

If there is Online Only content that cannot be converted to a Word processing format, you may have to click the Supplemental Files icon on the menu bar in your Reviewer Center to access.

Machine-Learning Model for Mortality Prediction in Patients with Community-acquired Pneumonia: Development and validation study

Journal:	CHEST
Manuscript ID	CHEST-22-0131.R3
Article Type:	Original Research
Date Submitted by the Author:	n/a
Complete List of Authors:	cilloniz, catia; Biomedical Research Networking Centres in Respiratory Diseases (Ciberes) Barcelona, Department of Pneumology Mogensen, Mads Lause; Treat Systems ApS Ward, Logan; Treat Systems Pericas, Juan Manuel; Hospital Clínic de Barcelona, Infectious Diseases; Vall d'Hebron Institut de Recerca, Méndez, Raúl; Hospital Universitari i Politecnic La Fe, Pneumology Gabarrús , Albert ; Hospital Clínic of Barcelona Ferrer, Miquel; Department of Pneumology, Institut Clinic de Respiratori - Hospital Clinic of Barcelona,Institut d'Investigacions Biomèdiques August Pi i Sunyer (IDIBAPS), University of Barcelona (UB), ICREA Academia award, Ciber de Enfermedades Respiratorias (Ciberes, CB06/06/0028) GARCIA-VIDAL, CAROLINA; Hospital Clinic de Barcelona, Infectious diseases Menéndez, Rosario; IIS/Hospital Universitario y Politécnico La Fe. CIBER Enfermedades Respiratorias (CIBERES), Pneumology Torres, Antoni; Biomedical Research Networking Centres in Respiratory Diseases (Ciberes) Barcelona, Department of Pneumology; ICREA
Keywords:	PNEUMONIA, artificial intelligence, machine learning, COMMUNITY- ACQUIRED PNEUMONIA
	1

SCHOLARONE[™] Manuscripts

CHEST

2		
3	1	Word counts abstract: 275
4 5	2	Word counts text: 3006
6	3	
7 8	4	Machine-Learning Model for Mortality Prediction in Patients with Community-
9 10 11	5	acquired Pneumonia: Development and validation study
12 13	6	Running title: Machine-Learning for Mortality Prediction in CAP
14 15 16	/ 8	Catia Cilloniz, PhD ^{1*} , Logan Ward, PhD ^{2*} , Mads Lause Mogensen, PhD ^{2*} , Juan M
17 18	9	Pericàs, MD, PhD ^{3,4*} , Raúl Méndez, MD,PhD ⁵ , Albert Gabarrús, MSc ¹ , Miquel Ferrer,
19 20 21	10	MD, PhD ¹ , Carolina Garcia-Vidal, MD, PhD ⁶ , Rosario Menendez, MD, PhD ⁵ , Antoni
22 23	11	Torres, MD, PhD ^{1.6}
24 25 26	12	¹ Department of Pneumology, Hospital Clinic of Barcelona; August Pi i Sunyer
27 28	13	Biomedical Research Institute–IDIBAPS, University of Barcelona; Biomedical Research
29 30 31	14	Networking Centres in Respiratory Diseases (Ciberes) Barcelona, Spain.
32 33	15	² Treat Systems ApS, Aalborg, Denmark
34 35 36	16	³ Department of Infectious Diseases, Hospital Clinic of Barcelona, Barcelona, Spain.
37	17	⁴ Liver Unit, Vall d'Hebron University Hospital, Vall d'Hebron Institute for Research
38 39	18	(VHIR), Barcelona, Spain.
40 41 42	19	⁵ Department of Pneumology, Hospital La Fe de Valencia, Valencia, Spain
42 43 44	20	⁶ ICREA, Barcelona, Spain
45 46	21	*Equal contribution
47	22	
48 49 50	23	Corresponding author: Prof. Antoni Torres (primary) or Dr. Catia Cillóniz (alternative)
50 51 52	24	Department of Pulmonary Medicine, Hospital Clinic of Barcelona
53 54	25	C/ Villarroel 170, 08036 Barcelona, Spain
55 56 57	26	Tel: (+34) 93-227-5779, fax: (+ 34) 93-227-9813
58 59 60	27	Email: <u>atorres@clinic.cat</u> (primary) or <u>catiacilloniz@yahoo.com</u> (alternative)

28 Summary conflict of interest statements:

The authors declare that they have no conflicts of interest with the study. Dr Cillóniz is the recipient of the 2018 SEPAR fellowship and a grant from the *Fondo de Investigación Sanitaria* (PI19/00207). RM has received honoraria for lectures from Pfizer.

Funding information:

This study was supported by *CIBER de Enfermedades Respiratorias* (CIBERES CB06/06/0028), and by 2009 Support to Research Groups of Catalonia 911, IDIBAPS. The founders of the study had no role in the study design, data collection, analysis, or interpretation, writing of the report, or decision to submit for publication. SEPAR integrated respiratory infections program.

