Supplementary appendix: Methods, Tables and Figures

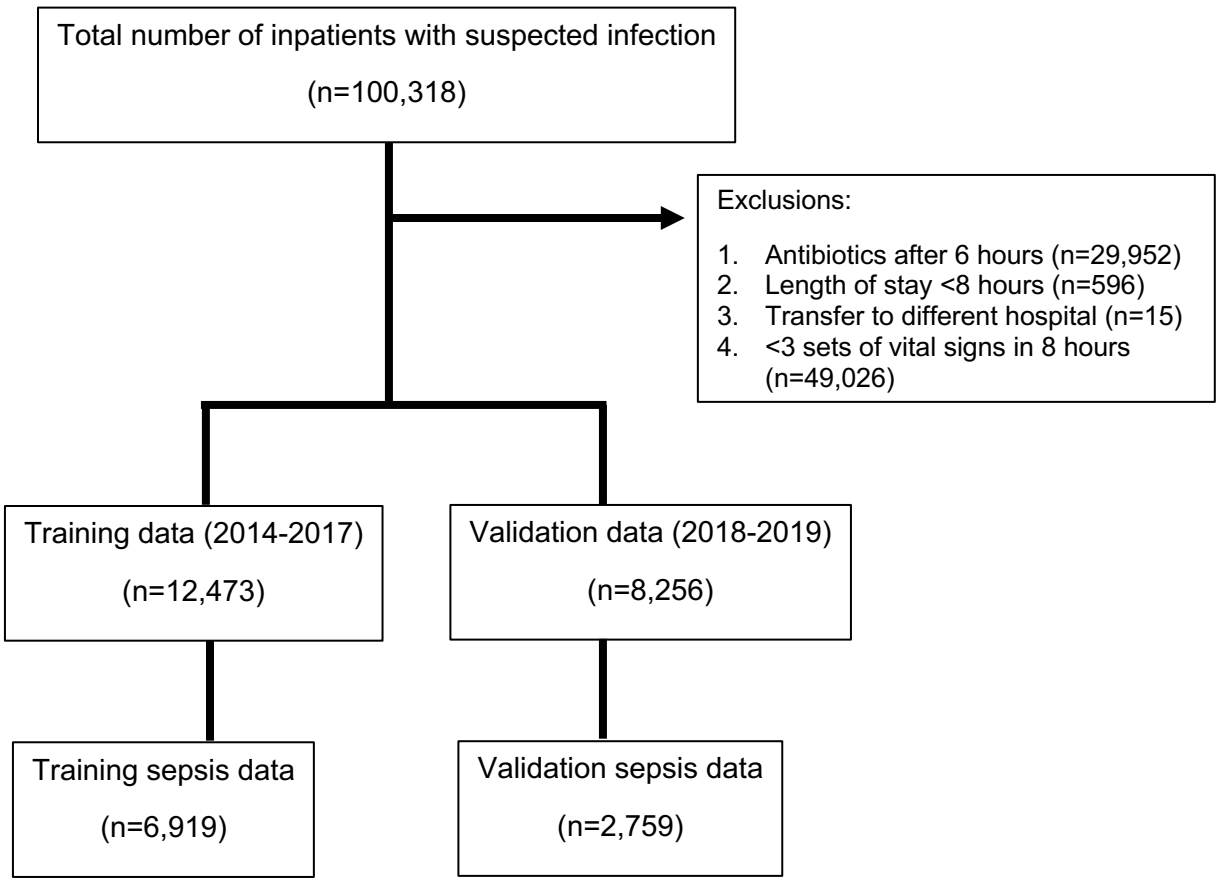
Table of Contents

Supplementary Methods

Simulation of precision enrollment in the SMART trial

A simulation study was performed to test the feasibility of application of the vitals trajectory model in precision enrollment of patients in the Isotonic Solutions and Major Adverse Renal Events Trial (SMART) study. The simulated clinical trial was designed to test the effect of balanced crystalloids versus saline in two subgroups: 1) Group D patients, and 2) all other patients. Group D was chosen as a separate group in this post-hoc analysis given the benefit from balanced crystalloids observed in the heterogeneity of treatment effect analyses. An early stopping strategy was incorporated to stop the trial if there was a significant mortality difference observed in Group D on interim analysis, as saline could cause harm to this subphenotype. We sampled patients from the SMART sepsis cohort in random order, and at each vital sign check, the study patients were assigned a probability of membership to Group D (calculated using mean squared distance from the Group D trajectory converted to a probability distribution with the softmax function). Three varying thresholds of probability of Group D membership were tested that would serve as a trigger for assignment to the Group D cohort (0.5,0.75, and 0.90). If the patient crossed the specified probability threshold prior to requiring intravenous fluids, they were assigned to Group D, and “randomized” to either balanced crystalloids or saline. Across 1000 simulations at each probability threshold, we evaluated the number of patients that we would need to randomize prior to detecting a significant enough benefit from balanced crystalloids in Group D to warrant early stopping of the clinical trial (the early stopping benchmark was set at a mortality difference >10% at a p-value of <0.05 with interim data review after every 100 randomized patients).

