

Methods for Phenotyping Adult Patients in Sepsis and Septic Shock: A Scoping Review

OBJECTIVE: Despite its heterogeneous phenotypes, sepsis or life-threatening dysfunction in response to infection is often treated empirically. Identifying patient subgroups with unique pathophysiology and treatment response is critical to the advancement of sepsis care. However, phenotyping methods and results are as heterogeneous as the disease itself. This scoping review evaluates the prognostic capabilities and treatment implications of adult sepsis and septic shock phenotyping methods.

DATA SOURCES: Medline and Embase.

STUDY SELECTION: We included clinical studies that described sepsis or septic shock and used any clustering method to identify sepsis phenotypes. We excluded conference abstracts, literature reviews, comments, letters to the editor, and in vitro studies. We assessed study quality using a validated risk of bias tool for observational cohort and cross-sectional studies.

DATA EXTRACTION: We extracted population, methodology, validation, and phenotyping characteristics from 17 studies.

DATA SYNTHESIS: Sepsis phenotyping methods most frequently grouped patients based on the degree of inflammatory response and coagulopathy using clinical, nongenomic variables. Five articles clustered patients based on genomic or transcriptomic data. Seven articles generated patient subgroups with differential response to sepsis treatments. Cluster clinical characteristics and their associations with mortality and treatment response were heterogeneous across studies, and validity was evaluated in nine of 17 articles, hindering pooled analysis of results and derivation of universal truths regarding sepsis phenotypes, their prognostic capabilities, and their associations with treatment response.

CONCLUSIONS: Sepsis phenotyping methods can identify high-risk patients and those with high probability of responding well to targeted treatments. Research quality was fair, but achieving generalizability and clinical impact of sepsis phenotyping will require external validation and direct comparison with alternative approaches.

KEY WORDS: cluster analyses; infections; machine learning; risk assessment; sepsis; septic shock

Sepsis is an acute, heterogeneous disease that affects 1.7 million adults in the United States annually, contributes to more than 250,000 deaths, and results in chronic complications in up to 70% of patients (1–3). Characterized by excessive inflammation, metabolic dysfunction, and immunosuppression, sepsis is defined as life-threatening organ dysfunction caused by a dysregulated host response to infection (4). Treatment is often empiric and involves fluid resuscitation, broad-spectrum antibiotics, source control of infection, and vasopressor support. Few treatments are targeted toward underlying dysregulation of the host immune response, coagulopathy, and vasculopathy. However, immunostimulants (5, 6), anticoagulants (7, 8), and anti-inflammatory agents (9, 10) have been trialed with only mixed results, which may reflect the multiple phenotypes within sepsis (11).

Han Li, BS¹

Asena Markal, BS²

Jeremy A. Balch, MD³

Tyler J. Loftus, MD^{3,4}

Philip A. Efron, MD^{3,5}

Tezcan Ozrazgat-Baslanti, PhD^{1,4,5}

Azra Bihorac, MD^{1,4,5}

Copyright © 2022 The Authors. Published by Wolters Kluwer Health, Inc. on behalf of the Society of Critical Care Medicine. This is an open-access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/CCE.0000000000000672

Some immunologic interventions, such as antitumor necrosis factor α (9), interleukin-1 receptor antagonists (10), and immunoglobulins (6), which demonstrated no benefit in sepsis broadly, have shown improved mortality in patients stratified to certain clinical or laboratory values.

These findings have led investigators to cluster sepsis into phenotypes responsive to targeted therapies. Phenotyping methods use clustering algorithms to group patients based on routine clinical, laboratory, genomic, or transcriptomic variables with goal of predicting outcomes or guiding treatment. These clustering algorithms attempt to derive meaningful relationships from data in an unsupervised manner, facilitating the discovery of novel subphenotypes with reduced human bias. However, clustering techniques vary drastically in the literature (12, 13). For example, the more commonly used clustering methods, such as latent class analysis or hierarchical clustering, differ in terms of statistical approach, assumptions about data heterogeneity and independence, and prognostic capabilities. Clustering methods frequently cannot identify the optimal number of clusters that can be derived from any given dataset, so researchers may choose from a variety of optimal-cluster-number algorithms that further drive heterogeneity in the field. Clinically, previously reported clustering methods vary significantly in their baseline patient populations. Sepsis can be defined using multiple criteria across multiple infectious organisms, sites of infection, preexisting comorbidities, and severities. Clusters derived from select datasets with older sepsis criteria may differ significantly from clusters derived from a wide selection of trial and observation cohorts and more modern sepsis criteria. Critically, there is an absence of cross-evaluation or comparison among existing phenotyping methods to ensure consistency and generalizability. Sepsis phenotyping is a novel, rapidly evolving field and would thus particularly benefit from a scoping review evaluating the available evidence.

Accordingly, we conducted a scoping review with the objective of describing and comparing sepsis phenotyping methods by the clustering mechanism employed, prognostic capabilities, degrees of validation, potential impact for racial inequities in sepsis, and implications for sepsis treatment. As a secondary objective, we aimed to evaluate research quality and risk of bias to identify limitations of the field. By aggregating and evaluating available sepsis phenotyping methods, we aim to provide clinicians and researchers with a broad perspective

on the current state of sepsis phenotyping with implication for the development of targeted sepsis treatment.

MATERIALS AND METHODS

Study Eligibility and Search Strategy

We performed screening, selection, data extraction, and quality assessment according to the Joanna Briggs Institute methodology for scoping reviews and the Preferred Reporting Items for Systematic Reviews extension for Scoping Reviews (14). The review protocol for search strategy, selection criteria, data extraction, and data assessment was developed prior to performing article search on Covidence (Melbourne, Australia), an online tool for review management, but is not publicly registered. Prior to this review, a preliminary search of PubMed, Medline, and Embase was conducted, and no published systematic reviews or scoping reviews were identified.