40 Notation of prior abstract publication/presentation:

- 41 None

1 2		
3	43	Key words list:
4 5	44	Community-acquired pneumonia
6 7	45	Machine learning
8 9	46	Pneumonia
10 11	47	Artificial intelligence
12 13	48	Mortality prediction
14	49	Abbreviation list:
15	50	ML: machine learning
17 18	51	AI: artificial intelligence
19 20	52	CAP: community-acquired pneumonia
21 22	53	CPN: causal probabilistic network
23 24	54	SeF: SepsisFinder
25	55	SeF-ML: SeF adapted model
20	56	ATS/IDSA: American Thoracic Society/Infectious Diseases Society of America
28 29	57	SOFA: Sequential Organ Failure Assessment
30 31	58	qSOFA: quick Sequential Organ Failure Assessment
32 33	59	ED: emergency department
34 35	60	MAP: median arterial pressure
36 27	61	SBP: systolic blood pressure
38	62	DBP: diastolic blood pressure
39 40	63	CRP: C-reactive protein
41 42	64	ARDS indicates acute respiratory distress syndrome
43 44	65	AUC-ROC: area under the receiver operating characteristic curve
45 46	66	CI: confidence intervals
47	67	Q1: first quartile
48 49	68	Q3: third quartile
50 51	69	FiO ₂ , fraction of inspired oxygen
52 53	70	IDSA/ATS: Infectious Diseases Society of America/American Thoracic Society
54 55	71	PaO ₂ : partial pressure of arterial oxygen
56 57	72	PSI, pneumonia severity index
58	73	Q1: first quartile
60	74	Q3: third quartile

75 Abstract

Background: Artificial intelligence tools and techniques such as machine learning (ML) are increasingly seen as a suitable manner to increase the prediction capacity of currently available clinical tools, including prognostic scores. However, studies evaluating the efficacy of ML methods in enhancing the predictive capacity of existing scores for community-acquired pneumonia (CAP) are limited. We aimed to apply and validate a causal probabilistic network (CPN) model to predict mortality in patients with CAP.

Research question: Is a CPN model able to predict mortality in patients with CAP
 better than the commonly-used severity scores?

Study design and methods: Derivation-validation retrospective study conducted in two Spanish University hospitals. The ability to predict 30-day mortality of a CPN designed to predict mortality in sepsis (SeF) and adapted for CAP (SeF-ML) was assessed and compared to other scoring systems (PSI, SOFA, qSOFA and CURB-65). The SeF models are proprietary software. Differences between ROC curves were assessed using De Long's method for correlated ROC curves.

- **Results:** The derivation cohort comprised 4,531 patients whilst the validation cohort had 1,034 patients. In the derivation cohort, the AUC of SeF-ML, CURB-65, SOFA, PSI and qSOFA were 0.801, 0.759, 0.671, 0.799 and 0.642, respectively, for 30-day mortality prediction. In the validation study, the AUC of SeF-ML was 0.826, concordantly with the AUC (0.801) in the derivation data (p=0.51). The AUC of SeF-ML was significantly higher than those of CURB-65 (0.764, p=0.03) and qSOFA (0.729, p=0.005). However, it did not differ significantly from PSI (0.830, p=0.92) and SOFA (0.771, p=0.14).
- Interpretation: SeF-ML shows potential for improving mortality prediction amongst
 patients with CAP using structured health data. Additional external validation studies
 should be conducted to support generalisability.

Introduction

1

CHEST

Community-acquired pneumonia (CAP) remains a major cause of morbidity and

mortality worldwide, with children and elderly being the most affected population^{1,2}.

Early-risk stratification of severity and adequate antimicrobial therapy are critical to

improving CAP outcomes^{1,3}. Scoring systems such as Pneumonia Severity Index (PSI)⁴

and CURB-65 score⁵ have been widely used to stratify patients with CAP and support

clinical decision-making processes³. However, these scores have several limitations

that hinder their capacity as clinical decision-making supporting tools, e.g., low positive

and negative predictive values in predicting the need of intensive care, or

Artificial intelligence (AI) tools and techniques such as machine learning (ML) are

increasingly seen as a suitable manner to increase the prediction capacity of currently

available tools in infectious diseases, e.g., sepsis, antimicrobial resistance and COVID-

19^{6,7}. Within respiratory medicine, the main applications of AI and ML have included

the interpretation of thoracic imaging, lung pathology slides and physiologic data such

as pulmonary function tests⁸. Nonetheless, studies evaluating the efficacy of ML

methods in enhancing the predictive capacity of existing scores for CAP are limited^{9,10}.

We aimed to apply a causal probabilistic network (CPN) model previously used in

sepsis (SeF) ^{11–13} to predict 30-day mortality in patients with CAP, comparing the

accuracy of this model to that of the established clinical scores. In addition, we

pursued validating the ML model in CAP using a large cohort of patients with CAP.

underestimated severity in certain age groups.