Supplementary Figure 1 – Consort diagram.



All adult patients admitted to four hospitals in the Emory Healthcare system, between January 2014 and December 2019, with suspicion for infection, and who received antibiotics within 6 hours of hospitalization were eligible for study inclusion. Patients who died or were discharged in the first 8 hours were excluded. Patients who were transferred to other hospitals were excluded. Patients with fewer than 3 sets of documented vital signs in the first 8 hours of hospitalization were excluded. The final study cohort of patients with suspected infection included 20,729 patients, with 12,473 patients in the training cohort (admitted between 2014 and 2017) and 8,256 patients in the validation cohort (admitted between 2018 and 2019). For the primary analysis of comparing clinical characteristics and outcomes, the cohort was further narrowed to patients meeting Sepsis-3 criteria (operationalized as SOFA≥2 in the first 24 hours), with 6,919 patients with sepsis in the training data and 2,759 patients with sepsis in the validation data.

Supplementary Table 1 – Comparison of vitals trajectory models.

K-cluster model	BIC	Lowest group membership
Training		
2	-371297.39	38%
3	-358499.35	27%
4	-351863.11	13%
5	-346232.07	3.8%
6	-341945.27	3.5%

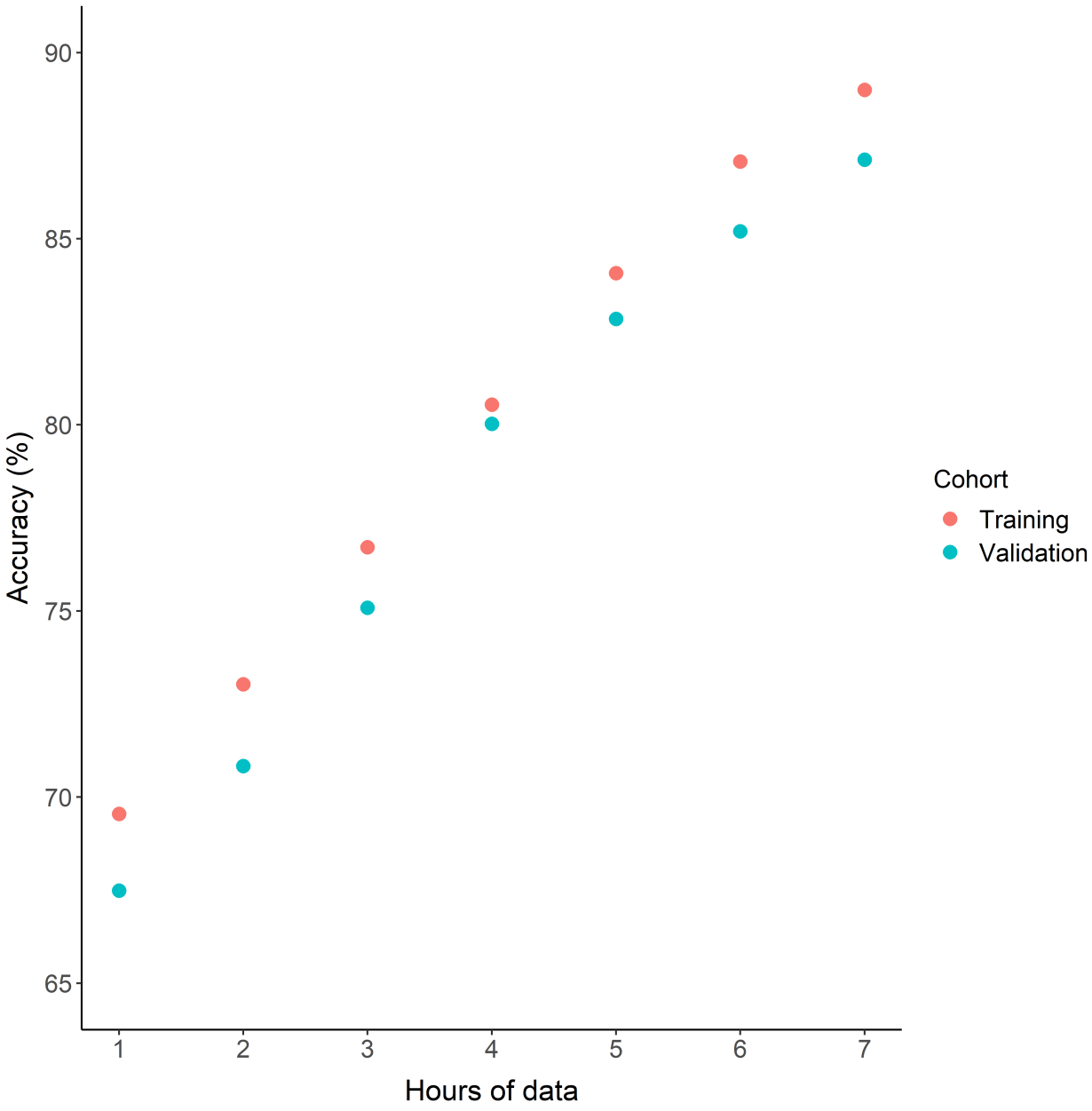
Group based trajectory modeling (GBTM) is a specialized application of finite mixture modeling and is used to identify groups of individuals following similar trajectories of variables over time. Using GBTM, we tested two, three, four, five, and six-group models. The resulting models with varying numbers of subphenotypes were compared using the following criteria: 1) Bayesian Information Criteria (BIC), and 2) Subgroup distribution – If one or more subphenotypes in a model contained less than five percent of the cohort, the model was considered to have less clinical utility and was not eligible for selection. The 4-group trajectory model had the highest BIC of the models with adequate subgroup distribution (the 5-group and 6-group models resulted in group memberships of less than 5% of the cohort).

