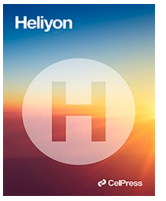
Heliyon 10 (2024) e28520

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



Research article

Identification and validation of sepsis subphenotypes using time-series data

Chenxiao Hao [a], Rui Hao [b], Huiying Zhao [a,*], Yong Zhang [c,**], Ming Sheng [c], Youzhong An [a]

[a] Department of Critical Care Medicine, Peking University People's Hospital, Beijing, 100044, China
[b] School of Computer Science, Beijing University of Posts and Telecommunications, Haidian District, Beijing, 100876, China
[c] BNRist, DCST, RIIT, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:
Sepsis
Subclasses
Clustering
Time-series data
Dynamic time warping

ABSTRACT

Purpose: The recognition of sepsis as a heterogeneous syndrome necessitates identifying distinct subphenotypes to select targeted treatment.
Methods: Patients with sepsis from the MIMIC-IV database (2008–2019) were randomly divided into a development cohort (80%) and an internal validation cohort (20%). Patients with sepsis from the ICU database of Peking University People's Hospital (2008–2022) were included in the external validation cohort. Time-series k-means clustering analysis and dynamic time warping was performed to develop and validate sepsis subphenotypes by analyzing the trends of 21 vital signs and laboratory indicators within 24 h after sepsis onset. Inflammatory biomarkers were compared in the ICU database of Peking University People's Hospital, whereas treatment heterogeneity was compared in the MIMIC-IV database.
Findings: Three sub-phenotypes were identified in the development cohort. Type A patients (N = 2525, 47%) exhibited stable vital signs and fair organ function, type B (N = 1552, 29%) was exhibited an obvious inflammatory response and stable organ function, and type C (N = 1251, 24%) exhibited severely impaired organ function with a deteriorating tendency. Type C demonstrated the highest mortality rate (33%) and levels of inflammatory biomarkers, followed by type B (24%), whereas type A exhibited the lowest mortality rate (11%) and levels of inflammatory biomarkers. These subphenotypes were confirmed in both the internal and external cohorts, demonstrating similar features and comparable mortality rates. In type C patients, survivors had significantly lower fluid intake within 24 h after sepsis onset (median 2891 mL, interquartile range (IQR) 1530–5470 mL) than that in non-survivors (median 4342 mL, IQR 2189–7305 mL). For types B and C, survivors showed a higher proportion of indwelling central venous catheters ($p < 0.05$).
Conclusion: Three novel phenotypes of patients with sepsis were identified and validated using time-series data, revealing significant heterogeneity in inflammatory biomarkers, treatments, and consistency across cohorts.

Abbreviations: DTW, dynamic time warping; MIMIC-IV, Medical Information Mart for Intensive Care IV; SSE, sum of squared errors; CRP, C-reactive protein; PCT, procalcitonin; IL, interleukin; IQR, interquartile range; LOS, length of stay; SD, standard deviation.

* Corresponding author. No 11, Southern Xizhimen Road, Beijing, China.
** Corresponding author. No. 30 Shuangqing Road, Beijing, China.
E-mail addresses: zhaohuiying@pkuph.edu.cn (H. Zhao), zhangyong05@tsinghua.edu.cn (Y. Zhang).

https://doi.org/10.1016/j.heliyon.2024.e28520
Received 25 August 2023; Received in revised form 10 March 2024; Accepted 20 March 2024
Available online 25 March 2024

1. Introduction

Sepsis is a serious systemic infection that can lead to multiple organ dysfunction and death [1]. In the last decade, sepsis incidence was 437 per 100,000 people, severe sepsis incidence was 270 per 100,000 people, in-hospital mortality rate of sepsis was 17%, and in-hospital mortality rate of severe sepsis was 26%, resulting in an estimated 48.9 million confirmed cases contributing to approximately 11 million deaths annually worldwide [2]. Morbidity and mortality associated with sepsis increase annually, making it a growing global public health concern [3].

Sepsis is not a specific disease; rather, it is an infection-induced syndrome. The clinical manifestations of sepsis are diverse owing to different pathogens and pathogenic mechanisms. Recently, exploring the subphenotypes of sepsis and the heterogeneity of its treatment has become a prominent focus in sepsis-related research. Most studies on sepsis subtypes are based on cross-sectional clustering [4–13]. Several recent studies have used time-series data for cluster classification in sepsis, examining vital sign [14], temperature [15], and SOFA score trajectories [16]. However, the clustering variables are relatively few and do not reflect the overall condition of patients. A summary of previous studies on sepsis subtypes is presented in Supplementary Material: Table S1. Considering the complex and rapidly evolving characteristics of sepsis, cluster analysis based on multi-systematic and comprehensive time-series data of easily accessible clinical indicators is more consistent with the characteristics of sepsis, providing better classification [17,18].

Owing to the heterogeneity of sepsis, there is no one-size-fits-all approach to its treatment [19]. Exploring precision treatment for sepsis based on its subtypes has become a new research direction [6,7,9,14,20].

The objectives of this study were to (1) discern and validate the subphenotypes of patients with sepsis based on time-series data of vital signs and laboratory measurements within the first 24 h of onset, (2) analyze the clinical features and outcomes of the sepsis subphenotypes, and (3) explore the therapeutic heterogeneity of early fluid resuscitation and indwelling central venous catheters

