

Received: 2026-03-10 | Accepted: 2026-05-18 | Published online: 2026-05-29

Abstract

Background: Sepsis prediction models using electronic health records (EHRs) are often challenged by temporal biases from irregular data entry and severe class imbalance. This study develops a novel deep learning (DL) framework to address these specific challenges for accurate and early sepsis detection.

Methods: We propose a Feature-Wise Multi-Head Self-Attention Transformer (FW-MHSA-former) with an Adaptive Balance-Preserving Ensemble (ABPE). FW-MHSA-former mitigates temporal bias by applying self-attention across medical features to model correlations directly. ABPE resolves class imbalance by partitioning the majority class to train multiple models on balanced datasets, aggregating predictions via weighted voting. The framework was retrospectively validated on the MIMIC-IV dataset; Kaplan-Meier analysis assessed survival outcomes. Visualizing the attention-derived feature correlation matrix enhances interpretability.

Results: The proposed framework achieved a peak Area Under the Receiver Operating Characteristic Curve (AUROC) of 0.94. At this optimal performance, the model demonstrated a recall of 0.90 and an accuracy of 0.87. We compared our method with three classical models and three advanced attention/Transformer-based models. The proposed approach yielded consistently superior performance across all evaluation metrics, including accuracy and F1 score. The KM analysis confirmed that the model effectively stratified patients into high- and low-risk cohorts with statistically significant differences in survival outcomes ($p < 0.001$).

Conclusions: The proposed framework effectively and robustly predicts early sepsis. By addressing timestamp irregularities and class imbalance, it achieves superior accuracy and provides an interpretable tool to enhance clinical decision support in critical care.

Keywords: Sepsis prediction; Transformer; Electronic health records; Deep learning; Interpretable model

Introduction

Sepsis remains a major global health challenge, characterized by high incidence and mortality rates [1-3]. A substantial body of research has established that early and accurate detection is critical for improving patient outcomes [4]. However, the clinical presentation of sepsis often overlaps with other conditions, presenting a diagnostic challenge even for experienced clinicians. Moreover, this challenge is compounded by the underlying immune dysregulation characteristic of sepsis, involving pathways such as cGAS-STING that drive inflammation and organ injury [5]. In response, various screening and predictive scoring systems have been developed, with traditional tools like the Systemic Inflammatory Response Syndrome (SIRS) criteria, Sequential Organ Failure Assessment (SOFA), and quick SOFA (qSOFA) being widely implemented. Despite their utility, these systems exhibit several intrinsic limitations. They pri-

marily rely on static physiological and laboratory parameters, inadequately capturing the complex, non-linear relationships within patient data. Furthermore, their dependence on single-point-in-time measurements fails to account for the dynamic evolution of a patient's clinical state [6-8]. These conventional methods also frequently impose stringent sample inclusion criteria, which can limit their generalizability across diverse patient populations.

The widespread adoption of electronic health records (EHRs) has created unprecedented opportunities to apply advanced computational algorithms to enhance clinical decision-making. Consequently, numerous studies have leveraged EHRs data to improve sepsis-related outcomes [9-14]. Machine learning (ML) techniques, in particular, allow for the identification of latent patterns in complex patient data, offering valuable guidance for diagnosis and treatment [15-16]. Early ML applications in sepsis research often employed multivariate models like Cox

¹ The First People's Hospital of Kashi, Kashi 844000, Xinjiang Uygur Autonomous Region, China

² Southwest Petroleum University, Chengdu 610599, China

† These authors contributed equally to this work.

* Corresponding Author.

proportional-hazards and logistic regression (LR). For instance, Oami et al. [17] utilized these models to demonstrate that intensive care unit (ICU) admission significantly reduced mortality risk in sepsis patients. Similarly, Zohar et al. [18] identified hyperglycemia as an independent predictor of poor outcomes. While powerful, these statistical models are limited in their ability to handle high-dimensional, dynamic features, especially when the temporal evolution of these features is critical. To address this, methods capable of modeling temporal dynamics have gained prominence, with recurrent neural networks (RNNs) and their variants emerging as a leading approach.

Recent research has produced various sophisticated deep learning (DL) models for sepsis prediction. Shashikumar et al. [10] proposed DeepAISE, which uses a gated recurrent unit (GRU) to process time series inputs. Das et al. [19] introduced an architecture that combines an attention mechanism with a bidirectional long short term memory (Bi-LSTM) and a convolutional neural network (CNN). Other notable examples include DEAR [20], which enhances a GRU base with a two-layer attention mechanism, and SSP [21], which pairs a LSTM network with a CNN. Although these models have advanced predictive performance, they face challenges. First, they are susceptible to manual time-stamping biases in EHRs. Our analysis of the MIMIC-IV dataset identified over 160,000 laboratory records with identical specimen collection and result output times, which is inconsistent with facts. Secondly, the inherent class imbalance in sepsis prediction tasks often impacts model training and optimization. Finally, "black box" nature of many DL models hinders understanding their decision-making logic, thus impacting their clinical credibility.

To address these limitations, we propose the Feature-Wise Multi-Head Self-Attention Transformer (FW-MHSA-former) for early sepsis prediction and an Adaptive Balance-Preserving Ensemble (ABPE) training strategy. The FW-MHSA-former eschews conventional recurrent processing, instead embedding medical features as tokens and encoding temporal relationships within its attention mechanism. This design enhances robustness against the temporal inconsistencies and recording inaccuracies prevalent in EHRs data. To combat data imbalance, the ABPE method automatically partitions the majority (negative) class into multiple subsets. Each subset is then combined with the entire minority (positive) class to form a balanced sub-dataset for training an independent model instance. This ensemble approach mitigates the effects of class imbalance while preserving the statistical integrity of the original dataset. Finally, to enhance model interpretability, we visualize the attention-derived correlation matrix from the FW-MHSA-former, providing clinicians with medically plausible insights into the model's predictions.

The main contributions of this study are as follows:

- We propose the FW-MHSA-former, a novel architecture for early sepsis prediction. By mapping temporal information into its hidden dimensions, the model mitigates the impact of temporal irregularities and timestamp inaccuracies common in EHRs data while effectively extracting informative representations.
- We propose ABPE, an ensemble training strategy that constructs multiple balanced training sub-datasets. By training independent models on these datasets and aggregating their predictions via a weighted vote, this approach effectively addresses class imbalance without discarding

data.

- We conduct a retrospective study on the large-scale MIMIC-IV dataset, demonstrating that our proposed method outperforms both classic ML and DL approaches in sepsis prediction, thereby validating its effectiveness and potential for clinical application.

Methods

Dataset Description

The data for this study were sourced from the Medical Information Mart for Intensive Care (MIMIC)-IV database (version 2.2) [22]. MIMIC-IV is a large-scale, publicly available, and de-identified database developed through a collaboration between the Massachusetts Institute of Technology (MIT) and Beth Israel Deaconess Medical Center (BIDMC). It contains comprehensive clinical records from patients admitted to the ICU at BIDMC between 2008 and 2019.

Our model utilizes a curated set of features extracted from this database, including demographic data, vital signs, and an extensive panel of laboratory measurements. In addition, we incorporated several engineered features, such as interaction terms between age and blood pressure. From a physiological perspective, arterial blood pressure (ABP) is a core marker of circulatory dysfunction and a key feature of sepsis progression in the Sepsis-3 diagnostic criteria [23]. The prognostic value of blood pressure abnormalities is age-dependent: elderly patients have impaired vascular elasticity, weakened autonomic compensatory capacity, and reduced hemodynamic reserve in response to infection-induced vasodilation and circulatory collapse. Identical ABP values correspond to different sepsis risk across age groups, and this non-linear synergistic effect cannot be effectively captured by age or blood pressure alone. The inclusion of these engineered features is supported by prior research demonstrating their utility in improving sepsis prediction models [24-26]. A complete list of the features used as model inputs is provided in Table 1.

Table 1. List of demographic, vital signs, laboratory measurements, and engineered features used in the sepsis prediction model.

Category	Features
Demographic	Age
Vital signs	ABPm, ABPs, ABPD, HR, Temp, O ₂ , Sat, RR, FiO ₂
Laboratory	WBC, INR, PT, BUN, Ven pH, Art pH, Creat, Ven CO ₂ , Art CO ₂ , Ven O ₂ , Art O ₂ Sat, AG, Hgb, Hct, PTT, AST, ALP, Glu, K, Mg, CL, Trop T, Fib, NH ₃ , HCO ₃ , Plt, Lactate, WB Cl, BG
Engineered	ABPm × Age, ABPs × Age, ABPD × Age, HR × Age, P/F Ratio

Definition of Sepsis Onset and Study Population

The identification of sepsis in this study followed the most recent Sepsis-3 criteria [27]. A case was labeled as sepsis when both a confirmed or suspected infection and a SOFA score increase of at least two points were observed.

In this study, each instance of ICU care was treated as a dis-

tinct sample, given that patients may have multiple ICU admissions. The samples were selected according to the following criteria:

- The patient was over the age of 18;
- The patient is not septic within the initial look-back hours following ICU admission but becomes septic at the hour of look-back + prediction time, or the patient never develops sepsis during their ICU stay;
- The number of missing records is no more than half of the look-back;
- The rate of missing values for the ABPm feature did not exceed 95%. ABPm was selected as the anchor for data quality control because it is a core indicator of circulatory function and organ perfusion, and hypotension is a key component of the Sepsis-3 definition for sepsis and septic shock. A missing rate over 95% indicates nearly no valid hemodynamic records in the observation window, which would introduce severe noise and reduce model reliability.

Compared to other studies, the inclusion criteria of this study are more tolerant of data sparsity and have no direct requirements on the frequency of feature measurement, making our work more applicable and conducive to promoting clinical application.

In the context above, the look-back refers to the time interval from which data is collected prior to the current moment, while the prediction time denotes the future interval following the current moment for which the model makes its prediction. The look-back and prediction time are both user-defined parameters that can be adjusted according to the specific requirements of the study.

Data Preprocessing

The requisite data were extracted from the MIMIC-IV dataset and subsequently adjusted to an hourly resolution by resetting minute and second fields to zero in each record. Admittedly, this hourly aggregation may smooth subtle short-term physiological variations. To further verify model robustness, we supplemented a sensitivity analysis at a 30-minute temporal resolution for comparison.

Following temporal resampling, outlier treatment was conducted. Any values less than zero were set to zero, and for each medical feature, any values exceeding three standard deviations from the mean were also set to zero. Additionally, cases involving standard deviation were handled as follows: if the standard deviation exceeded twice the mean, values greater than the mean were set to zero; if the standard deviation was between one-tenth and twice the mean, values outside five standard deviations from the mean were set to zero; and if the standard deviation was less than one-tenth of the mean, no additional processing was applied. Instead of utilizing non-original data imputation techniques, the missing data were preserved by replacing them with a value of zero. To evaluate imputation effects and justify zero-filling, we compared it against forward-filling, k-NN, and mean imputation. Using identical model settings, data splits, and metrics, we isolated the imputation strategy as the sole variable.

Negative samples spanned the full ICU stay, while positive samples included records from ICU admission up to the pre-sepsis prediction window. Only samples with a time span exceeding the sum of the look-back window and the prediction time window were retained. In the event that the records within

the look-back period did not encompass the entire range of hours, missing records were populated with zeros. Ultimately, samples exhibiting over 95% missing values for ABPm were excluded. The data preprocessing steps are summarized in Figure 1.

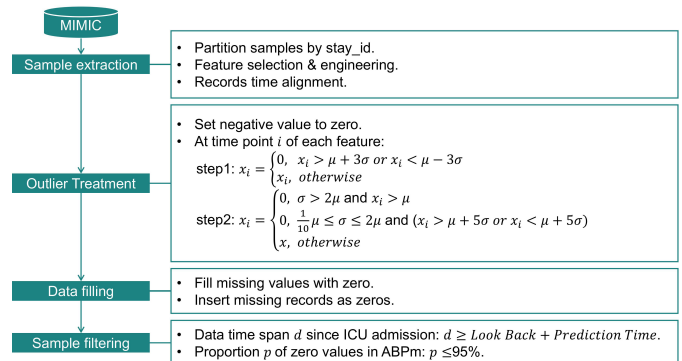


Figure 1. Data preprocessing steps for the MIMIC-IV dataset. The original data is first aligned to the hourly level, followed by outlier treatment and missing value handling.

Feature-Wise Multi-Head Self-Attention Transformer (FW-MHSA-former)

In the canonical Transformer architecture, the encoder is responsible for converting the input sequence into high-dimensional representations, while the decoder is tasked with generating the target sequence [28]. Given that sepsis prediction is a classification task rather than a sequence generation task, we have elected to employ an encoder-only Transformer architecture. The inherent manual time-stamping biases in EHRs, when combined with the alignment of records to the hour-level during preprocessing, result in the occurrence of timestamp shifts. Such discrepancies in recorded time undermine the reliability of time-step-based analysis. Therefore, In contrast to most Transformer-based models designed for time-series analysis, our approach tokenizes medical features instead of individual time points, thereby encapsulating the temporal dimension as a latent variable. This perspective enables a feature-centric representation while implicitly modeling temporal dependencies. In this configuration, each medical feature is regarded as an independent token, shifting the model's focus from cross-temporal correlations to cross-feature correlations. This mitigates the direct adverse effects of timestamp biases. Furthermore, cross-feature correlation analysis allows for the identification of interrelationships among medical features within the model's decision-making process and the assessment of each feature's influence on the model's predictions. This interpretability is intrinsic to the model and does not rely on external interpretation methods such as SHapley Additive exPlanations (SHAP), thereby providing a more direct explanatory basis for clinical decision-making.

This chapter provides a detailed introduction to the structure of the FW-MHSA-former. Figure 2 illustrates the overall structure of the proposed model.

Embedding

Embedding is a technique that maps high-dimensional or symbolic data (such as text or categorical information) into a lower-dimensional vector space [29]. This approach is widely used in fields such as Natural Language Processing (NLP)

and recommendation systems. In this study, we employ an embedding layer to map the input medical features into a continuous high-dimensional space. Specifically, the embedding layer transforms the temporal dimension, determined by the look-back size, into a higher-dimensional space with d_{model} . This transformation enables subsequent network layers to capture correlations among medical features in this enriched, high-dimensional space. In this study, the embedding operation is implemented through linear projection. Given the input data $X = x_1, \dots, x_t \in \mathbf{R}^{(T \times N)}$, where T represents the historical observations over Y hours, and N is the number of medical features. The embedding layer maps each feature into a d_{model} -dimensional space. The transformation process can be expressed as follows:

$$H^0 = \text{Dropout}(XW_E + b_E) \in \mathbf{R}^{T \times d_{\text{model}}} \quad (1)$$

where $W_E \in \mathbf{R}^{N \times d_{\text{model}}}$ is the weight matrix of the embedding layer, $b_E \in \mathbf{R}^{d_{\text{model}}}$ is the bias vector, and $H^0 \in \mathbf{R}^{T \times d_{\text{model}}}$ is the output of the embedding layer. To prevent model overfitting, the embedded result is processed through a dropout layer, where a certain percentage of neurons is randomly dropped with a fixed probability.

Encoder

The canonical Transformer architecture employs an encoder-decoder structure. However, ongoing research has introduced both encoder-only and decoder-only architectures. The decoder-only architecture generates outputs in an autoregressive manner, rendering it well-suited for generative tasks such as text generation [30-31] and speech synthesis [32-33]. In contrast, the encoder-only architecture processes the entire input sequence in parallel, thereby enabling faster training and inference compared to decoder-only architectures. It is particularly well-suited for tasks such as classification [34-35], feature extraction [36-37], and information retrieval [38-39].

Sepsis prediction is essentially a binary classification task, where the objective is to determine whether a patient is likely to develop sepsis based on the input sequence. Consequently,

we have adopted an encoder-only architecture in our study. This approach is in accordance with the latest advancements in computer vision (CV) [40-42] and time series analysis (TSA) [43-44], where encoder-only architectures have demonstrated substantial effectiveness in extracting meaningful patterns and making predictions from structured data.

The fundamental function of the encoder is to perform deep feature extraction and model the contextual relationships within the input sequence [45]. The encoder is typically composed of multiple stacked layers with identical structures. Each layer comprises two key submodules: a FW-MHSA mechanism and a Feedforward neural network (FFN). These modules are integrated via serial residual connections and layer normalization, ensuring that essential original features are retained while accelerating model convergence. The encoder's operation can be formalized as follows:

$$H^L = \text{LayerNorm}(H^{L-1} + \text{MHSA}(H^{L-1})) \quad (2)$$

$$H^L = \text{LayerNorm}(H^L + \text{FFN}(H^L)) \quad (3)$$

where L denotes the L -th encoder layer, and $H^L \in \mathbf{R}^{T \times d_{\text{model}}}$ represents the output of that layer. The FW-MHSA mechanism, represented by MHSA in Equation (2), captures global dependencies within the input sequence, while the FFN applies further nonlinear transformations to the extracted features.

Feature-Wise Multi-Head Self-Attention

The attention mechanism enables models to concentrate on the most salient parts of the input data, thereby markedly enhancing processing efficiency and performance [46]. The advent of the Transformer architecture saw the introduction of self-attention mechanisms, marking a significant shift in DL from traditional RNNs and CNNs to more efficient, parallel processing architectures. To further enhance the representational capacity of self-attention, the Transformer also introduced multi-head self-attention. This approach applies multiple attention heads in parallel, thereby improving the model's ability to capture dependencies and global information across different

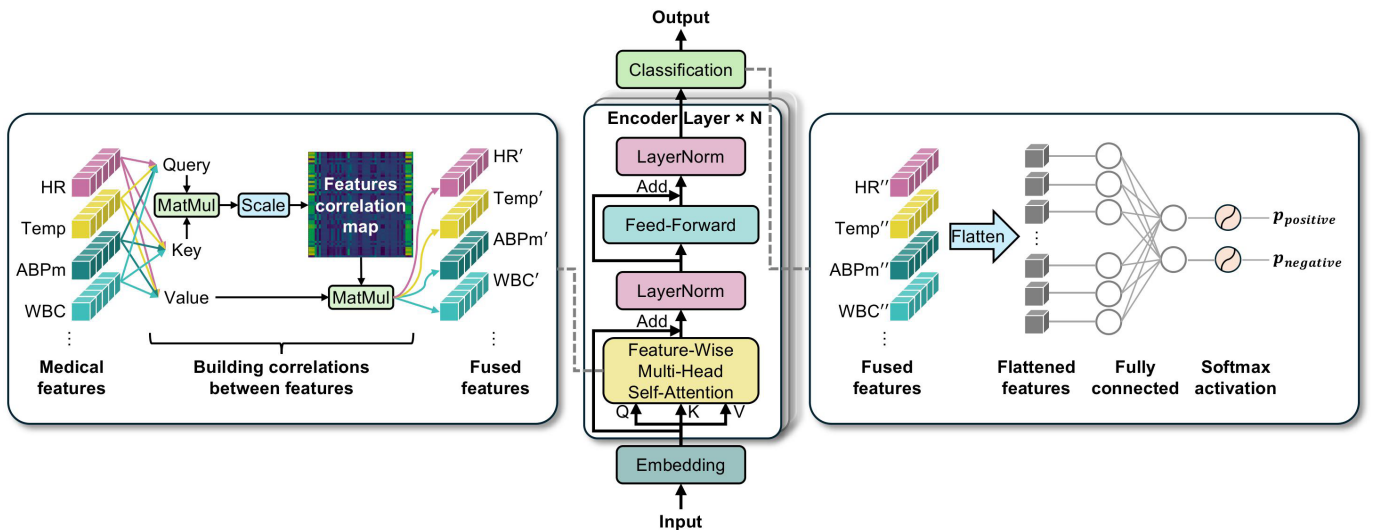


Figure 2. The architecture of the proposed FW-MHSA-former. Each encoder layer includes an input embedding layer, a Feature-Wise Multi-Head Self-Attention layer, a feed-forward neural network, and two residual connections with layer normalization. The output from the final encoder layer is fed into the fully connected layer for classification.

positions within the input sequence. In the proposed FW-MH-SA-former, the multi-head self-attention mechanism is used to compute attention across input medical features in multiple feature subspaces simultaneously, allowing the model to capture more diverse dependencies between those features.