Supplementary Table 2 - Importance of individual vital signs in the subphenotype model.

Vital sign	Agreement with full model (%)
Respiratory rate	35
Temperature	38
Heart rate	49
Diastolic blood pressure	48
Systolic blood pressure	52

The importance of each vital sign was assessed by the agreement in classification between the one-variable model and the full model (i.e., with all variables), with higher agreement suggesting higher importance of the one-variable. Respiratory rate alone had 35% agreement, temperature 38%, heart rate 49%, diastolic blood pressure 48% and systolic blood pressure 52%.

Supplemental Figure 2 – Accuracy of shorter period of vitals measurement compared to 8-hour gold standard model



Presented is the accuracy of subphenotype classification with varying hours of vitals data (1 to 7 hours) compared to the full 8-hour vitals trajectory model. There was reasonable agreement with vital signs from the first hour, but the agreement significantly improved with each additional hour of vitals data. In the training cohort, one hour of vital signs resulted in 69.6% agreement with the full 8-hour model; four hours resulted in 80.5% agreement; seven hours resulted in 89.0%.

Supplementary Table 3 - Comparison of clinical characteristics between subphenotypes in all patients with suspected infection in the training and validation cohorts.

	Training cohort (2014-2017)					**Validation cohort (2018-2019)**				
Characteristics	A	B	C	D	P-value	A	B	C	D	P-value
N	3483	1578	4044	3368		2231	1073	2698	2254	
Age, years	54	57	67	67	<0.001	55	56	66	65	<0.001
Sex, male	48	54	54	50	<0.001	49	58	57	50	<0.001
Race					<0.001					<0.001
Black	45	61	32	27		44	58	34	31	
White	47	34	61	66		46	37	58	61	
Other	7.6	5.1	6.9	7.1		10	5.1	8	8.2	
Hispanic	5	2.9	3.4	3.3	0.02	5.5	2.8	3.9	5	<0.001
Comorbidities										
CHF	20	30	27	29	<0.001	19	27	25	29	<0.001
Pulmonary	27	26	27	28	0.7	23	22	22	22	0.9
Hypertension	54	87	76	62	<0.001	53	85	75	60	<0.001
DM	28	41	38	31	<0.001	28	44	38	31	<0.001
CKD	23	48	35	34	<0.001	23	42	35	32	<0.001
Liver	13	11	11	15	<0.001	14	13	13	15	0.09
Cancer	9.8	4.2	5	7.5	<0.001	13	6.2	6	9.5	<0.001
Outcomes										
ICU	27	22	15	25	<0.001	28	23	15	25	<0.001
Dialysis	6.6	23	8.9	8.4	<0.001	6.2	18	10	8.3	<0.001
Ventilator	9.6	8.6	8.5	8	0.1	8.9	9.2	7.6	8.4	0.3
Vasopressors	13	9.7	11	16	<0.001	16	11	12	17	<0.001
Inotropes	2.3	0.9	2.8	3.7	<0.001	2.5	1.1	1.6	4.3	<0.001
Mortality	2.5	1.6	1.7	2.8	0.003	2.6	2.1	1.7	3	0.02

Definition of abbreviations: *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*.