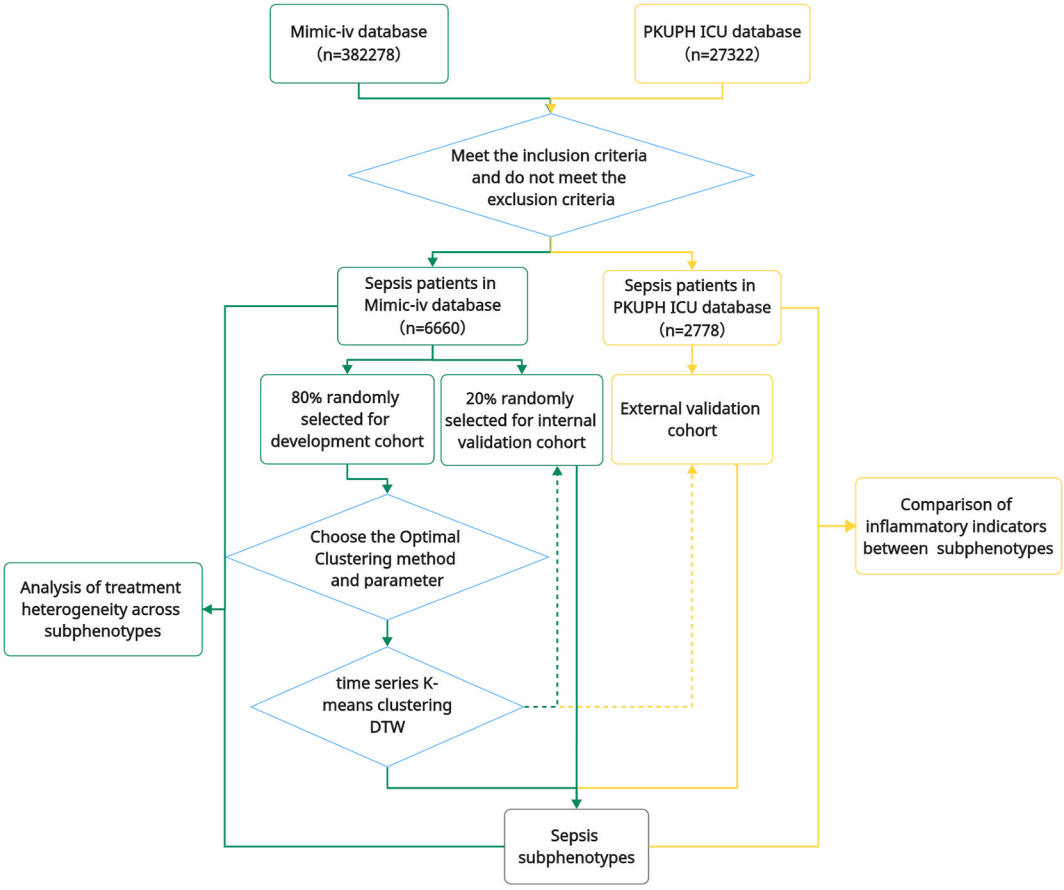


Fig. 1. Workflow of the study. Patients with sepsis were selected in the Mimic-iv and PKUPH ICU databases according to the definition of sepsis 3.0. Patients in the Mimic-iv database were randomly divided 80%/20% into the development cohort and internal validation cohort, respectively. The PKUPH ICU database was used as an external validation cohort. The optimal clustering method and parameter, i.e., time series k-means clustering, were explored and established in the development cohort. The same clustering method and variables were applied in the validation cohorts to verify its reproducibility. We finally obtained three subphenotypes in both validation cohorts, and their characteristics corresponded to those of the development cohort. We compared the inflammatory biomarkers between three subphenotypes in the PKUPH ICU database and analyzed treatment heterogeneity across subphenotypes in the Mimic-iv database.

between subphenotypes and the mechanisms underlying the different effects.

2. Materials and methods

2.1. Study design

We conducted a retrospective, observational study of patients with sepsis in the Medical Information Mart for Intensive Care IV (MIMIC-IV) and ICU databases of Peking University People's Hospital (hereinafter referred to as the PKUPH ICU database), whose overall workflow is illustrated in Fig. 1. Our goal was to develop and validate sepsis subphenotypes of patients in the ICU according to their trends in clinically accessible vital signs and laboratory measures using time series k-means clustering and dynamic time warping (DTW). We subsequently explored the heterogeneity of inflammatory biomarker levels and treatment effects among the three subphenotypes in two separate databases.

2.2. Study cohort

The MIMIC-IV database, collectively developed by the Massachusetts Institute of Technology, Harvard Medical School, and other institutions, is a large-scale, multimodal, open-source repository. It contains data from approximately 380,000 patients gathered between 2008 and 2019, covering comprehensive information on all patients during hospitalization, including clinical and laboratory data, medication and surgical information, and vital sign data [21,22]. The MIMIC-IV database was randomly divided 80%/20% into a development cohort and an internal validation cohort, respectively. The external validation cohort (PKUPH ICU database) stored comprehensive information on all patients admitted to the PKUPH ICU from 2008 to 2022. According to the sepsis 3.0 definition of coexisting infection and organ dysfunction, patients meeting the inclusion criteria were identified as having sepsis if: (i) they had a SOFA score $\geq$2 within 24 h after admission to the ICU; (ii) antibiotics were administered within 24 h after admission to the ICU; and (iii) microbiological examination was conducted within 24 h after admission to the ICU. If there were multiple hospitalizations or admissions to the ICU ward during hospitalization, the first ICU admission was considered. Patients with one of these variables completely missing were excluded.

2.3. Ethics statement

The study was approved by Peking University People's Hospital (2023PHB099-001) and performed in accordance with the principles of the Declaration of Helsinki.

2.4. Clustering variables

According to the definition of sepsis as fatal organ insufficiency caused by the dysregulation of the host response to infection [23], our study included indicators of infection, organ function parameters, and vital signs as clustering variables. In total, 21 clinical variables were included: heart rate (HR), respiratory rate (RR), mean arterial pressure, body temperature (T), white blood cell (WBC) count, platelet count, hemoglobin, serum creatinine (Scr), blood urea nitrogen, total bilirubin (tBil), pH, lactic acid (lac), prothrombin time (PT), activated partial prothrombin time, albumin, blood potassium, blood sodium, glucose, blood chloride, carbon dioxide partial pressure, and oxygenation index (pao2/fio2 ratio, p/f ratio).

The time at which the diagnostic criteria for sepsis were met was defined as the onset time for each patient. The first 6 h after onset was divided into hourly intervals, and from 6 to 24 h, it was divided into 2-h intervals. The worst values of the 21 variables within each time interval were recorded and compared among the three subphenotypes. The PKUPH ICU and MIMIC-IV databases had different degrees of missing data (Supplementary Material: Table S2). Missing values were imputed using Self-Attention Imputation for Time Series (Supplementary Material: Appendix 1).

2.5. Time series k-means clustering and dynamic time warping

Time series k-means clustering was applied to the 21 clustering variables in both the development and validation cohorts. Time-series k-means clustering is an extension of the k-means clustering algorithm, which is primarily used to divide data points with similar time patterns into distinct categories. The goal of the algorithm is to divide these time-series data points into k different classes, such that the time-series data points within the same class have similar temporal patterns. The distance between the two time-series data points was calculated using DTW (Supplementary Material: Appendix 2). The selection of the number of clusters, k, was based on the following three indicators: (1) the sum of squared errors (SSE), a metric used to evaluate the similarity between data within a group; the higher the intragroup similarity, the lower the SSE; (2) entropy, a metric used to evaluate the distribution between groups; the more even the distribution of groups, the higher the entropy; and (3) subgroup distribution, ensuring that each group contained >5% of the cohort.

2.6. Association of subphenotypes with mortality and inflammatory biomarkers

The primary prognostic outcome was 28-day mortality. Inflammatory biomarkers from the first 24 h after onset were compared

among the three subphenotypes: C-reactive protein (CRP), procalcitonin (PCT), and interleukin (IL)-6. For patients with multiple measurements of a laboratory indicator, the maximum value recorded within the first 24 h after onset was used.

2.7. Heterogeneity of treatment effect among subphenotypes

In accordance with the recommendations of the SSC sepsis guidelines and considering the completeness of medical practice records in the MIMIC-IV database, two therapeutic measures widely used in clinical practice were selected for this study: indwelling central venous catheter and early fluid resuscitation. Each sub-phenotype was categorized into survival and non-survival groups. The proportions of the survival and non-survival groups within each subphenotype were compared for the use of indwelling venous catheters. Additionally, the volumes of fluids administered at 1, 3, 6, and 24 h of illness were compared.

