ORIGINAL

Development and validation of novel sepsis subphenotypes using trajectories of vital signs

Sivasubramanium V. Bhavani[1,2,10*] , Matthew Semler[3], Edward T. Qian[3], Philip A. Verhoef[4,5], Chad Robichaux[6], Matthew M. Churpek[7,8] and Craig M. Coopersmith[2,9]

Abstract

Purpose: Sepsis is a heterogeneous syndrome and identification of sub-phenotypes is essential. This study used trajectories of vital signs to develop and validate sub-phenotypes and investigated the interaction of sub-phenotypes with treatment using randomized controlled trial data.

Methods: All patients with suspected infection admitted to four academic hospitals in Emory Healthcare between 2014–2017 (training cohort) and 2018–2019 (validation cohort) were included. Group-based trajectory modeling was applied to vital signs from the first 8 h of hospitalization to develop and validate vitals trajectory sub-phenotypes. The associations between sub-phenotypes and outcomes were evaluated in patients with sepsis. The interaction between sub-phenotype and treatment with balanced crystalloids versus saline was tested in a secondary analysis of SMART (Isotonic Solutions and Major Adverse Renal Events Trial).

Results: There were 12,473 patients with suspected infection in training and 8256 patients in validation cohorts, and 4 vitals trajectory sub-phenotypes were found. Group A ($N = 3483$, 28%) were hyperthermic, tachycardic, tachypneic, and hypotensive. Group B ($N = 1578$, 13%) were hyperthermic, tachycardic, tachypneic (not as pronounced as Group A) and hypertensive. Groups C ($N = 4044$, 32%) and D ($N = 3368$, 27%) had lower temperatures, heart rates, and respiratory rates, with Group C normotensive and Group D hypotensive. In the 6,919 patients with sepsis, Groups A and B were younger while Groups C and D were older. Group A had the lowest prevalence of congestive heart failure, hypertension, diabetes mellitus, and chronic kidney disease, while Group B had the highest prevalence. Groups A and D had the highest vasopressor use ($p < 0.001$ for all analyses above). In logistic regression, 30-day mortality was significantly higher in Groups A and D ($p < 0.001$ and $p = 0.03$, respectively). In the SMART trial, sub-phenotype significantly modified treatment effect ($p = 0.03$). Group D had significantly lower odds of mortality with balanced crystalloids compared to saline (odds ratio (OR) 0.39, 95% confidence interval (CI) 0.23–0.67, $p < 0.001$).

Conclusion: Sepsis sub-phenotypes based on vital sign trajectory were consistent across cohorts, had distinct outcomes, and different responses to treatment with balanced crystalloids versus saline.

Keywords: Sepsis, Vital signs, Sub-phenotypes, Intravenous fluids, Phenotypes

*Correspondence: sbhava2@emory.edu
[10] Division of Pulmonary, Allergy, Critical Care & Sleep Medicine, Emory University School of Medicine, 615 Michael St., Atlanta, GA 30322, USA Full author information is available at the end of the article

Introduction

Sepsis is a heterogeneous syndrome characterized by a dysregulated host response to infection that results in 270,000 deaths and $60 billion in hospital costs in the United States of America (USA) annually [1–3]. Decades of clinical trials have not identified therapies that consistently benefit patients with sepsis overall [4]. It

![Springer logo]

is hypothesized that these negative trials occur due to between-patient variability in response to treatment [5]. Thus, subtyping the heterogeneous syndrome of sepsis into distinct "physiological states of interest" may lead to precision therapies targeted toward treatable traits [6].

Prior studies have identified between two to six sepsis sub-phenotypes by applying unsupervised approaches, such as k-means and latent class analysis, to clinical and biomarker data from electronic health records and randomized control trials [7–11]. Most studies have identified sub-phenotypes using static measurements of biomarkers or vital signs. However, sepsis is a dynamic process with biological and physiological responses that evolve over minutes to hours [12–14]. This "temporal instability" of sepsis suggests that static snapshots of labs and vitals may not identify sub-phenotypes that are consistent over time [15–19]. Using longitudinal data may more precisely identify sepsis sub-phenotypes that differ in clinical characteristics, outcomes, and responses to treatment [20–24]. Additionally, routine bedside measurements such as vital signs may allow for increased feasibility in precision enrollment for clinical trials.

Most sub-phenotype models are developed using data from patients who present to the emergency department (ED) with established organ dysfunction and sepsis. However, this excludes informative data from infected patients who will go on to develop sepsis later in the hospitalization. To develop a sub-phenotype model for prospective implementation in the ED, the model would ideally be generalizable and applicable to all patients presenting with infection, because at point of presentation it would be uncertain which patients will go on to meet formal sepsis criteria. This study was designed with prospective implementation in mind, and thus we chose to develop and validate the sepsis sub-phenotype model in all patients presenting with suspected infection, and subsequently evaluated the performance of the model specifically in patients with sepsis.

The objectives of this study were to: (1) develop and validate dynamic sub-phenotypes in patients with infection using longitudinal vital signs (temperature, heart rate, respiratory rate, and blood pressure) measured within the first 8 h of hospital presentation; (2) evaluate the clinical characteristics and outcomes of the vitals trajectory sub-phenotypes in patients with sepsis; (3) test whether sub-phenotype modifies the effect of treatment with balanced crystalloids versus saline on mortality in patients with sepsis using data from a randomized controlled trial (RCT) [25].

Take-home message

In a multi-center retrospective analysis of 20,729 hospitalized patients with suspected infection, four subphenotypes were identified following distinct trajectories of vital signs over the first 8 hours of hospitalization. These vitals trajectory subphenotypes had distinct baseline clinical characteristics, lab abnormalities, organ dysfunction profiles, and outcomes. In a secondary analysis of a randomized control trial of balanced crystalloids versus saline, subphenotype significantly modified the effect of intravenous fluid on mortality. The four sepsis subphenotypes based on universally available vital signs may have different responses to treatment and could inform precision enrollment for clinical trials.

Methods

Study design

The first part of the study was a retrospective, observational study performed in the Emory Healthcare system to develop and validate vitals trajectory sub-phenotypes. The second part of the study was a secondary analysis of patient-level data from the Isotonic Solutions and Major Adverse Renal Events Trial (SMART) performed at Vanderbilt University Medical Center. The retrospective, observational study was approved by the Emory Institutional Review Board (IRB) with waiver of informed consent, and the secondary analysis of de-identified data from the RCT was considered by the Emory IRB to be non-human subject research (STUDY00001815).