Presented are the comparison of demographics, comorbidities, and outcomes between subphenotypes 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 subphenotypes through chi-squared or ANOVA testing, as appropriate.

Supplementary Table 4 - Comparison of clinical characteristics between subphenotypes in patients with suspected infection and SIRS criteria.

Characteristics	Group A	Group B	Group C	Group D	P-value
N	3153 (37.9)	1239 (14.9)	1862 (22.4)	2067 (24.8)	
Age, years	54 (38-68)	56 (44-68)	66 (53-78)	66 (51-77)	<0.001
Sex, male	1512 (48)	683 (55.1)	995 (53.4)	1035 (50.1)	<0.001
Race					<0.001
Black	1455 (46.1)	768 (62)	698 (37.5)	634 (30.7)	
White	1451 (46)	412 (33.3)	1033 (55.5)	1292 (62.5)	
Other	247 (7.8)	59 (4.8)	131 (7)	141 (6.8)	
Ethnicity, Hispanic	162 (5.1)	35 (2.8)	61 (3.3)	66 (3.2)	0.01
Comorbidities					
CHF	630 (20)	383 (30.9)	523 (28.1)	566 (27.4)	<0.001
Pulmonary	853 (27.1)	339 (27.4)	530 (28.5)	581 (28.1)	0.7
Hypertension	1684 (53.4)	1075 (86.8)	1392 (74.8)	1262 (61.1)	<0.001
Diabetes mellitus	863 (27.4)	509 (41.1)	693 (37.2)	628 (30.4)	<0.001
Renal disease	734 (23.3)	576 (46.5)	638 (34.3)	693 (33.5)	<0.001
Liver disease	391 (12.4)	140 (11.3)	196 (10.5)	322 (15.6)	<0.001
Metastatic cancer	306 (9.7)	57 (4.6)	115 (6.2)	173 (8.4)	<0.001
Outcomes					
ICU	897 (28.4)	303 (24.5)	326 (17.5)	613 (29.7)	<0.001
Dialysis	215 (6.8)	290 (23.4)	167 (9)	175 (8.5)	<0.001
Ventilator	317 (10.1)	114 (9.2)	169 (9.1)	178 (8.6)	0.3
Vasopressors	428 (13.6)	120 (9.7)	199 (10.7)	351 (17)	<0.001
Inotropes	71 (2.3)	10 (0.8)	31 (1.7)	71 (3.4)	<0.001
Mortality	85 (2.7)	22 (1.8)	36 (1.9)	63 (3)	0.04