2.8. Statistical analyses

All statistical analyses were performed using R software version 4.2.1 for Windows (R Foundation for Statistical Computing) and STATA/MP version 16.0. The R packages used were mice, survival, and random forest. Survival analysis for 28 d was performed using Kaplan–Meier curves. Statistical significance was set at $p < 0.05$, and all tests were 2-tailed. Detailed descriptions of the statistical analyses are provided in Supplementary Material: Appendix 3.

3. Results

3.1. Cohort characteristics

The development cohort consisted of 5328 patients with sepsis, with a median age of 66 years (interquartile range (IQR) [54,78]), including 2952 males (55.4%). The 28-day mortality rate was 19.8%, and the median ICU length of stay (LOS) was 3.4 d (IQR [1.9–6.9]). The mean baseline SOFA and Elixhauser scores obtained within the first 24 h after ICU admission were 7.03 (standard deviation (SD): 3·9) and 20.79 (SD: 17.0), respectively. The internal validation cohort included 1332 patients with sepsis, with a median age of 66 years (IQR [55,77]), including 738 males (55.4%). Its 28-day mortality rate was 18.6%, and its median ICU LOS was 3.3 d (IQR [1·9–7·3]). The mean baseline SOFA and Elixhauser scores obtained within the first 24 h after ICU admission were 6.94 (SD: 4·1) and 20·77 (SD: 16·2), respectively. The external validation cohort from the PKUPH ICU database included 2778 patients with sepsis, and the demographic distributions were similar to those of the development cohort, which included 1484 males (54.3%). The median patient age was 63 years (IQR [50–73]). The mortality rate was 19.3%, and the median ICU LOS was 4.0 d (IQR [1.1–10·7]). The mean baseline SOFA and Elixhauser scores obtained within the first 24 h after ICU admission were 5.31 (SD: 2.2) and 8.02 (SD: 6.4), respectively.

3.2. Sepsis subphenotypes

By analyzing 21 clustering variables using the time-series k-means clustering method, we concluded that the three-subphenotype model was the best-fitting result (labeled types A, B, and C). The three-subphenotype model exhibited high entropy, a high SSE inflection point, and an appropriate distribution in both the development and validation cohorts (Supplementary Material: Appendix 4, Fig. S1). The importance of each cluster variable is presented in Supplementary Material: Fig. S2. The clinical features of the three subphenotypes differed between the development and validation cohorts (Fig. 2). In the development cohort, patients with type A sepsis (N = 2525, 47%) were exhibited stable vital signs and indicators of organ function. Patients with type B sepsis (N = 1552, 29%)

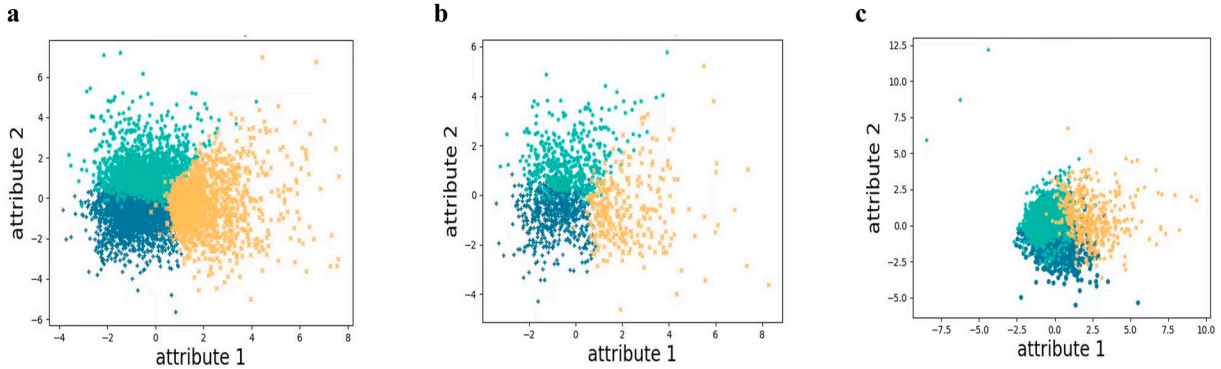


Fig. 2. Visualization of phenotype assignments. (A) Development cohort. (B) Internal validation cohort. (C) External validation cohort. The three subphenotypes were visualized in each of the three cohorts by means of principal component analysis. The horizontal and vertical axes are the two directions in which the variance of the data is highest in the new coordinate system.

exhibited a severe inflammatory reaction, primarily manifested as elevated HR, T, RR, and WBC count, which tended to stabilize within 24 h. Patients with type C sepsis (N = 1251, 23%) exhibited severely impaired organ function, primarily manifested as elevated Scr, tBil, PT, and lactate levels, with a tendency to progressively deteriorate within 24 h from the onset time (Fig. 3, Supplementary Material: Fig. S6). The clinical features of the three subphenotypes in the validation cohort were generally consistent with those in the development cohort (Supplementary Material: Figs. S3–5). Chord diagrams (Fig. 4, Supplementary Material: Appendix 5) show the differences between the subphenotypes for abnormal clinical biomarkers. Type A exhibited the least inflammatory response and impairment of organ function, whereas type C demonstrated the most severe impairment of hepatic, cardiovascular, hematopoietic, pulmonary, and renal functions. In the development cohort, type B exhibited the most severe inflammatory response.