Study cohort

In the observational study, all adult patients admitted though the ED to four hospitals in the Emory Healthcare system between January 2014 and December 2019 were eligible for study inclusion. Patients with evidence of suspected infection were identified in the electronic health record using a combination of antibiotic administration (oral or parenteral) and body fluid culture collection (blood, urine, or cerebrospinal fluid) [8], with the cohort further limited to patients who received antibiotics within 6 h of presentation. Patients who died or were discharged in the first 8 h were excluded since study analyses used vital signs from the first 8 h of hospital presentation. Patients transferred to other hospitals were excluded for incomplete outcome information. Finally, patients with less than 3 complete sets of vital signs in the first 8 h were excluded, because development of the trajectory model requires multiple longitudinal measurements. The multi-center cohorts of patients with suspected infection were partitioned into training and validation cohorts by year of admission, with admissions between 2014 and 2017 partitioned to training and admissions between 2018 and 2019 partitioned to validation. The temporally distinct validation cohort was designed to emulate prospective

implementation of the sub-phenotype model across the four hospitals to test generalizability. Given the objective to emulate prospective implementation, where we would not have the knowledge of whether the patient will go on to meet formal sepsis criteria, we developed and validated the sub-phenotype model in all-comers presenting to the ED with suspected infection.

Measurement of vital signs

Oral temperature, heart rate, respiratory rate, systolic and diastolic blood pressure from the first 8 h of presentation to the hospital were included. Erroneous recordings of vital signs were identified and excluded as per prior publications [26]. The vitals data were split into eight 1-h blocks of time. For missing vital signs (i.e., no vital signs measured at a particular hour), no imputation process was employed because the group-based trajectory model (GBTM) algorithm outlined below handles missing data through likelihood estimation. If multiple measurements were available in a 1-h period, the mean measurement was used for analysis. All vital signs in the training and validation cohort were standardized to the mean and standard deviation of that vital sign in that cohort.

Group-based trajectory model development and validation

The GBTM algorithm was applied to vitals data in the training cohort. GBTM is an application of finite mixture modeling and is used to identify groups of individuals following similar trajectories of variables over time [27, 28]. The algorithm computes the underlying coefficients for the polynomial functions describing the trajectories of the vital signs over time for each of the groups. Prior to fitting the model, we used likelihood ratio testing to determine the best-fit polynomial shape for each vital sign (i.e., linear vs. quadratic). Using GBTM, we pre-specified the selected polynomial shapes for each vital sign, and tested two, three, four, five, and six-group models. The resulting models with varying numbers of sub-phenotypes were compared using: (1) Bayesian Information Criteria (BIC)—a metric of how well the model fits the data, with penalization for increasing complexity, and (2) Subgroup distribution—if one or more sub-phenotypes contained less than five percent of the cohort, the model was not eligible for selection. After the optimal model was selected, patients were assigned the sub-phenotype for which they had the highest probability of group membership. Model goodness-of-fit was assessed by ensuring the average posterior probability of group membership was ≥ 70% for all sub-phenotypes. The agreement between the full 8-h model and parsimonious models with 1 to 7 h of vitals data were evaluated to test whether fewer vital sign measurements

were sufficient for classification. The full 8-h training model was then applied to the validation data: The sub-phenotypes are defined by a set of five unique polynomial functions describing each vital sign as a function of time from presentation to the hospital (i.e., Temperature $= \beta_0 + \beta_1 *Time + \beta_2 *Time^2$). As done in prior work, patients in the validation cohort were classified to the sub-phenotype trajectory that resulted in the lowest mean squared error [22, 23].

Model performance in patients with sepsis in training and validation cohorts

Patients with sepsis were identified from the overall cohort of infected patients using Sequential Organ Failure Assessment (SOFA) score ≥ 2 in the first 24 h of admission. Differences in demographics, comorbidities, and clinical characteristics between sub-phenotypes in patients with sepsis were compared using analysis of variance (ANOVA) or chi-squared tests, as appropriate. The sepsis sub-phenotypes were evaluated for association with the need for renal replacement therapy, mechanical ventilation, vasopressors, inotropes, and 30-day hospital mortality. The above analyses were repeated on (1) the overall cohort of patients with infection, (2) patients with suspicion of infection meeting ≥ 2 Systemic Inflammatory Response Syndrome (SIRS) criteria in the first 24 h, and (3) patients with suspicion of infection and fewer than 3 sets of vital signs in the first 8 h.

Association of sub-phenotypes with laboratory values

Laboratory values from the first 24 h of hospitalization were compared between sub-phenotypes: white blood cell (WBC) count, absolute neutrophils, absolute lymphocytes, neutrophil-to-lymphocyte ratio, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), hemoglobin, platelets, international normalized ratio (INR), serum creatinine, albumin, total bilirubin, B-Natriuretic Peptide (BNP), and lactic acid. If a patient had multiple measurements of a lab, the maximum value from the first 24 h of hospitalization was used (for hemoglobin, platelets, and albumin, the minimum value was used). No imputation process was used for missing labs. Lab values were compared between sub-phenotypes using ANOVA. Given the pre-specified set of 14 lab tests, all tests of significance were corrected for multiple testing using Bonferroni correction. We tested whether the addition of laboratory values to the GBTM model significantly changed sub-phenotype assignments compared to the vitals-only model.

Association of sub-phenotypes with mortality

The primary prognostic outcome was 30-day hospital mortality. Logistic regression was performed to evaluate

the association between sub-phenotypes and 30-day mortality adjusting for age, sex, race, and comorbidities (congestive heart failure, chronic pulmonary disease, diabetes mellitus, hypertension, chronic kidney disease, liver disease, and metastatic cancer). The most prevalent sub-phenotype in the overall cohort served as the reference group.

Heterogeneity of treatment effect for intravenous fluids

The SMART trial was a cluster-randomized cluster-crossover trial that compared balanced crystalloids versus saline in critically ill adults admitted to Vanderbilt University Medical Center [25]. In this secondary analysis of the SMART trial, the vitals trajectory model was applied to vital signs from the first 8 h of hospitalization for study patients enrolled with a diagnosis of sepsis. We excluded patients transferred from other hospitals and those who were transferred to the intensive care unit (ICU) more than 72 h after hospital presentation. The vitals used for sub-phenotype classification were restricted to vitals measured prior to administration of study fluid (given that fluid itself could alter the trajectory of vital signs). Vital signs were standardized to the mean and standard deviation of the original training data and the study patients were classified to the sub-phenotype trajectory that resulted in the lowest mean squared error [22, 23]. Clinical characteristics and outcomes were compared between the sub-phenotypes.