In the FW-MHSA-former, the output of the embedding layer, $H^0 \in \mathbf{R}^{T \times d_{model}}$, consists of T tokens, where each token represents a medical feature and is transformed as a vector in $\mathbf{R}^{1 \times d_{model}}$. In the feature-wise multi-head self-attention mechanism, each token is linearly transformed into a query Q , key K , and value V as follows:

$$Q = H^{L-1} W_Q, K = H^{L-1} W_K, V = H^{L-1} W_V \quad (4)$$

where $W_Q \in \mathbf{R}^{d_{model} \times d_K}$, $W_K \in \mathbf{R}^{d_{model} \times d_K}$, and $W_V \in \mathbf{R}^{d_{model} \times d_V}$ are learnable projection matrices. Multiple sets of W_Q , W_K , and W_V are used to execute these transformations, yielding distinct Q , K , and V matrices for each attention head. For each head, the similarity between the query and key is calculated, and Softmax is applied to generate attention weights. These weights are then utilized to aggregate the value vectors, resulting in the output for the i -th head:

$$\begin{aligned} head_i &= \text{Attention}(QW_Q^{(i)}, KW_K^{(i)}, VW_V^{(i)}) \\ &= \text{softmax}\left(\frac{(QW_Q^{(i)}, KW_K^{(i)})^T}{\sqrt{d_K}}\right)(VW_V^{(i)}) \end{aligned} \quad (5)$$

The outputs from all heads are concatenated, and a linear transformation is applied to generate the final multi-head self-attention output:

$$MHSA(H^{L-1}) = \text{Concat}(head_1, \dots, head_h) W_O \quad (6)$$

where W_O is a projection matrix that maps the concatenated multi-head attention outputs back to the original d_{model} -dimensional space.

Classification Head

After traversing multiple encoder layers, the input sequence, represented by $H^L \in \mathbf{R}^{T \times d_{model}}$, is encoded into a high-dimensional feature representation that is rich in contextual information. The subsequent step is to transform this representation into a two-dimensional vector, which represents the probability for sepsis positive and sepsis negative. To simplify the high-dimensional feature representation, all features are first flattened into a one-dimensional vector:

$$X = \text{Flatten}(H^L) \in \mathbf{R}^{T \times d_{model}} \quad (7)$$

The flattening operation concatenates the features across T time steps into the vector X , thereby providing a unified input for the subsequent classification layer. This vector is then passed through a linear transformation layer, which maps it into the classification space. In this linear layer, the input features are multiplied by the weight matrix $W_{cls} \in \mathbf{R}^{(T \times d_{model}) \times 2}$, thereby generating the classification output y' :

$$y' = xW_{cls} + b_{cls} \in \mathbf{R}^2 \quad (8)$$

where $b_{cls} \in \mathbf{R}^2$ is the bias term. At this stage of the process,

the model compresses the high-dimensional feature representation into the output dimension required for the classification task. To convert the output of the linear layer into a probability distribution, the Softmax function is applied:

$$\hat{y}_i = \frac{e^{y_i}}{\sum_{j=1}^2 e^{y_j}}, i = 1, 2 \quad (9)$$

The Softmax function normalizes the raw scores generated by the linear layer into a probability distribution, thereby representing the likelihood that the input sequence belongs to each class. In the context of binary classification tasks related to sepsis prediction, the output of the Softmax function comprises two probability values, each corresponding to the predicted likelihood for sepsis positive or sepsis negative.

Adaptive Balance-Preserving Ensemble (ABPE)

We propose an ABPE to avoid the issue of data imbalance in sepsis prediction. In the preliminary phase, the negative sample set is divided into multiple subsets, with each subset containing a sample size as close as possible to that of the positive sample set. Each subset of negative samples is then combined with the positive sample set to create multiple balanced sub-datasets. This even division strategy ensures the original dataset is preserved in its integrity.

The FW-MHSA-former is instantiated and trained on each sub-dataset. The area under the receiver operating characteristic curve (AUROC) score from each model serves as its respective weight. For a given prediction sample, the risk probabilities predicted by each model are weighted according to these scores. The final prediction is then derived using a weighted voting approach. The framework of the proposed ABPE and the ratio for dividing the training, validation, and test sets are illustrated in Figure 3.

Statistical Analysis

We compared baseline demographic, clinical, and laboratory characteristics between the septic and non-septic cohorts. The normality of continuous variables was evaluated using the Shapiro-Wilk test. As the data were predominantly non-normally distributed, continuous variables, such as HR and ABP, are reported as medians with interquartile ranges (IQRs), and differences between groups were assessed using the Mann-Whitney U test. Categorical variables are reported as frequencies and percentages, with inter-group comparisons performed using the Chi-squared test or Fisher's exact test, as appropriate.

To assess the prognostic performance of our FW-MHSA-former model, Kaplan-Meier (KM) survival analysis was conducted. Patients were stratified into high-risk and low-risk categories based on risk scores generated by the model. The log-rank test was then used to determine the statistical significance of any differences observed between the KM survival curves of these two groups. For all statistical tests, a two-sided $p < 0.05$ was considered the threshold for statistical significance.

Evaluation Metrics

Model performance was evaluated using accuracy, recall, the F1 score, and AUROC. Accuracy measures the overall rate of correct classifications. Recall assesses the model's ability to identify all true sepsis cases, which is critical for minimizing

missed diagnoses. The F1 score, as the harmonic mean of precision and recall, provides a balanced evaluation, especially for imbalanced datasets. Finally, the AUROC represents the model's capacity to distinguish between sepsis and non-sepsis patients. These metrics are defined as follows:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+TN+FP+FN)} \quad (10)$$

$$\text{Recall} = \frac{TP}{(TP+FN)} \quad (11)$$

$$\text{F1 score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

where TP (true positive) refers to the number of correctly predicted sepsis cases, TN (true negative) is the number of correctly predicted non-sepsis cases, FP (false positive) is the number of non-sepsis cases incorrectly predicted as sepsis, and FN (false negative) is the number of sepsis cases incorrectly predicted as non-sepsis.

Results

Early Sepsis Prediction and Survival Analysis

To evaluate the model's performance in the task of early sepsis prediction, we selected accuracy, recall, F1 score, and AUROC. We designed experiments with different combinations of look-back (5h, 10h, 15h, 20h) and prediction time (1h, 3h, 6h, 12h) to analyze the difference in model performance across different configurations. Table 2 presents the results of these metrics across different configurations.

Experimental results demonstrate that the optimal predictive performance across all horizons is consistently achieved with a 20-hour look-back window. Notably, the AUROC for a 1-hour lead time reached 0.94, superior to that of the 5, 10 and 15-

hour configurations. A monotonic attenuation in discriminative power was observed as the prediction horizon extended; specifically, within the 20-hour window, the AUROC declined from 0.94 to 0.87 as the horizon increased from 1 to 12 hours. This phenomenon is congruent with clinical trajectories, where the stochastic nature of sepsis increases forecasting complexity over longer intervals. Furthermore, the results suggest that extended temporal depth within the look-back window provides

Table 2. The performance of the proposed method, evaluated on an independent test set for early sepsis prediction under different look-back and prediction time.

Prediction configuration		Metrics			
Look-back	Prediction time	Accuracy	Recall	F1 score	AUROC
5	1	0.86	0.83	0.85	0.93
5	3	0.83	0.78	0.82	0.89
5	6	0.79	0.72	0.77	0.87
5	12	0.76	0.70	0.74	0.83
10	1	0.83	0.82	0.83	0.91
10	3	0.83	0.81	0.82	0.90
10	6	0.81	0.76	0.80	0.86
10	12	0.75	0.72	0.74	0.82
15	1	0.86	0.88	0.86	0.92
15	3	0.83	0.82	0.83	0.91
15	6	0.79	0.74	0.78	0.88
15	12	0.79	0.77	0.78	0.86
20	1	0.87	0.90	0.87	0.94
20	3	0.86	0.84	0.86	0.93
20	6	0.79	0.84	0.86	0.93
20	12	0.80	0.83	0.80	0.87

Note: the "look-back" refers to the historical period from which patient data is used to inform the model, the "prediction time" is the lead interval for forecasting.

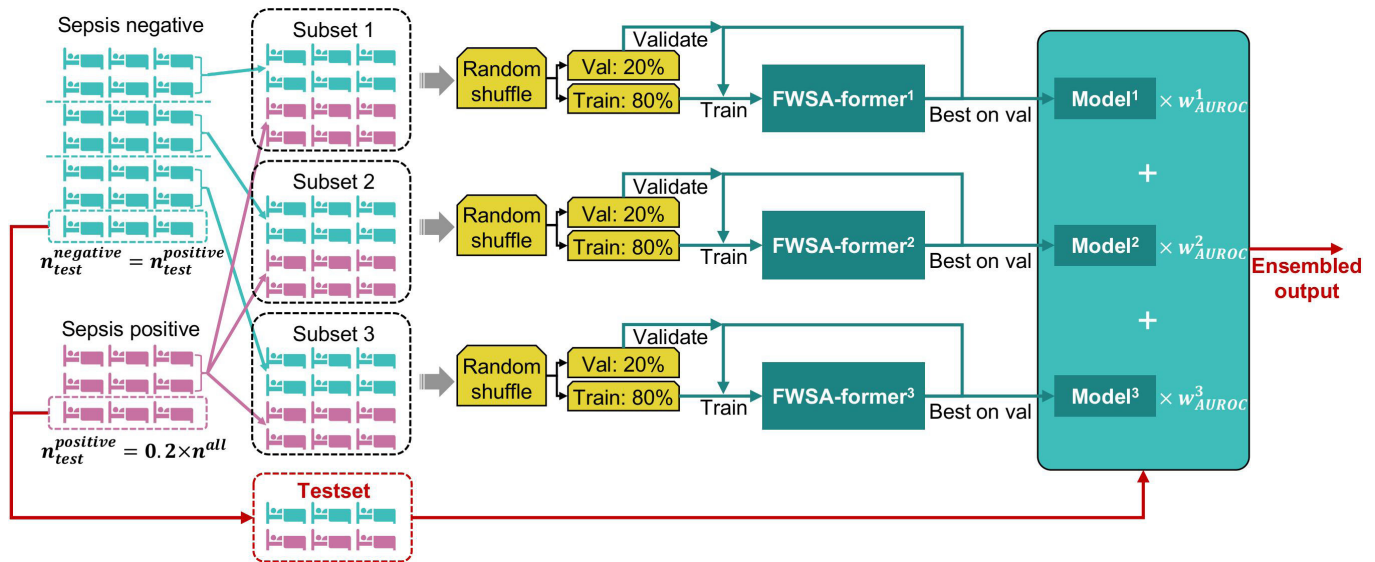


Figure 3. Overview of the proposed ABPE framework, illustrating the creation of balanced sub-datasets, model training, and the weighted voting mechanism for final prediction.

critical contextual features that augment short-term predictive stability.

To assess clinical applicability, KM survival analysis was used to stratify patients into high- and low-risk groups based on model-predicted probabilities. Figure 4 displays survival curves across various look-back and prediction windows. The results demonstrate that the model consistently distinguishes risk groups, with significantly lower survival in the high-risk cohorts. Log-rank tests confirmed all comparisons reached $p < 0.001$, indicating statistically significant survival differences.

Sensitivity Analysis of Temporal Resolution

We compared 1-hour and 30-minute resolutions to assess how temporal aggregation affects performance. Table 3 details the 12-hour prediction results by look-back window, with all other quantitative data provided in Appendix Table A.

The results show that the proposed FW-MHSA-former maintains consistent and competitive performance at the higher 30-minute temporal resolution. Across all experimental configurations, all evaluation metrics exhibited only minor fluctuations between the two resolutions, with AUROC differing by at most 0.02. This demonstrates that our model is robust to changes in time granularity and is not overly dependent on coarse hourly aggregation, further verifying the reliability and generalizability of the developed framework.

Table 3. Performance Comparison of the Proposed Model Under 1-hour and 30-minute Temporal Resolutions at a Fixed 12-hour Prediction Horizon, Stratified by Look-back Window.

Resolution	Accuracy	Recall	F1 score	AUROC
5-h look-back				
30min	0.75	0.69	0.73	0.83
1h	0.76	0.70	0.74	0.83
Δ	0.01	0.01	0.01	0.00
10-h look-back				
30min	0.74	0.70	0.73	0.80
1h	0.75	0.72	0.74	0.82
Δ	0.01	0.02	0.01	0.02
15-h look-back				
30min	0.77	0.77	0.77	0.84
1h	0.79	0.77	0.78	0.86
Δ	0.02	0.00	0.01	0.02
20-h look-back				
30min	0.78	0.78	0.78	0.85
1h	0.80	0.80	0.80	0.85
Δ	0.02	0.02	0.02	0.00

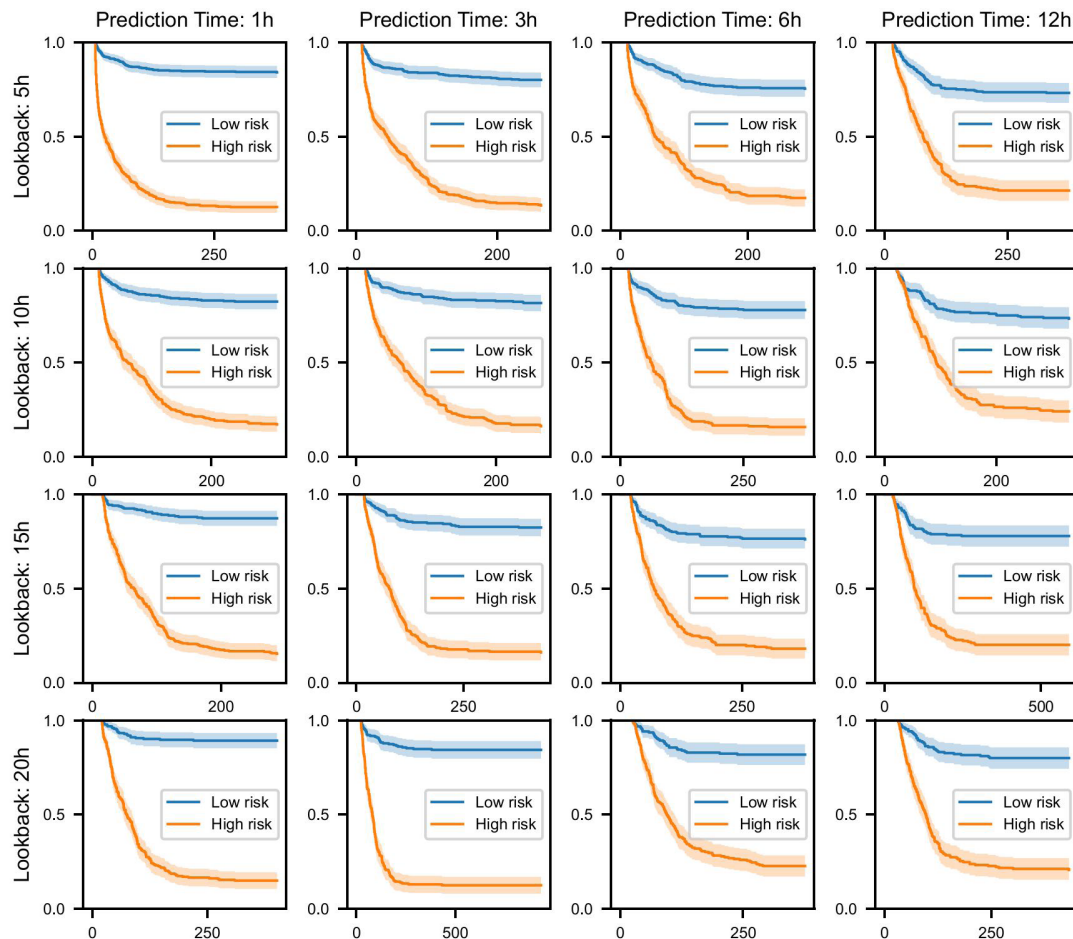


Figure 4. Kaplan-Meier survival curves comparing high-risk and low-risk sepsis groups across different look-back (5h, 10h, 15h, 20h) and prediction time (1h, 3h, 6h, 12h). The x-axis represents the survival time in hours, while the y-axis represents the survival rate. The shaded areas around each curve represent the 95% confidence intervals.

Comparison of Missing Value Imputation Strategies

Table 4 compares the performance of four missing value imputation strategies under a 20-hour look-back window; complete results for all settings are provided in Appendix Table B, C, D. Across all evaluated configurations, the zero-filling strategy consistently achieved the highest AUROC, outperforming other methods. Notably, all four imputation methods showed consistent performance trends: model AUROC increased with the extension of the look-back window and decreased with the extension of the prediction time. This verified that the zero-filling approach used in this study does not introduce obvious artificial bias or distort the physiological distribution of vital signs and laboratory features. On the contrary, the zero-filling method better preserved the original temporal characteristics of the EHRs data, enabling the model to capture the progressive physiological deterioration trajectory of sepsis more effectively.

Comparison with classic models

Six models were selected to benchmark the proposed method. These include three classical approaches: Adaptive Boosting (AdaBoost), Long Short-Term Memory (LSTM), and Random Forest (RF). The other three are advanced attention-based or

Table 4. Performance Comparison of Four Missing Value Imputation Strategies Under 20-Hour Look-Back Window Across All Prediction Time Configurations.

Method	Accuracy	Recall	F1 score	AUROC
1-h prediction				
Forward filling	0.81	0.78	0.81	0.89
K-NN imputation	0.83	0.82	0.83	0.91
Mean imputation	0.83	0.82	0.83	0.91
Zero filling	0.87	0.90	0.87	0.94
3-h prediction				
Forward filling	0.82	0.79	0.81	0.91
K-NN imputation	0.84	0.82	0.84	0.91
Mean imputation	0.82	0.81	0.82	0.91
Zero filling	0.86	0.84	0.86	0.93
6-h prediction				
Forward filling	0.74	0.76	0.74	0.81
K-NN imputation	0.75	0.74	0.75	0.81
Mean imputation	0.76	0.75	0.76	0.82
Zero filling	0.79	0.83	0.80	0.87
12-h prediction				
Forward filling	0.75	0.76	0.75	0.81
K-NN imputation	0.74	0.73	0.74	0.81
Mean imputation	0.74	0.76	0.75	0.82
Zero filling	0.80	0.80	0.80	0.85

Note: Forward filling: Missing values are filled with the last valid observation from the same patient; any remaining missing values are further imputed using the training-set mean; K-NN imputation: Missing values are imputed based on the average of the 5 nearest neighbors with complete feature profiles. Mean imputation: Missing values are filled with the mean of the corresponding feature derived from the training dataset; Zero filling: The baseline strategy in this study, which replaces missing values and out-of-range negative values with 0.