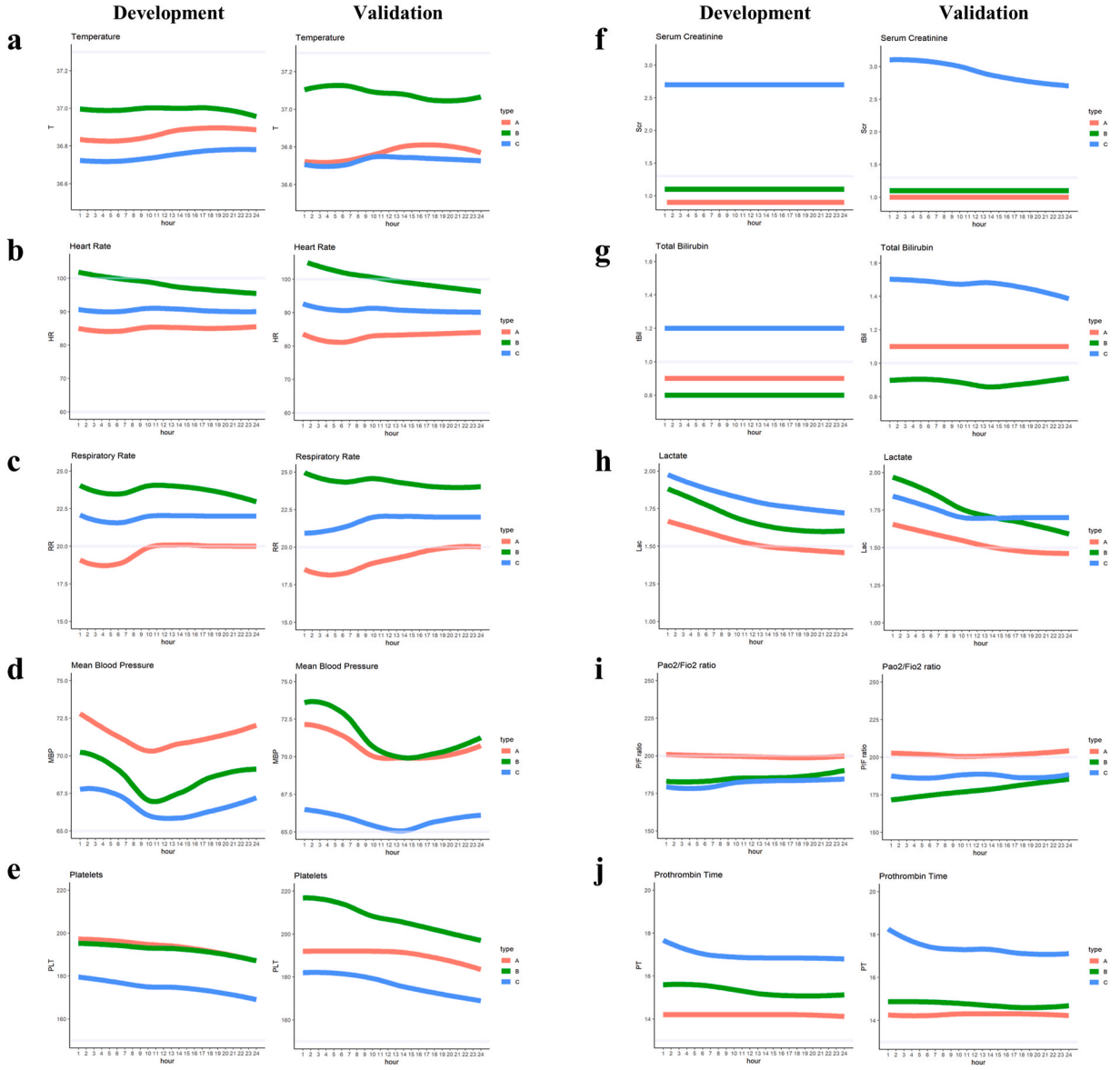


Fig. 3. Comparison of clinical characteristics among the three subphenotypes in the development cohort and internal validation cohort. (A) Temperature. (B) Heart rate. (C) Respiratory rate. (D Mean blood pressure. (E) Platelets. (F) Serum creatinine. (G) Total bilirubin. (H) Lactate. (I) PaO2/FiO2 ratio. (J) Prothrombin time. Type A patients were characterized by the stability of vital signs and organ function indicators. Type B patients were characterized by a severe inflammatory reaction, mainly manifested as a higher heart rate, body temperature, respiratory rate and white blood cell count, but the trend tended to be steady within 24 h. Type C patients were characterized by severely impaired organ function, mainly manifested by elevated levels of serum creatinine, total bilirubin, prothrombin time, and lactate, with a gradual deterioration trend within 24 h of onset.

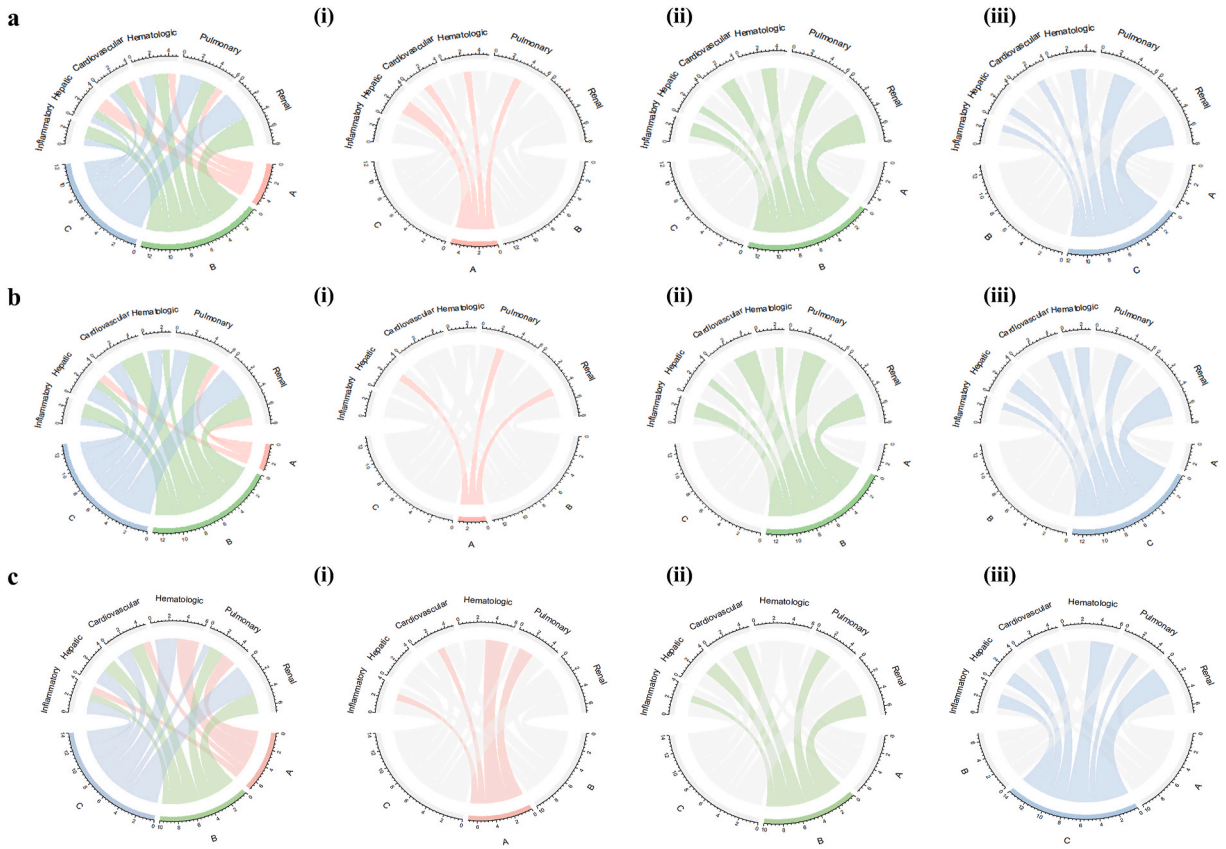


Fig. 4. Chord diagrams showing the differences in subphenotypes in terms of abnormal clinical biomarkers. (A) Abnormal biomarkers vs. all subphenotypes in the development cohort; i: abnormal biomarkers vs. type A; ii: abnormal biomarkers vs. type B; iii: abnormal biomarkers vs. type C; (B) Abnormal biomarkers vs. all subphenotypes in internal validation cohort; i: abnormal biomarkers vs. type A; ii: abnormal biomarkers vs. type B; iii: abnormal biomarkers vs. type C; (C) Abnormal biomarkers vs. all subphenotypes in external validation cohort; i: abnormal biomarkers vs. type A; iii: abnormal biomarkers vs. type B; iii: abnormal biomarkers vs. type C.