Mr. Li and Ms. Markal searched Embase and Medline using a keyword and Emtree subject heading search for criteria related to the following: sepsis, septic shock, phenotyping, and mortality published from database inception up until June 30, 2021, for each database. Detailed search criteria are detailed in **Supplemental Table 1** (<http://links.lww.com/CCX/A963>). After removing duplicates, Mr. Li and Ms. Markal independently screened articles by an initial title and abstract screen, followed by full-text review for inclusion criteria, with disagreements resolved by a third investigator (T.J.L.). We included clinical studies that described sepsis, septic shock, or severe sepsis in hospitalized patients. All clustering methods, including but not limited to hierarchical, centroid-based, distribution-based, and density-based methods, were allowed so long as the method stratified patients into distinct groups. Studies were included if they identified sepsis by any of the Sepsis-1, -2, or -3 criterion. Articles concerning septic shock were included if they used any of the above criteria so long as they included signs of organ malperfusion with the administration of vasopressors despite adequate fluid resuscitation. Because of the predicted dearth of articles describing differential treatment responses, we included select articles with special attention to specific treatment modalities even if they included a minority of nonseptic patients. Conference abstracts/proceedings, literature reviews, comments, letters to the editor, simulation studies, studies not in English, or in vitro studies were excluded. To improve

the generalizability of our findings, studies investigating only the pediatric population or an adult subpopulation, namely, pregnancy or sepsis from any single microbiological infection (e.g. only *Pseudomonas aeruginosa* sepsis patients), were also excluded. To include a broad range of findings, study eligibility was not limited by infection site, years considered by patient dataset, publication status, or viral or bacterial origin of sepsis. We performed a backward citation search of relevant works to identify additional studies with included studies from this search requiring approval of three independent investigators (H.L., A.M., T.J.L.).

Data Extraction and Quality Assessment

A data extraction and quality assessment form was created with the review and consensus of three independent investigators. Mr. Li and Ms. Markal independently performed data extraction and tabulation in duplicate, with conflicts resolved by consensus. Investigators were not blinded to study author or journal. Extracted variables are shown as they are described in the extraction form in **Supplemental Table 2** (<http://links.lww.com/CCX/A963>). Following extraction, data were synthesized qualitatively by Mr. Li and Ms. Markal by dividing tabulated variables into the subsections of study characteristics, patient populations, clustering methods and findings, treatment response, and race and ethnicity inclusion. Mr. Li and Ms. Markal independently performed quality assessment in duplicate using the U.S. National Heart, Lung, and Blood Institute risk-of-bias tool for observational cohort and cross-sectional studies. With this tool, each study was given a final quality rating of good, fair, or poor after evaluating 14 possible sources of bias (**Supplemental Table 3**, <http://links.lww.com/CCX/A963>). Disagreements were resolved through consensus.

RESULTS

Study Characteristics

Three hundred sixty articles were identified in our initial search with an additional 19 articles identified through manual reverse citation search. Twenty-nine full-text articles were assessed for eligibility, and 17 studies were included (**Fig. 1**) (12, 13, 15–29). Articles were published from 2016 to 2021 (**Supplemental Table 4**, <http://links.lww.com/CCX/A963>).

For risk of bias, six articles were deemed as good-quality papers (13, 19, 22–25), eight were fair (12, 15, 16, 18, 20, 21, 28, 29), and three were poor (**Fig. 2**; **Supplemental Table 3**, <http://links.lww.com/CCX/A963>) (17, 26, 27). Studies at high risk of bias were most frequently without sensitivity analysis and external validation, possessed a conflict of interest in study funding, or focused primarily on a novel clustering methods rather than clinical applications. All articles clearly stated their research question, defined the population, evaluated patients within a reasonable timeframe to see differences in outcome, and clearly defined independent and dependent variables. No articles blinded outcome assessors to the exposure status of participants or justified sample sizes. Seven articles investigated clustering variables more than once over time (15, 16, 21, 23, 25–27), and 12 articles adjusted for key confounding variables (13, 15, 16, 18–20, 22–25, 28, 29). Loss-to-follow-up measurements were not applicable toward 12 articles given the retrospective nature of these analyses (12, 15–21, 25–27, 29). Given that identified patient subgroups were nominal classifications in 15 articles, the risk of bias field evaluating whether the outcome associated with increasing levels of the independent variable in a dose-dependent fashion was only applicable to two studies (25, 26).

All articles were either retrospective, prospective, or a mixed retrospective analysis and prospective cohort study (**Table 1**). Two articles retrieved patients from the Medical Information Mart for Intensive Care-III database (16, 21), and two articles used data from the Gene Expression Omnibus (GEO) and ArrayExpress (28, 29). There were no other overlaps in datasets. Other datasets included the Vasopressin vs Norepinephrine as Initial Therapy in Septic Shock, Activated Protein C Worldwide Evaluation in Severe Sepsis, A Controlled Comparison of Eritoran in Severe Sepsis, Genetic and Inflammatory Markers of Sepsis, and Protocol-Based Care for Early Septic Shock trials, as well as the HemoSepsis study, the HemoPred study, electronic ICU, and the Prognostic Hemodynamic Profiling in the Acutely Ill Emergency Department Patient International Registry.

Although most articles defined sepsis explicitly in their inclusion criteria, two articles were included despite the fact that their cohorts did not entirely comprise sepsis patients (**Table 1**) (18, 23). Han et al (18) specified inclusion criteria as blood cultures ordered within 24 hours of admission combined with 4 consecutive days of antibiotics, IV antibiotics within 24 hours