Transformer-based architectures: DuETT [47], the LSTM-Transformer hybrid, and Medformer [48]. All models were evaluated on the same test set and assessed in accordance with the evaluation metrics described in Evaluation metrics. Given that shorter look-back increases the difficulty of prediction while allowing the model to initiate sepsis prediction earlier after ICU admission, and that longer prediction times provide more response time for healthcare professionals, two representative experimental scenarios were selected for comparison. Table 5 presents the accuracy, recall, F1 score, and AUROC results for each model under a fixed 5-hour look-back with different prediction times (1h, 3h, 6h, and 12h). Table 6 illustrates the corresponding performance for distinct look-back hours (5h, 10h, 15h, and 20h) with a fixed 12-hour prediction time. The experimental results demonstrate that our method exhibits superior overall performance.

Visualization and interpretation analysis

In order to enhance the interpretability of the model, we employed the self-attention mechanism of the Transformer for sepsis prediction, encoding medical features (such as HR, ABP, WBC, etc.) as tokens and calculating correlations between these features through the attention mechanism. Figure 5 displays the correlation map from the model's final layer, obtained by averaging across all attention heads. Each cell represents the attention score when the query is sent from the corresponding row to the corresponding column. It can be observed that ABP (including ABPm, ABPs, and ABPd) exhibit a noticeable response to almost every feature, which is reasonable given the relatively higher density of ABP data. Additionally, certain features exhibit strong uniform response patterns, including (1) ABPm, ABPs, and ABPd, (2) HR and O2 Sat, and (3) Ven CO2 and Ven O2, among others. These relationships align with known medical correlations. While Figure 5 shows relatively low pairwise attention weights between core laboratory markers including Lactate, Creatinine, and Bilirubin, this does not indicate the model's neglect

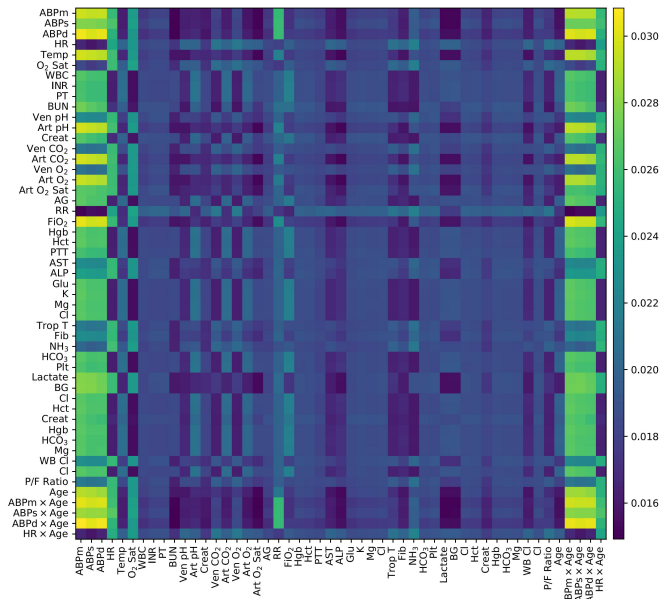


Figure 5. Correlation map of the FW-MHSA-former. Each cell represents the attention score when the query is sent from the corresponding row to the corresponding column.

Table 5. Comparison of performance metrics between our proposed method and baselines across different prediction times with a 5-hour look-back. The evaluation was conducted on the same test set, which was balanced for positive and negative samples.

Method	Accuracy	Recall	F1 score	AUROC
1-h prediction				
Adaboost	0.79	0.62	0.74	0.90
LSTM	0.79	0.64	0.76	0.90
RF	0.75	0.53	0.68	0.91
DuETT	0.85	0.81	0.82	0.92
LSTM Transformer	0.85	0.80	0.82	0.92
Medformer	0.85	0.82	0.83	0.93
FW-MHSA-former (Ours)	0.86	0.83	0.85	0.93
3-h prediction				
Adaboost	0.71	0.44	0.60	0.85
LSTM	0.74	0.52	0.67	0.87
RF	0.65	0.32	0.48	0.89
DuETT	0.81	0.75	0.81	0.88
LSTM Transformer	0.80	0.74	0.81	0.86
Medformer	0.81	0.76	0.80	0.88
FW-MHSA-former (Ours)	0.83	0.78	0.82	0.89
6-h prediction				
Adaboost	0.61	0.26	0.41	0.82
LSTM	0.63	0.29	0.44	0.80
RF	0.54	0.08	0.15	0.80
DuETT	0.78	0.69	0.72	0.85
LSTM Transformer	0.76	0.71	0.76	0.83
Medformer	0.78	0.71	0.74	0.83
FW-MHSA-former (Ours)	0.79	0.72	0.77	0.87
12-h prediction				
Adaboost	0.54	0.11	0.19	0.77
LSTM	-	-	-	0.67
RF	0.50	0.00	0.01	0.75
DuETT	0.75	0.63	0.68	0.83
LSTM Transformer	0.74	0.67	0.71	0.82
Medformer	0.74	0.69	0.73	0.82
FW-MHSA-former (Ours)	0.76	0.70	0.74	0.83

Note: The "-" denotes undefined values due to no positive predictions, which can occur when the model fails to identify any minority class instances.

of organ dysfunction— a key point of the Sepsis-3 definition centered on organ dysfunction [49]. Instead, this phenomenon is attributed to the sparse and irregular measurement of these laboratory markers in clinical practice, which leads to the lack of strong static correlations between them. Importantly, the model does not rely solely on high-frequency vital signs; rather, it effectively captures the dynamic temporal changes of these core laboratory markers that are associated with organ failure, and integrates them with vital signs and engineered features (e.g., P/F Ratio, ABP×Age) that directly reflect organ function. This integration fully aligns with the Sepsis-3 definition [49], confirming that the model has not overlooked the core multi-organ failure logic of sepsis.

The correlation map illustrates the relationships between medical features, with each column representing the attention scores when responding to queries from other features. The higher the attention score, the more information from the queried feature is incorporated into the output of the querying feature. The more a feature is incorporated by other features, the greater its influence on the model's final output. By summing the attention scores in each column of the correlation map, we calculated an importance score, indicating the contribution of each medical feature to the model's predictions. These importance scores were ranked in descending order, as shown in Figure 6.

Table 6. Comparison of performance metrics between our proposed method and baselines across different prediction times with a 12-hour prediction. The evaluation was conducted on the same test set, which was balanced for positive and negative samples.

Method	Accuracy	Recall	F1 score	AUROC
20-h look-back				
Adaboost	0.63	0.31	0.46	0.82
LSTM	0.64	0.31	0.46	0.76
RF	0.50	0.01	0.01	0.82
DuETT	0.78	0.75	0.77	0.84
LSTM Transformer	0.79	0.78	0.79	0.85
Medformer	0.78	0.75	0.77	0.83
FW-MHSA-former (Ours)	0.80	0.80	0.80	0.85
15-h look-back				
Adaboost	0.58	0.21	0.33	0.80
LSTM	-	-	-	0.66
RF	0.51	0.01	0.22	0.77
DuETT	0.75	0.73	0.75	0.86
LSTM Transformer	0.77	0.75	0.77	0.85
Medformer	0.76	0.71	0.76	0.84
FW-MHSA-former (Ours)	0.79	0.77	0.78	0.86
10-h look-back				
Adaboost	0.56	0.15	0.25	0.79
LSTM	-	-	-	0.55
RF	-	-	-	0.79
DuETT	0.74	0.71	0.73	0.81
LSTM Transformer	0.72	0.69	0.71	0.80
Medformer	0.72	0.70	0.72	0.80
FW-MHSA-former (Ours)	0.75	0.72	0.74	0.82
5-h look-back				
Adaboost	0.54	0.11	0.19	0.77
LSTM	-	-	-	0.67
RF	0.50	0.00	0.01	0.75
DuETT	0.75	0.69	0.73	0.82
LSTM Transformer	0.75	0.65	0.71	0.82
Medformer	0.76	0.68	0.70	0.83
FW-MHSA-former (Ours)	0.76	0.70	0.74	0.83

Note: The "-" denotes undefined values due to no positive predictions, which can occur when the model fails to identify any minority class instances.

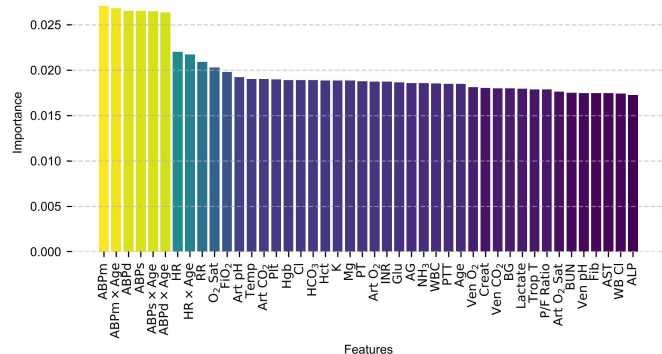


Figure 6. Feature importance ranking based on the sum of attention scores in the correlation map. The y-axis represents the medical feature importance score, while the x-axis displays the corresponding feature names.

Discussions

This study introduces and validates a novel DL framework that synergizes a FW-MHSA-former with an ABPE strategy. The architecture was specifically engineered to overcome two fundamental challenges inherent in EHRs data: the management of temporally irregular observations and the profound class imbalance typical of medical datasets. Evaluated retrospectively on the comprehensive, publicly accessible MIMIC-IV dataset, the model demonstrated robust performance. Across a range of configurations, varying the look-back window from 5 to 20 hours and the prediction horizon from 1 to 12 hours, the model consistently yielded an AUROC exceeding 0.80. It achieved a peak AUROC of 0.94 when utilizing a 20-hour look-back window for a 1-hour prediction horizon. Furthermore, in comparative analyses, our framework outperformed six conventional ML and DL models.

A detailed investigation into the model's performance characteristics revealed crucial temporal dynamics. Experimental results demonstrate that the model's predictive performance scales with the extension of the look-back window, peaking at 20 hours, although performance remains comparable between the 15-hour and 20-hour windows specifically at a 12-hour lead time. This finding suggests that physiological data within the 20-hour window preceding sepsis onset encapsulates the most critical information for risk prediction, as longer temporal sequences better capture the complete trajectory of progressive clinical deterioration. Such insight provides an empirical basis for determining optimal data acquisition windows in clinical settings. Furthermore, while the observed decay in performance as the lead time extends aligns with the established principle that predictive uncertainty increases with temporal distance, the model maintains robust discriminative capacity (AUROC > 0.82) even at a 12-hour lead time. This underscores its potential to establish a clinically meaningful window for early intervention.

An analysis of the model's interpretability confirmed that it learned patterns highly congruent with established clinical pathophysiology. Feature importance analysis revealed that the highest-weighted predictors were engineered interaction terms, namely ABPm×Age, ABPd×Age, and ABPs×Age. This finding is significant, as it indicates that the prognostic value of blood pressure in sepsis is non-linear and substantially

modulated by age. This interaction term more effectively captures the heightened hemodynamic vulnerability of older patients during infectious stress than individual metrics alone—a nuance frequently missed by traditional linear models. Following these were foundational vital signs such as HR, RR, Temp, and O2 Sat, which are the cornerstones of rapid screening tools like the qSOFA. This alignment demonstrates that the model independently learned decision-making criteria consistent with expert clinical consensus. Moreover, laboratory markers reflecting organ dysfunction and tissue hypoperfusion—including Plt, Creat, and lactate—also received high importance scores, which is consistent with the Sepsis-3 definition centered on the SOFA score. Thus, the model functions not as an inscrutable 'black box' but as a system whose decision-making logic is largely transparent and capable of augmenting our understanding of sepsis pathophysiology.

While our model demonstrated exemplary performance in a retrospective context, its ultimate clinical value is contingent upon prospective validation. Kaplan-Meier survival analysis confirmed the model's ability to stratify patients into high- and low-risk cohorts with statistically significant differences in survival outcomes ($p < 0.001$), laying a critical foundation for its translation into clinical decision support tools. Notably, with a stable AUROC of 0.82–0.86 across all look-back configurations, the 12-hour prediction window fits well into routine ICU workflows based on 12-hour nursing shift cycles, providing reliable evidence to assist real-time clinical scheduling and hierarchical patient management. This lead time aligns perfectly with routine patient assessment and handover procedures: the model can be run once per shift at handover to generate a 12-hour sepsis risk forecast for the entire on-duty period, enabling proactive rather than reactive care. For patients stratified as high-risk, this 12-hour lead time enables concrete, guideline-aligned interventions: upgrading from 4-hourly routine vital sign monitoring to 1-hourly continuous hemodynamic monitoring, pre-emptive completion of blood culture and inflammatory biomarker testing, and advance preparation of vascular access and fluid resuscitation protocols. This fully addresses the core clinical requirement of the Surviving Sepsis Campaign guidelines, which mandate initiation of evidence-based care within 1 hour of sepsis recognition, by providing a sufficient window for risk mitigation before clinical deterioration. Furthermore, the model's robust risk stratification enables optimized ICU resource allocation: low-risk patients (with consistently negative predictions) can avoid unnecessary invasive monitoring and frequent laboratory testing, while clinical resources are prioritized to the high-risk cohort identified by the model.

Nevertheless, this study is subject to several limitations. First, we standardized the time series by aggregating raw records to a uniform hourly resolution, which may smooth high-frequency physiological fluctuations and subtle acute physiological changes. To mitigate this limitation and evaluate the influence of temporal granularity, we supplemented a sensitivity analysis using a 30-minute interval resolution. Experimental results demonstrated that our model yields stable and comparable performance at both 1-hour and 30-minute temporal resolutions, confirming that the prediction capability is not substantially compromised by hourly aggregation and supporting the robustness of the proposed method.

Second, the model was developed and validated exclusively using data from the MIMIC-IV database. Although large and of

high quality, MIMIC-IV represents a single academic medical center. Therefore, the model's performance and generalizability must be confirmed through external validation on datasets from different healthcare systems and patient populations.

Conclusions

This study introduces a novel DL architecture that integrates a FW-MHSA-former with an ABPE for the early prediction of sepsis. The proposed framework exhibits superior predictive performance, enabling the robust stratification of patients into high- and low-risk cohorts, thereby enhancing clinical decision support and optimizing patient management strategies. A critical finding is the model's exceptional robustness and reliability, as evidenced by its sustained high-fidelity predictive accuracy across diverse and challenging look-back and forecast time configurations. These results underscore the framework's significant clinical utility and its potential to improve patient outcomes through more timely and precise.

Abbreviations

electronic health records: EHRs; deep learning: DL; Feature-Wise Multi-Head Self-Attention Transformer: FW-MHSA-former; Adaptive Balance-Preserving Ensemble: ABPE; Medical Information Mart for Intensive Care IV: MIMIC-IV; Kaplan-Meier: KM; Area Under the Receiver Operating Characteristic Curve: AUROC; machine learning: ML; Systemic Inflammatory Response Syndrome: SIRS; Sequential Organ Failure Assessment: SOFA; quick SOFA: qSOFA; intensive care unit: ICU; logistic regression: LR; recurrent neural network: RNNs; gated recurrent unit: GRU; bidirectional long short term memory: Bi-LSTM; convolutional neural network: CNN; Massachusetts Institute of Technology: MIT; Beth Israel Deaconess Medical Center: BIDMC; SHapley Additive Explanations: SHAP; natural language processing: NLP; computer vision: CV; time series analysis: TSA; Feedforward neural network: FFN; inter-quartile ranges: IQRs; true positive: TP; true negative: TN; false positive: FP; false negative: FN; Adaptive Boosting: AdaBoost; Random Forest: RF; Support Vector Classifier: SVC.

Author Contributions

Mireaye Abudurexiti: Conceptualization, Investigation, Supervision, Writing – Original Draft. Jianshu Wang: Methodology, Visualization, Validation, Software. Pengfei Zhang: Data Curation, Investigation, Supervision. Fangfang Liu: Resources, Project Administration. Zhiqiang Jia: Funding Acquisition, Supervision, Investigation, Writing – Review and Editing.

Acknowledgments

Not Applicable.

Funding Information

This study was supported by the Xinjiang Uygur Autonomous Region Health Science and Technology Program (2025001CG-ZHYDXM653124504), the "Tianshan Talent" High-Level Medical and Health Personnel Training Program (TSYC202401B223) and Kashgar Prefecture Science and Technology Plan Program (KS2024022).

Ethics Approval and Consent to Participate

This study uses data from the MIMIC-IV database, which consists of anonymized records that do not require individual patient consent.

Competing Interests

The authors declare no conflicts of interest relevant to this article.

Data Availability

The MIMIC-IV dataset utilized in this study originates from the Beth Israel Deaconess Medical Center (BIDMC) and is managed by the Massachusetts Institute of Technology's Computer Science and Artificial Intelligence Laboratory (MIT CSAIL). Given that the dataset contains sensitive patient health information, researchers must apply for access and complete a Data Use Agreement (DUA) along with Collaborative Institutional Training Initiative (CITI) certification to ensure ethical usage and protection of privacy. Consequently, the raw data is not publicly available; however, authorized researchers may obtain access through the official MIMIC-IV portal at <https://mimic.physionet.org/>.