3.3. Patient characteristics compared across subphenotypes

The patient characteristics varied among the subphenotypes (Table 1, Fig. 5, and Supplementary Material: Fig. S6). A pairwise comparison of the three subphenotypes showed that baseline SOFA and Elixhauser comorbidity scores were significantly higher in type B and C patients than those in type A patients ($p = 0.000$). The prevalence of comorbidities significantly differed among subphenotypes ($p = 0.000$): patients with type C had a higher prevalence of diabetes mellitus (26.8%), heart failure (46.5%), chronic kidney disease (53.1%), and chronic obstructive pulmonary disease (8.7%); however, they exhibited a lower prevalence of hypertension (20.9%). The 28-day mortality rate was significantly different among the subphenotypes ($p = 0{\cdot}000$): 10.5% for type A, 24.3% for type B, and 33.1% for type C.

Among the 1322 patients in the internal validation cohort and 2778 patients in the external validation cohort, the relative distributions of demographics, comorbidities, and outcomes of subphenotypes were similar to those in the development cohort. In the internal validation cohort, the mortality rate was 9·9% for type A, 21·3% for type B and 33·8% for type C ($p = 0{\cdot}001$). In the external validation cohort, the mortality rates of the subphenotypes were 13·9% for type A, 19·4% for type B and 39·1% for type C ($p = 0.000$).

3.4. Association of subphenotypes with inflammatory biomarkers

Inflammatory biomarker levels within 24 h of onset were significantly different among the three subphenotypes in the external validation cohort. The CRP level was the highest in type B, lowest in type A, and intermediate in type C ($p < 0.05$). PCT and IL-6 levels were the lowest in type A, intermediate in type B, and highest in type C ($p < 0.05$) (Table 2).

3.5. Heterogeneity of treatment effect

In the MIMIC-IV database, the three subphenotypes were categorized into survival and non-survival groups. The proportion of indwelling central venous catheters and total fluid intake at 24 h after onset significantly differed between the survival and non-

Table 1
Patient characteristics among subphenotypes in the development cohort and the two validation cohorts.

	Development cohort				Internal validation cohort				External validation cohort			
	A	B	C	p value	A	B	C	p value	A	B	C	p value
N(%)	2525(47%)	1552(29%)	1251(23%)		621(47%)	430(32%)	281(21%)		1311(47%)	1074(39%)	393(14%)	
Age(IQR)	66[54,77]	65[55,76]	65[54,76]	0.0457	67[56,78]	63[54.74]	65[54,77]	0.001	66[55,75]	59[42,70]	59[45,71]	<0.001
Male(%)	1429(46.5%)	779(50.1%)	744(59.4%)	<0.001	346(55.7%)	233(54.1%)	159(56.5%)	<0.001	56.5	49.9	52.4	0.002
SOFA(IQR)	5 [3,7]	7 [4,10]	9 [6,12]	<0.001	5 [3,7]	7 [4,10]	9 [6,13]	<0.001	4 [3,6]	5 [3,7]	7 [4,10]	<0.001
Elixhauser(IQR)	13 [5,24]	18 [8,25]	28[19,39]	<0.001	15[6.26]	17 [8,26]	30[17.40]	<0.001	6 [2,13]	5 [1,13]	7 [2,15]	<0.001
Hypertension(%)	1096(43.4%)	598(38.5%)	262(20.9%)	<0.001	255(41.0%)	175(40.6%)	59(20.9%)	<0.001	42.3	32.1	38.4	<0.001
Diabetes(%)	387(15.3%)	216(13.9%)	336(26.8%)	<0.001	91(14.6%)	80(18.6%)	54(19.2%)	<0.001	21.5	18.5	23.1	0.080
Heart failure(%)	769(30.4%)	447(28.8%)	582(46.5%)	<0.001	202(32.5%)	124(18.8%)	126(44.8%)	<0.001	14.4	16.1	23.9	<0.001
CKD(%)	501(19.8%)	346(22.6%)	665(53.1%)	<0.001	144(23.1%)	86(20.0%)	137(48.7%)	<0.001	8.4	9.3	33.3	<0.001
Tumor(%)	28(1.1%)	13(0.8%)	24(1.9%)	0.020	10(1.6%)	1(0.2%)	4(1.4)	0.028	51.2	35.5	23.6	0.020
COPD(%)	150(5.9%)	81(5.2%)	109(8.7%)	<0.001	40(6.4%)	47(10.9%)	23(8.1)	<0.001	1.5	0.7	2.0	0.093
Mortality(%)	10.5	24.3	33.1	<0.001	9.9	21.3	33.8	<0.001	13.9	19.4	39.1	<0.001
ICU LOS(IQR)	3.1[1.9,6.0]	3.8[2.0,7.9]	3.7[2.0,7.9]	<0.001	3.0[1.8,6.7]	3.9[2.0,8.4]	3.3[2.0,7.5]	0.001	2.8[0.9,8.3]	5.0[1.7,11.0]	7.2[3.2,16.9]	<0.001

IQR: Interquartile range. SOFA: Sequential Organ Failure Assessment. CKD: Chronic kidney disease. COPD: chronic obstructive pulmonary disease.