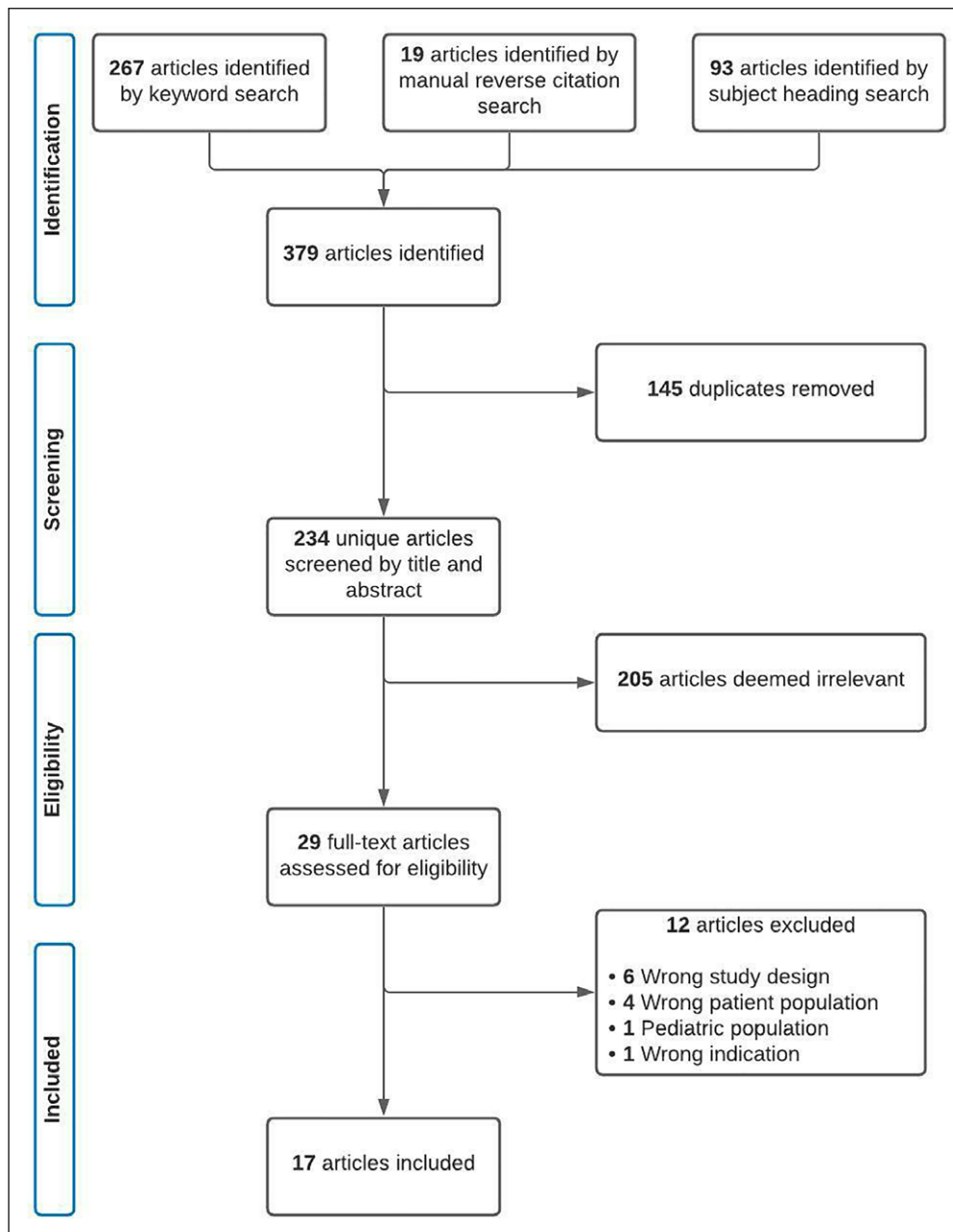


Figure 1. Preferred Reporting Items for Systematic Review diagram of study selection.

of admission, and initiation of antibiotics. Although septic patients were not the primary population of study, 67% of the study population was diagnosed with sepsis. This study was included because of its analysis of preferential response to antibiotic delay. Bhavani et al (15) used probable infection defined with similar criteria, but later validated their phenotyping approach in a patient population defined by Sepsis-3 criteria. Four articles used Sepsis-3 guidelines to identify sepsis (15, 16, 21, 25), four articles used Sepsis-2 guidelines (12, 17, 20, 24), and one article used Sepsis-1 guidelines (22).

(12, 13, 15–17, 19, 21, 23, 25–27). Transcriptomic variables were used in four articles (22, 24, 28, 29). Genomic variables were used in one article (20), and response to antibiotic delays was used in the final article (18). The most common clustering method was K-means, used in six articles (Table 2) (13, 19, 24, 27–29). Other methods included hierarchical clustering and density-based spatial clustering of applications with noise, consensus clustering, causal forests, subgraph augment nonnegative matrix factorization, latent class analysis, spectral clustering, hierarchical clustering based on

Some studies used a mix of sepsis-3 and -2 criteria (13) or sepsis -2 and -1 criteria (19). The remaining articles either retrieved data from multiple studies using different sepsis definitions, as in ArrayExpress or GEO, or used separate definitions for sepsis (18, 23, 26–29). Exclusion criteria aimed to eliminate cases with missing or non-sensical data in seven articles (16, 19–21, 23, 27, 29) and to removing confounding patient populations in 10 articles (12, 15, 17, 18, 21, 22, 24, 27–29). Exclusion criteria were not stated in three articles (13, 25, 26). The inclusion criteria, exclusion criteria, population descriptions, and possible conflicts of interest are included in full in Supplemental Table 4 (<http://links.lww.com/CCX/A963>).

Clustering Methods and Findings

Clinical variables, including age, sex, and temperature, were used to cluster patients in 11 articles (Table 2)

Author	Risk of Bias Final Assessment*	Main driver behind assessment
Gerl 2019 (12)	●	No sensitivity analysis. However, echocardiography clustering variables are not easily apparent from sepsis presentation alone and SOFA scores similar between clusters.
Seymour 2019 (13)	●	Extensive validation across multiple cohorts, adjustment for multiple possible confounders
Bhavani 2020 (15)	●	Further validation of previously discovered and validated clusters, but with a small sample size.
Ding 2021 (16)	●	Sensitivity analyses performed but without external validation.
Gårdlund 2018 (17)	●	No sensitivity analysis for SOFA and APACHE scores. No internal or external validation. Since many clusters are defined by clinical, diagnosable diseases, clusters do not add much value over clinician diagnoses.
Han 2021 (18)	●	Sensitivity analyses performed but without external validation.
Kudo 2021 (19)	●	Assessed multiple datasets with a 28-d follow-up period with external validation and sensitivity analyses.
Scicluna 2017 (20)	●	Validated across multiple sites of infection, patient ages, and locations.
Sharafoddini 2021 (21)	●	Uses cluster analysis to identify biomarkers for sepsis severity, rather than proving the clusters as true subphenotypes, <i>i.e.</i> the authors focus on highly variable biomarkers between clusters rather than cluster validity.
Antcliffe 2019 (22)	●	Further development of previously discovered and validated clusters using transcriptomics.
Bhavani 2019 (23)	●	Sensitivity analysis and external validation performed, clustering variable was measured over time and with a large sample size.
Davenport 2016 (24)	●	Sensitivity analysis and external validation performed.
Liu 2020 (25)	●	Authors control for multiple confounders and carefully delineate how clusters evolve over time
Mayhew 2018 (26)	●	Focuses on the clustering method and its ability to handle EMR data as opposed to validating the sepsis subphenotypes it discovers.
Nowak 2016 (27)	●	Authors funded by manufacturer of device used to monitor patients. K-means clustering was defined manually.
Sweeney 2018 (28)	●	The authors accounted for several confounding factors but the inherent study design of using heterogeneous public datasets introduces bias.
Zhang 2020 (29)	●	The authors compare their model against several previous clustering models for sepsis with some success. However, the heterogeneous nature of the databases and exclusion of genes (exact number unspecified) during dataset merging introduces bias.