The primary outcome (30-day hospital mortality) in each sub-phenotype was compared between balanced crystalloid and saline treatment arms adjusting for pre-specified baseline covariates [25, 29]. Heterogeneity of treatment effect (HTE) was tested in a model including the baseline covariates, sub-phenotypes, treatment assignment, and interaction terms between the sub-phenotype and treatment assignment. P-values for HTE were calculated using likelihood ratio test between the nested model without interaction terms and full model with interaction terms. A simulation study was performed to test the feasibility of application of the vitals trajectory model in real-time precision enrollment in the SMART trial (Supplementary Methods).

Results

The retrospective cohort included 20,729 patients with suspected infection who received antibiotics within 6 h of hospital presentation, with 12,473 patients in the training cohort (admitted between 2014 and 2017) and 8256 patients in the validation cohort (admitted between 2018 and 2019) (Supplementary Fig. 1). The 4-group trajectory model had the highest BIC of the models with adequate subgroup distribution (Supplementary Table 1) and had an average posterior probability of group membership > 90% for all sub-phenotypes. The importance of individual vitals in the final trajectory model are presented (Supplementary Table 2).

Group A patients ($N = 3483$, 28%) were hyperthermic, tachycardic, and tachypneic and were relatively hypotensive. Group B ($N = 1578$, 13%) were also hyperthermic, tachycardic and tachypneic, but not as pronounced as Group A, and were hypertensive. Group C ($N = 4044$, 32%) and Group D ($N = 3368$, 27%) had lower temperatures, heart rates, and respiratory rates, with Group C being normotensive and Group D being the most hypotensive sub-phenotype. The trajectory shapes in the validation cohort were similar to the training cohort (Fig. 1). Baseline vital signs alone had 69.6% accuracy in predicting sub-phenotype membership compared to the full 8-h model, but each additional hour of vital signs significantly increased accuracy (Supplementary Fig. 2).

In the training cohort, there were 6,919 patients meeting criteria for sepsis (defined as SOFA ≥ 2 in the first 24 h). Groups A and B were younger with a median age of 58 years, while Groups C and D were older with a median age of 70 and 69 years, respectively ($p < 0.001$). Baseline prevalence of comorbidities were significantly different between sub-phenotypes ($p < 0.001$): Group A had the lowest prevalence of congestive heart failure (26%), hypertension (59%), diabetes mellitus (31%), and chronic kidney disease (29%), but the highest prevalence of metastatic cancer (10%). Group B had the highest prevalence of congestive heart failure (40%), hypertension (91%), and chronic kidney disease (57%). Admission to the ICU was higher in Groups A, B and D (39%, 33%, and 33%) than in Group C (22%) ($p < 0.001$). Requirement of renal replacement therapy during hospitalization was highest in Group B ($p < 0.001$). Vasopressor use was higher in Groups A and D, and inotrope use was highest in Group D ($p < 0.001$). 30-day hospital mortality was significantly different between sub-phenotypes ($p = 0.02$): 3.8% for Group A, 2.2% Group B, 2.4% Group C, and 3.7% Group D (Table 1).

Among the 2,759 patients with sepsis in the validation cohort, the relative distribution of demographics,

(See figure on next page.)
Fig. 1 Group-based trajectory modeling of vital signs in the training and validation cohorts. Using group-based trajectory modeling, four vitals trajectory sub-phenotypes were identified in the training and validation cohorts. Group A patients ($N = 3483$, 28%) were hyperthermic, tachycardic, and tachypneic and had relatively lower systolic and diastolic blood pressure. Group B ($N = 1578$, 13%) were also hyperthermic, tachycardic and tachypneic, but not as pronounced as Group A, and were hypertensive. Group C ($N = 4044$, 32%) and Group D ($N = 3368$, 27%) had lower temperatures, heart rates, and respiratory rates, with Group C normotensive and Group D being the most hypotensive sub-phenotype
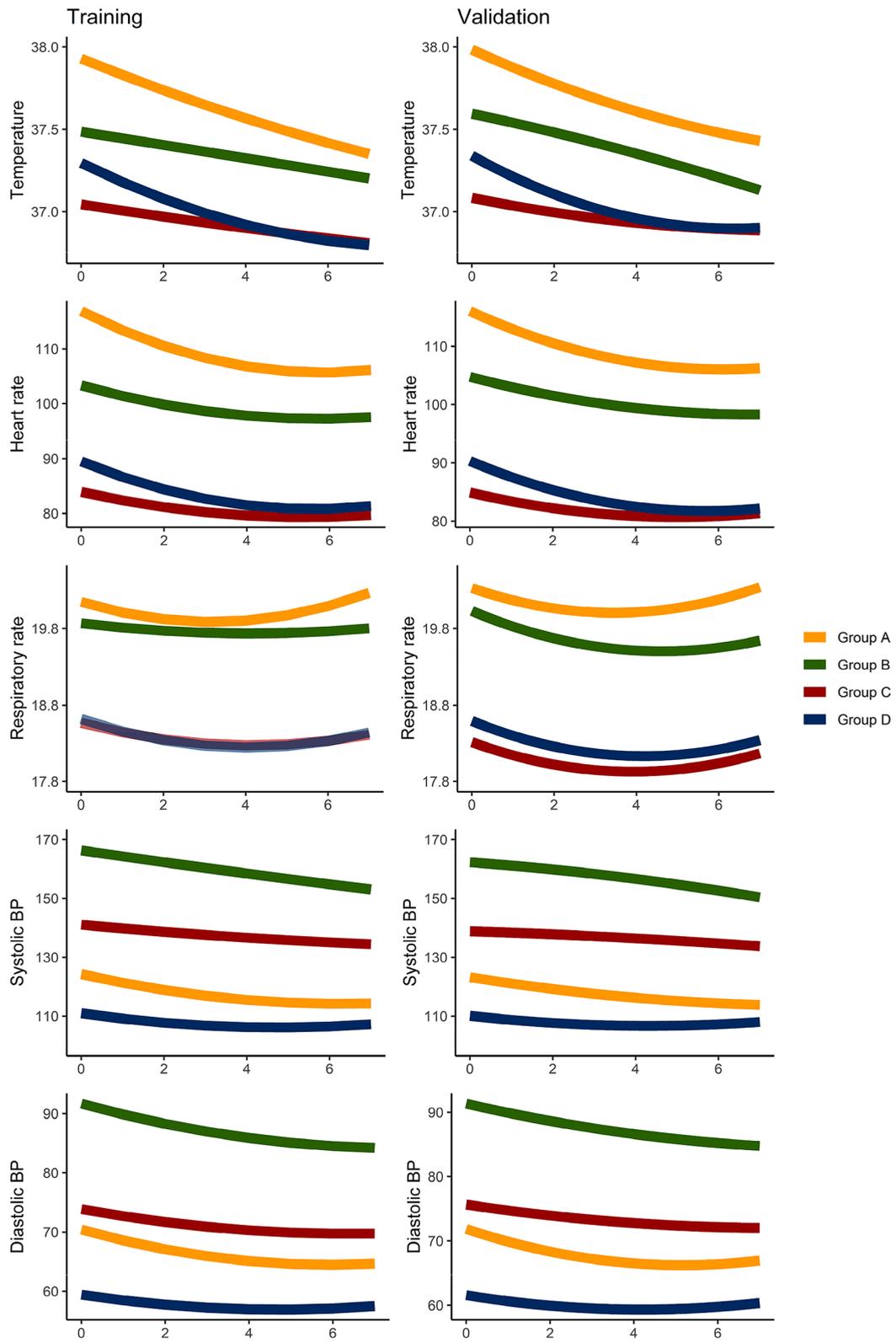
Fig. 1 (See legend on previous page.)