References

- [1] Duan, Y., Huo, J., Chen, M., Hou, F., Yan, G., Li, S., et al. (2023). Early prediction of sepsis using double fusion of deep features and handcrafted features. *Applied intelligence* (Dordrecht, Netherlands), 1–17. Advance online publication. <https://doi.org/10.1007/s10489-022-04425-z>
- [2] Fleuren, L. M., Klausch, T. L. T., Zwager, C. L., Schoonmade, L. J., Guo, T., Roggeveen, L. F., et al. (2020). Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy. *Intensive care medicine*, 46(3), 383–400. <https://doi.org/10.1007/s00134-019-05872-y>
- [3] Reyna, M. A., Josef, C. S., Jeter, R., Shashikumar, S. P., Westover, M. B., Nemat, S., et al. (2020). Early Prediction of Sepsis From Clinical Data: The PhysioNet/Computing in Cardiology Challenge 2019. *Critical care medicine*, 48(2), 210–217. <https://doi.org/10.1097/CCM.0000000000004145>
- [4] Lauritsen, S. M., Kalør, M. E., Kongsgaard, E. L., Lauritsen, K. M., Jørgensen, M. J., Lange, J., et al. (2020). Early detection of sepsis utilizing deep learning on electronic health record event sequences. *Artificial intelligence in medicine*, 104, 101820. <https://doi.org/10.1016/>

- j.artmed.2020.101820
- [5] Wang, Y., Lu, X., Tan, J., He, J., Yin, H., Chen, P., et al. (2025). The cGAS-STING Pathway: Insights into Regulatory Mechanisms, Disease Dysregulation, and Therapeutic Development. *Life Conflux*, 2(1), e300. <https://doi.org/10.71321/dr57c347>
 - [6] Bomrah, S., Uddin, M., Upadhyay, U., Komorowski, M., Priya, J., Dhar, E., et al. (2024). A scoping review of machine learning for sepsis prediction- feature engineering strategies and model performance: a step towards explainability. *Critical care (London, England)*, 28(1), 180. <https://doi.org/10.1186/s13054-024-04948-6>
 - [7] Deng, H. F., Sun, M. W., Wang, Y., Zeng, J., Yuan, T., Li, T., et al. (2021). Evaluating machine learning models for sepsis prediction: A systematic review of methodologies. *iScience*, 25(1), 103651. <https://doi.org/10.1016/j.isci.2021.103651>
 - [8] Delahanty, R. J., Alvarez, J., Flynn, L. M., Sherwin, R. L., & Jones, S. S. (2019). Development and Evaluation of a Machine Learning Model for the Early Identification of Patients at Risk for Sepsis. *Annals of emergency medicine*, 73(4), 334–344. <https://doi.org/10.1016/j.annemergmed.2018.11.036>
 - [9] Yan, M. Y., Gustad, L. T., & Nytrø, Ø. (2022). Sepsis prediction, early detection, and identification using clinical text for machine learning: a systematic review. *Journal of the American Medical Informatics Association: JAMIA*, 29(3), 559–575. <https://doi.org/10.1093/jamia/ocab236>
 - [10] Nemati, S., Holder, A., Razmi, F., Stanley, M. D., Clifford, G. D., & Buchman, T. G. (2018). An Interpretable Machine Learning Model for Accurate Prediction of Sepsis in the ICU. *Critical care medicine*, 46(4), 547–553. <https://doi.org/10.1097/CCM.0000000000002936>
 - [11] Shashikumar, S. P., Josef, C. S., Sharma, A., & Nemati, S. (2021). DeepAISE - An interpretable and recurrent neural survival model for early prediction of sepsis. *Artificial intelligence in medicine*, 113, 102036. <https://doi.org/10.1016/j.artmed.2021.102036>
 - [12] Lauritsen, S. M., Kristensen, M., Olsen, M. V., Larsen, M. S., Lauritsen, K. M., Jørgensen, M. J., et al. (2020). Explainable artificial intelligence model to predict acute critical illness from electronic health records. *Nature communications*, 11(1), 3852. <https://doi.org/10.1038/s41467-020-17431-x>
 - [13] Mao, Q., Jay, M., Hoffman, J. L., Calvert, J., Barton, C., Shimabukuro, D., et al. (2018). Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU. *BMJ open*, 8(1), e017833. <https://doi.org/10.1136/bmjopen-2017-017833>
 - [14] Tang, X., Zheng, D., Kebede, G. S., Li, Z., Li, X., Lu, C., et al. (2023). An automatic segmentation framework of quasi-periodic time series through graph structure. *Applied Intelligence*, 53(20), 23482–23499. <https://doi.org/10.1007/s10489-023-04814-y>
 - [15] van de Sande, D., van Genderen, M. E., Huiskens, J., Gommers, D., & van Bommel, J. (2021). Moving from bytes to bedside: a systematic review on the use of artificial intelligence in the intensive care unit. *Intensive care medicine*, 47(7), 750–760. <https://doi.org/10.1007/s00134-021-06446-7>
 - [16] Lai, H., Wu, G., Zhong, Y., Chen, G., Zhang, W., Shi, S., et al. (2023). Red blood cell distribution width improves the prediction of 28-day mortality for patients with sepsis-induced acute kidney injury: A retrospective analysis from MIMIC-IV database using propensity score matching. *Journal of intensive medicine*, 3(3), 275–282. <https://doi.org/10.1016/j.jointm.2023.02.005>
 - [17] Oami, T., Imaeda, T., Nakada, T. A., Abe, T., Takahashi, N., Yamao, Y., et al. (2023). Mortality analysis among sepsis patients in and out of intensive care units using the Japanese nationwide medical claims database: a study by the Japan Sepsis Alliance study group. *Journal of intensive care*, 11(1), 2. <https://doi.org/10.1186/s40560-023-00650-x>
 - [18] Zohar, Y., Zilberman Itskovich, S., Koren, S., Zaidenstein, R., Marchaim, D., & Koren, R. (2021). The association of diabetes and hyperglycemia with sepsis outcomes: a population-based cohort analysis. *Internal and emergency medicine*, 16(3), 719–728. <https://doi.org/10.1007/s11739-020-02507-9>
 - [19] Das, P.P., Wiese, L., Mast, M., Böhnke, J., Wulff, A., Marschollek, M., et al. (2024). An attention-based bidirectional LSTM-CNN architecture for the early prediction of sepsis. *International Journal of Data Science and Analytics*, 20, 1841 - 1855. <https://doi.org/10.1007/s41060-024-00568-z>
 - [20] Zhao, Y., Wu, Y., Liu, M., Cai, X., Zhang, Y., Yuan, X. (2022). DEAR: Dual-Level Self-attention GRU for Online Early Prediction of Sepsis. In: Zhao, X., Yang, S., Wang, X., Li, J. (eds) *Web Information Systems and Applications*. WISA 2022. *Lecture Notes in Computer Science*, vol 13579. Springer, Cham. https://doi.org/10.1007/978-3-031-20309-1_37
 - [21] Rafiei, A., Rezaee, A., Hajati, F., Gheisari, S., & Golzan, M. (2021). SSP: Early prediction of sepsis using fully connected LSTM-CNN model. *Computers in biology and medicine*, 128, 104110. <https://doi.org/10.1016/j.compbiomed.2020.104110>
 - [22] Johnson, A. E. W., Bulgarelli, L., Shen, L., Gayles, A., Shammout, A., Horng, S., et al. (2023). MIMIC-IV, a freely accessible electronic health record dataset. *Scientific data*, 10(1), 1. <https://doi.org/10.1038/s41597-022-01899-x>
 - [23] Seymour, C. W., Liu, V. X., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., et al. (2016). Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 762–774. <https://doi.org/10.1001/jama.2016.0288>
 - [24] Seymour, C. W., Kennedy, J. N., Wang, S., Chang, C. H., El-liott, C. F., Xu, Z., et al. (2019). Derivation, Validation, and Potential Treatment Implications of Novel Clinical Phenotypes for Sepsis. *JAMA*, 321(20), 2003–2017. <https://doi.org/10.1001/jama.2019.5791>
 - [25] Gyawali, B., Ramakrishna, K., & Dhamoon, A. S. (2019). Sepsis: The evolution in definition, pathophysiology, and management. *SAGE open medicine*, 7, 2050312119835043. <https://doi.org/10.1177/2050312119835043>
 - [26] Camacho-Cogollo, J. E., Bonet, I., Gil, B., & Iadanza, E. (2022). Machine Learning Models for Early Prediction of Sepsis on Large Healthcare Datasets. *Electronics*, 11(9), 1507. <https://doi.org/10.3390/electronics11091507>
 - [27] Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The Third

- International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801–810. <https://doi.org/10.1001/jama.2016.0287>
- [28] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems*, 5998–6008. <https://doi.org/10.48550/arXiv.1706.03762>
- [29] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv*, 1301.3781. <https://arxiv.org/abs/1301.3781>
- [30] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., et al. (2020). Language models are few-shot learners. *arXiv*, 2005.14165. <https://arxiv.org/abs/2005.14165>
- [31] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI*.
- [32] Wang, X., Thakker, M., Chen, Z., Kanda, N., Eskimez, S.E., Chen, S., et al. (2023). SpeechX: Neural Codec Language Model as a Versatile Speech Transformer. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 32, 3355–3364. <https://doi.org/10.1109/TASLP.2024.3419418>
- [33] Yang, D., Tian, J., Tan, X., Huang, R., Liu, S., Chang, X., et al. (2023). UniAudio: An Audio Foundation Model Toward Universal Audio Generation. *ArXiv*, 2310.00704. <https://arxiv.org/abs/2310.00704>
- [34] Chen, C.-F., Fan, Q., & Panda, R. (2021). CrossViT: Cross-attention multi-scale vision transformer for image classification. *arXiv*, 2103.14899. <https://arxiv.org/abs/2103.14899>
- [35] Xin, C., Liu, Z., Zhao, K., Miao, L., Ma, Y., Zhu, X., et al. (2022). An improved transformer network for skin cancer classification. *Computers in biology and medicine*, 149, 105939. <https://doi.org/10.1016/j.combiomed.2022.105939>
- [36] Dutta, P., Sathi, K. A., Hossain, M. A., & Dewan, M. A. A. (2023). Conv-ViT: A Convolution and Vision Transformer-Based Hybrid Feature Extraction Method for Retinal Disease Detection. *Journal of imaging*, 9(7), 140. <https://doi.org/10.3390/jimaging9070140>
- [37] Wang, L., Fang, S., Meng, X., & Li, R. (2022). Building extraction with vision transformer. *IEEE Transactions on Geoscience and Remote Sensing*, 60, 1–11. <https://doi.org/10.1109/TGRS.2022.3186634>
- [38] Bhopale, A. P., & Tiwari, A. (2024). Transformer based contextual text representation framework for intelligent information retrieval. *Expert Systems with Applications*, 238, 121629. <https://doi.org/10.1016/j.eswa.2023.121629>
- [39] Kim, Y., Bang, S., Sohn, J., & Kim, H. (2022). Question answering method for infrastructure damage information retrieval from textual data using bidirectional encoder representations from transformers. *Automation in Construction*, 134, 104061. <https://doi.org/10.1016/j.autcon.2021.104061>
- [40] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv*, 2010.11929. <https://arxiv.org/abs/2010.11929>
- [41] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV), 9992–10002. <https://doi.org/10.48550/arXiv.2103.14030>
- [42] Liu, Z., Hu, H., Lin, Y., Yao, Z., Xie, Z., Wei, Y., et al. (2022). Swin transformer V2: Scaling up capacity and resolution. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 11999–12009. <https://doi.org/10.1109/CVPR52688.2022.01170>
- [43] Nie, Y., Nguyen, N. H., Sinthong, P., & Kalagnanam, J. (2022). A time series is worth 64 words: Long-term forecasting with transformers. *arXiv*. <https://doi.org/10.48550/arXiv.2211.14730>
- [44] Liu, Y., Hu, T., Zhang, H., Wu, H., Wang, S., Ma, L., et al. (2023). iTransformer: Inverted transformers are effective for time series forecasting. *arXiv*. <https://doi.org/10.48550/arXiv.2310.06625>
- [45] Mo, Y., Wu, Q., Li, X., & Huang, B. (2021). Remaining useful life estimation via transformer encoder enhanced by a gated convolutional unit. *J Intell Manuf* 32, 1997–2006. <https://doi.org/10.1007/s10845-021-01750-x>
- [46] Guo, M.-H., Xu, T.-X., Liu, J.-J., Liu, Z.-N., Jiang, P.-T., Mu, T.-J., et al. (2022). Attention mechanisms in computer vision: A survey. *Computational Visual Media*, 8(3), 331–368. <https://doi.org/10.1007/s41095-022-0271-y>
- [47] Labach, A., Pokhrel, A., Huang, X., Zuberi, S., Yi, S., Volkovs, M., et al. (2023). DuETT: Dual Event Time Transformer for Electronic Health Records. *arXiv*, 2304.13017. <https://doi.org/10.48550/arXiv.2304.13017>
- [48] Wang, Y., Huang, N., Li, T., Yan, Y., Zhang, X. (2024). Medformer: A Multi-Granularity Patching Transformer for Medical Time-Series Classification. *arXiv*, 2405.19363. <https://doi.org/10.48550/arXiv.2405.19363>
- [49] Singer, M., Deutschman, C. S., Seymour, C. W., Shankar-Hari, M., Annane, D., Bauer, M., et al. (2016). The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). *JAMA*, 315(8), 801–810. <https://doi.org/10.1001/jama.2016.0287>

Clinicopathological Analysis of Succinate Dehydrogenase-Deficient Renal Cell Carcinoma

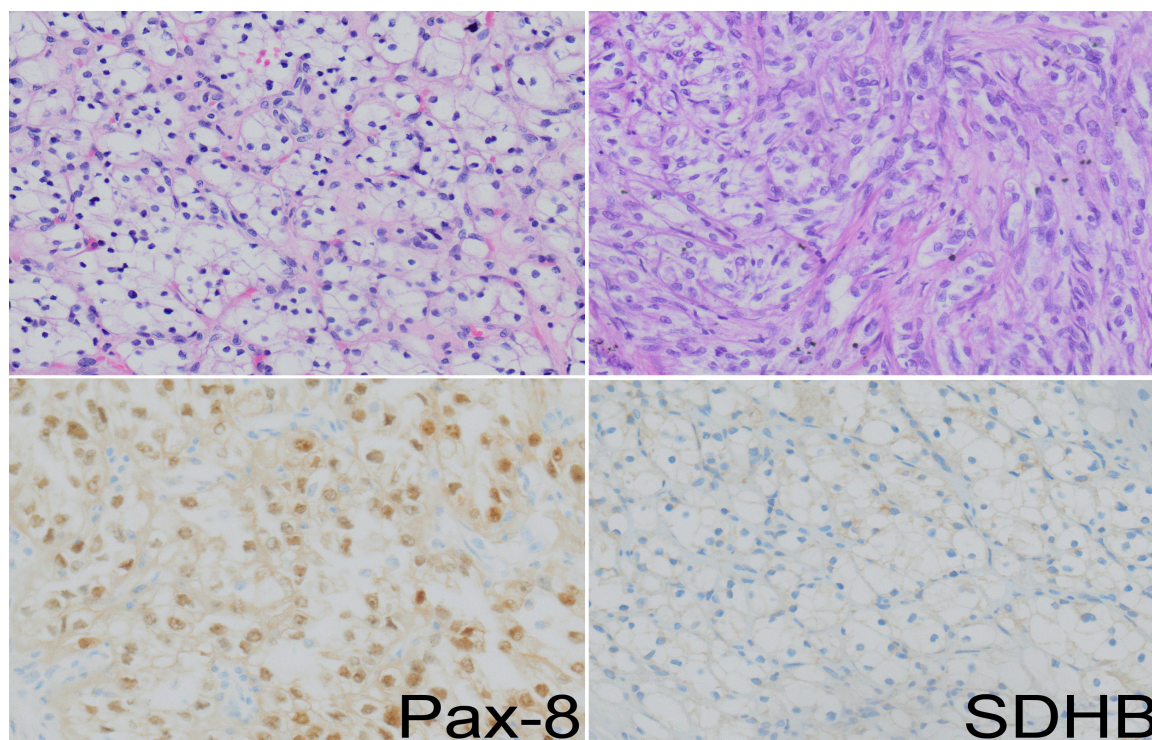
Authors

Shizhi Hu, Ruling Zhang, Shujun Li, Jing Ma, Heqin Zhan

Correspondence

zhanhq@ahmu.edu.cn (H. Zhan)

Graphical Abstract



This study analyzed the clinicopathological and molecular characteristics of 14 cases of succinate dehydrogenase-deficient renal cell carcinoma (SDH-RCC), among which 4 cases were high nuclear grade. The results revealed the typical histopathological and immunohistochemical features of SDH-RCC, further enriched the case database of SDH-RCC, and provided a certain pathological basis for subsequent research on SDH-RCC.

<https://doi.org/10.71321/3zc3th83>

© 2026 The Author(s). Published by Life Conflux Press Limited. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Clinicopathological Analysis of Succinate Dehydrogenase-Deficient Renal Cell Carcinoma

Shizhi Hu¹, Ruling Zhang¹, Shujun Li¹, Jing Ma², Heqin Zhan^{1,2*}

Received: 2026-03-21 | Accepted: 2026-05-15 | Published online: 2026-05-30

Abstract

Objective: To investigate the clinicopathological features, immunophenotype, key diagnostic points and prognosis of succinate dehydrogenase (SDH)-deficient renal cell carcinoma in order to provide a reference for its clinical diagnosis, treatment and pathological diagnosis.

Materials and Methods: The retrospective analysis was performed on 14 cases of SDH-deficient renal cell carcinoma, with a summary and evaluation of their clinicopathological, morphological, immunohistochemical features and prognosis.

Results: Among the 14 patients, 10 were male and 4 were female. The age ranged from 43 to 78 years, with a median age of 52 years. 9 tumors were located in the left kidney and 5 in the right kidney. Tumor size ranged from 1.5 cm to 8.0 cm. Microscopically, the tumor cells exhibited diverse architectural patterns, including predominantly solid and nested growth, occasional microcystic and cystic dilatation, and papillary architecture. The tumor cells had abundant cytoplasm, which was vacuolated or weakly eosinophilic. Immunohistochemical results showed that all 14 tumor cells were positive for vimentin and PAX-8, and negative for SDHB, CK7, CD117, and CA IX. The Ki-67 proliferation index ranged from 5% to 20%. Among the 14 patients, 1 was lost to follow-up, and 13 were followed up. Of the 13 followed patients, 12 were alive and 1 died.

Conclusion: SDH-deficient renal cell carcinoma is a relatively rare subtype of renal cell carcinoma. Patients with low nuclear grade are generally younger, whereas those with high nuclear grade are mostly middle-aged or elderly and prone to metastasis. Loss of SDHB expression by immunohistochemical staining is helpful for the diagnosis of this type of tumor.

Keywords: Succinate dehydrogenase-deficient renal cell carcinoma; Histology; Immunohistochemistry; Diagnosis; Prognosis

Introduction

Succinate dehydrogenase (SDH) is a respiratory enzyme complex composed of four protein subunits, SDHA, SDHB, SDHC, and SDHD, and is localized to the inner mitochondrial membrane [1]. Its primary function is to catalyze the conversion of succinate to fumarate and facilitate electron transport in the respiratory chain [2]. SDH-deficient renal cell carcinoma (SDH-RCC) is a relatively rare subtype of RCC, which was first identified in 2004. It was officially classified as an independent subtype of RCC by the World Health Organization (WHO) in 2016. In 2022, it was formally defined as RCC with loss of SDHB expression and recognized as a distinct subtype of RCC [3-4]. As a rare subtype of RCC (accounting for 0.05%-0.2% of all RCC), SDH-RCC has been relatively uncommon reported in China, with limited relevant research available. Consequently, clinicians and pathologists lack sufficient familiarity with this entity, leading to a high risk of misdiagnosis. To improve the understanding of this subtype of RCC, we collected a total of

14 cases of SDH-RCC and summarized their clinicopathological and immunohistochemical features.

Materials and Methods

Case Collection

A total of 14 cases of SDH-RCC diagnosed between 2018 and 2024 were collected from the First and Second Affiliated Hospitals of Anhui Medical University, all of which met the diagnostic criteria for SDH-RCC. Detailed clinicopathological data and follow-up information were collected for each case. All histological slides were independently reviewed in a blind manner by two senior pathologists, and the microscopic pathological features of each specimen were documented in detail.