*Green: low risk of bias. Yellow: medium risk of bias. Red: high risk of bias

Figure 2. Study risk of bias assessments and reasoning. APACHE = Acute Physiology And Chronic Health Evaluation, EMR = Electronic Medical Record, SOFA = Sequential Organ Failure Assessment.

principal components, group-based trajectory modeling, or a novel composite mixture model. Methods of imputation were not stated in nine articles (15, 20, 22–25, 27, 28) and otherwise included median imputation, linear approximation, iterative principal component analysis, predictive mean matching, multiple imputation with chained equations, and random forests.

We classified each article's clusters based on the principle differences between clusters. For example, an article with one cluster defined by elevated inflammatory markers and another defined by anti-inflammatory markers would have a principle difference in inflammation. In seven articles, clusters differed primarily in clinical or cellular markers for immune response, inflammation, and coagulopathy (Table 2) (13, 15, 20, 22, 24, 28, 29).

Five of these seven studies clustered cases on a transcriptomic or genomic approach (20, 22, 24, 28, 29). In three articles, clusters differed across a wide array of nonspecific characteristics such as patient age, sex, comorbidities, and vital signs, leaving mortality as the principle difference between clusters (16, 21, 26). Other clusters stratified patients based on delayed antibiotic response, thrombomodulin response, risk of progression to septic shock, hemodynamic state, infection site, disease timeline, and temperature trajectory.

All studies identified clusters that predicted clinical outcomes, that is, mortality (**Supplemental Table 5**, <http://links.lww.com/CCX/A963>); however, only seven studies identified clusters that predicted treatment response (**Table 3**). Of these studies, six articles included sensitivity analyses (13, 18, 19, 22, 25, 29), and four

TABLE 1.
Study Cohort Characteristics

Reference	Study Design	Total Number of Participants	Data Source	Inclusion Criteria (Sepsis Criteria/Admission Type)
Geri et al (12)	Cohort	360	Two prospective databases from 12 different ICUs	Sepsis-2 criteria for septic shock/ICU admission
Seymour et al (13)	Retrospective analysis and a prospective cohort	20,189	Twelve hospitals in the University of Pittsburgh Medical Center, GenIMS study, A Controlled Comparison of Ertoran in Severe Sepsis trial, PROWESS trial, and Protocol-Based Care for Early Septic Shock trial	For Sepsis Endotyping in Emergency Care cohorts, sepsis-3 criteria/ED hospital admission/within first 6 hr of ED presentation. For GenIMS cohort, sepsis-2 criteria/ED hospital admission/within 1 hr of ED presentation. For randomized clinical trials, sepsis-2 criteria for severe sepsis and septic shock
Bhavani et al (15)	Cohort	120 patients with septic shock, 88	University of Chicago Medicine between 2013 and 2019	Adults/sepsis-3 criteria for septic shock or one positive <i>Staphylococcus aureus</i> blood culture/ICU admission
Ding and Luo (16)	Retrospective analysis	5,782	MIMIC-III	Adults/sepsis-3 criteria/ICU admission/within 24 hr of ICU admission
Gårdlund et al (17)	Cohort	1,696	PROWESS shock clinical trial	Sepsis-2 criteria for septic shock with vasopressors given within 4 hr of each other and <24 hr
Han et al (18)	Retrospective analysis	60,817 (44,018 with sepsis or septic shock)	Two tertiary-care medical centers and four community hospitals	Patients with blood culture orders, 4 consecutive days of antibiotics (or until discharge, if <4 d), IV antibiotics within 24 hr of admission, and early initiation of antibiotics IV antibiotics identified by The Severe Sepsis and Septic Shock Management Bundle guidelines
Kudo et al (19)	Retrospective analysis	3,694	The Japan Septic Disseminated Intravascular Coagulation study, Tohoku Sepsis Registry, and the Focused Outcomes Research in Emergency Care in Acute Respiratory Distress Syndrome, Sepsis, and Trauma sepsis study	Adults/sepsis-1 and -2 criteria for septic shock/ICU admission
Sciicluna et al (20)	Cohort	306	Two ICUs in the Netherlands 2011–2012, 29 ICUs, a nursing home in the Netherlands, U.S. ICUs	Sepsis-2 criteria with expected length of stay >24 hr/ICU admission

(Continued)

TABLE 1. (Continued).
Study Cohort Characteristics

Reference	Study Design	Total Number of Participants	Data Source	Inclusion Criteria (Sepsis Criteria/Admission Type)
Sharafoddini et al (21)	Cohort	5,539	MIMIC-III	Adults/sepsis-3 criteria/ICU admission
Antcliffe et al (22)	Post hoc analysis	176	The Vasopressin vs Norepinephrine as Initial Therapy in Septic Shock Trial	Adults/sepsis-1 criteria for septic shock/ICU admission
Bhavani et al (23)	Cohort	12,413	University of Chicago Medicine 2008–2016 and Loyola University Medical Center 2006–2017	Patients with a blood culture order, at least 4 consecutive days of antibiotics, and antibiotics received within the first 24 hr of the first procured vital sign, with the first day of antibiotics required to be given within 48 hr before or after the blood culture order/hospital admission
Davenport et al (24)	Cohort	265	265 adult patients admitted to U.K. ICUs	Sepsis-2 criteria or community-acquired pneumonia/within <2 d of hospital admission/ICU admission
Liu et al (25)	Retrospective analysis	41,368	The electronic ICU database	Sepsis-3 criteria/ICU admission
Mayhew et al (26)	Cohort	53,659	Kaiser Permanente Northern California	Suspected or confirmed infection and sepsis diagnosis and length of stay of at least 12 hr/ED hospital admission
Nowak et al (27)	Cohort	127	Prognostic Hemodynamic Profiling in the Acutely Ill Emergency Department Patient International Registry	Patients with suspected sepsis symptoms of acute onset (<3 d), blood culture, or lactate orders and confirmed as sepsis/ED admission/Nexfin device monitoring
Sweeney et al (28)	Retrospective analysis	700	14 GEO and ArrayExpress databases	Primary admission for bacterial sepsis/hospital or ICU admission
Zhang et al (29)	Retrospective analysis	685	12 GEO and ArrayExpress databases	Adult/sepsis/ICU admission