Table 1 Comparison of clinical characteristics between sub-phenotypes of patients with sepsis in the training and validation cohorts

	Training cohort (2014–2017)					Validation cohort (2018–2019)				
	A	*B*	*C*	*D*	*p* value	*A*	*B*	*C*	*D*	*p* value
N (%)	2024 (29)	787 (11)	1847 (27)	2261 (33)		877 (32)	290 (11)	609 (22)	983 (36)	
Characteristics										
Age, years	58	58	70	69	< 0.001	62	61	71	70	< 0.001
Sex, male	50	57	56	51	< 0.001	51	59	55	50	0.03
Race					< 0.001					< 0.001
Black	43	61	33	27		40	54	31	27	
White	50	34	60	67		51	39	62	66	
Other	7.2	4.8	6.9	6.6		9.7	6.6	7.4	7.1	
Hispanic	4.4	2.7	2.8	2.9	0.2	5.1	2.1	3.4	4.5	0.05
Comorbidities										
CHF	26	40	36	36	< 0.001	30	39	38	38	0.001
Pulmonary	32	32	32	32	0.9	30	34	36	30	0.04
Hypertension	59	91	81	69	< 0.001	58	87	83	67	< 0.001
DM	31	42	42	36	< 0.001	33	48	39	35	< 0.001
CKD	29	57	46	41	< 0.001	28	41	38	38	< 0.001
Liver	16	14	14	18	0.001	15	9.7	13	15	0.1
Cancer	10	4.4	4.5	7.1	< 0.001	12	7.6	6.1	9.3	0.001
Outcomes										
ICU	39	33	22	33	< 0.001	52	46	33	44	< 0.001
Dialysis	8.8	30	13	11	< 0.001	7.6	16	13	12	< 0.001
Ventilator	15	12	12	10	< 0.001	18	23	17	15	0.009
Vasopressors	19	12	14	20	< 0.001	28	17	22	29	< 0.001
Inotropes	3.4	1.1	2.8	4.5	< 0.001	4.8	1.7	2.3	4.5	0.01
Mortality	3.8	2.2	2.4	3.7	0.02	4.7	3.8	2.3	5.2	0.04

Presented are the comparison of demographics, comorbidities, and outcomes between sub-phenotypes A, B, C, and D in training and validation cohorts. Age is presented as median, and all other values are presented as percentages. Inotropes are defined as dobutamine and milrinone. Mortality represents 30-day hospital mortality. *p* values signify the results of comparisons between sub-phenotypes through chi-squared or ANOVA testing, as appropriate

CHF congestive heart failure, *Pulmonary* chronic pulmonary disease, *DM* diabetes mellitus, *CKD* chronic kidney disease, *Liver* chronic liver disease, *Cancer* metastatic cancer, *ICU* intensive care unit

comorbidities, and outcomes of the sub-phenotypes were similar to the training cohort (Table 1). In the validation cohort, admission to ICU was higher in Groups A, B, and D compared to Group C (*p* < 0.001); requirement of renal replacement therapy was highest in Group B (*p* < 0.001); vasopressor use was higher in Groups A and D (*p* < 0.001). Mortality for the sub-phenotypes was 4.7% for Group A, 3.8% Group B, 2.3% Group C, and 5.2% Group D (*p* = 0.04). The relative distribution of sub-phenotype characteristics were also similar in the overall cohort of infected patients and across varying inclusion criteria (Supplementary Tables 3–5).

Association of sub-phenotypes with laboratory values
Several laboratory markers were significantly different between the 4 sub-phenotypes in both training and validation cohorts. In the training cohort, Group A had the highest WBC count, neutrophil-to-lymphocyte ratio,

and lactic acid levels (*p* < 0.001) (Fig. 2); Group B had the highest serum creatinine and BNP (*p* < 0.001); Groups A and D had lower albumin (*p* < 0.001); Group D had the lowest hemoglobin (*p* < 0.001). The relative distributions of laboratory values were similar in the validation cohort (Supplementary Fig. 3, Supplementary Tables 6 & 7). The addition of laboratory markers to clustering did not significantly change sub-phenotype membership (Supplementary Table 8).