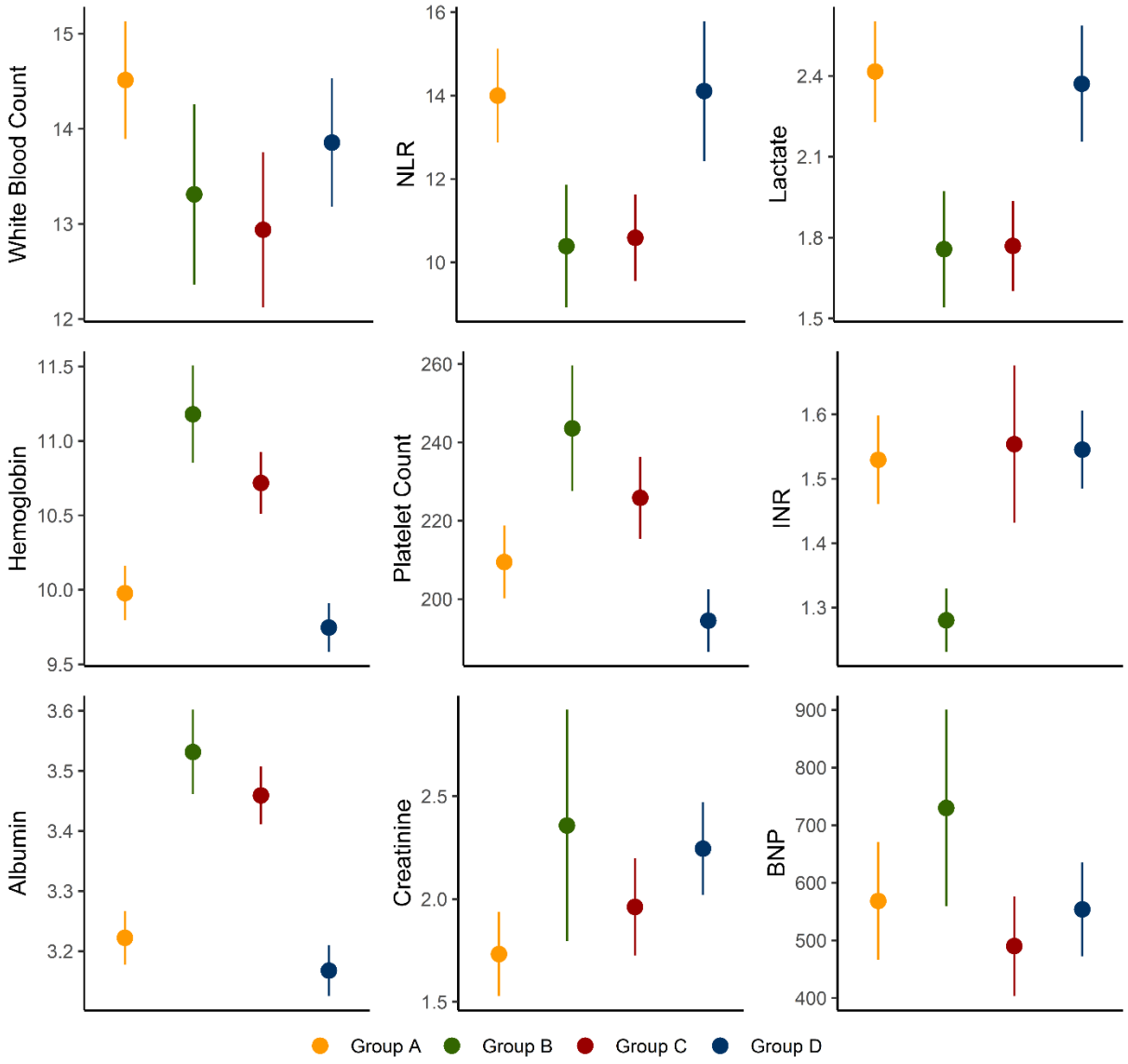
Definition of abbreviations: *CHF – congestive heart failure; ICU – intensive care unit*.
Age is presented as median (IQR) and all other values are presented as n (%). Inotropes defined as dobutamine and milrinone. Mortality represents 30-day hospital mortality. P-values signify the results of comparisons between subphenotypes through chi-squared or ANOVA testing, as appropriate.

Supplementary Table 5 - Comparison of clinical characteristics between subphenotypes in patients with suspected infection and fewer than 3 sets of vital signs.

Characteristics	Group A	Group B	Group C	Group D	P-value
N	5187 (26.0)	2779 (13.9)	6566 (32.9)	5443 (27.2)	
Age, years	58 (41-70)	58 (46-70)	67 (54-78)	66 (53-78)	<0.001
Sex, male	2545 (49.1)	1431 (51.5)	3414 (52)	2636 (48.4)	<0.001
Race					<0.001
Black	2228 (43)	1634 (58.8)	2383 (36.3)	1687 (31)	
White	2589 (49.9)	1014 (36.5)	3810 (58)	3413 (62.7)	
Other	370 (7.1)	131 (4.7)	373 (5.7)	343 (6.3)	
Hispanic Ethnicity	172 (3.3)	66 (2.4)	173 (2.6)	141 (2.6)	0.02
Comorbidities					
CHF	1331 (25.7)	884 (31.8)	1939 (29.5)	1965 (36.1)	<0.001
Pulmonary disease	1520 (29.3)	760 (27.3)	1836 (28)	1670 (30.7)	0.002
Hypertension	2916 (56.2)	2405 (86.5)	5049 (76.9)	3553 (65.3)	<0.001
Diabetes Mellitus	1456 (28.1)	1129 (40.6)	2509 (38.2)	1771 (32.5)	<0.001
Renal Disease	1338 (25.8)	1225 (44.1)	2392 (36.4)	2015 (37)	<0.001
Liver disease	754 (14.5)	333 (12)	868 (13.2)	970 (17.8)	<0.001
Metastatic cancer	567 (10.9)	138 (5)	324 (4.9)	398 (7.3)	<0.001
Outcomes					
ICU	1822 (35.1)	697 (25.1)	1365 (20.8)	1767 (32.5)	<0.001
Dialysis	360 (6.9)	470 (16.9)	585 (8.9)	569 (10.5)	<0.001
Ventilator	961 (18.5)	380 (13.7)	891 (13.6)	920 (16.9)	<0.001
Vasopressors	1080 (20.8)	370 (13.3)	1094 (16.7)	1353 (24.9)	<0.001
Inotropes	197 (3.8)	45 (1.6)	231 (3.5)	349 (6.4)	<0.001
Mortality	270 (5.2)	62 (2.2)	178 (2.7)	307 (5.6)	<0.001