HE Staining and Immunohistochemical Staining

All specimens were obtained via surgical operation, including radical nephrectomy and partial nephrectomy. The specimens

1 Department of Pathology, The Second Affiliated Hospital of Anhui Medical University, 230601, Hefei, China

2 Department of Pathology, School of Basic Medical Sciences, Anhui Medical University, 230032 Hefei, China

* Corresponding Author.

were fixed in formalin, embedded in paraffin, and subjected to routine deparaffinization and rehydration prior to HE staining. Histomorphological features were examined under a light microscope. Immunohistochemical staining was performed using the EnVision two-step method. The primary antibodies included SDHB (MXB Biotechnologies, clone 21A11AE7), vimentin (MXB Biotechnologies, clone MX034), PAX-8 (MXB Biotechnologies, clone EP298), CK7 (MXB Biotechnologies, clone OV-TL12/30), CD117 (MXB Biotechnologies, clone YR145), CA IX (MXB Biotechnologies, rabbit polyclonal antibody), and Ki-67 (Roche, clone 30-9).

Follow-Up

Follow-up information was collected through medical record review and telephone follow-up. The recorded parameters included tumor recurrence, metastasis, and patient survival status.

Result

Clinical Information

Detailed clinical information of the 14 patients was summarized (Table 1). The patients ranged in age from 43 to 78 years, with the mean age of 54.4 years and the median age of 52 years. Regarding gender, there were 10 males and 4 females, with a marked male predominance and a male-to-female ratio of 2.5:1. With respect to tumor location, 9 tumors were located in the left kidney and 5 in the right kidney. As for the reasons for initial presentation, 8 patients were asymptomatic and detected incidentally during physical examination, while the remaining patients presented with varying degrees of discomfort, including ipsilateral abdominal pain, lumbodorsal pain, abnormal liver function, and ipsilateral inguinal hernia. All patients underwent abdominal CT or MRI, which revealed a space-occupying lesion in the affected kidney (Figures 1A and 1B).

Gross Findings

The maximum tumor diameter ranged from 1.5 cm to 8.0 cm in the 14 patients. Among them, 9 tumors were obtained from partial nephrectomy specimens, and 5 from radical nephrecto-

my specimens of the affected kidneys.

The cut surfaces of the tumors exhibited variable colors, predominantly gray-white, gray-yellow, and gray-brown. Regarding tumor texture, cases 1, 4, 9, 11 displayed a cystic-solid cut surface, and the remaining cases showed solid and slightly firm cut surfaces. Most tumors were well-demarcated with pushing growth borders. Only cases 3, 13, and 14 showed invasion of the renal capsule and perirenal adipose tissue.

Histologic Features

Most tumors were well-circumscribed, with pushing borders into adjacent tissue and formation of a fibrous capsule. Focally, a lobular architecture was noted. Invasion into surrounding adipose tissue was identified in 3 cases. Entrapped non-neoplastic renal tubules were visible at the periphery of the tumors.

Tumor cells were arranged in solid sheets, nests, and tubules (Figures 2A, 2B), with prominent sinusoidal vasculature within tumor nests. Microcystic or multicystic structures were present in some cases (Figure 2C). In addition, focal micropapillary architecture was observed in 1 case (case 13; Figure 2D), and papillary structures were identified in 1 case (case 14; Figure 2E).

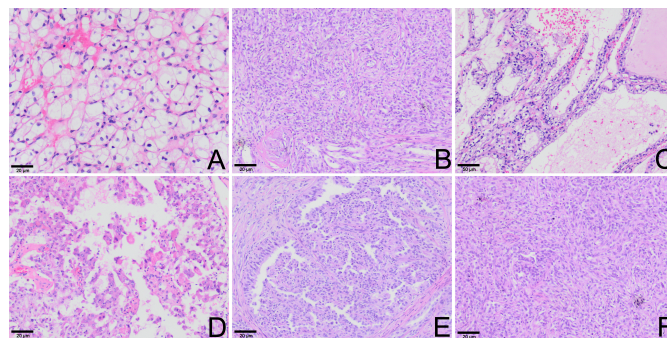


Figure 2. Histological features of SDH-RCC. (A) (B) Tumor cells are arranged in solid sheets, nests, and tubules (HE×400). (C) Multicystic structures are present within the lesion (HE×200). (D) Micropapillary structures are identified in some cases (HE×400). (E) Papillary structures are observed in a minority of cases (HE×400). (F) Short spindle cell proliferation is seen in high-grade tumors (HE×400).

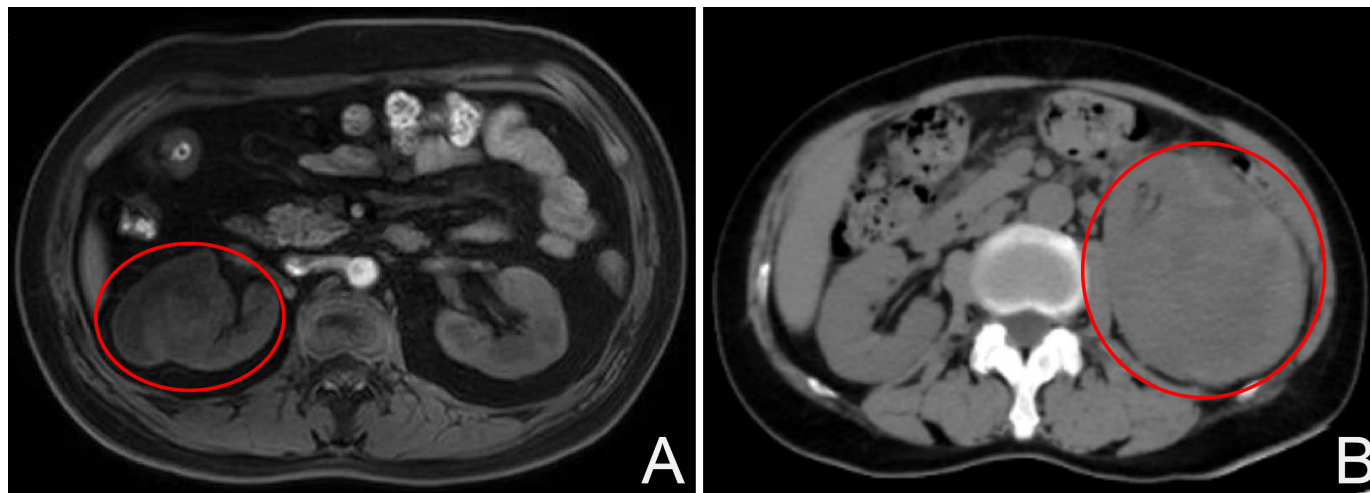


Figure 1. Imaging features of SDH-RCC. (A) CT shows a right renal mass; (B) CT shows a left renal mass.

With respect to cytomorphology, tumor cells exhibited abundant eosinophilic cytoplasm, with vacuolar changes noted focally in some cases. Focal short spindle-cell morphology was present in 1 case (case 14; [Figure 2F](#)). Nuclei were round to oval with fine, evenly distributed chromatin and mild nuclear atypia. A small number of cases showed more marked nuclear atypia, characterized by enlarged nuclei, coarse chromatin, occasional prominent nucleoli, and focal necrosis.

Of the 14 cases included in this study, 10 were classified as low-grade [International Society of Urological Pathology (ISUP)/WHO grade 1-2], and 4 as high-grade, including 3 cases of [ISUP/WHO grade 3] and 1 case of [ISUP/WHO grade 4].

Immunohistochemical Staining

Immunohistochemical staining was performed in all 14 cases ([Table 2](#)). The results showed that tumor cells in all cases were positive for PAX-8 ([Figure 3A](#)) and vimentin ([Figure 3B](#)). All

cases were negative for SDHB ([Figures 3C, D](#)), while non-neoplastic renal tubules adjacent to the tumor were positive. In addition, CK7, CA IX, and CD117 were all negative. The Ki-67 proliferation index ranged from 5% to 30%. All low-grade tumors exhibited a low Ki-67 index, ranging from 5% to 15% ([Figure 3E](#)), whereas the Ki-67 index of high-grade tumors was 20–30% ([Figure 3F](#)).

Follow-Up Results

Follow-up was conducted in all 14 patients, with a follow-up duration of 2-7 years. One patient was lost to follow-up, 1 patient was dead, and the remaining 12 patients were alive at the last follow-up. In addition, extrarenal metastasis occurred in 3 patients (cases 8, 13 and 14). Case 8 was found to have lung metastases during examination upon seeking medical attention for fever and cough 11 months after surgery, and case 13 was diagnosed with lung metastases during routine physical

Table 1. Clinicopathological information of 14 patients.

Number	Age	Sex	Location (kidney)	Size	Reason for consultation	Surgical approach	Cut surface of the mass	Metastasis status	Follow-up
1	50	Male	Right	2.5cm×2.0cm×2.0cm	Physical examination	Partial nephrectomy	Yellowish-gray, multiloculated cystic	None	Alive
2	53	Female	Right	5.0cm×4.0cm×1.5cm	Lumbar disc herniation	Radical nephrectomy	Grayish-white and grayish-red, slightly firm in texture	None	Lost to follow-up
3	78	Female	Right	7.5cm×5.0cm×4.5cm	Physical examination	Radical nephrectomy	Variegated appearance	None	Alive
4	48	Female	Right	6.5cm×5.0cm×3.5cm	Physical examination	Partial nephrectomy	Multiloculated cystic	None	Alive
5	48	Male	Left	6.0cm×5.0cm×4.0cm	Physical examination	Partial nephrectomy	Yellowish-gray and grayish-brown, firm in texture	None	Alive
6	43	Male	Right	3.2cm×3.2cm×2.8cm	Physical examination	Partial nephrectomy	Grayish-white and yellowish-gray, firm in texture	None	Alive
7	65	Male	Left	2.5cm×2.0cm×2.0cm	Physical examination	Partial nephrectomy	Yellowish-gray and grayish-brown	None	Alive
8	59	Female	Left	7.5cm×7.0cm×6.0cm	Left lower abdominal pain	Radical nephrectomy	Yellowish-gray and grayish-white, with a thin capsule	Pulmonary metastasis	Alive
9	49	Male	Left	3.5cm×3.0cm×1.5cm	Rapid weight loss, gastric discomfort	Partial nephrectomy	Grayish-white and grayish-brown, cystic and solid	None	Alive
10	56	Male	Left	5.0cm×5.0cm×4.0cm	Physical examination	Partial nephrectomy	Variegated, relatively well-circumscribed	None	Alive
11	50	Male	Left	4.5cm×3.2cm×2.7cm	Physical examination	Partial nephrectomy	Yellowish-gray and grayish-red, hemorrhagic cystic degeneration	None	Alive
12	58	Male	Left	1.5cm×1.5cm×1.2cm	Abnormal liver function	Partial nephrectomy	Yellowish-gray, with a capsule	None	Alive
13	53	Male	Right	8.0cm×7.0cm×7.0cm	Right inguinal hernia	Radical nephrectomy	Grayish-brown, medium in texture	Pulmonary metastasis	Alive
14	51	Male	Left	4.0cm×3.5cm×3.5cm	Low back pain with fever	Partial nephrectomy	Grayish-white, firm in texture	Lumbar spine metastasis	Death

examination 5 years after surgery. Case 14 developed bone metastasis and has died.

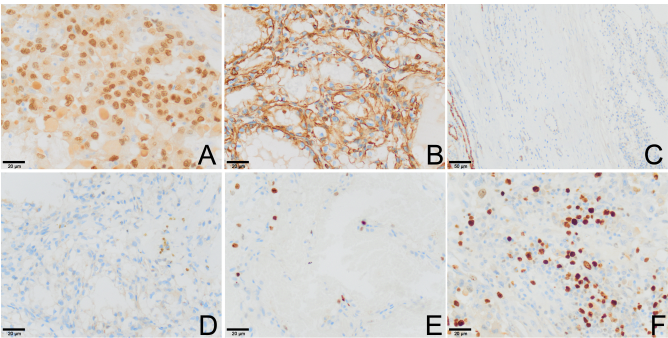


Figure 3. Immunohistochemical features of SDH-RCC. (A) Tumor cells show nuclear positivity for PAX-8 (EnVision×400). (B) Tumor cells are positive for Vimentin (EnVision×400). (C) SDHB staining: positive normal renal tubules (lower left), negative tumor cells (upper right) (EnVision×200). (D) SDHB staining is negative in tumor tissue (EnVision×400). (E) Low nuclear grade with low Ki-67 proliferation index (EnVision×400). (F) High nuclear grade with high Ki-67 proliferation index (EnVision×400).

Table 2. Results of immunohistochemical staining.

Num- ber	Immunohistochemical staining items						
	SDHB	PAX-8	Vimentin	CK7	CD117	CAIX	Ki – 67
1	-	+	+	-	-	-	8%
2	-	+	+	-	-	-	10%
3	-	+	+	-	-	-	20%
4	-	+	+	-	-	-	10%
5	-	+	+	-	-	-	15%
6	-	+	+	-	-	-	8%
7	-	+	+	-	-	-	8%
8	-	+	+	-	-	-	25%
9	-	+	+	-	-	-	5%
10	-	+	+	-	-	-	10%
11	-	+	+	-	-	-	8%
12	-	+	+	-	-	-	7%
13	-	+	+	-	-	-	20%
14	-	+	+	-	-	-	25%

Note: - : negative; + : positive.

Discussion

SDH is a key enzyme in the tricarboxylic acid cycle that catalyzes the conversion of succinate to fumarate, and it is composed of four distinct subunits, including SDHA, SDHB, SDHC, and SDHD [1]. Mutations in any of the SDH subunits can lead to loss of normal function of the complex, thereby contributing to the development of associated tumors. In 2004, Vanharanta et al. first identified familial genetic characteristics of SDHB gene mutations in 3 relatively young patients with RCC, and suggested a potential correlation between these mutations and the development of early-onset RCC. Subsequently, due to the limited number of cases, this type of RCC was not classified separately. It was not until 2016 that the WHO officially classified SDH-RCC as an independent subtype of RCC [5].

SDH-RCC is an extremely rare subtype of RCC, accounting for only 0.05%-0.2% of all RCC. It has a wide age distribution, ranging from 14 to 76 years, and shows a male predominance [3]. In this study, a total of 14 SDH-RCC patients were collected, with an age range of 43-78 years, including 10 males and 4 females. These findings were generally consistent with the age and gender distribution characteristics of SDH-RCC reported in the literature. Patients with SDH-RCC harbor mutations in the SDHA, SDHB, SDHC and SDHD genes, among which SDHB mutations are the most common [6-8]. In addition, germline mutations in SDH genes can cause patients to develop gastrointestinal stromal tumors, pheochromocytoma/paraganglioma, and pituitary adenomas [1]. No clinical records of these associated neoplasms were identified in the 14 patients included in this study. Histologically, the tumors are mostly well-circumscribed, with variable cut surface colors ranging from gray-yellow to gray-red [9]. Microscopically, tumor cells are predominantly arranged in solid and nested patterns, often accompanied by the formation of microcystic, multicystic and tubular structures. Papillary structures may occasionally be observed. Tumor cells show abundant eosinophilic cytoplasm with vacuolar changes, and the nuclei are usually bland [9]. In the present series of 14 cases, typical solid sheet-like and nested structures were identified in all cases, and papillary structures were found in case 14. The cytomorphology was generally consistent with that described in the literature, except that focal short spindle-cell morphology was displayed in case 14. 10 cases were low-grade nuclear grade, and 4 cases were high-grade nuclear grade. In addition, SDH-RCC shows overlapping histomorphological features with other types of RCC, including clear cell RCC, papillary RCC and chromophobe RCC [1]. Sarcomatoid transformation may occur in high-grade SDH-RCC. Regarding treatment and prognosis, most SDH-RCC are low-grade tumors with weak invasiveness and rare metastasis. The long-term metastasis rate is approximately 11%, and the overall prognosis is generally favorable. Treatment is generally performed with partial nephrectomy [6, 9]. In contrast, high-grade SDH-RCC, including tumors with sarcomatoid change, necrosis, and high nuclear grade, is relatively rare but the invasiveness of the tumor is significantly increased, and the tumor is prone to metastasis with a metastatic rate of up to 70%, resulting in a relatively poor prognosis. Radical nephrectomy is usually performed for these high-risk tumors [10]. In the present study, among the 14 patients, 1 patient was lost to follow-up. Of the remaining 13 patients, 12 were alive and 1 had died at the last follow-up. Notably, all 3 patients who developed extrarenal metastasis presented high nuclear-grade. Differential diagnosis: 1. Clear cell RCC is the most common subtype, accounting for the highest incidence among all RCC. Tumor cells are arranged in solid nests or acinar structures, with a characteristic network of thin-walled blood vessels in the stroma [11]. The cells are large and polygonal, with abundant clear cytoplasm or eosinophilic cytoplasm in high-grade tumors [11]. Immunohistochemically, tumor cells are positive for CA IX, without loss of SDHB expression [9]. 2. Chromophobe RCC is the third most common subtype of RCC. Microscopically, tumor cells are mostly arranged in solid sheets and nests, with occasional tubular or trabecular patterns [12-13]. The tumor is composed of two cell types. One is small eosinophilic cells with granular cytoplasm and the other is large

polygonal cells with flocculent cytoplasm and characteristic plant-like cell membranes, showing wrinkled nuclei with inconspicuous nucleoli [13]. Immunohistochemically, tumor cells are positive for CK7 and CD117, without loss of SDHB [13]. 3. Oncocytoma is a benign renal tumor. Microscopically, tumor cells are densely arranged in nests, islands, and acini. The tumor stroma is often loosely edematous, accompanied by myxoid or collagenized changes, and a characteristic central scar composed of collagen fibers and compressed blood vessels may be present [14-15]. Tumor cells are round or cuboidal with indistinct cell borders, moderate to abundant cytoplasm containing dense eosinophilic granules, small nuclei, and inconspicuous nucleoli [14-15]. Immunohistochemically, tumor cells are positive for CD117 and negative for CK7, without loss of SDHB expression [9].

Conclusion

In summary, SDH-RCC is a rare subtype of RCC primarily caused by abnormal expression of the SDHB gene, which further leads to dysfunction of SDH. Among affected patients, those with low-grade SDH-RCC have a relatively more favorable prognosis but still require long-term follow-up and monitoring, whereas those with high-grade SDH-RCC are at a high risk of recurrence and metastasis. Currently, surgical resection remains the first-line treatment for SDH-RCC. Given the complex morphological features of SDH-RCC tumors, SDHB immunohistochemistry can assist in the diagnosis of this disease. In addition, it is recommended that family members of affected patients may undergo concurrent clinical counseling and screening to identify potential hereditary risks.

Abbreviations

SDH: Succinate dehydrogenase; SDH-RCC: Succinate dehydrogenase-deficient renal cell carcinoma; WHO: World Health Organization; ISUP: International Society of Urological Pathology.

Author Contribution

H.Q.Z.: Conception and design of the study, data analysis, and writing and critical review of the manuscript. **S.Z.H.** and **R.L.Z.:** Data collection and analysis, and draft of the manuscript. **S.J.L.** and **J.M.:** Data collection and analysis, and editing of the manuscript. All authors have approved the final manuscript.

Acknowledgements

Not Applicable.

Funding Information

This study was supported by the Research Fund of Anhui Institute of Translational Medicine (No. 2023zhxy-C05).

Ethics Approval and Consent to Participate

Not Applicable.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

The data used in this study are available from the corresponding author upon reasonable request.