Association of sub-phenotypes with mortality
In logistic regression adjusting for age, demographics, and comorbidities, with Group C as the reference group, 30-day mortality was significantly higher in Group A (Training cohort: odds ratio (OR) 1.96, 95% confidence interval (CI) 1.32–2.91, *p* < 0.001; Validation cohort: OR 2.50, 95% CI 1.31–4.77, *p* = 0.005). Mortality was also significantly higher in Group D (Training cohort: OR 1.54,
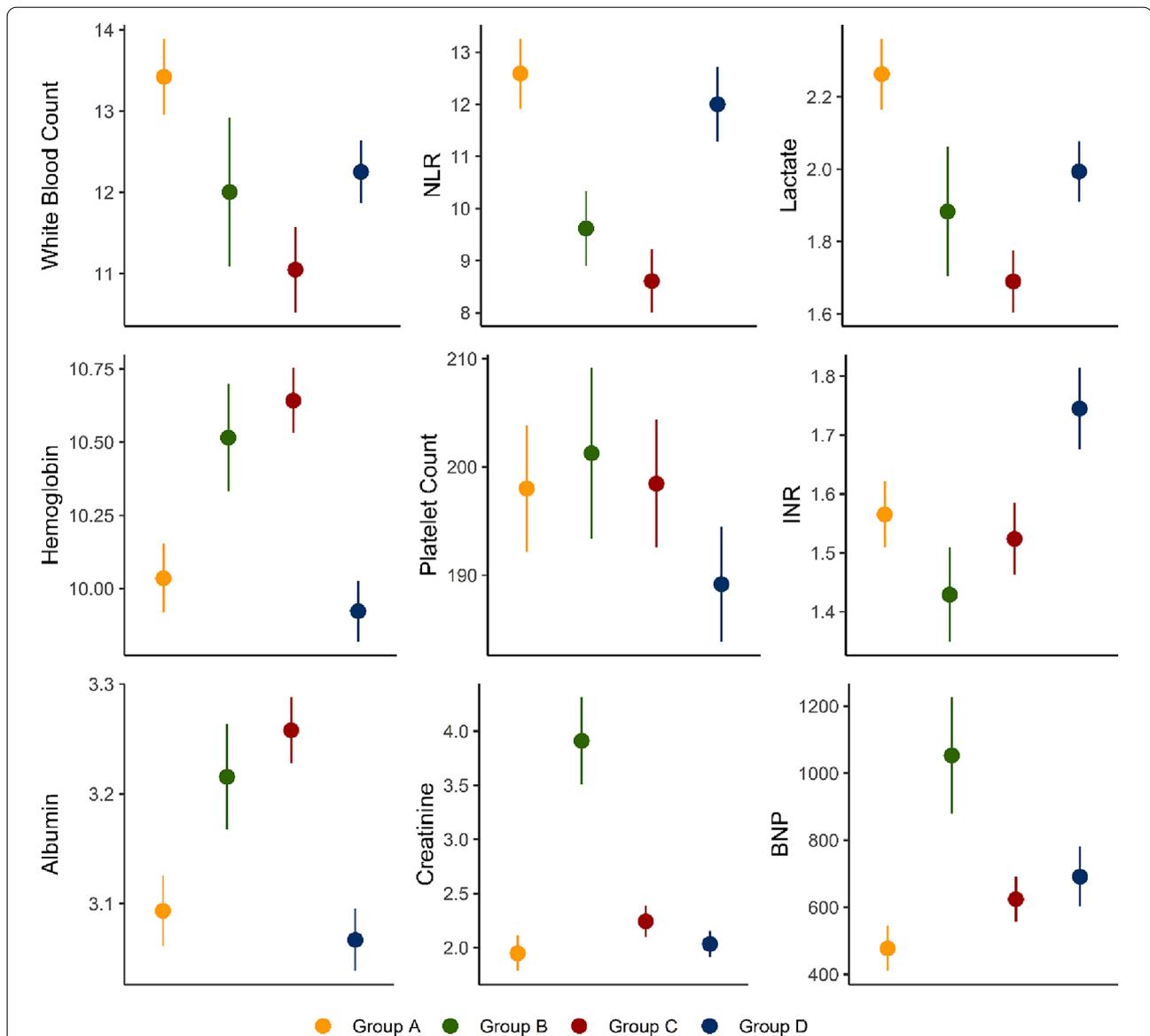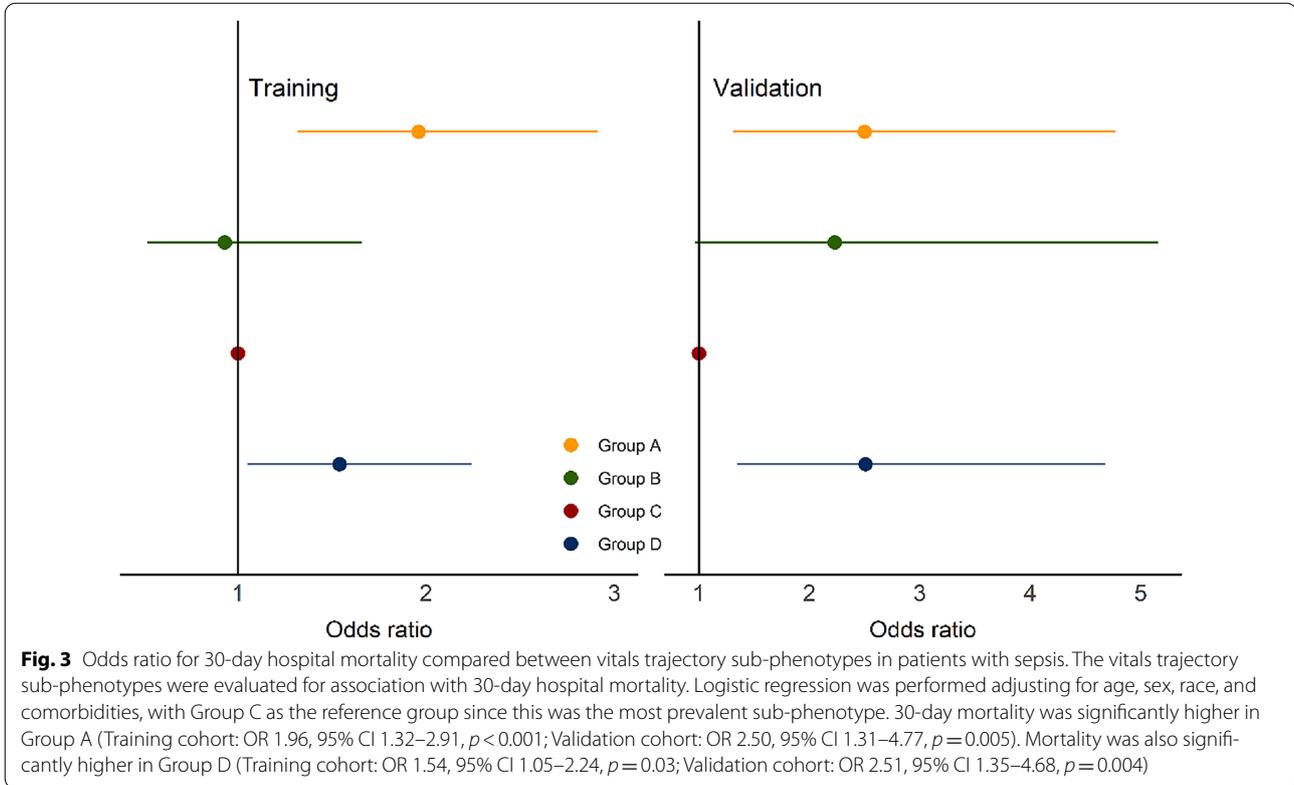
Fig. 2 Laboratory values compared between vitals trajectory sub-phenotypes in sepsis patients in the training cohort. Laboratory values (most abnormal values in the first 24 h of hospitalization) were compared between the vitals trajectory sub-phenotypes using ANOVA testing. All laboratory values presented were significantly different between sub-phenotypes after multiple testing correction either in the training or validation cohorts. Group A had the highest white blood cell count, neutrophil-to-lymphocyte ratio, and lactic acid levels ($p < 0.001$). Group B had the highest creatinine and BNP levels ($p < 0.001$). Group D had the lowest hemoglobin ($p < 0.001$). These relative distributions of labs were similar in the validation cohort. *NLR* neutrophil-to-lymphocyte ratio, *INR* international normalized ratio, *BNP* Brain natriuretic peptide

95% CI 1.05–2.24, $p = 0.03$; Validation cohort: OR 2.51, 95% CI 1.35–4.68, $p = 0.004$) (Fig. 3).