Definition of abbreviations: *CHF – congestive heart failure; ICU – intensive care unit*.
Age is presented as median (IQR) and all other values are presented as n (%). Inotropes defined as dobutamine and milrinone. Mortality represents 30-day hospital mortality. P-values signify the results of comparisons between subphenotypes through chi-squared or ANOVA testing, as appropriate.

Supplementary Figure 3 - Laboratory values compared between subphenotypes in the validation cohort.



Laboratory values (most abnormal values in the first 24 hours of hospitalization) were compared between the vitals trajectory subphenotypes using ANOVA testing. All labs presented were significantly different between the 4 subphenotypes when corrected for multiple testing in either the training or validation cohort. Group A had the highest neutrophil-to-lymphocyte ratio and lactic acid levels ($p<0.001$). Group B had the highest creatinine levels ($p=0.002$). Group D had the lowest albumin and hemoglobin level ($p<0.001$).

Definition of abbreviations: NLR – neutrophil to lymphocyte ratio; INR – international normalized ratio; BNP – Brain natriuretic peptide.

Supplementary Table 6 - Laboratory values compared between subphenotypes in patients with sepsis in the training cohort.

Lab	Group A	Group B	Group C	Group D	P-value	Missing
White Blood Count	13.5 (9.8)	12 (12.2)	11.1 (10.3)	12.3 (8.1)	<0.001	17.8
Absolute Neutrophils	10.4 (6.3)	8.8 (5.3)	8 (5.1)	9.5 (6.2)	<0.001	38.7
Absolute Lymphocyte	1.2 (0.9)	1.2 (0.9)	1.4 (3.4)	1.2 (1)	0.01	38.7
NLR	12.6 (11.8)	9.6 (8.1)	8.6 (10)	12 (13.1)	<0.001	38.7
C-Reactive Protein	128 (110)	116 (121)	88 (90)	101 (91)	0.02	94.3
ESR	54.2 (35.1)	71 (31.1)	59.4 (34.8)	56.8 (38.1)	0.09	94.7
Hemoglobin	10 (2.5)	10.5 (2.4)	10.6 (2.2)	9.9 (2.2)	<0.001	17.8
Platelet Count	198 (121)	201 (1043)	198 (116)	189 (113)	0.04	17.8
INR	1.6 (1.1)	1.4 (0.9)	1.5 (1)	1.7 (1.3)	<0.001	29.9
Creatinine	1.9 (2.7)	3.9 (4.3)	2.2 (2.5)	2 (2.2)	<0.001	47.5
Total bilirubin	1.6 (2.5)	1.2 (2)	1.6 (3.8)	1.8 (4.1)	0.005	16.1
Albumin	3.1 (0.7)	3.2 (0.7)	3.3 (0.6)	3.1 (0.7)	<0.001	4.7
B-Natriuretic Peptide	479 (744)	1042 (1525)	631 (798)	689 (1058)	<0.001	76.3
Lactic Acid	2.3 (1.8)	1.9 (1.9)	1.7 (1.1)	2 (1.4)	<0.001	37.6

Definition of abbreviations: *NLR – neutrophil to lymphocyte ratio; ESR – erythrocyte sedimentation rate; INR – international normalized ratio*. All values presented mean (standard deviation) and missingness is presented as percentage of patients in the training cohort with the laboratory marker missing in the first 24 hours of hospitalization. P-values signify the results of comparisons between subphenotypes through ANOVA testing. The level of significance after correction for multiple testing was $p<0.004$.

Supplementary Table 7 - Laboratory values compared between subphenotypes in patients with sepsis in the validation cohort.