ED = emergency department, GenIMS = Genetic and Inflammatory Markers of Sepsis, GEO = gene expression omnibus, MIMIC-III = Medical Information Mart for Intensive Care-III, PROWESS = Activated Protein C Worldwide Evaluation in Severe Sepsis.

were validated in an external cohort (13, 19, 22, 29). Four articles were rated as good quality (13, 19, 22, 25), and three were fair in quality (12, 18, 29). Two articles demonstrated differential cluster response to hydrocortisone (22, 29). Three articles identified clusters predictive of antibiotic or IV fluid response (12, 18, 25). One article by Kudo et al (19) investigated cluster response to thrombomodulin, and one by Seymour et al (13) predicted treatment response to eritoran, early goal-directed therapy, and recombinant-activated

protein C using simulation studies. The two articles predicting hydrocortisone response used gene expression as the clustering variable (22, 29), whereas the remaining five used clinical variables.

Out of all 17 articles, 11 studies performed sensitivity analyses (13, 16, 18–20, 22–25, 28, 29), and nine studies validated their findings in a separate cohort or had their findings validated by a previous study (Supplemental Table 5, <http://links.lww.com/CCX/A963>) (13, 15, 19, 20, 22–24, 28, 29).

TABLE 2.
Clustering Methods and Findings

Reference	Type of Clustering Variables	Clustering Algorithm/Method	Data Imputation Method	No. of Clusters	Principle Difference Between Clusters
Geri et al (12)	11 clinical variables: echocardiographic and hemodynamic markers	Hierarchical clustering on principal components	Iterative principal component analysis	5	Hemodynamic state
Seymour et al (13)	29 clinical variables: demographics, vital signs, and inflammatory markers	K-means	Multiple imputation with chained equations	4	Inflammation and coagulation
Bhavani et al (15)	1 clinical variable: body temperature	Clusters previously discovered	NA	4	Inflammation
Ding and Luo (16)	34 clinical variables: vital signs and blood markers	Subgraph augmented nonnegative matrix factorization	Linear approximation from values from previous time intervals	3	Mortality
Gårdlund et al (17)	46 clinical variables: demographics, vital signs, infection site, and prior history	Latent class analysis	Creation of a monotone missingness pattern using the Markov chain Monte Carlo method Linear regression for continuous variables/ logistic regression for categorical variables	6	Infection site and disease timeline
Han et al (18)	Treatment response	Causal forests	Median imputation	2	Response to antibiotic delays
Kudo et al (19)	6 clinical variables: coagulation markers	K-means	Random forest method	4	Response to thrombomodulin
Scicluna et al (20)	Genomic	Unsupervised consensus clustering	NA	4	Immunity
Sharafoddini et al (21)	36 clinical variables: demographics, vital signs, blood and renal markers, interventions, and <i>International Classification of Diseases, 9th Edition</i> code	Hierarchical clustering and Density-Based Spatial Clustering of Application with Noise	Predictive mean matching imputation	12	Mortality
Antcliffe et al (22)	Transcriptomic	None, previously described	NA	2	Immunocompetence vs suppression
Bhavani et al (23)	1 clinical variable: body temperature	Group-based trajectory modeling	NA	4	Temperature trajectories
Davenport et al (24)	Transcriptomic	K-means	NA	2	Immune response

(Continued)

**TABLE 2. (Continued).
Clustering Methods and Findings**

Reference	Type of Clustering Variables	Clustering Algorithm/Method	Data Imputation Method	No. of Clusters	Principle Difference Between Clusters
Liu et al (25)	28 clinical variables: vital signs and blood markers	Spectral clustering	NA	4	Risk of progression to septic shock
Mayhew et al (26)	14 clinical variables: demographics and vital signs	Novel composite mixture model; PAM algorithm	No missing values	20	Mortality
Nowak et al (27)	Other: Hemodynamic measurements (stroke volume, cardiac index, systemic vascular resistance, etc.) as measured	K-means with manual identification of optimal clusters	NA	3	Cardiac index and systemic vascular resistance index
Sweeney et al (28)	Transcriptomic	The COmbined Mapping of Multiple cLUsteriNg ALgorithms using K-means and PAM	NA	3	Inflammation, coagulopathy
Zhang et al (29)	Transcriptomic	K-means	Excluded missing gene symbols.	2	Immunosuppression

NA = not applicable, PAM = Partition Around Medoids.

Race in Sepsis Phenotypes and Outcome Prediction

Race and ethnicity were included as input clustering variables in two articles but did not vary between clusters (18, 21). Race and ethnicity differed significantly between clusters in one article that clustered patients based on temperature trajectories, but the article did not include race or ethnicity as an initial clustering variable (15). Specifically, fewer African Americans were in the hyperthermic phenotypes.

DISCUSSION

Sepsis is a heterogeneous disease with high morbidity and mortality, but a few targeted treatments (2–5). The 30- and 90-day mortalities associated with sepsis, even with such treatments, remain as high as 24% and 32%, respectively, in Western nations (30). To allow for more effective treatment, many competing methods have classified sepsis into distinct phenotypes. However, methods phenotyping sepsis are also heterogeneous, and there are no resources aggregating and evaluating

existing methods for clinical impact. Accordingly, we performed a scoping review evaluating sepsis phenotyping methods and their applicability to clinical practice. Our findings show that sepsis phenotyping methods frequently identify subgroups at higher risk for poor outcome in a potentially rapid fashion, even after accounting for sepsis severity or preexisting risk factors. However, the robustness of sepsis phenotypes could benefit from validation across multiple datasets and in determining treatment response.