References

- [1] Wang G, & Rao P. (2018). Succinate Dehydrogenase-Deficient Renal Cell Carcinoma: A Short Review. *Arch Pathol Lab Med*, 142(10), 1284-1288. <https://doi.org/10.5858/arpa.2017-0199-RS>
- [2] Dalla Pozza E, Dando I, Pacchiana R, Liboi E, Scupoli MT, Donadelli M, et al. (2020). Regulation of succinate dehydrogenase and role of succinate in cancer. *Semin Cell Dev Biol*, 98, 4-14. <https://doi.org/10.1016/j.semcdb.2019.04.013>
- [3] Wu J, Wang Y, Yang L, Wang Y, Hu P, Fan L, et al. (2025). New Insights Into Succinate Dehydrogenase-Deficient Renal Cell Carcinoma. *Arch Pathol Lab Med*, 10.5858/arpa.2025-0233-OA. <https://doi.org/10.5858/arpa.2025-0233-OA>
- [4] Moch H, Amin MB, Berney DM, Comp  rat EM, Gill AJ, Hartmann A, et al. (2022). The 2022 World Health Organization Classification of Tumours of the Urinary System and Male Genital Organs—Part A: Renal, Penile, and Testicular Tumours. *European Urology*, 82(5), 458-468. <https://doi.org/10.1016/j.eururo.2022.06.016>
- [5] Moch H, Cubilla AL, Humphrey PA, Reuter VE, & Ulbright TM. (2016). The 2016 WHO Classification of Tumours of the Urinary System and Male Genital Organs-Part A: Renal, Penile, and Testicular Tumours. *Eur Urol*, 70(1), 93-105. <https://doi.org/10.1016/j.eururo.2016.02.029>
- [6] Gill AJ, Hes O, Papathomas T,   edivcov   M, Tan PH, Agaimy A, et al. (2014). Succinate dehydrogenase (SDH)-deficient renal carcinoma: a morphologically distinct entity: a clinicopathologic series of 36 tumors from 27 patients. *Am J Surg Pathol*, 38(12), 1588-1602. <https://doi.org/10.1097/pas.0000000000000292>
- [7] Kama   T, Higashi S, Murakami S, Arai K, Namatame T, Kijima T, et al. (2021). Single nucleotide variants of succinate dehydrogenase A gene in renal cell carcinoma. *Cancer Sci*, 112(8), 3375-3387. <https://doi.org/10.1111/cas.14977>
- [8] Sun X, Wang G, Huang Z, Li P, Yang B, Wang T, et al. (2023). Succinate Dehydrogenase Defects Giant Renal Cell Carcinoma. *Urol Int*, 107(8), 819-822. <https://doi.org/10.1159/000531059>
- [9] Pan X, Wei Y, Sui X, Yin X, Zheng L, Zeng H, et al. (2024).

- [Succinate Dehydrogenase-Deficient Renal Cell Carcinoma: Clinicopathological Analysis of 11 Cases]. *Sichuan Da Xue Xue Bao Yi Xue Ban*, 55(5), 1099-1106. <https://doi.org/10.12182/20240960101>
- [10] Wang XT, Wang X, Zhang RS, Cheng K, Xia QY, & Rao Q. (2022). [Succinate dehydrogenase-deficient renal cell carcinoma:a clinicopathological, ultrastructural and molecular analysis]. *Zhonghua Bing Li Xue Za Zhi*, 51(1), 12-16. <https://doi.org/10.3760/cma.j.cn112151-20210823-00590>
- [11] Nezami BG, & MacLennan GT. (2025). Clear Cell Renal Cell Carcinoma: A Comprehensive Review of its Histopathology, Genetics, and Differential Diagnosis. *Int J Surg Pathol*, 33(2), 265-280. <https://doi.org/10.1177/10668969241256111>
- [12] Henske EP, Cheng L, Hakimi AA, Choueiri TK, & Braun DA. (2023). Chromophobe renal cell carcinoma. *Cancer Cell*, 41(8), 1383-1388. <https://doi.org/10.1016/j.ccell.2023.07.006>
- [13] Garje R, Elhag D, Yasin HA, Acharya L, Vaena D, & Dahmouh L. (2021). Comprehensive review of chromophobe renal cell carcinoma. *Crit Rev Oncol Hematol*, 160, 103287. <https://doi.org/10.1016/j.critrevonc.2021.103287>
- [14] Williams GM, & Lynch DT. (2025). Renal Oncocytoma. In *StatPearls*. StatPearls Publishing Copyright © 2025, StatPearls Publishing LLC.
- [15] Baranovska VV, Romanenko A, & Zakhartseva LM. (2020). Histological differential diagnostics of renal oncocytoma. *Exp Oncol*, 42(3), 233-237. <https://doi.org/10.32471/exp-oncology.2312-8852.vol-42-no-3.14968>

Spontaneous Reduction in ART-Conceived Twin Pregnancies: Risk Factors, Perinatal Outcomes, and a Prediction Model

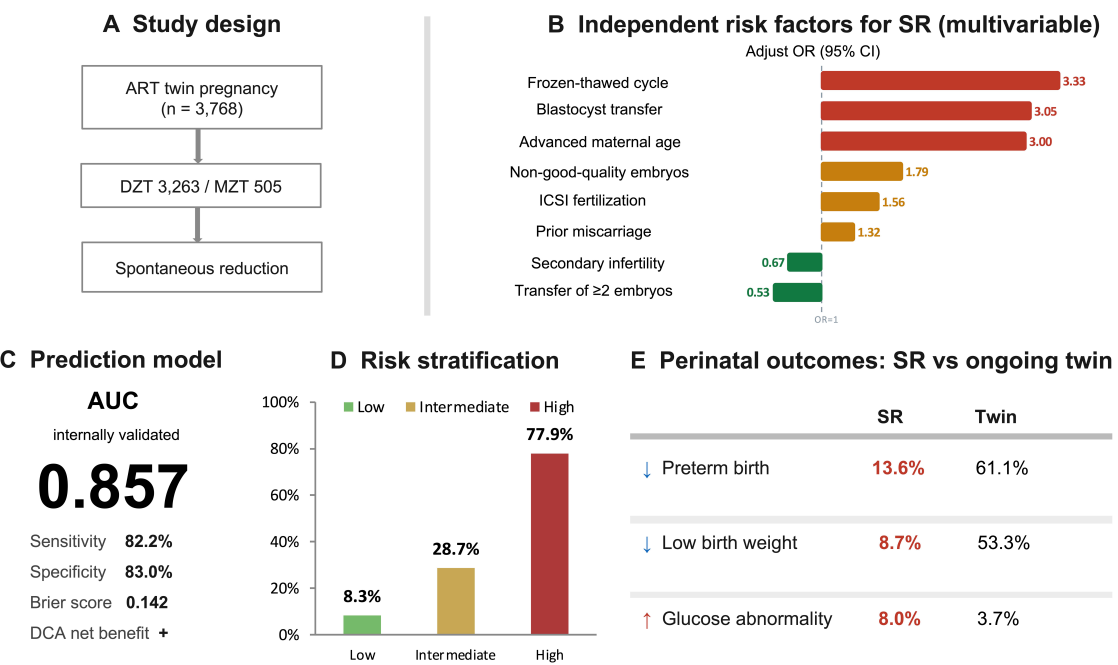
Authors

Shuqi Cai, Zhaolian Wei

Correspondence

weizhaolian@ahmu.edu.cn (Z. Wei)

Graphical Abstract



Spontaneous Reduction in ART-Conceived Twin Pregnancies: Risk Factors, Perinatal Outcomes, and a Prediction Model

Shuqi Cai¹, Zhaolian Wei^{1,2,3*}

Received: 2026-05-30 | Accepted: 2026-05-31 | Published online: 2026-05-31

Abstract

Objective: To investigate the risk factors for spontaneous reduction (SR) in assisted reproductive technology (ART)-conceived twin pregnancies and compare perinatal outcomes between SR and non-SR pregnancies, and to develop and validate an SR risk prediction model in the dizygotic twin (DZT) cohort.

Methods: This single-center retrospective cohort study enrolled patients who achieved clinical twin pregnancies following ART at the Reproductive Center of the First Affiliated Hospital of Anhui Medical University between January 2015 and December 2024. Participants were classified into monozygotic twins (MZT, n=505) and dizygotic twins (DZT, n=3,263) based on the number of gestational sacs and fetal heartbeats detected on early pregnancy ultrasound, and each cohort was further divided into SR and Non-SR groups. Baseline characteristics, treatment parameters, and perinatal outcomes were compared between groups. Univariable and multivariable logistic regression analyses were performed with SR as the dependent variable, incorporating nine core predictors. A predictive model for SR was constructed in the DZT cohort and internally validated; model performance was evaluated using the receiver operating characteristic (ROC) curve with area under the curve (AUC), calibration curve, Brier score, and decision curve analysis (DCA). Risk stratification was performed by tertiles of predicted probability.

Results: In the DZT cohort, the SR group had significantly older maternal age, a higher proportion of advanced maternal age (≥ 35 years), and higher rates of blastocyst-stage transfer, intracytoplasmic sperm injection (ICSI), and frozen-thawed embryo transfer cycles than the Non-SR group (all $P < 0.05$). No statistically significant differences were observed in any of these characteristics in the MZT cohort. Perinatal outcome trends were consistent across both cohorts: the SR group had later gestational age at delivery, lower rates of preterm birth and low birth weight, but higher rates of glucose metabolism abnormalities and lower rates of premature rupture of membranes. Multivariable logistic regression identified advanced maternal age (aOR=3.00), history of previous miscarriage (aOR=1.32), blastocyst-stage transfer (aOR=3.05), non-good-quality embryos (aOR=1.79), ICSI (aOR=1.56), and frozen-thawed cycles (aOR=3.33) as independent risk factors for SR, whereas secondary infertility (aOR=0.67) and transfer of ≥ 2 embryos (aOR=0.53) were associated with a reduced risk. The DZT-SR prediction model achieved an AUC of 0.857 on internal validation, with acceptable calibration and favorable net clinical benefit on DCA. SR incidence increased progressively across low-, intermediate-, and high-risk strata at 8.3%, 28.7%, and 77.9%, respectively.

Conclusion: SR in ART-conceived twin pregnancies is closely associated with maternal age and multiple embryo- and cycle-related factors. The DZT-SR prediction model, built on nine routinely available clinical variables, demonstrated good internal validation performance and may serve as a practical tool for early pregnancy risk stratification and individualized follow-up planning.

Keywords: Assisted reproductive technology; Twin pregnancy; Spontaneous reduction; Logistic regression; Prediction model

Introduction

According to the World Health Organization (WHO), approximately one in six people worldwide experiences infertility at some point during their lifetime [1]. The widespread adoption of assisted reproductive technology (ART) has substantially improved pregnancy outcomes for affected couples; however, the use of ovarian stimulation and multiple embryo trans-

fer has led to a higher rate of iatrogenic multiple pregnancy compared with natural conception, making it one of the most common complications associated with ART [2-3]. Multiple pregnancy carries considerable maternal and neonatal risks: compared with singleton pregnancy, the risks of preterm birth, preterm birth before 32 weeks, stillbirth, and neonatal death are increased approximately 6-, 13-, 5-, and 7-fold, respectively, and the rates of severe maternal morbidity and mortality are

1 Department of Obstetrics and Gynecology, First Affiliated Hospital of Anhui Medical University, No 218 Jixi Road, Hefei, China.

2 Anhui Province Key Laboratory of Reproductive Health and Genetics, Anhui Medical University, No 81 Meishan Road, Hefei 230032, China.

3 NHC Key Laboratory of study on abnormal gametes and reproductive tract (Anhui Medical University), No 81 Meishan Road, Hefei 230032, China.

* Corresponding Author.

also markedly elevated [4-5]. In response, reproductive medicine societies worldwide have issued guidelines recommending a reduction in the number of embryos transferred and encouraging the adoption of elective single embryo transfer (eSET) to reduce multiple pregnancy rates [2-3, 6-7].

The routine use of first-trimester ultrasound has made vanishing twin syndrome (VTS) — or spontaneous reduction (SR) — increasingly recognizable in clinical practice. SR refers to the arrest of development of one or more embryos in an intrauterine multiple pregnancy that has already been confirmed clinically. In ART-conceived twin pregnancies, the reported incidence of SR ranges from approximately 12% to 38% [8-10]. Its etiology remains incompletely understood, but proposed contributing factors include abnormal embryonic development, suboptimal placental implantation, and an unfavorable intrauterine environment [9, 11-12]; maternal age, chorionicity, and certain ART-related procedures may also play a role [13-16], though the available evidence is limited and findings across studies are inconsistent.

The impact of SR on pregnancy outcomes also warrants attention. Perinatal outcomes in pregnancies that undergo SR and continue as singletons appear to fall somewhere between those of pregnancies that were singletons from the outset and those that remain twin pregnancies, but existing studies are limited by small sample sizes and heterogeneous findings [8, 10]. Furthermore, the clinical patterns and implications of SR may differ between monozygotic twins (MZT) and dizygotic twins (DZT), yet systematic stratified analyses are lacking, and data from Chinese populations remain scarce.

Against this background, the present study compared baseline characteristics, treatment parameters, and perinatal outcomes between SR and non-SR pregnancies in both MZT and DZT cohorts, examined the risk factors associated with SR, and developed and internally validated an SR risk prediction model within the DZT cohort, with the aim of informing early pregnancy risk assessment and follow-up management.

Materials and Methods

Study population

This was a single-center retrospective cohort study. We enrolled patients who achieved clinical pregnancy following ART at the Reproductive Center of the First Affiliated Hospital of Anhui Medical University between January 2015 and December 2024, and whose twin pregnancy was confirmed by transvaginal ultrasound in the first trimester. Participants were classified into MZT and DZT based on the number of gestational sacs and fetal heartbeats on early ultrasound, and each cohort was divided into an SR group and a Non-SR group according to whether spontaneous reduction occurred during first-trimester follow-up.

Definitions and diagnostic criteria

Twin pregnancy was confirmed approximately 30 days after embryo transfer by transvaginal ultrasound showing either two separate gestational sacs or a single sac containing two fetal heartbeats. MZT was defined as a single gestational sac with two fetal heartbeats, and DZT as two gestational sacs each with a fetal heartbeat. SR was defined as a spontaneous decrease in the number of fetal heartbeats from two to one

on routine follow-up ultrasound in a previously confirmed twin pregnancy; the date of the first ultrasound documenting this change was recorded as the time of SR. Non-SR was defined as the absence of any such reduction in fetal heartbeats throughout the first trimester, with subsequent delivery of one or two live infants.

Inclusion and exclusion criteria

Patients were included if they: (1) achieved clinical pregnancy following ART with first-trimester ultrasound confirmation of a twin pregnancy (MZT or DZT); and (2) were followed through to delivery with complete obstetric and neonatal records.

Patients were excluded if they: (1) underwent multifetal pregnancy reduction; (2) had recurrent miscarriage or structural abnormalities of the reproductive tract; (3) had the pregnancy terminated due to systemic illness, uterine scarring, or personal reasons; (4) had pre-existing hypertension or diabetes mellitus before pregnancy; (5) conceived through donor oocytes, donor sperm, intrauterine insemination, or preimplantation genetic testing cycles; (6) were in a consanguineous relationship; (7) had a severe psychiatric disorder; or (8) had incomplete clinical records.

Embryo quality grading

Cleavage-stage embryos (D2/D3) were assessed morphologically according to the Istanbul Consensus and the ESHRE-ALPHA scoring criteria [17-19], taking into account blastomere number, fragmentation, and symmetry. On day 3, embryos meeting specific score combinations — such as 8CF4S1 — were classified as good-quality. Blastocysts (D5/D6) were graded using the Gardner-Schoolcraft system [18-23]. Day-5 blastocysts expanded to stage ≥ 3 with inner cell mass (ICM) and trophoctoderm (TE) both graded A or B (i.e., AA, AB, BA, or BB) were considered good-quality; for day-6 blastocysts, the same grading criteria applied with a minimum expansion stage of 4. When all transferred embryos were good-quality, the cycle was categorized as the "good-quality" group; if any transferred embryo was below this threshold, the cycle was assigned to the "non-good-quality" group.

Study variables

Baseline and treatment characteristics collected included maternal age, body mass index (BMI), history of previous miscarriage, infertility type (primary or secondary), cycle type (fresh or frozen-thawed), fertilization method (conventional IVF or ICSI), embryonic developmental stage at transfer (cleavage-stage D2/D3 or blastocyst D5/D6), number of embryos transferred (1 or ≥ 2), embryo quality (good-quality or non-good-quality), and the underlying causes of infertility.

Perinatal outcomes included mode of delivery, gestational age at delivery, neonatal birth weight, presence of any low birth weight (any LBW, defined as at least one infant weighing $< 2,500$ g), and pregnancy complications including hypertensive disorders of pregnancy, glucose metabolism abnormalities, and premature rupture of membranes.

Nine variables were entered into the logistic regression analyses as core predictors: advanced maternal age (≥ 35 years), maternal BMI, infertility type, history of previous miscarriage, embryonic developmental stage, number of embryos transferred, embryo quality, fertilization method, and cycle type.

Table 1. Baseline and treatment characteristics of the DZT group.

Variable	Non-SR(n=2456)	SR(n=807)	P value
Maternal age (years)	29.99 ± 3.67	31.86 ± 4.25	< 0.001
BMI(kg/m ²)	22.35 ± 3.14	22.61 ± 3.26	0.052
No. of previous miscarriages	0.45 ± 0.80	0.51 ± 0.84	0.084
Advanced maternal age (≥35 years), n (%)			< 0.001
No	2163 (88.1)	594 (73.6)	
Yes	293 (11.9)	213 (26.4)	
Infertility type, n (%)			0.104
Primary infertility	1332 (54.2)	464 (57.5)	
Secondary infertility	1124 (45.8)	343 (42.5)	
History of previous miscarriage, n (%)			0.305
No	1643 (66.9)	524 (64.9)	
Yes	813 (33.1)	283 (35.1)	
Infertility factors, n (%)			
Pelvic factor	1585 (64.5)	403 (49.9)	< 0.001
Tubal factor	1451 (59.1)	316 (39.1)	< 0.001
Endometriosis	143 (5.8)	46 (5.7)	0.891
Ovulatory disorder / ovarian factor	566 (23.0)	135 (16.7)	< 0.001
Uterine factor	110 (4.5)	45 (5.6)	0.206
Other factors	190 (7.7)	280 (34.7)	< 0.001
No female factor	365 (14.9)	40 (5.0)	< 0.001
Male factor infertility	675 (27.5)	336 (41.6)	< 0.001
Embryonic developmental stage, n (%)			< 0.001
D2/D3 (cleavage stage)	317 (12.9)	36 (4.4)	
D5/D6 (blastocyst stage)	2139 (87.1)	771 (95.6)	
No. of embryos transferred, n (%)			0.019
1	42 (1.7)	25 (3.0)	
≥2	2414 (98.3)	782 (97.0)	
Embryo quality, n (%)			< 0.001
Good quality	2154 (87.7)	648 (80.3)	
Non-good quality	302 (12.3)	159 (19.7)	
Fertilization method, n (%)			< 0.001
IVF	1793 (73.0)	514 (63.7)	
ICSI	663 (27.0)	293 (36.3)	
Cycle type, n (%)			< 0.001
Fresh cycle	550 (22.4)	67 (8.3)	
Frozen-thawed cycle	1906 (77.6)	740 (91.7)	

Note: Continuous variables are presented as mean ± SD; categorical variables as n (%). Pelvic factors include pelvic inflammatory disease and adhesions; tubal factors include tubal obstruction, hydrosalpinx, and prior tubal surgery; ovulatory disorder/ovarian factors include polycystic ovary syndrome, ovulatory dysfunction, and diminished ovarian reserve; uterine factors include fibroids, adenomyosis, intrauterine adhesions, and prior uterine surgery; other factors include thyroid disorders and systemic autoimmune conditions; male factor infertility includes oligozoospermia, asthenozoospermia, teratozoospermia, and azoospermia.