Heterogeneity of treatment effect for intravenous fluids
Of the 1641 patients with sepsis in the SMART trial, 368 were excluded for not having documented vital signs prior to fluid administration, 285 for transferring from other hospitals, and 154 for hospitalization longer than 72 h prior to trial enrollment. Enrollment criteria for the

SMART trial cohort are compared to the observational cohorts in Supplementary Table 9. The 834 study patients were classified into the 4 sub-phenotypes: Group A ($N = 319$, 38%), Group B (93, 11%), Group C (100, 12%), and Group D (322, 39%). Consistent with our primary results, Group A was the youngest, while Group D was the oldest (median 53 years vs 63 years, $p < 0.001$). Group A had the lowest prevalence of congestive heart failure, hypertension, and chronic kidney disease. Groups B and

Fig. 3 Odds ratio for 30-day hospital mortality compared between vitals trajectory sub-phenotypes in patients with sepsis. The vitals trajectory sub-phenotypes were evaluated for association with 30-day hospital mortality. Logistic regression was performed adjusting for age, sex, race, and comorbidities, with Group C as the reference group since this was the most prevalent sub-phenotype. 30-day mortality was significantly higher in Group A (Training cohort: OR 1.96, 95% CI 1.32–2.91, $p < 0.001$; Validation cohort: OR 2.50, 95% CI 1.31–4.77, $p = 0.005$). Mortality was also significantly higher in Group D (Training cohort: OR 1.54, 95% CI 1.05–2.24, $p = 0.03$; Validation cohort: OR 2.51, 95% CI 1.35–4.68, $p = 0.004$)

D had high baseline prevalence of congestive heart failure and chronic kidney disease. 30-day hospital mortality was significantly different between the sub-phenotypes ($p = 0.02$): Group A (21%), Group B (14%), Group C (22%), and Group D (28%) (Supplementary Tables 10 and 11).
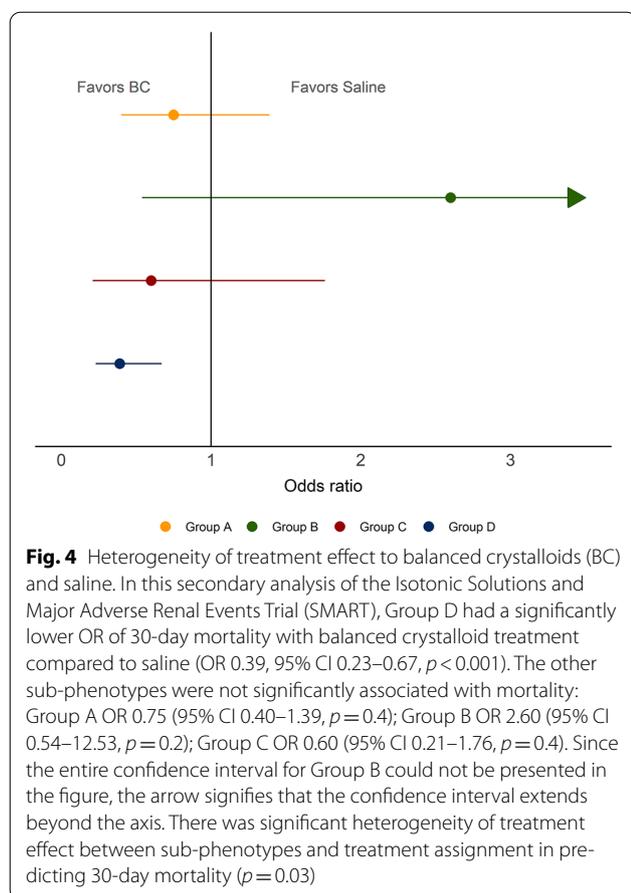
In logistic regression, Group D had a significantly lower OR of 30-day mortality with balanced crystalloids compared to saline (OR 0.39, 95% CI 0.23–0.67, $p < 0.001$) (Fig. 4). There was significant HTE between sub-phenotype and treatment assignment in predicting 30-day mortality ($p = 0.03$). In a sensitivity analysis limited to patients with a complete set of 3 vital signs before fluid administration ($N = 335$), there remained significant HTE ($p = 0.04$), with lower OR of mortality with balanced crystalloids in Group D (OR 0.35, 95% CI 0.15–0.79, $p = 0.01$).

Across 1000 simulation experiments of real-time classification of study patients, 607 experiments detected a significant enough mortality benefit from balanced crystalloids in Group D to warrant early stopping of the clinical trial (mortality difference > 10% at a p value of < 0.05) (Supplementary Table 12).

Discussion

We present novel dynamic sub-phenotypes developed and validated using universally available vital signs from a broad cohort of patients presenting to the ED with suspected infection. The sub-phenotypes demonstrate strong generalizability in patients with sepsis in the training cohort and a temporally distinct validation cohort, and in critically ill patients with sepsis in an RCT cohort from a different institution. The sub-phenotypes have distinct baseline characteristics, lab abnormalities, and patterns of organ dysfunction. Finally, the sub-phenotypes demonstrate significantly different responses to balanced crystalloids and saline, suggesting these sub-phenotypes could play a role in the precision medicine approach to sepsis.

Whether a patient with infection will go on to develop organ dysfunction and sepsis is often uncertain at presentation to the ED. The vitals trajectory sub-phenotypes were developed on a broad cohort of patients presenting with suspected infection so that the model could be applied to patients with or without established sepsis on presentation. Using this generalizable sub-phenotype model, we found robust results with similar distribution of sub-phenotype clinical characteristics and outcomes in patients with infection, patients with sepsis, patients

Fig. 4 Heterogeneity of treatment effect to balanced crystalloids (BC) and saline. In this secondary analysis of the Isotonic Solutions and Major Adverse Renal Events Trial (SMART), Group D had a significantly lower OR of 30-day mortality with balanced crystalloid treatment compared to saline (OR 0.39, 95% CI 0.23–0.67, $p < 0.001$). The other sub-phenotypes were not significantly associated with mortality: Group A OR 0.75 (95% CI 0.40–1.39, $p = 0.4$); Group B OR 2.60 (95% CI 0.54–12.53, $p = 0.2$); Group C OR 0.60 (95% CI 0.21–1.76, $p = 0.4$). Since the entire confidence interval for Group B could not be presented in the figure, the arrow signifies that the confidence interval extends beyond the axis. There was significant heterogeneity of treatment effect between sub-phenotypes and treatment assignment in predicting 30-day mortality ($p = 0.03$)

meeting SIRS criteria, patients with sparse vital signs data, and ICU patients with sepsis.