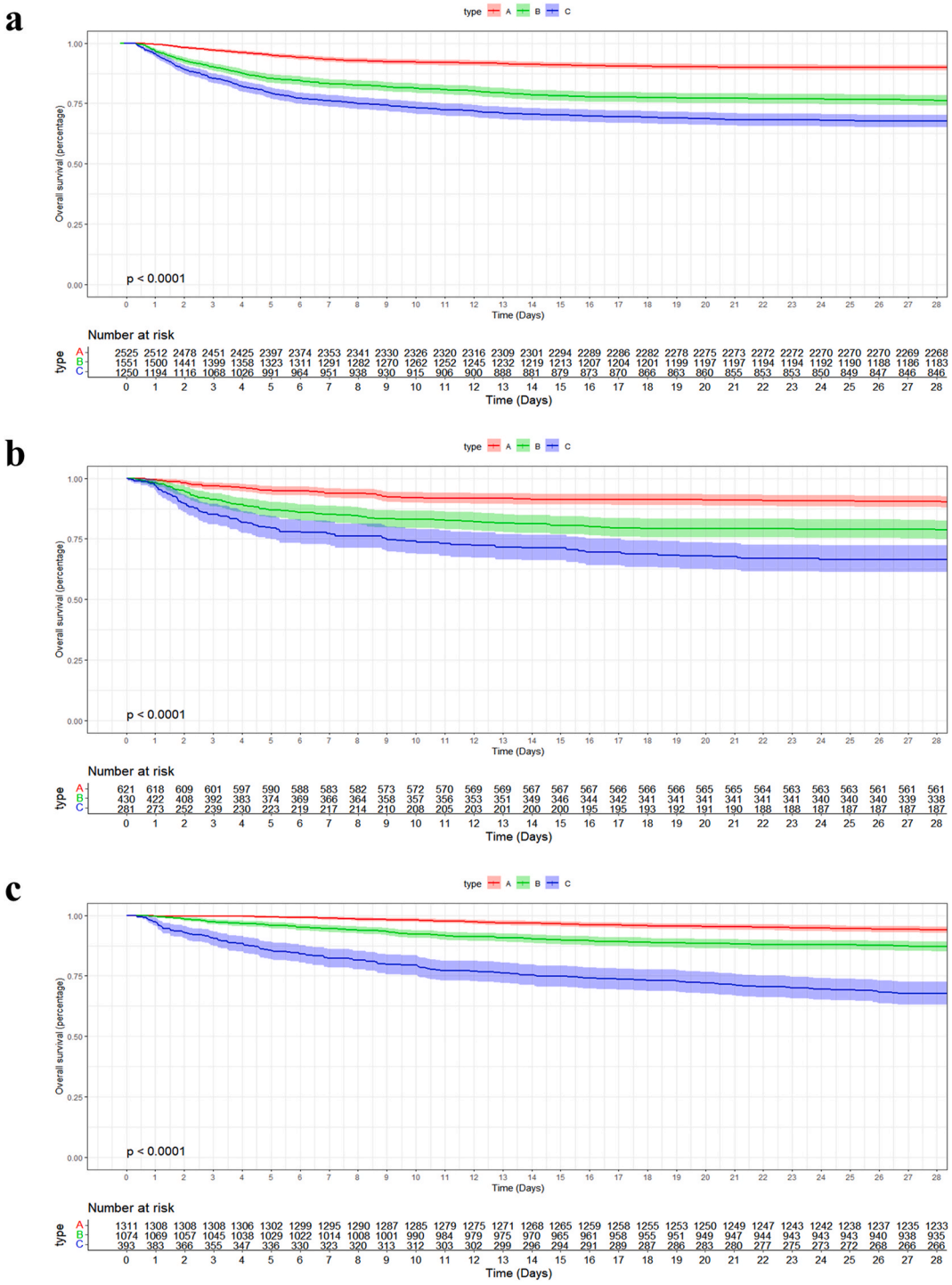


Fig. 5. Kaplan-Meier survival curves of the three subphenotypes in the development cohort. (A), internal validation cohort (B), and external validation cohort(C).

survival groups, regardless of the sepsis phenotype ($p < 0.05$). Within type A, there were no significant differences in the total fluid intake within 24 h of onset or in the proportion of patients with indwelling central venous catheters between the survival and non-survival groups. Within type B, a higher proportion of patients in the survival group than in the non-survival group were administered an indwelling central venous catheter within 3 h of onset (20.2% vs. 14.0%, $p = 0.003$). Within type C, a higher proportion of patients in the survival group than in the non-survival group were administered an indwelling central venous catheter within 24 h of

Table 2
Comparison of inflammatory biomarkers between different subphenotypes.

	A	B	C	type	P value
CRP [mg/dl, M [IQR]]	79.03[137.33,139.34]	106.30[53.80,175.75]	99.25[52.55,175.63]	A vs. B	0.000
				A vs. C	0.000
				B vs. C	0.443
PCT [ng/ml, M [IQR]]	0.29[0.13,5.08]	0.58[0.12,6.74]	2.34[0.17,16.47]	A vs. B	0.216
				A vs. C	0.000
				B vs. C	0.000
IL-6, M[IQR]]	96.60[24.50,454.80]	111.805[39.45,439.82]	198.47[90.12,908.48]	A vs. B	0.229
				A vs. C	0.013
				B vs. C	0.098

IQR: Interquartile range. CRP: C-reactive protein. PCT: Procalcitonin. IL-6: interleukin-6.

onset (16.0% vs. 10.1%, p = 0.002). The total fluid intake within 24 h of onset in the type C survival group was significantly lower than that in the non-survival group (2890.67 mL vs. 4342.13 mL, p = 0.000) (Table 3).

4. Discussion

We identified and validated the differences among the three subphenotypes of patients with sepsis. Sepsis subphenotypes were reproducible in the MIMIC-IV and PKUPH ICU databases. The three sub-phenotypes had different baseline SOFA and Elixhauser scores, diverse clinical features, inflammatory biomarkers, and mortality. Treatment heterogeneity among the three sub-phenotypes included variations in early fluid resuscitation and the use of indwelling central venous catheters.

Sepsis is a heterogeneous disease that rapidly progresses. Although most previous studies on sepsis subtypes have been based on cross-sectional data [4–13], time-series data for clustering was used in this study. Time-series data can accurately and quickly diagnose and categorize sepsis by reflecting dynamic physiological changes in patients, thereby facilitating the provision of appropriate treatment recommendations.

Sepsis refers to an imbalance in the immune system triggered by infection, leading to impaired organ function that can be life-threatening. It manifests when there is an imbalance between immune activation and immunosuppression pathways [27,24]. In general, early proinflammatory activation in response to invading pathogens or danger signals is associated with the onset of multiorgan failure and early death, whereas immunosuppression is associated with the reactivation of latent infection and delayed death. The degree of immune activation and immunosuppression, along with their impact on the immune response associated with sepsis, varies among individual patients [25,26,28].

Exploring the diverse clinical manifestations of patients may reveal differences in the intrinsic pathogenesis of the three subphenotypes. Type C patients exhibited significantly higher PCT and IL-6 levels in than type A and B patients, whereas the CRP levels in type B patients were significantly higher than those in type A patients. This suggested that different subphenotypes of sepsis involve different pathways of inflammatory response regulation and varying degrees of balance between immune activation and immune suppression *in vivo*. In patients with type C sepsis, excessive immune activation and severe immunosuppression result in the production of numerous pro-inflammatory cytokines and reactive oxygen species that impair organ function and lead to the abnormal apoptosis of cells in different organs, multiple organ dysfunction syndrome, and death [29]. In patients with type B sepsis, organ function remains unimpaired despite a severe inflammatory response, resulting in a lower mortality rate than that in type C patients. In patients with type A sepsis, immune activation and immunosuppression in the body reach a dynamic balance, effectively clearing the infection without damaging organ function. Consequently, this group exhibited the lowest mortality rate.

Owing to the heterogeneity of sepsis, there is no universal approach for its treatment [19]. Exploring the precision treatment of sepsis based on its sepsis subtypes has become a new direction in clinical research [6,7,9,14,20]. Treatment heterogeneity was observed among the three subphenotypes regarding early fluid resuscitation and the use of indwelling central venous catheters.