Statistical analysis

Statistical analyses were performed using IBM SPSS Statistics 25, and the prediction model was built and validated in RStudio (v4.5.1). Continuous variables with approximately normal distributions are presented as mean ± standard deviation (SD) and compared between groups using the independent-samples t-test (or Welch's t-test when variances were unequal). Categorical variables are presented as counts and percentages and compared using the Pearson chi-square test; Fisher's exact test or the continuity-corrected chi-square test was used when any expected cell count was less than 5. All tests were two-sided, and $P < 0.05$ was considered statistically significant. Risk factors for SR were examined by univariable and multivariable logistic regression, with effect sizes expressed as odds ratios (OR) with 95% confidence intervals (95% CI).

An SR risk prediction model was developed in the DZT cohort and internally validated using a 70:30 training-to-test split. Discrimination was assessed by the receiver operating characteristic (ROC) curve and the area under the curve (AUC); the optimal classification threshold was determined by the Youden index, from which sensitivity, specificity, and related metrics were derived. Calibration was evaluated using a calibration curve and the Brier score. Clinical utility was assessed by decision curve analysis (DCA). Risk stratification was performed by dividing the test-set predicted probabilities into tertiles, and SR incidence was compared across the resulting low-, intermediate-, and high-risk groups.

Results

Study population

A total of 3,768 patients with ART-conceived twin pregnancies were included: 3,263 in the DZT cohort (SR group, $n=807$; Non-SR group, $n=2,456$) and 505 in the MZT cohort (SR group, $n=62$; Non-SR group, $n=443$). The DZT cohort underwent full comparative analysis, logistic regression, and prediction model development; findings from the MZT cohort are presented descriptively.

Baseline and treatment characteristics in the DZT cohort

Baseline and treatment characteristics are shown in Table 1. Several differences were observed between the SR and Non-SR groups.

Regarding continuous variables, maternal age was significantly higher in the SR group than in the Non-SR group (31.86 ± 4.25 vs. 29.99 ± 3.67 years, $P < 0.001$), and BMI was also slightly higher (22.61 ± 3.26 vs. 22.35 ± 3.14 kg/m², $P = 0.052$). The number of previous miscarriages did not differ significantly between groups (0.51 ± 0.84 vs. 0.45 ± 0.80 , $P = 0.084$).

Among categorical variables, the proportion of women of advanced maternal age (≥35 years) was significantly higher in the SR group (26.4% vs. 11.9%, $P < 0.001$). Infertility type (pri-

mary vs. secondary) was similar between groups ($P = 0.104$). A history of previous miscarriage did not differ significantly between groups (SR 35.1% vs. Non-SR 33.1%, $P = 0.305$). The distribution of infertility causes also differed significantly, with notable differences in the proportions of pelvic, tubal, and ovarian factors (all $P < 0.001$ or $P = 0.001$), suggesting that the underlying cause of infertility may be related to SR risk. With respect to embryo and cycle characteristics, blastocyst-stage transfer (D5/D6) was more frequent in the SR group (95.6% vs. 87.1%, $P < 0.001$). The distribution of the number of embryos transferred also differed between groups, with ≥ 2 embryos being the predominant strategy in both ($P = 0.001$). The SR group had a higher rate of ICSI (36.3% vs. 27.0%, $P < 0.001$) and a markedly higher proportion of frozen-thawed cycles (91.7% vs. 77.6%, $P < 0.001$).

Risk factors for SR in the DZT cohort: logistic regression analysis

Univariable and multivariable logistic regression analyses were performed with SR as the outcome variable, incorporating the nine prespecified core predictors. The multivariable model was fitted using complete-case analysis.

Univariable analysis

On univariable analysis (Table 2), advanced maternal age (≥ 35 years) was significantly associated with SR (OR=2.65, 95% CI: 2.17–3.23, $P < 0.001$). Each 1 kg/m² increase in BMI was associated with a modest increase in SR risk (OR=1.03, 95% CI: 1.00–1.05, $P = 0.048$). History of previous miscarriage and infertility type did not reach statistical significance or showed unstable effect estimates in univariable analysis. Transfer of ≥ 2 embryos was associated with a lower likelihood of SR (OR=0.55, 95% CI: 0.33–0.91, $P = 0.021$). Non-good-quality embryos were associated with higher SR risk (OR=1.67, 95% CI: 1.35–2.06, $P < 0.001$), as were ICSI compared with conventional IVF (OR=1.54, 95% CI: 1.30–1.82, $P < 0.001$) and frozen-thawed cycles compared with fresh cycles (OR=2.95, 95% CI: 2.03–4.29, $P < 0.001$).

Multivariable analysis

After adjustment for all nine predictors (Table 2), advanced maternal age remained an independent risk factor for SR (aOR=3.00, 95% CI: 2.40–3.76, $P < 0.001$). BMI was no longer significant in the multivariable model (aOR=1.01, 95% CI: 0.99–1.04, $P = 0.344$). A history of previous miscarriage was independently associated with increased SR risk (aOR=1.32, 95% CI: 1.06–1.64, $P = 0.013$). Secondary infertility was inversely associated with SR (aOR=0.67, 95% CI: 0.54–0.83, $P < 0.001$). Embryonic developmental stage was independently associated with SR, with blastocyst-stage (D5/D6) transfer showing a substantially higher risk than cleavage-stage transfer (aOR = 3.05, 95% CI: 1.92–4.82, $P < 0.001$). Transfer of ≥ 2 embryos remained associated with lower SR risk (aOR=0.53, 95% CI: 0.31–0.90, $P = 0.019$). Non-good-quality embryos (aOR=1.79, 95% CI: 1.42–2.26, $P < 0.001$), ICSI (aOR=1.56, 95% CI: 1.30–1.88, $P < 0.001$), and frozen-thawed cycles (aOR=3.33, 95% CI: 2.23–4.99, $P < 0.001$) were each independently associated with increased SR risk; frozen-thawed cycle type showed the largest effect size among all predictors in the model.

Development and validation of the DZT-SR prediction model

A multivariable logistic prediction model was constructed using the nine core predictors described above. The regression equation is as follows:

$$\text{logit}(p) = -4.072 + 1.304 \times (\text{advanced maternal age} \geq 35 \text{ years}) + 0.056 \times \text{BMI} - 0.077 \times (\text{secondary infertility}) + 0.439 \times (\text{history of previous miscarriage}) + 1.243 \times (\text{blastocyst transfer D5/D6}) - 1.119 \times (\geq 2 \text{ embryos transferred}) - 0.621 \times (\text{non-good-quality embryos}) + 0.317 \times (\text{ICSI}) + 3.016 \times (\text{frozen-thawed cycle})$$

$p = 1 / [1 + \exp(-\text{logit}(p))]$ represents the predicted probability of SR.

Discrimination

ROC curve analysis showed good discriminative performance for the model (Figure 1): AUC was 0.868 in the training set and 0.857 in the test set. Using the Youden index, the optimal clas-

Table 2. Factors associated with SR: univariate and multivariable logistic regression

Variable	OR (95% CI)	P (univariable)	aOR (95% CI)	P value (multivariable)
Advanced maternal age (≥ 35 vs < 35 years)	2.65 (2.17-3.23)	< 0.001	3.00 (2.40-3.76)	< 0.001
Maternal BMI (per 1 kg/m ² increase)	1.03 (1.00-1.05)	0.048	1.01 (0.99-1.04)	0.344
Infertility type (secondary infertility)	0.87 (0.74-1.03)	0.104	0.67 (0.54-0.83)	< 0.001
Embryonic stage (D5/D6 blastocyst)	3.23 (2.29-4.57)	< 0.001	3.05 (1.92-4.82)	< 0.001
History of previous miscarriage (yes)	1.09 (0.92-1.29)	0.305	1.32 (1.06-1.64)	0.013
No. of embryos transferred (≥ 2)	0.55 (0.33-0.91)	0.021	0.53 (0.31-0.90)	0.019
Embryo quality (non-good quality)	1.67 (1.35-2.06)	< 0.001	1.79 (1.42-2.26)	< 0.001
Fertilization method (ICSI)	1.54 (1.30-1.82)	< 0.001	1.56 (1.30-1.88)	< 0.001
Cycle type (frozen-thawed cycle)	2.95 (2.03-4.29)	< 0.001	3.33 (2.23-4.99)	< 0.001

sification threshold was 0.501, at which the model achieved a sensitivity of 0.822 and a specificity of 0.830 in the test set. The positive predictive value (PPV) was 0.750, the negative predictive value (NPV) was 0.883, and overall accuracy was 0.827.

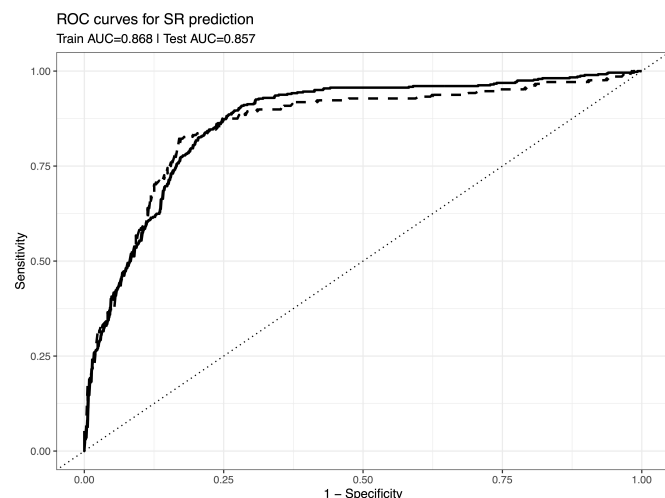


Figure 1. ROC curves of the DZT-SR prediction model in the training and test sets.

ROC curves for the spontaneous reduction (SR) prediction model in dizygotic twin (DZT) pregnancies, shown separately for the training set (solid line) and the test set (dashed line). The area under the curve (AUC) was 0.868 in the training set and 0.857 in the test set. The optimal classification threshold determined by the Youden index was 0.501, at which the model achieved a sensitivity of 0.822 and a specificity of 0.830 in the test set. The diagonal reference line represents a non-discriminating classifier.

Calibration

The calibration curve for the test set showed good overall agreement between predicted probabilities and observed SR rates across risk strata, with the curve tracking close to the 45° reference line and no obvious systematic over- or under-estimation across most of the risk range (Figure 2). The Brier score was 0.142, indicating that the overall probability prediction error was within an acceptable range.

Clinical net benefit

Decision curve analysis showed that, across a reasonable range of threshold probabilities, the net benefit of using the model exceeded both the "treat all" and "treat none" reference strategies (Figure 3). This suggests that applying the model to identify high-risk patients and guide follow-up decisions could help avoid unnecessary interventions while maintaining the ability to detect truly high-risk individuals

Risk stratification

To make the model more applicable in practice, predicted probabilities in the test set were divided into tertiles to form low-, intermediate-, and high-risk groups. Observed SR rates increased markedly across strata: 8.3% (15/181) in the low-risk group, 28.7% (52/181) in the intermediate-risk group, and 77.9% (141/181) in the high-risk group (Figure 4).

Perinatal outcomes in the DZT cohort

Perinatal outcome analyses were restricted to live births with complete delivery records. Compared with the Non-SR group, the SR group had a significantly later gestational age at delivery (38.42 ± 1.99 vs. 36.10 ± 2.04 weeks, $P < 0.001$) and a marked-

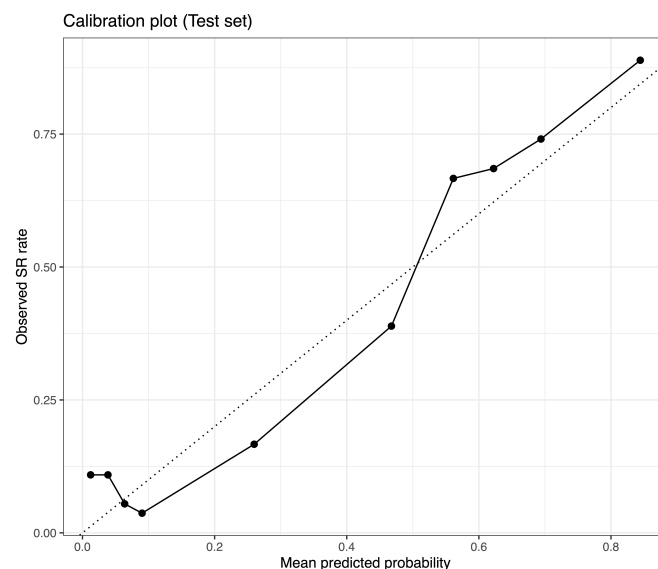


Figure 2. Calibration curve of the DZT-SR prediction model in the test set.

The calibration curve plots predicted probability of SR against the observed SR rate across risk strata in the test set. The diagonal dashed line represents perfect calibration (predicted probability = observed rate). Points falling close to this reference line indicate good agreement between model predictions and actual outcomes. The Brier score was 0.142, indicating acceptable overall prediction error.

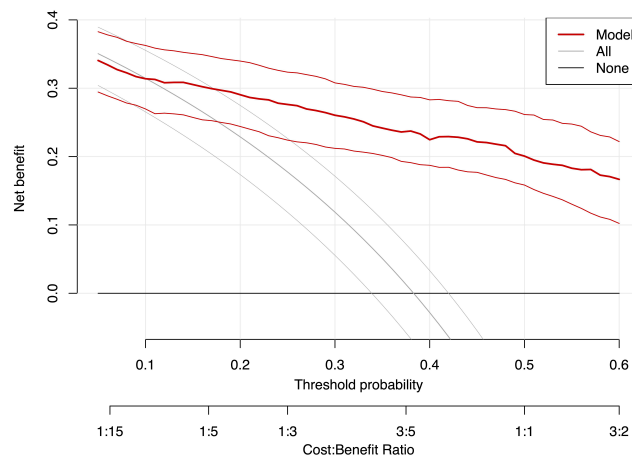


Figure 3. Decision curve analysis (DCA) for the DZT-SR prediction model.

Decision curve analysis evaluating the clinical net benefit of the SR prediction model across a range of threshold probabilities in the test set. The model curve (solid line) is compared with two reference strategies: treating all patients as high risk ("Treat all", gray line) and treating no patients as high risk ("Treat none", horizontal line at zero). The model provides greater net benefit than either reference strategy across a clinically meaningful range of threshold probabilities, supporting its potential utility for guiding individualized follow-up decisions in early pregnancy.

ly lower rate of preterm birth (13.6% vs. 61.1%, $P<0.001$). The smaller of the two birth weights was significantly higher in the SR group ($P<0.001$), and the proportion of pregnancies with any low birth weight infant (any LBW, $<2,500$ g) was substantially lower (8.7% vs. 53.3%, $P<0.001$). Birth weight discordance of $\geq 20\%$ (discord20) was observed in 15.6% of Non-SR deliveries; as most SR pregnancies resulted in singleton deliveries, this measure was effectively zero in the SR group ($P<0.001$). Overall rates of any pregnancy complication were similar between groups (21.2% vs. 20.8%, $P=0.816$). The rate of hypertensive disorders of pregnancy was slightly lower in the SR group but did not reach significance (6.7% vs. 8.8%, $P=0.060$). Glucose metabolism abnormalities were more frequent in the SR group (8.0% vs. 3.7%, $P<0.001$), whereas premature rupture of membranes was less common (2.2% vs. 5.1%, $P<0.001$). Detailed perinatal outcomes are presented in Table 3.

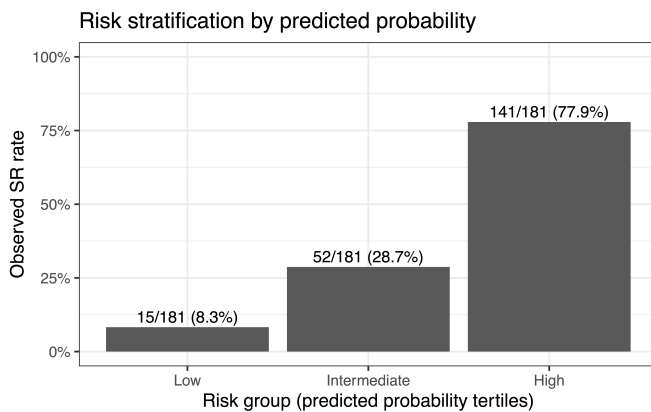


Figure 4. Risk stratification by predicted probability tertiles in the test set.

Bar chart showing the observed SR incidence across three risk strata defined by tertiles of the model-predicted probability in the test set ($n=543$). The low-risk group (first tertile) had an SR rate of 8.3% (15/181), the intermediate-risk group (second tertile) 28.7% (52/181), and the high-risk group (third tertile) 77.9% (141/181). The stepwise increase in SR incidence across strata confirms the model's risk stratification ability.

Baseline and treatment characteristics in the MZT cohort

The MZT cohort was smaller, comprising 443 women in the Non-SR group and 62 in the SR group. Among continuous variables, maternal age was slightly higher in the SR group but the difference did not reach significance (31.73 ± 4.45 vs. 30.76 ± 4.28 years, $P=0.098$); BMI ($P=0.887$) and number of previous miscarriages ($P=0.131$) were also similar between groups. Among categorical variables, the proportion of women with advanced maternal age was higher in the SR group (25.8% vs. 17.4%), though again without statistical significance ($P=0.109$). Infertility type, history of previous miscarriage, and most infertility-related factors including pelvic and tubal causes showed no significant differences. One exception was uterine factor infertility, which was more common in the SR group (14.5% vs. 6.3%, $P=0.033$). Embryonic developmental stage, number of embryos transferred, embryo quality, fertilization method, and cycle type were all comparable between groups (all $P>0.05$). Full details are provided in Table 4.

Table 3. Comparison of perinatal outcomes in the dizygotic twin group.

Variable	Non-SR	SR	P value
Gestational age at delivery (weeks)	36.10 $\pm$ 2.04	38.42 $\pm$ 1.99	< 0.001
Preterm birth, n (%)	1500 (61.1)	109 (13.6)	< 0.001
Cesarean delivery, n (%)	2326 (94.7)	572 (70.9)	< 0.001
Smaller birth weight (g)	2398.70 $\pm$ 462.32	3284.94 $\pm$ 596.27	< 0.001
Any LBW (<2500 g), n (%)	1309 (53.3)	70 (8.7)	< 0.001
Birth weight discordance $\geq 20\%$, n (%)	383 (15.6)	—	—
Any complication, n (%)	511 (20.8)	171 (21.2)	0.816
Hypertensive disorders of pregnancy, n (%)	216 (8.8)	54 (6.7)	0.060
Glucose metabolism abnormality, n (%)	91 (3.7)	64 (8.0)	< 0.001
Premature rupture of membranes, n (%)	125 (5.1)	17 (2.2)	< 0.001

Note: Continuous variables are presented as mean $\pm$ SD; categorical variables as n (%). Analyses were restricted to live births with complete delivery records. LBW, low birth weight; any LBW, at least one infant with birth weight <2500 g; birth weight discordance $\geq 20\%$, defined as (larger birth weight – smaller birth weight) / larger birth weight ≥ 0.20 .