An ideal patient phenotyping algorithm must impact clinical decision-making, use commonly and rapidly obtained variables, and provide additional value to current Sequential Organ Failure Assessment (SOFA) and Acute Physiology And Chronic Health Evaluation (APACHE) scores to classify disease severity. To impact clinical practice, an algorithm needs to either identify groups at higher risk of harm from routine sepsis treatment or identify groups poised to benefit from investigational treatments. We identified seven articles that discovered phenotypes that preferentially respond to certain treatments, four of which evaluated

TABLE 3.
Characteristics of Studies That Predict Treatment Response

Reference	Treatment	Principle Differences Between Clusters	Notable Clustering Variables	Sensitivity Analysis	External Validation
Geri et al (12)	IV fluids	Hemodynamic state	Transesophageal echocardiography measurements	No	No
Seymour et al (13)	Eritoran, early goal-directed therapy, and recombinant activated protein C	Inflammation and coagulation	Creatinine, blood urea nitrogen, liver function tests, erythrocyte sedimentation rate, albumin, lactate, bicarbonate, troponin	Yes	Yes
Han et al (18)	Antibiotics	Response to antibiotic delays	Lactate, age, and sex	Yes	No
Kudo et al (19)	Thrombomodulin	Response to thrombomodulin	Fibrin degradation products, D-dimer	Yes	Yes
Antcliffe et al (22)	Hydrocortisone	Immunocompetence vs suppression	Gene expression	Yes	Other: Clustering method previously discovered so no validation was performed
Liu et al (25)	Early antibiotics, fluids, or vasopressors	Risk of progression to septic shock	Lactate, systolic blood pressure, cardiac SOFA, creatinine, Glasgow Coma Scale, Pao ₂ , urine output, Resp SOFA	Yes	No
Zhang et al (29)	Hydrocortisone	Immunosuppression	Gene expression	Yes	Yes

SOFA = Sequential Organ Failure Assessment.

investigational therapies. Most studies identified clusters that separated patients primarily on an immunocompetent to immunocompromised spectrum. Antcliffe et al (22) and Zhang et al (29) separately discovered immunocompetent and immunocompromised clusters of septic patients from whole blood transcriptomic data, in which only the immunocompetent group received mortality benefit from hydrocortisone. Seymour et al (13) discovered four clusters with varying degrees of inflammation, coagulopathy, and mortality and found that simulating an increased proportion of the phenotype with highest inflammatory biomarkers resulted in decreased benefit in trials for eritoran and early goal-directed therapy. Likewise, increasing the proportion of the least inflammatory phenotype predicted reduced harm, except for the trial for recombinant protein C, where benefit decreased instead.

Kudo et al (19) identified four clusters with varying levels of coagulopathy and found that only the phenotype with the greatest degree of coagulopathy had reduced 28-day and inhospital mortalities following thrombomodulin administration. Three articles evaluated routine sepsis management, identifying patients at high risk from delays or who responded appropriately to fluid or vasopressor resuscitation. Geri et al (12) used echocardiographic variables to cluster patients into five hemodynamic groups, only one of which—defined by a persistently hypovolemic state with low preload and low cardiac index despite an increased left ventricular function—responded to IV fluids. Liu et al (25) grouped patients based on risk of progression to septic shock, finding that higher risk scores reflected disease refractory to fluids, antibiotics, or vasopressors, which may be mitigated by earlier risk stratification and treatment. Similarly, Han et al (18) identified

a subgroup of patients defined by substantially greater lactate levels that responded particularly poorly to delays in antibiotics.

Sepsis progresses rapidly. Antibiotics and IV fluids are ideally administered within the first hour of suspected sepsis (31). Thus, an ideal patient stratification algorithm would include biomarkers and clinical data routinely retrieved and collected at presentation to the ICU. Of the articles that identified treatment-responsive phenotypes, the Seymour et al (13), Kudo et al (19), Han et al (18), and Liu et al (25) groups used clinical and laboratory values routinely and rapidly obtained at presentation, such as lactate, creatinine, age, and sex. Bhavani et al (23) classified patients solely based on temperature trajectories, values that can be automatically and cheaply obtained.

Finally, a patient stratification algorithm should be validated across multiple patient populations, add prognostic value above that already obtained by SOFA or APACHE III scores, and be generalizable to modern septic patients. Of the seven articles predicting treatment, four performed both sensitivity analyses accounting for comorbidities or higher SOFA/APACHE III scores and external validation in one or more validation cohorts. Out of all included articles, Seymour et al (13) included the greatest number of distinct cohorts in their external validation. Scicluna et al (20), Sweeney et al (28), and Zhang et al (29) also evaluated their algorithms in validation cohorts. Zhang et al (29) also validated their findings against the findings by Davenport et al (24) in predicting mortality. To be generalizable to current septic populations, sepsis phenotyping methods should use modern datasets and define sepsis exacting, up-to-date criteria. A majority of studies used either outdated (i.e., sepsis-2 or -1 criterion), heterogeneous (amalgamating data from multiple studies in sepsis), or unspecified (defined by clinical suspicion or diagnosis at discharge) sepsis criterion, and only five articles used some form of sepsis-3 criteria to derive sepsis phenotypes.

Racial biases drive inequities in sepsis care. Black and Hispanic patients have a 1.1–1.7 times higher occurrence rate of severe sepsis compared with White patients, American Indians have a mortality 1.6 times greater than the national average, and Asian patients have a mortality 18% greater than White patients (32). We identify two studies that cluster based on race but without racial differences in clusters (18, 21). The

Journal of the American Medical Association Guidance on Reporting Race and Ethnicity recommend that the reporting of race and ethnicity, if undertaken, should be done with consistency and in conjunction with other sociodemographic determinants of health to minimize the possibility of introducing racial inequity into research (33). Reporting race and ethnicity in sepsis phenotyping methods may serve to further research transparency and provide insight to the generalizability of cohort findings. Including race and ethnicity as clustering variables may theoretically allow subphenotypes to better account for effects of structural racism and sociodemographic factors on sepsis prognosis. However, doing so also increases the risk of false generalization, oversimplification of racial inequities, and racially biased interpretations and medical decisions (33). Race is a social construct, and since a primary objective of sepsis phenotyping is to identify targeted sepsis treatments based on biological and pathologic differences, including race and ethnicity in clustering algorithms may confound phenotypes and further drive racial inequities in sepsis.