Early fluid resuscitation is an important measure for the treatment of sepsis. Whether adequate or restrictive fluid resuscitation should be performed remains controversial [30]. The severity of the illness prompts rapid fluid administration, supported by a by Rivers et al. which showed that an early goal-directed pattern of massive fluid resuscitation in patients with severe sepsis and septic shock within the first 6 h of resuscitation is associated with an improved prognosis [31]. However, fluid overload can have deleterious consequences, prolonging mechanical ventilation and increasing mortality in critically ill patients, especially in those that have sepsis [32]. A conservative resuscitative fluid strategy, in contrast to liberal fluid therapy or standard care, is associated with an increased number of ventilator-free days and a reduced length of ICU stay, with no considerable impact on mortality [33]. However, Meyhoff et al. found that fluid restriction did not reduce mortality in patients with septic shock [34]. In this study, the 24-h total fluid intake of type C patients in the survival group was significantly lower than that of patients in the non-survival group. Therefore, in clinical practice, for patients with a poor general condition and organ dysfunction, limiting fluid resuscitation is advisable to avoid volume overload and mitigate potential harm to the body. The potential benefits of volume expansion related to increased cardiac output and oxygen delivery must be balanced against the risk of exacerbated pulmonary and tissue edema [35].

In addition, among type B and C patients, the proportion of those with early indwelling central venous catheters was significantly higher in the survival group than that in the non-survival group. This suggested that, in clinical practice, the early placement of central venous catheters is advisable in patients with relatively unstable conditions. In sepsis treatment, continuous assessment of the

Table 3
Heterogeneity of treatment effects among the three subphenotypes.

Type	A			B			C		
Treatments	survival	nonsurvival	P value	survival	nonsurvival	P value	survival	nonsurvival	P value
Indwelling CVC in 1 h (%)	165(5.8%)	23(7.0%)	0.403	179(11.8%)	41(8.7%)	0.060	101(9.8%)	31(6.0%)	0.012
Indwelling CVC in 3 h (%)	192(6.8%)	28(8.5%)	0.247	213(14.0%)	47(10.0%)	0.022	113(11.0%	36(7.0%	0.013
Indwelling CVC in 6 h (%)	210(7.4%)	29(8.8%)	0.369	239(15.8%)	52(11.0%)	0.011	123(12.0%)	42(8.2%)	0.024
Indwelling CVC in 24 h (%)	280(9.9%)	41(12.5%)	0.147	306(20.2%)	66(14.0%)	0.003	164(16.0%)	52(10.1%	0.002
Liquid input in 1 h[ml] M (IQR)	110.00[27.86,290.93]	113.82[25.17,230.28]	0.621	212.83[62.39,783.33]	200.00[60,72.209.00]	0.343	131.61[28.16,336.48]	156.81[60.57,430.74]	0.009
Liquid input in 3 h[ml] M (IQR)	284.00[25.64,772.77]	374.75[113.83,815.46]	0.112	614.82 [182.50,1464.25]	666.02 [227.67,1413.28]	0.98	353.38[49.33,1025.00]	514.57 [200.78,1177.92]	0.000
Liquid input in 6 h[ml] M (IQR)	664.92 [190.00,1467.03]	757.43 [311.92,1451.24]	0.638	1361.33 [561.56,2679.62]	1384.67 [587.33,2801.49]	0.919	825.09 [287.94,1899.61]	1192.73 [512.44,2464.73]	0.000
Liquid input in 24 h [ml]M (IQR)	3002.50 [1650.00,5052.29]	2768.47 [1275.00,4762.29]	0.209	4388.06 [2496.01,6844.82]	4711.21 [2620.42,7703.40]	0.289	2890.67 [1529.80,5470.29]	4342.13 [2189.32,7305.07]	0.000

Among patients with type A, there was no significant difference in the proportion with indwelling central venous catheters or in the fluid intake over 24 h between the nonsurvival group and the survival group. In type B patients, the proportion of patients who had indwelling central venous catheters within 24 h in the survival group was significantly higher than that in the nonsurvival group ($p < 0.05$). There were no significant differences in fluid intake over the 24 h between the survivors and nonsurvivors. Among patients with type C, the proportion with early indwelling central venous catheters in the survival group was significantly higher than that in the nonsurvival group ($p < 0.05$), and the 24-h fluid intake of the nonsurvival group was significantly higher than that of the survival group ($p < 0.05$).

hemodynamic status of patients with sepsis is essential to optimize perfusion pressure and overall blood flow, aiming to restore and optimize tissue perfusion. As both hypovolemia and hypervolemia are associated with poor prognosis [32,36], the assessment of hemodynamic status, including blood volume and fluid responsiveness, remains a key challenge in the management of septic shock. Previously, central venous pressure and central venous oxygen saturation were important observational indicators in goal-directed therapy [31]. Although the latest version of the SSC guidelines no longer emphasizes repeated observations of the above metrics to adjust therapy, central venous catheterization remains essential in addressing how massive fluid resuscitation is performed and when to stop it, which are the two central questions in hemodynamic management. The results of this study recommends early placement of a central venous catheter in patients with sepsis unless the patient's general condition is relatively stable and organ function indicators are favorable.

This study had some limitations. First, the exploration of sepsis subphenotypes was performed only at the clinical level. Although common clinical indicators related to infection were analyzed, the study did not explore their pathophysiology, genomics, or molecular mechanisms. This limitation prevents the combination of external clinical manifestations with the intrinsic genetic characteristics of patients with different subphenotypes, preventing the formulation of personalized therapeutic regimens targeting the immune response *in vivo*. Second, owing to the lack of data related to treatment measures in the PKUPH ICU database, this study did not validate the findings of treatment heterogeneity. Additionally, this study was retrospective; therefore, biases in data collection and selection may have occurred during the construction of the two databases and the screening of patients with sepsis, potentially affecting the accuracy and reliability of the data. In addition, this study failed to determine a causal relationship between different treatment measures and prognosis. Prospective studies are needed to verify the stability of the three sepsis subphenotypes and the effect of different therapeutic measures on prognosis.

5. Conclusions

Three novel sepsis subphenotypes were identified and validated using time-series data collected within 24 h after sepsis onset. These subphenotypes exhibited varying levels of inflammatory biomarkers and exhibited distinct responses to treatment with early fluid resuscitation and indwelling central venous catheters. This study demonstrated reproducibility and stability across different databases, offering valuable insights into the treatment of patients with sepsis and providing individualized treatment recommendations based on different subphenotypes. There are some limitations in the research, primarily in the lack of prospective validation and insightful comparisons of pathophysiologic mechanisms between different subphenotypes. Future studies are required to confirm the stability of the three sepsis subphenotypes and assess the impact of various therapeutic measures on patient prognosis.

Data availability statement

Deidentified data are available from the corresponding author, HZ, upon request. The contact information for HZ is included on the title page.