Perinatal outcomes in the MZT cohort

In the MZT cohort, the SR group delivered at a significantly later gestational age (37.49 ± 3.22 vs. 35.63 ± 1.78 weeks, $P<0.001$) and had a markedly lower rate of preterm birth (30.6% vs. 83.5%, $P<0.001$). The cesarean delivery rate was also significantly lower in the SR group (71.0% vs. 96.8%, $P<0.001$). The smaller birth weight was higher in the SR group ($3,063.33\pm 716.92$ vs. $2,251.06\pm 508.81$ g, $P<0.001$), and the rate of any LBW was substantially lower (17.7% vs. 65.2%, $P<0.001$). The overall rate of any pregnancy complication did not differ between groups (24.2% vs. 20.5%, $P=0.508$), nor did rates of hypertensive disorders of pregnancy ($P=0.339$) or premature rupture of membranes ($P=0.501$). Glucose metabolism abnormalities could not be formally tested due to the small number of events. Perinatal outcomes for the MZT cohort are summarized in Table 5.

Discussion

This study analyzed risk factors for SR and compared perinatal outcomes in a large single-center cohort of ART-conceived twin pregnancies, and developed and internally validated an SR prediction model in the DZT cohort. The findings suggest that SR is not a random occurrence but is closely tied to maternal age, embryo characteristics, and treatment cycle factors.

Interpretation of risk factors for SR

Advanced maternal age (≥ 35 years) was the strongest independent risk factor for SR in this study ($aOR=3.00$), consistent with previous reports describing older maternal age in VTS populations [9, 11]. The biological rationale is reasonably straightforward: as women age, the rate of oocyte aneuploidy

risers and the proportion of chromosomally abnormal embryos increases, raising the likelihood of early embryonic arrest. In a twin pregnancy, this may manifest as one embryo failing to develop further. Older age may also be accompanied by reduced endometrial receptivity, which could further compromise early placental establishment.

Frozen-thawed cycle type had the largest effect size in the

Table 4. Baseline and treatment characteristics of the monozygotic twin group.

Variable	Non-SR (n = 443)	SR (n = 62)	P value
Maternal age (years)	30.76 ± 4.28	31.73 ± 4.45	0.098
BMI(kg/m ²)	22.49 ± 3.25	22.43 ± 2.76	0.887
No. of previous miscarriages	0.44 ± 0.72	0.69 ± 1.29	0.131
Advanced maternal age (≥35 years), n (%)			0.109
No	366 (82.6)	46 (74.2)	
Yes	77 (17.4)	16 (25.8)	
Infertility type, n (%)			0.301
Primary infertility	224 (50.6)	27 (43.5)	
Secondary infertility	219 (49.4)	35 (56.5)	
History of previous miscarriage, n (%)			0.518
No	297 (67.0)	39 (62.9)	
Yes	146 (33.0)	23 (37.1)	
Infertility factors, n (%)			
Pelvic factor	239 (54.0)	28 (45.2)	0.222
Tubal factor	153 (34.5)	20 (32.3)	0.723
Endometriosis	41 (9.3)	7 (11.3)	0.609
Ovulatory disorder / ovarian factor	81 (18.3)	9 (14.5)	0.468
Uterine factor	28 (6.3)	9 (14.5)	0.033
Other factors	55 (12.4)	9 (14.5)	0.641
No female factor	46 (10.4)	10 (16.1)	0.177
Male factor infertility	100 (22.6)	18 (29.0)	0.260
Embryonic developmental stage, n (%)			0.332
D2/D3 (cleavage stage)	18 (4.0)	4 (6.5)	
D5/D6 (blastocyst stage)	425 (96.0)	58 (93.5)	
No. of embryos transferred, n (%)			0.434
1	266 (60.1)	34 (54.8)	
≥2	177 (39.9)	28 (45.2)	
Embryo quality, n (%)			0.652
Good quality	401 (90.5)	55 (88.7)	
Non-good quality	42 (9.5)	7 (11.3)	
Fertilization method, n (%)			0.128
IVF	307 (69.3)	37 (59.7)	
ICSI	136 (30.7)	25 (40.3)	
Cycle type, n (%)			0.392
Fresh cycle	80 (18.1)	14 (22.6)	
Frozen-thawed cycle	363 (81.9)	48 (77.4)	

Note: Continuous variables are presented as mean ± SD; categorical variables as n (%). Infertility factor categories are defined as in Table 1.

Table 5. Comparison of perinatal outcomes in the monozygotic twin group.

Variable	Non-SR	SR	P value
Gestational age at delivery (weeks)	35.63 ± 1.78	37.49 ± 3.22	< 0.001
Preterm birth, n (%)	370 (83.5)	19 (30.6)	< 0.001
Cesarean delivery, n (%)	429 (96.8)	44 (71.0)	< 0.001
Smaller birth weight (g)	2251.06 ± 508.81	3063.33 ± 716.92	< 0.001
Any LBW (<2500 g), n (%)	289 (65.2)	11 (17.7)	< 0.001
Birth weight discordance ≥20%, n (%)	53 (12.0)	—	—
Any complication, n (%)	91 (20.5)	15 (24.2)	0.508
Hypertensive disorders of pregnancy, n (%)	40 (9.0)	3 (4.8)	0.339
Glucose metabolism abnormality, n (%)	27 (6.1)	3 (4.8)	—
Premature rupture of membranes, n (%)	21 (4.7)	1 (1.6)	0.501

model (aOR=3.33). These cycles are often used when fresh transfers have previously failed or when endometrial preparation is needed, and the patients selected for them may already carry a higher background risk of early pregnancy complications. Beyond patient selection, individual variation in embryo-endometrial synchrony during frozen-thawed transfer may affect the stability of early placentation [24]. Blastocyst-stage transfer (aOR=3.05) was also associated with higher SR risk.

Two explanations seem plausible: first, blastocysts implant more efficiently and are more likely to be identified as twin pregnancies on the initial ultrasound, making subsequent SR easier to detect on follow-up; second, patients selected for blastocyst transfer may represent more complex treatment histories that independently carry higher SR risk [18, 22].

Non-good-quality embryos (aOR=1.79) and ICSI (aOR=1.56) were both independently associated with SR, pointing toward embryo developmental potential as a common underlying theme. Embryos of lower morphological quality may have limited capacity to sustain development after implantation. For ICSI, the association may partly reflect the characteristics of the patient population — particularly those with severe male factor infertility — rather than the procedure itself [18, 22]. A history of previous miscarriage (aOR=1.32) may reflect underlying endometrial dysfunction that compromises early implantation stability.

The inverse association between transferring ≥2 embryos and SR risk (aOR=0.53) requires some clarification. This study enrolled only patients with confirmed twin pregnancies, so the comparison concerns who among established twin pregnancies is more likely to undergo SR, not who is more likely to form a multiple pregnancy in the first place. When two embryos are transferred, clinicians may select a better-matched embryo combination, or follow-up management for these patients may differ in ways that contribute to greater twin pregnancy stability. BMI showed a weak association with SR on univariable analysis but lost significance after adjustment, suggesting it may act through shared pathways with age or cycle strategy rather than as an independent contributor. Secondary infertility was associated with lower SR risk (aOR=0.67), possibly because these patients have had at least one prior live birth and

may have a more favorable baseline endometrial environment, though the underlying mechanism warrants further study.

Perinatal outcomes after SR

The SR group delivered at a later gestational age and had substantially lower rates of preterm birth and low birth weight compared with the ongoing twin group. This is broadly consistent with published systematic reviews showing that perinatal outcomes in SR-survivor pregnancies tend to be better than those in continuing twin pregnancies, while remaining somewhat worse than in pregnancies that were singletons from the start [9-10]. Early exposure to a twin implantation environment and the more complex placental bed remodeling that accompanies it may have lasting effects on placental function in the surviving fetus. The higher proportions of advanced maternal age and frozen-thawed cycles in the SR group may also independently contribute to perinatal risk. The higher rate of glucose metabolism abnormalities in the SR group (8.0% vs. 3.7%) alongside the lower rate of premature rupture of membranes (2.2% vs. 5.1%) suggests that the complication profile of SR-survivor pregnancies may have some distinctive features that deserve further investigation.

Prediction model performance and clinical implications

The DZT-SR prediction model was built entirely from variables available through routine clinical assessment. Internal validation yielded an AUC of 0.857, with sensitivity and specificity both exceeding 0.82 at the optimal threshold. Calibration and DCA net benefit were both acceptable. After stratifying by predicted probability tertiles, observed SR rates rose from 8.3% in the low-risk group to 77.9% in the high-risk group, providing a practical basis for tailoring follow-up intensity in early pregnancy. Women in the low-risk group can be managed with standard follow-up, while those in the high-risk group may benefit from shorter ultrasound intervals, closer monitoring for gestational sac development and intrauterine hematoma, and additional psychological support. The model should be used as a risk stratification aid rather than a definitive diagnostic tool, and clinical judgment remains essential.

Limitations and future directions

Several limitations should be noted. First, the retrospective single-center design introduces the potential for selection and information bias. Second, SR was not subclassified by whether only a gestational sac disappeared or whether cardiac activity had already been detected before arrest, which may affect incidence estimates. Third, data on chorionicity, ovarian stimulation protocol, luteal support regimen, first-trimester intrauterine hematoma, and embryo genetic testing results were unavailable, limiting mechanistic interpretation. Fourth, the MZT cohort was too small to support reliable multivariable modeling. Fifth, only internal validation was performed, and the model's generalizability needs to be tested in external cohorts. Future work could address these gaps through multicenter prospective studies with standardized SR definitions and follow-up protocols; by incorporating chorionicity, intrauterine hematoma, and embryo genetic information to explore biological mechanisms; by externally validating the model and translating it into more accessible formats such as a nomogram or online calculator; and by accumulating larger MZT samples to enable a more rigorous comparison of SR patterns and perina-

tal outcome profiles between MZT and DZT pregnancies.

Conclusion

This study found that SR in ART-conceived twin pregnancies is independently associated with advanced maternal age, history of previous miscarriage, blastocyst-stage transfer, non-good-quality embryos, ICSI, and frozen-thawed cycles, while secondary infertility and transfer of ≥ 2 embryos were associated with lower SR risk. Perinatal outcomes in SR-survivor pregnancies were better than those in ongoing twin pregnancies overall, though the higher rate of glucose metabolism abnormalities points to a somewhat distinct complication profile. The DZT-SR prediction model, built on nine routinely available clinical variables, showed good internal validation performance (AUC=0.857) and clear risk stratification ability, and may support individualized follow-up planning in early pregnancy. Multicenter prospective studies and external validation are needed to strengthen the evidence base and improve the model's broader applicability.

Abbreviations

SR: Spontaneous reduction; ART: Assisted reproductive technology; DZT: Dizygotic twins; MZT: Monozygotic twins; ROC: Receiver operating characteristic; AUC: Area under the curve; DCA: Decision curve analysis; ICSI: Intracytoplasmic sperm injection; aOR: Adjusted odds ratio; WHO: World Health Organization; eSET: Elective single embryo transfer; VTS: Vanishing twin syndrome; BMI: Body mass index; IVF: In vitro fertilization; ICM: Inner cell mass; TE: Trophoctoderm; LBW: Low birth weight; SD: Standard deviation; OR: Odds ratio; CI: Confidence interval; PPV: Positive predictive value; NPV: Negative predictive value.

Author Contributions

Zhaolian Wei: Conceptualization, Project administration, Resources, Supervision, Writing – review & editing. Shuqi Cai: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft.

Acknowledgements

The authors would like to thank the staff of the Reproductive Center of the First Affiliated Hospital of Anhui Medical University for their assistance with data collection and clinical follow-up.

Funding Information

Not Applicable.

Ethics Approval and Consent to Participate

This study was approved by the Committee on Medical Ethics of the First Affiliated Hospital of Anhui Medical University (Reference number: Quick-PJ 2018-07-05). The study was conducted in accordance with the Declaration of Helsinki and relevant ethical guidelines for biomedical research involving human subjects. As this was a retrospective study based on routinely collected clinical data, the requirement for individual informed consent was waived by the ethics committee.

Competing Interests

The authors declare that they have no existing or potential commercial or financial relationships that could create a conflict of interest at the time of conducting this study.

Data Availability

Not Applicable.

References

- [1] Organization WH. (2023). Infertility prevalence estimates, 1990–2021. World Health Organization. <http://digitallibrary.un.org/record/4012749>
- [2] Multiple gestation associated with infertility therapy: an American Society for Reproductive Medicine Practice Committee opinion. (2012). *Fertility and Sterility*, 97(4), 825–834. <https://doi.org/https://doi.org/10.1016/j.fertnstert.2011.11.048>
- [3] Guidance on the limits to the number of embryos to transfer: a committee opinion. (2021). *Fertility and Sterility*, 116(3), 651–654. <https://doi.org/https://doi.org/10.1016/j.fertnstert.2021.06.050>
- [4] Multifetal Gestations: Twin, Triplet, and Higher-Order Multifetal Pregnancies: ACOG Practice Bulletin, Number 231. (2021). *Obstet Gynecol*, 137(6), e145–e162. <https://doi.org/10.1097/aog.0000000000004397>
- [5] Santana DS, Cecatti JG, Surita FG, Silveira C, Costa ML, Souza JP, et al. (2016). Twin Pregnancy and Severe Maternal Outcomes: The World Health Organization Multicountry Survey on Maternal and Newborn Health. *Obstetrics & Gynecology*, 127(4), 631–641. <https://doi.org/10.1097/aog.0000000000001338>
- [6] Alteri A, Arroyo G, Baccino G, Craciunas L, De Geyter C, Ebner T, et al. (2024). ESHRE guideline: number of embryos to transfer during IVF/ICSI†. *Hum Reprod*, 39(4), 647–657. <https://doi.org/10.1093/humrep/deae010>
- [7] Gerris JM. (2005). Single embryo transfer and IVF/ICSI outcome: a balanced appraisal. *Hum Reprod Update*, 11(2), 105–121. <https://doi.org/10.1093/humupd/dmh049>
- [8] Sukur YE, Altun T, & Pal L. (2012). Predictors of spontaneous reduction in multiple pregnancies conceived following assisted reproductive technology. *Eur J Obstet Gynecol Reprod Biol*, 162(2), 174–177. <https://doi.org/10.1016/j.ejogrb.2012.02.031>
- [9] Zhou F, Li Y, Wang H, Liu S, & Wang X. (2020). Perinatal outcomes in vanishing twin pregnancies following assisted reproductive technology (ART) - a systematic review and meta-analysis. *J Perinat Med*, 48(7), 639–647. <https://doi.org/10.1515/jpm-2020-0088>
- [10] Sun L, Jiang LX, & Chen HZ. (2017). Obstetric outcome of vanishing twins syndrome: a systematic review and meta-analysis. *Arch Gynecol Obstet*, 295(3), 559–567. <https://doi.org/10.1007/s00404-017-4289-9>
- [11] Carlson K, Mikes BA. Vanishing Twin Syndrome. [Updated 2025 Nov 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK563220/>
- [12] Ge Y, Lai S, Li X, Shi J, Ma C, & Zhao J. (2022). Vanishing twin syndrome is associated with first-trimester intrauterine hematoma in twin pregnancies after in vitro fertilization. *Front Endocrinol (Lausanne)*, 13, 1062303. <https://doi.org/10.3389/fendo.2022.1062303>
- [13] Vitthala S, Gelbaya TA, Brison DR, Fitzgerald CT, & Nardo LG. (2009). The risk of monozygotic twins after assisted reproductive technology: a systematic review and meta-analysis. *Hum Reprod Update*, 15(1), 45–55. <https://doi.org/10.1093/humupd/dmn045>
- [14] Busnelli A, Dallagiovanna C, Reschini M, Paffoni A, Fedele L, & Somigliana E. (2019). Risk factors for monozygotic twinning after in vitro fertilization: a systematic review and meta-analysis. *Fertil Steril*, 111(2), 302–317. <https://doi.org/10.1016/j.fertnstert.2018.10.025>
- [15] Hviid KVR, Malchau SS, Pinborg A, & Nielsen HS. (2018). Determinants of monozygotic twinning in ART: a systematic review and a meta-analysis. *Hum Reprod Update*, 24(4), 468–483. <https://doi.org/10.1093/humupd/dmy006>
- [16] Kanter JR, Boulet SL, Kawwass JF, Jamieson DJ, & Kissin DM. (2015). Trends and correlates of monozygotic twinning after single embryo transfer. *Obstet Gynecol*, 125(1), 111–117. <https://doi.org/10.1097/aog.0000000000000579>
- [17] Lundin K, & Ahlström A. (2015). Quality control and standardization of embryo morphology scoring and viability markers. *Reprod Biomed Online*, 31(4), 459–471. <https://doi.org/10.1016/j.rbmo.2015.06.026>
- [18] The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. (2011). *Hum Reprod*, 26(6), 1270–1283. <https://doi.org/10.1093/humrep/der037>
- [19] Coticchio G, Ahlström A, Arroyo G, Balaban B, Campbell A, De Los Santos MJ, et al. (2025). The Istanbul Consensus update: a revised ESHRE/ALPHA consensus on oocyte and embryo static and dynamic morphological assessment(† ±). *Reprod Biomed Online*, 50(6), 104955. <https://doi.org/10.1016/j.rbmo.2025.104955>
- [20] Schoolcraft WB, Gardner DK, Lane M, Schlenker T, Hamilton F, & Meldrum DR. (1999). Blastocyst culture and transfer: analysis of results and parameters affecting outcome in two in vitro fertilization programs. *Fertil Steril*, 72(4), 604–609. [https://doi.org/10.1016/s0015-0282\(99\)00311-8](https://doi.org/10.1016/s0015-0282(99)00311-8)
- [21] De los Santos MJ, Apter S, Coticchio G, Debrock S, Lundin K, Plancha CE, et al. (2016). Revised guidelines for good practice in IVF laboratories (2015). *Hum Reprod*, 31(4),

- 685–686. <https://doi.org/10.1093/humrep/dew016>
- [22] Gardner DK, Schoolcraft WB, Wagley L, Schlenker T, Stevens J, & Hesla J. (1998). A prospective randomized trial of blastocyst culture and transfer in in-vitro fertilization. *Hum Reprod*, 13(12), 3434–3440. <https://doi.org/10.1093/humrep/13.12.3434>
- [23] Gardner DK, Lane M, Stevens J, Schlenker T, & Schoolcraft WB. (2000). Blastocyst score affects implantation and pregnancy outcome: towards a single blastocyst transfer. *Fertil Steril*, 73(6), 1155–1158. [https://doi.org/10.1016/s0015-0282\(00\)00518-5](https://doi.org/10.1016/s0015-0282(00)00518-5)
- [24] Committee Opinion No. 719: Multifetal Pregnancy Reduction. (2017). *Obstet Gynecol*, 130(3), e158–e163. <https://doi.org/10.1097/aog.0000000000002302>