Recommendations

From the 17 articles studied, it appears that one major limitation on the growth and scientific rigor of the field is the tendency for articles to use similar statistical approaches applied to parallel, limited patient cohorts, rather than develop more robustly validated methods consistent across multiple cohorts. Given that the “prototypical” approach derives phenotypes from one or two cohorts, without utilizing a separate cohort for external validation or evaluating preferential treatment response across clusters, it can be difficult for clinicians to decide if given prototypical clusters are “real” or artifacts from distinct datasets. For example, one clustering algorithm may generate different clusters of sepsis, with distinct biomarker and clinical profiles, if applied to two separate cohorts. Even well-validated competing subphenotypes can be difficult to evaluate, as it was rare for articles to cross-validate their subphenotypes with competing clustering methods. Furthermore, since most articles used clinical variables (e.g., heart rate) for clustering, which are individually nonspecific toward underlying pathophysiology, it can also be difficult for researchers to extrapolate precise pathophysiological mechanisms underlying subphenotypes with prototypical methods. Instead of generating new subphenotypes

with each dataset, sepsis phenotyping methods should focus on extensively validating a single set of septic subphenotypes across multiple datasets. Identifying whether subphenotypes respond differentially to certain treatments may serve the dual purpose of providing utility to clinicians while also demonstrating that subphenotypes reflect underlying pathophysiology. Thus, the most promising approach in our review involved extensive validation and sensitivity analyses combined with delineation of preferential treatment response and comparison with other subphenotyping methods.

Practically, although all phenotyping methods were able to predict mortality, fewer methods did so in a manner that may change clinical practice, providing little advantage over existing sepsis severity scores or reflecting underlying sepsis pathophysiology. Predicting effective treatment, performing external validation, and evaluating phenotypes after controlling for established risk scores are under used, but critical, factors in deriving clinically impactful phenotypes. Since methods described similar phenotypes despite different clustering methods and variables, the most common one being immunocompromised versus immunocompetent states, it may be beneficial to future phenotyping methods to compare prognostic capabilities against competing methods that identify similar phenotypes. Finally, current phenotyping methods would benefit from evaluating phenotype treatment response to treatments targeting multiple posited pathogenesis pathways. Patient phenotypes would be more robust and clinically applicable if they each predict response or nonresponse to therapies targeting coagulopathy, immune response, and inflammation, rather than only one aspect of sepsis pathogenesis.

Limitations

Our study has two major limitations. It is possible that our search criteria missed articles that would have otherwise been included. We attempted a comprehensive search of the literature by performing two searches of two databases and by manual reverse citation, searching until additional citation searches revealed no additional eligible articles. Our search criteria were developed with expert review by two independent investigators (T.J.L., T.O.-B.). However, inclusion of additional databases, such as PubMed, Web of Science, or Cochrane, may have revealed additional articles not covered by our search criteria. Second, our search criteria were intentionally broad to encompass as much

of the field as possible, according to the Peters et al (14) framework for scoping reviews. We thus included several articles that investigated populations encompassing more than only sepsis patients. We are transparent which articles included patients with infection but not sepsis, but this discrepancy in population should be noted in result interpretation.

CONCLUSIONS

Sepsis phenotyping can identify high-risk patients and those with greater likelihood of responding to treatment. The generalizability and clinical impact of the field will require increased external validation and cross-validation analyses against existing risk stratification algorithms.

1 Division of Nephrology, Hypertension, and Renal Transplantation, Department of Medicine, University of Florida College of Medicine, Gainesville, FL.

2 University of Florida College of Medicine, Gainesville, FL.

3 Department of Surgery, University of Florida College of Medicine, Gainesville, FL.

4 Precision and Intelligent Systems in Medicine (Prisma[®]), University of Florida, Gainesville, FL.

5 Sepsis and Critical Illness Research Center, University of Florida, Gainesville, FL.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website (<http://journals.lww.com/ccejjournal>).

Mr. Li was supported by a training grant from the National Institutes of Health (NIH T35HL007489). Dr. Loftus was supported by the National Institute of General Medical Sciences (NIGMS) of the NIH under Award Number K23GM140268. Dr. Ozrazgat-Baslanti was supported by K01 DK120784, R01 DK123078, and R01 DK121730 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIH/NIDDK), R01 GM110240 from the NIH/NIGMS, R01 EB029699 from the National Institute of Biomedical Imaging and Bioengineering (NIH/NIBIB), and R01 NS120924 from the National Institute of Neurological Disorders and Stroke (NIH/NINDS). Dr. Bihorac was supported R01 GM110240 from the NIH/NIGMS, R01 EB029699 and R21 EB027344 from the NIH/NIBIB, R01 NS120924 from the NIH/NINDS, and R01 DK121730 from the NIH/NIDDK. The remaining authors have disclosed that they do not have any potential conflicts of interest.