CRediT authorship contribution statement

Chenxiao Hao: Writing – original draft, Software, Formal analysis, Data curation. **Rui Hao:** Software, Methodology. **Huiying Zhao:** Writing – review & editing, Validation, Funding acquisition, Formal analysis, Conceptualization. **Yong Zhang:** Methodology, Funding acquisition, Conceptualization. **Ming Sheng:** Methodology, Conceptualization. **Youzhong An:** Supervision.

Declaration of competing interest

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.

Acknowledgements

This research was funded by the National Natural Science Foundation of China (grant no. 82202366), National Key R&D Program of China (2020AAA0109603), Clinical Medicine Plus X - Young Scholars Project, Peking University, the Fundamental Research Funds for the Central Universities (PKU2022LCXQ031), and Wu Jieping Medical Foundation Runze Fund for Critical Care Medicine (NO320.6750.2022-2-34).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e28520.

References

[1] C. Fleischmann, et al., Assessment of global incidence and mortality of hospital-treated sepsis. Current estimates and limitations, Am. J. Respir. Crit. Care Med. 193 (3) (2016) 259–272.
[2] C. Rhee, et al., Incidence and trends of sepsis in US hospitals using clinical vs claims data, 2009-2014, JAMA 318 (13) (2017) 1241–1249.
[3] K.E. Rudd, et al., Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study, Lancet 395 (10219) (2020) 200–211.
[4] C.W. Seymour, et al., Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis, JAMA 321 (20) (2019) 2003–2017.
[5] W.A. Knaus, R.D. Marks, New phenotypes for sepsis: the promise and problem of applying machine learning and artificial intelligence in clinical research, JAMA 321 (20) (2019) 1981–1982.
[6] Z. Zhang, et al., Identification of subclasses of sepsis that showed different clinical outcomes and responses to amount of fluid resuscitation: a latent profile analysis, Crit. Care 22 (1) (2018) 347.
[7] P. Ma, et al., Individualized resuscitation strategy for septic shock formalized by finite mixture modeling and dynamic treatment regimen, Crit. Care 25 (1) (2021) 243.
[8] H. Zhao, et al., Revising host phenotypes of sepsis using microbiology, Front. Med. 8 (2021) 775511.
[9] Z. Zhang, et al., Deep learning-based clustering robustly identified two classes of sepsis with both prognostic and predictive values, EBioMedicine 62 (2020) 103081.
[10] B.P. Scicluna, et al., Classification of patients with sepsis according to blood genomic endotype: a prospective cohort study, Lancet Respir. Med. 5 (10) (2017) 816–826.
[11] H.R. Wong, et al., Developing a clinically feasible personalized medicine approach to pediatric septic shock, Am. J. Respir. Crit. Care Med. 191 (3) (2015) 309–315.
[12] J.O. Yang, et al., Whole blood transcriptomics identifies subclasses of pediatric septic shock, Crit. Care 27 (1) (2023).
[13] Z. Zhang, et al., Exploring disease axes as an alternative to distinct clusters for characterizing sepsis heterogeneity, Intensive Care Med. 49 (11) (2023) 1349–1359.
[14] S.V. Bhavani, et al., Development and validation of novel sepsis subphenotypes using trajectories of vital signs, Intensive Care Med. 48 (11) (2022) 1582–1592.
[15] S.V. Bhavani, et al., Identifying novel sepsis subphenotypes using temperature trajectories, Am. J. Respir. Crit. Care Med. 200 (3) (2019) 327–335.
[16] Z. Xu, et al., Sepsis subphenotyping based on organ dysfunction trajectory, Crit. Care 26 (1) (2022) 197.
[17] P. Sinha, N.J. Meyer, C.S. Calfee, Biological phenotyping in sepsis and acute respiratory distress syndrome, Annu. Rev. Med. 74 (1) (2023) 457–471.
[18] R.B.E. van Amstel, et al., Uncovering heterogeneity in sepsis: a comparative analysis of subphenotypes, Intensive Care Med. 49 (11) (2023) 1360–1369.
[19] C.W. Seymour, et al., Precision medicine for all? Challenges and opportunities for a precision medicine approach to critical illness, Crit. Care 21 (1) (2017) 257.
[20] J.L. Vincent, The coming era of precision medicine for intensive care, Crit. Care 21 (Suppl 3) (2017) 314.
[21] A. Johnson, et al., MIMIC-IV, a freely accessible electronic health record dataset, Sci. Data 10 (1) (2023) 1.
[22] A.L. Goldberger, et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals, Circulation 101 (23) (2000) E215–E220.
[23] M. Singer, et al., The third international consensus definitions for sepsis and septic shock (Sepsis-3), JAMA 315 (8) (2016) 801–810.
[24] H. Yadav, R. Cartin-Ceba, Balance between hyperinflammation and immunosuppression in sepsis, Semin. Respir. Crit. Care Med. 37 (1) (2016) 42–50.
[25] R.S. Hotchkiss, G. Monneret, D. Payen, Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy, Nat. Rev. Immunol. 13 (12) (2013) 862–874.
[26] T. van der Poll, M. Shankar-Hari, W.J. Wiersinga, The immunology of sepsis, Immunity 54 (11) (2021) 2450–2464.
[27] R.S. Hotchkiss, et al., Sepsis and septic shock, Nat. Rev. Dis. Prim. 2 (2016) 16045.
[28] J. Cohen, The immunopathogenesis of sepsis, Nature 420 (6917) (2002) 885–891.
[29] B. Alikiaii, et al., The role of phytochemicals in sepsis: a mechanistic and therapeutic perspective, Biofactors 47 (1) (2021) 19–40.
[30] L.A. McIntyre, J.C. Marshall, Intravenous fluids in septic shock - more or less? N. Engl. J. Med. 386 (26) (2022) 2518–2519.
[31] E. Rivers, et al., Early goal-directed therapy in the treatment of severe sepsis and septic shock, N. Engl. J. Med. 345 (19) (2001) 1368–1377.
[32] J.H. Boyd, et al., Fluid resuscitation in septic shock: a positive fluid balance and elevated central venous pressure are associated with increased mortality, Crit. Care Med. 39 (2) (2011) 259–265.
[33] J.A. Silversides, et al., Conservative fluid management or deresuscitation for patients with sepsis or acute respiratory distress syndrome following the resuscitation phase of critical illness: a systematic review and meta-analysis, Intensive Care Med. 43 (2) (2017) 155–170.
[34] T.S. Meyhoff, et al., Restriction of intravenous fluid in ICU patients with septic shock, N. Engl. J. Med. 386 (26) (2022) 2459–2470.
[35] J. Benes, et al., Fluid therapy: double-edged sword during critical care? BioMed Res. Int. 2015 (2015) 729075.
[36] P.E. Marik, Iatrogenic salt water drowning and the hazards of a high central venous pressure, Ann. Intensive Care 4 (2014) 21.