Lab	Group A	Group B	Group C	Group D	P-value	Missing
White Blood Count	14.5 (8.5)	13.3 (7.4)	12.9 (9)	13.9 (9.4)	0.01	17.7
Absolute Neutrophils	11 (6.6)	9.8 (5.9)	9.5 (4.9)	10 (6)	<0.001	32.0
Absolute Lymphocyte	1.2 (0.9)	1.4 (1.2)	1.4 (1.5)	1.2 (1.1)	0.06	32.0
NLR	13.9 (14)	10.4 (10)	10.5 (11)	14.2 (21)	<0.001	32.0
C-Reactive Protein	177 (123)	151 (128)	104 (86)	143 (101)	0.002	92.0
ESR	65.5 (34.9)	63.3 (36.1)	57.3 (38.4)	68.8 (40.7)	0.5	93.8
Hemoglobin	10 (2.5)	11.2 (2.6)	10.7 (2.3)	9.7 (2.3)	<0.001	18.4
Platelet Count	210 (128)	244 (126)	226 (117)	195 (111)	<0.001	18.1
INR	1.5 (0.9)	1.3 (0.4)	1.6 (1.2)	1.5 (0.8)	0.001	16.0
Creatinine	1.7 (2.3)	2.4 (3.7)	2.0 (2.3)	2.2 (2.6)	0.005	45.2
Total bilirubin	1.4 (2.5)	1.1 (3.9)	1.1 (2.2)	1.6 (3.3)	0.1	37.4
Albumin	3.2 (0.7)	3.5 (0.6)	3.5 (0.6)	3.2 (0.7)	<0.001	1.6
B-Natriuretic Peptide	566 (919)	722 (1034)	488 (703)	550 (783)	0.07	64.0
Lactic Acid	2.4 (2.1)	1.8 (1.3)	1.8 (1.2)	2.4 (2.2)	<0.001	44.9

Definition of abbreviations: *NLR – neutrophil to lymphocyte ratio; ESR – erythrocyte sedimentation rate; INR – international normalized ratio*. All values presented mean (standard deviation) and missingness is presented as percentage of patients in the validation cohort with the laboratory marker missing in the first 24 hours of hospitalization. P-values signify the results of comparisons between subphenotypes through ANOVA testing. The level of significance after correction for multiple testing was $p<0.004$.

Supplementary Table 8 - Effect of additional laboratory variables to subphenotype model classification

Laboratory value	Agreement with vitals model (%)
WBC	95.0
Anion gap	93.2
Bicarbonate	91.2
BUN	94.7
Creatinine	94.3
Total Bilirubin	90.5

Group-based trajectory modeling allows for modeling a maximum of 6 variables. Given the 5-variable vital signs model, the algorithm allows for the testing of 1 additional laboratory marker at a time. Presented are the results of fitting a new model by adding one variable at a time to the vital signs model: white blood cell count (WBC), anion gap, bicarbonate, blood urea nitrogen (BUN), creatinine, and total bilirubin. The agreement of the vitals and lab model was compared to the original vitals only model, with lower agreement suggesting the laboratory marker significantly changed the model. All laboratory markers resulted in >90% agreement with the original model, suggesting that the addition of labs did not significantly alter the subphenotype model.

Supplementary Table 9 - Inclusion and exclusion criteria for training, validation, sepsis, and RCT cohorts

	Training	Validation	Training sepsis	Validation sepsis	SMART RCT
Years	2014-2017	2018-2019	2014-2017	2018-2019	2015-2017
Setting	Emory	Emory	Emory	Emory	VUMC
Hospital location	Any	Any	Any	Any	ICU
Inclusion	1) Body fluid culture 2) Antibiotics in 6 hours	1) Body fluid culture 2) Antibiotics in 6 hours	1) Body fluid culture 2) Antibiotics in 6 hours 3) SOFA≥2 in 24 hours	1) Body fluid culture 2) Antibiotics in 6 hours 3) SOFA≥2 in 24 hours	1) ICD-10 diagnosis of sepsis
Exclusion	1) <3 sets of vital signs in 8 hours 2) Transfer to other hospital 3) Death/ discharge in 8 hours	1) <3 sets of vital signs in 8 hours 2) Transfer to other hospital 3) Death/ discharge in 8 hours	1) <3 sets of vital signs in 8 hours 2) Transfer to other hospital 3) Death/ discharge in 8 hours	1) <3 sets of vital signs in 8 hours 2) Transfer to outside hospital 3) Death/ discharge in 8 hours	1) No vital signs prior to fluids 2) Transfers from other hospitals 3) Enrollment >72 hours after presentation

Definition of abbreviations: *VUMC – Vanderbilt University Medical Center; ICU – intensive care unit; SMART - Isotonic Solutions and Major Adverse Renal Events Trial; RCT – randomized controlled trial; SOFA – Sequential Organ Failure Assessment.*

Supplementary Table 10 - Subphenotype characteristics in the Isotonic Solutions and Major Adverse Renal Events Trial (SMART).