Prior studies on subtyping acute respiratory distress syndrome (ARDS) and sepsis have identified between two to six sub-phenotypes, with some studies using transcriptomic data and others using clinical data from electronic health records and RCTs [7, 30]. Calfee and colleagues identified two consistent ARDS sub-phenotypes across multiple trials with differences in inflammatory biomarkers and responses to treatments [31]. In sepsis, there has been more variability in clinical sub-phenotypes. Seymour et al. identified four sub-phenotypes using labs and vitals, but these sub-phenotypes did not have significant interaction with treatments in RCTs [8]. Shankar-Hari et al. identified two sub-phenotypes in the VANISH trial and three sub-phenotypes in the LeoPARDS trial using clinical and biomarker data, with a consistent hyperinflammatory sub-phenotype associated with higher mortality, but without HTE [7]. Gardlund et al. reported six sub-phenotypes of septic shock using the PROWESS Shock Study without treatment effect differences between groups [11]. The lack of treatment differences in sepsis sub-phenotypes could

be due to lack of statistical power, clinical trial design (explanatory rather than pragmatic), or true lack of HTE [7, 32]. The distinctive strengths of our clinical sub-phenotypes are: (1) the use of routine bedside measurements, (2) the use of longitudinal measurements, and (3) the identification of HTE to the type of intravenous fluid.

Sepsis is a dynamic process characterized by rapidly evolving physiological responses that may not be adequately captured by static measurements [14, 33]. Compared to traditional static clustering methods, GBTM has the advantage of modeling the dynamic evolution of physiology over time. Additionally, GBTM sub-phenotypes are represented by clinically interpretable trajectories of clinical variables. When GBTM was applied to our cohort, we found that Group A were the hyperthermic, tachycardic, tachypneic, and hypotensive sub-phenotype. Group A had high lactate, WBC count, and neutrophil-to-lymphocyte ratio. Group B were hyperthermic, tachycardic, and tachypneic, but not to the extent of Group A. In addition, these patients were hypertensive and had high prevalence of baseline comorbidities including congestive heart failure and chronic kidney disease. Groups C and D had lower temperatures, heart rates, and respiratory rates. Group C was normotensive while Group D was the most hypotensive sub-phenotype. Groups A and D had higher odds ratio of mortality in both training and validation cohorts compared to Group C (considered the reference group as it was the predominant sub-phenotype).

One of the most common interventions in sepsis is intravenous fluids. There is evidence that dynamic vital signs could guide fluid resuscitation strategies [34–36]. In the secondary analysis of the SMART trial, we found significant heterogeneity in the treatment effect of balanced crystalloids versus saline across the sub-phenotypes. Group D (low temperature, heart rate and respiratory rate, and hypotensive) had the highest benefit from balanced crystalloids compared to saline, with significantly lower 30-day hospital mortality. In future clinical trials, Group D may be the target sub-phenotype for precision enrollment comparing balanced crystalloids and saline. Development of this trajectory model required three sets of vital signs over the first 8 h. However, the model can be applied in real-time without the entire 8 h for sub-phenotype classification. The feasibility of real-time application was demonstrated through a simulation study, in which patients in the SMART trial were assigned to Group D if they met a pre-specified probability threshold. We found that a majority of the simulated trials resulted in early stopping due to detection of a clinically and statistically significant mortality benefit from balanced crystalloids for Group D.

The study has several limitations. First, the retrospective nature of the study limits causal inferences. Second, the secondary analysis of the RCT may be underpowered to detect HTE. Third, the SALT-ED trial may serve as a complementary cohort to evaluate sub-phenotype modification of treatment effect in patients who were admitted to the wards instead of the ICU. Fourth, temporal sub-phenotypes may be modified by treatments within the sub-phenotyping window. Fifth, laboratory markers were collected as clinically indicated, resulting in missing data. Sixth, measurement variability in respiratory rates may occur due to method of measurement (bedside monitors versus visual measurements). Finally, these findings may not be generalizable to hospital-acquired infections.

The vitals trajectory sub-phenotypes with distinct clinical characteristics and organ dysfunction profiles capture the dynamic complexity of sepsis. These sub-phenotypes may respond differently to resuscitation and sepsis management strategies. Finally, the sub-phenotypes could be used for precision enrollment of future clinical trials and may play a role in the transition from the one-size-fits-all approach to the precision medicine approach to sepsis management.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1007/s00134-022-06890-z.

Author details
[1] Department of Medicine, Emory University, Atlanta, GA, USA. [2] Emory Critical Care Center, Atlanta, GA, USA. [3] Department of Medicine, Vanderbilt University, Nashville, TN, USA. [4] Department of Medicine, University of Hawaii John A. Burns School of Medicine, Honolulu, HI, USA. [5] Hawaii Permanente Medical Group, Honolulu, HI, USA. [6] Department of Biomedical Informatics, Emory University, Atlanta, GA, USA. [7] Department of Medicine, University of Wisconsin, Madison, WI, USA. [8] Department of Biostatistics and Medical Informatics, University of Wisconsin, Madison, WI, USA. [9] Department of Surgery, Emory University, Atlanta, GA, USA. [10] Division of Pulmonary, Allergy, Critical Care & Sleep Medicine, Emory University School of Medicine, 615 Michael St., Atlanta, GA 30322, USA.