The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

For information regarding this article, E-mail: abihorac@ufl.edu

REFERENCES

1. Prescott HC, Angus DC: Enhancing recovery from sepsis: A review. *JAMA* 2018; 319:62–75

2. Rhee C, Jones TM, Hamad Y, et al; Centers for Disease Control and Prevention (CDC) Prevention Epicenters Program: Prevalence, underlying causes, and preventability of sepsis-associated mortality in US acute care hospitals. *JAMA Netw Open* 2019; 2:e187571
3. Singer M, Deutschman CS, Seymour CW, et al: The third international consensus definitions for sepsis and septic shock (Sepsis-3). *JAMA* 2016; 315:801–810
4. van der Poll T, van de Veerdonk FL, Scicluna BP, et al: The immunopathology of sepsis and potential therapeutic targets. *Nat Rev Immunol* 2017; 17:407–420
5. Meisel C, Scheffold JC, Pschowski R, et al: Granulocyte-macrophage colony-stimulating factor to reverse sepsis-associated immunosuppression: A double-blind, randomized, placebo-controlled multicenter trial. *Am J Respir Crit Care Med* 2009; 180:640–648
6. Welte T, Dellinger RP, Ebel H, et al: Efficacy and safety of trimodulin, a novel polyclonal antibody preparation, in patients with severe community-acquired pneumonia: A randomized, placebo-controlled, double-blind, multicenter, phase II trial (CIGMA study). *Intensive Care Med* 2018; 44:438–448
7. Warren BL, Eid A, Singer P, et al; KyberSept Trial Study Group: Caring for the critically ill patient. High-dose antithrombin III in severe sepsis: A randomized controlled trial. *JAMA* 2001; 286:1869–1878
8. Wunderink RG, Laterre PF, Francois B, et al; CAPTIVATE Trial Group: Recombinant tissue factor pathway inhibitor in severe community-acquired pneumonia: A randomized trial. *Am J Respir Crit Care Med* 2011; 183:1561–1568
9. Panacek EA, Marshall JC, Albertson TE, et al; Monoclonal Anti-TNF: A Randomized Controlled Sepsis Study Investigators: Efficacy and safety of the monoclonal anti-tumor necrosis factor antibody F(ab)² fragment afelimomab in patients with severe sepsis and elevated interleukin-6 levels. *Crit Care Med* 2004; 32:2173–2182
10. Shakoory B, Carcillo JA, Chatham WW, et al: Interleukin-1 receptor blockade is associated with reduced mortality in sepsis patients with features of macrophage activation syndrome: Reanalysis of a prior phase III trial. *Crit Care Med* 2016; 44:275–281
11. Cavaillon JM, Singer M, Skirecki T: Sepsis therapies: Learning from 30 years of failure of translational research to propose new leads. *EMBO Mol Med* 2020; 12:e10128
12. Geri G, Vignon P, Aubry A, et al: Cardiovascular clusters in septic shock combining clinical and echocardiographic parameters: A post hoc analysis. *Intensive Care Med* 2019; 45:657–667
13. Seymour CW, Kennedy JN, Wang S, et al: Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* 2019; 321:2003–2017
14. Peters MD, Godfrey C, McInerney P, et al: Chapter 11: scoping reviews (2020 version). JBI Manual for Evidence Synthesis. JBI, 2020
15. Bhavani SV, Wolfe KS, Hrusch CL, et al: Temperature trajectory subphenotypes correlate with immune responses in patients with sepsis. *Crit Care Med* 2020; 48:1645–1653
16. Ding M, Luo Y: Unsupervised phenotyping of sepsis using nonnegative matrix factorization of temporal trends from a multivariate panel of physiological measurements. *BMC Med Inform Decis Mak* 2021; 21:95
17. Gårdlund B, Dmitrieva NO, Pieper CF, et al: Six subphenotypes in septic shock: Latent class analysis of the PROWESS shock study. *J Crit Care* 2018; 47:70–79
18. Han X, Spicer A, Carey KA, et al: Identifying high-risk subphenotypes and associated harms from delayed antibiotic orders and delivery. *Crit Care Med* 2021; 49:1694–1705
19. Kudo D, Goto T, Uchimido R, et al: Coagulation phenotypes in sepsis and effects of recombinant human thrombomodulin: An analysis of three multicentre observational studies. *Crit Care* 2021; 25:114
20. Scicluna BP, van Vught LA, Zwinderman AH, et al; MARS consortium: Classification of patients with sepsis according to blood genomic endotype: A prospective cohort study. *Lancet Respir Med* 2017; 5:816–826
21. Sharafoddini A, Dubin JA, Lee J: Identifying subpopulations of septic patients: A temporal data-driven approach. *Comput Biol Med* 2021; 130:104182
22. Antcliffe DB, Burnham KL, Al-Beidh F, et al: Transcriptomic signatures in sepsis and a differential response to steroids. From the VANISH randomized trial. *Am J Respir Crit Care Med* 2019; 199:980–986
23. Bhavani SV, Carey KA, Gilbert ER, et al: Identifying novel sepsis subphenotypes using temperature trajectories. *Am J Respir Crit Care Med* 2019; 200:327–335
24. Davenport EE, Burnham KL, Radhakrishnan J, et al: Genomic landscape of the individual host response and outcomes in sepsis: A prospective cohort study. *Lancet Respir Med* 2016; 4:259–271
25. Liu R, Greenstein JL, Fackler JC, et al: Spectral clustering of risk score trajectories stratifies sepsis patients by clinical outcome and interventions received. *Elife* 2020; 9:e58142
26. Mayhew MB, Petersen BK, Sales AP, et al: Flexible, cluster-based analysis of the electronic medical record of sepsis with composite mixture models. *J Biomed Inform* 2018; 78:33–42
27. Nowak RM, Reed BP, Nanayakkara P, et al: Presenting hemodynamic phenotypes in ED patients with confirmed sepsis. *Am J Emerg Med* 2016; 34:2291–2297
28. Sweeney TE, Azad TD, Donato M, et al: Unsupervised analysis of transcriptomics in bacterial sepsis across multiple datasets reveals three robust clusters. *Crit Care Med* 2018; 46:915–925
29. Zhang Z, Pan Q, Ge H, et al: Deep learning-based clustering robustly identified two classes of sepsis with both prognostic and predictive values. *EBioMedicine* 2020; 62:103081
30. Bauer M, Gerlach H, Vogelmann T, et al: Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019—results from a systematic review and meta-analysis. *Crit Care* 2020; 24:239
31. Rivers E, Nguyen B, Havstad S, et al; Early Goal-Directed Therapy Collaborative Group: Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med* 2001; 345:1368–1377
32. Sepsis Alliance: Racial Equity in Sepsis Care Matters. 2021. Available at: [https://www.sepsis.org/news/racial-equity-in-sepsis-care-matters/#:~:text=We%20know%2C%20for%20example%2C%20that,%25%20higher\)%5Bix%5D](https://www.sepsis.org/news/racial-equity-in-sepsis-care-matters/#:~:text=We%20know%2C%20for%20example%2C%20that,%25%20higher)%5Bix%5D.). Accessed January 3, 2022
33. Flanagin A, Frey T, Christiansen SL; AMA Manual of Style Committee: Updated guidance on the reporting of race and ethnicity in medical and science journals. *JAMA* 2021; 326:621–627