Characteristics	Group A	Group B	Group C	Group D	P-value
N	319	93	100	322	
Balanced crystalloids	165 (51.7)	39 (41.9)	59 (59)	150 (46.6)	0.06
Age	53 (37.5-66)	55 (45-68)	59.5 (52-69)	62.5 (50-71)	<0.001
Sex, male	173 (54.2)	63 (67.7)	50 (50)	164 (50.9)	0.03
Race					0.2
White	236 (74)	68 (73.1)	75 (75)	259 (80.4)	
Black	69 (21.6)	19 (20.4)	21 (21)	49 (15.2)	
Other	14 (4.4)	6 (6.5)	4 (4)	14 (4.3)	
Source of Admission					0.01
ED	257 (80.6)	70 (75.3)	70 (70)	277 (86)	
Wards	51 (16)	21 (22.6)	27 (27)	36 (11.2)	
OR	7 (2.2)	1 (1.1)	0 (0)	5 (1.6)	
Other area	4 (1.3)	1 (1.1)	3 (3)	4 (1.2)	
Comorbidities					
CHF	58 (18.2)	27 (29)	22 (22)	85 (26.4)	0.04
Pulmonary disease	61 (19.1)	20 (21.5)	25 (25)	90 (28)	0.06
Hypertension	152 (47.6)	61 (65.6)	64 (64)	204 (63.4)	<0.001
Diabetes mellitus	109 (34.2)	45 (48.4)	34 (34)	111 (34.5)	0.07
Renal disease	56 (17.6)	31 (33.3)	23 (23)	112 (34.8)	<0.001
Liver disease	56 (17.6)	17 (18.3)	21 (21)	84 (26.1)	0.06
Metastatic cancer	30 (9.4)	2 (2.2)	10 (10)	39 (12.1)	0.04
30-day mortality	66 (20.7)	13 (14)	22 (22)	90 (28)	0.02

Definition of abbreviations: *CHF – congestive heart failure; ED – emergency department; OR – operating room*. All values presented as n (%) or median (IQR). P-values signify the results of comparisons between subphenotypes through chi-squared or ANOVA testing, as appropriate.

Supplementary Table 11 - Subphenotype mortality rate in balanced crystalloids versus saline treatment arms.

Subphenotype	Treatment	N	30-day mortality
Group A	Balanced crystalloids	165	32 (19.4)
	Normal saline	154	34 (22.1)
Group B	Balanced crystalloids	39	8 (20.5)
	Normal saline	54	5 (9.3)
Group C	Balanced crystalloids	59	12 (20.3)
	Normal saline	41	10 (24.4)
Group D	Balanced crystalloids	150	30 (20)
	Normal saline	172	60 (34.9)

Mortality numbers presented as n (%).

Supplementary Table 12 - Simulation of precision enrollment in the SMART trial

Probability	Early stopping	Mortality difference	Overall enrollment	Group D enrollment
0.50	607	15.0	300 (200-400)	245 (160-333)
0.75	710	15.1	300 (200-400)	239 (154-326)
0.90	697	14.9	300 (200-400)	231 (152-315)

Presented are the results of the 1,000 simulation trials at three varying probability thresholds for Group D membership (0.50, 0.75, and 0.90). The number of trials of the 1000 simulations that resulted in early stopping by reaching a mortality difference of >10% at a p-value of <0.05 are presented. The mean mortality difference observed in Group D patients assigned to balanced crystalloids compared to saline are presented. The median and interquartile range of patients enrolled overall, and Group D patients enrolled prior to early trial stopping are presented. At a probability threshold of 0.5, out of 1000 simulation trials, 607 trials resulted in early stopping of enrollment of Group D patients after reaching the pre-defined mortality benefit from balanced crystalloids, with an average mortality difference of 15% stopping after a median of 300 patients were randomized overall (IQR 200-400), with 245 Group D patients (IQR 160-333). The remaining 393 trials did not meet the early stopping threshold and resulted in the recruitment of all 834 study patients, and the mortality rate in Group D was 8.6% lower with balanced crystalloids vs saline (p=0.008). At a probability threshold of 0.75, 710 trials resulted in early stopping of enrollment of Group D patients, with an average mortality difference of 15.1% stopping after a median of 300 patients were randomized overall (IQR 200-400), with 239 Group D patients (IQR 154-326). At a probability threshold of 0.90, 697 trials resulted in early stopping of enrollment of Group D patients, with an average mortality difference of 14.9% stopping after a median of 300 patients were randomized overall (IQR 200-400), with 231 Group D patients (IQR 152-315).