Declarations

Publisher's Note

References
1. Buchman TG et al (2020) Sepsis among medicare beneficiaries: 3 the methods, models, and forecasts of sepsis, 2012–2018. Critical Care Med 48:302–318. https://doi.org/10.1097/ccm.0000000000004225
2. Rhee C et al (2017) Incidence and trends of sepsis in US Hospitals using clinical vs claims data, 2009–2014. JAMA 318:1241–1249. https://doi.org/10.1001/jama.2017.13836
3. Singer M et al (2016) The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA 315:801–810. https://doi.org/10.1001/jama.2016.0287
4. Santacruz CA, Pereira AJ, Celis E, Vincent J-L (2019) Which multicenter randomized controlled trials in critical care medicine have shown reduced mortality? A systematic review. Crit Care Med 47:1680–1691. https://doi.org/10.1097/ccm.0000000000004000
5. Prescott HC, Calfee CS, Thompson BT, Angus DC, Liu VX (2016) Toward smarter lumping and smarter splitting: rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. Am J Respir Crit Care Med 194:147–155. https://doi.org/10.1164/rccm.201512-2544CP
6. Maslove DM et al (2022) Redefining critical illness. Nat Med 28:1141–1148. https://doi.org/10.1038/s41591-022-01843-x
7. Shankar-Hari M et al (2021) Defining phenotypes and treatment effect heterogeneity to inform acute respiratory distress syndrome and sepsis trials: secondary analyses of three RCTs. Efficacy Mech Eval 8:1–104
8. Seymour CW et al (2019) Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. JAMA 321:2003–2017. https://doi.org/10.1001/jama.2019.5791
9. Zhang Z, Zhang G, Goyal H, Mo L, Hong Y (2018) Identification of subclasses of sepsis that showed different clinical outcomes and responses to amount of fluid resuscitation: a latent profile analysis. Crit Care 22:347. https://doi.org/10.1186/s13054-018-2279-3
10. Knox DB, Lanspa MJ, Kuttler KG, Brewer SC, Brown SM (2015) Phenotypic clusters within sepsis-associated multiple organ dysfunction syndrome. Intensive Care Med 41:814–822. https://doi.org/10.1007/s00134-015-3764-7
11. Gårdlund B et al (2018) Six subphenotypes in septic shock: latent class analysis of the PROWESS Shock study. J Crit Care 47:70–79. https://doi.org/10.1016/j.jcrc.2018.06.012
12. Cazalis MA et al (2014) Early and dynamic changes in gene expression in septic shock patients: a genome-wide approach. Intensive Care Med Exp 2:20. https://doi.org/10.1186/s40635-014-0020-3
13. Maslove DM, Wong HR (2014) Gene expression profiling in sepsis: timing, tissue, and translational considerations. Trends Mol Med 20:204–213. https://doi.org/10.1016/j.molmed.2014.01.006
14. Namas RA, Vodovotz Y (2016) From static to dynamic: a sepsis-specific dynamic model from clinical criteria in polytrauma patients. Ann Transl Med 4:492. https://doi.org/10.21037/atm.2016.11.72
15. Reddy K et al (2020) Subphenotypes in critical care: translation into clinical practice. Lancet Respir Med 8:631–643. https://doi.org/10.1016/S2213-2600(20)30124-7
16. Kwan A, Hubank M, Rashid A, Klein N, Peters MJ (2013) Transcriptional instability during evolving sepsis may limit biomarker based risk stratification. PLoS ONE 8:e60501. https://doi.org/10.1371/journal.pone.0060501
17. Wong HR et al (2018) Endotype transitions during the acute phase of pediatric septic shock reflect changing risk and treatment response. Crit Care Med 46:e242–e249. https://doi.org/10.1097/ccm.0000000000002932
18. Hollen MK et al (2019) Myeloid-derived suppressor cell function and epigenetic expression evolves over time after surgical sepsis. Crit Care 23:355. https://doi.org/10.1186/s13054-019-2628-x
19. Sweeney TE, Shidham A, Wong HR, Khatri P (2015) A comprehensive time-course-based multicohort analysis of sepsis and sterile inflammation reveals a robust diagnostic gene set. Sci Transl Med 7:287ra271. https://doi.org/10.1126/scitranslmed.aaa5993

20. Bhavani SV et al (2019) Identifying novel sepsis subphenotypes using temperature trajectories. Am J Respir Crit Care Med. https://doi.org/10.1164/rccm.201806-1197OC

21. Bhavani SV, Huang ES, Verhoef PA, Churpek MM (2020) Novel temperature trajectory subphenotypes in COVID-19. Chest. https://doi.org/10.1016/j.chest.2020.07.027

22. Bhavani SV et al (2022) Coronavirus disease 2019 temperature trajectories correlate with hyperinflammatory and hypercoagulable subphenotypes. Crit Care Med 50:212–223. https://doi.org/10.1097/ccm.0000000000005397

23. Bhavani SV et al (2020) Temperature trajectory subphenotypes correlate with immune responses in patients with sepsis. Crit Care Med 48:1645–1653. https://doi.org/10.1097/ccm.0000000000004610

24 Yehya N et al (2021) Temperature trajectory sub-phenotypes and the immuno-inflammatory response in pediatric sepsis. Shock. https://doi.org/10.1097/shk.0000000000001906

25. Semler MW et al (2018) Balanced crystalloids versus saline in critically ill adults. N Engl J Med 378:829–839. https://doi.org/10.1056/NEJMoa1711584

26. Churpek MM, Zadravecz FJ, Winslow C, Howell MD, Edelson DP (2015) Incidence and prognostic value of the systemic inflammatory response syndrome and organ dysfunctions in ward patients. Am J Respir Crit Care Med 192:958–964. https://doi.org/10.1164/rccm.201502-0275OC

27. Nagin DS, Jones BL, Passos VL, Tremblay RE (2018) Group-based multi-trajectory modeling. Stat Methods Med Res 27:2015–2023. https://doi.org/10.1177/0962280216673085

28. Nagin DS, Odgers CL (2010) Group-based trajectory modeling in clinical research. Annu Rev Clin Psychol 6:109–138. https://doi.org/10.1146/annurev.clinpsy.121208.131413

29 Brown RM et al (2019) Balanced crystalloids versus saline in sepsis. A secondary analysis of the SMART clinical trial. Am J Respir Crit Care Med 200:1487–1495. https://doi.org/10.1164/rccm.201903-0557OC

30. Stanski NL, Wong HR (2020) Prognostic and predictive enrichment in sepsis. Nat Rev Nephrol 16:20–31. https://doi.org/10.1038/s41581-019-0199-3

31. Calfee CS et al (2014) Subphenotypes in acute respiratory distress syndrome: latent class analysis of data from two randomised controlled trials. Lancet Respir Med 2:611–620. https://doi.org/10.1016/S2213-2600(14)70097-9

32. Santhakumaran S et al (2019) Heterogeneity of treatment effect by baseline risk of mortality in critically ill patients: re-analysis of three recent sepsis and ARDS randomised controlled trials. Crit Care 23:156. https://doi.org/10.1186/s13054-019-2446-1

33. Mayaud L et al (2013) Dynamic data during hypotensive episode improves mortality predictions among patients with sepsis and hypotension. Crit Care Med 41:954–962. https://doi.org/10.1097/CCM.0b013e3182772adb

34. Gu Q, Prasad V, Heldt T (2019) Characterizing fluid response and sepsis progression in emergency department patients. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference https://doi.org/10.1109/embc.2019.8856521

35. Prasad V, Lynch JC, Filbin MR, Reisner AT, Heldt T (2019) Clustering blood pressure trajectories in septic shock in the emergency department. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference, 494–497, doi:https://doi.org/10.1109/embc.2019.8857191

36. Brown SM et al (2016) Multi-complexity measures of heart rate variability and the effect of vasopressor titration: a prospective cohort study of patients with septic shock. BMC Infect Dis 16:551. https://doi.org/10.1186/s12879-016-1896-1