2	
3	104
4	104
5	105
6	106
7	
8	107
9	
10	108
12	100
13	100
14	109
15	110
16	110
17	
18	111
19	
20 21	112
22	
23	113
24	
25	114
26	111
27	115
28	115
29 30	110
31	116
32	
33	117
34	
35	118
36	
3/	119
39	
40	120
41	
42	121
43	121
44 45	122
45 46	122
47	100
48	123
49	
50	124
51	
52	125
55 54	
55	126
56	
57	127
58	= /
59	128
60	120

129 Methods

130 Study design and participants

A derivation-validation retrospective study was performed when using an innovative CPN model^{12,13} to predict mortality in adult patients hospitalised with CAP. The rationale for using this new statistical approach is trying to diminish the effect of correlations and make better use of the variables, thus avoiding losing patients with missing information. We applied the guidelines provided by Leisman et al ¹⁴ for reporting of prediction models.

Institutional approval was provided by the IRBs of both University Hospital La Fe of
Valencia (EC2011/2019) and Hospital Clínic of Barcelona (HCP2009/5451), which
waived the need for informed consent due to the retrospective nature of the study.

140 Data were collected within the first 24 hours of admission. Collected data had similar 141 definitions for both cohorts, and harmonization between cohorts was elementary. All 142 data set were anonymously analyzed, and the study was performed following current 143 recommendation of the Declaration of Helsinki.

Definitions

CAP was diagnosed if new acute respiratory symptoms, signs, and compatible infiltrate(s) on chest x-ray were present. Severe CAP was defined according to the ATS/IDSA guidelines³. Prior antibiotic treatment was defined as the intake of antibiotics during the week before hospital admission. The appropriateness of empiric antibiotic treatment was determined according to multidisciplinary guidelines for the management of CAP¹⁵. Sepsis was defined as the presence of pneumonia and an increase of ≥ 2 points in the Sequential Organ Failure Assessment (SOFA) score¹⁶. We also calculated median arterial pressure (MAP) from systolic blood pressure (SBP) and

CHEST

diastolic blood pressure (DBP) as 1/3*SBP + 2/3*DBP. The ML technique used is based on CPN ^{11,12,17}. TREAT-Lab (Treat Systems ApS, Aalborg, Denmark) is a medical device software program that provides a risk assessment of patients with suspected infection (proprietary software). The aim of such software is to inform the use of additional or adjunct diagnostics. For example, the clinician (or clinical microbiologist) can use the risk score to identify high-risk patients, i.e., classified as those with high predicted probability of mortality, who may benefit from rapid diagnostics. Conversely, it may be used to identify low-risk patients who only receive standard of care. Customisable risk thresholds can be set for individual clinical installations depending on resources available and local practice requirements. The risk assessment model used within TREAT-Lab is the SeF CPN.

164 Patient selection, inclusion and exclusion criteria

165 Derivation cohort

We enrolled all consecutive adult patients with a CAP diagnosis in the Emergency Department of Hospital Clinic of Barcelona between January 2003 and December 2016. We excluded patients with severe immunosuppression due but not limited to human immunodeficiency virus (HIV) infection, active solid or haematologic malignancy treated with chemotherapy, oral corticosteroid treatment with at least 20 mg of prednisone (or equivalent) per day for at least two weeks, and treatment with other immunosuppressive drugs. We also excluded those with active tuberculosis or a confirmed alternative diagnosis.

174 Validation cohort

All consecutive patients admitted with CAP to Hospital Universitario y Politécnico La Fe(Valencia, Spain) between January 2012 and December 2018 were included. The

inclusion criteria were CAP diagnosis based on a new radiologic infiltrate with at least
two compatible clinical symptoms and age ≥18 years. Exclusion criteria were hospital
admission within the previous 15 days, immunosuppressive treatments and HIV
infection.

181 Data collection and evaluation

182 Derivation cohort

Demographic variables, comorbidities, and physiologic parameters were collected at the emergency department within 24 hours of admission. The PSI, CURB-65, qSOFA and SOFA score were calculated at admission^{4,5,18}. We recorded whether patients had specific complications, including multilobar infiltration, pleural effusions, acute respiratory distress syndrome (ARDS)¹⁹, septic shock²⁰ and acute renal failure²¹ during hospitalisation. All surviving patients were visited or contacted by telephone within 30 days of discharge; hospital records and the Catalunya Health Department database were reviewed at the 1-year mark. We included all available patients in this analysis. We also calculated MAP from SBP and DBP as 1/3*SBP + 2/3*DBP. We discretised PO2 and FiO2 as required by the model. Finally, we transformed creatinine, C-reactive protein (CRP), lactate, bilirubin and platelets through the natural logarithm.

194 Validation cohort

195 Demographic characteristics and comorbidities (diabetes mellitus, respiratory, heart, 196 liver, neurological, and renal diseases) were collected at time of admission. The 197 severity of disease at presentation was assessed with the PSI. Antibiotic treatment 198 before CAP diagnosis in the current episode was recorded. We included all available 199 patients in this analysis.

60

CHEST

200	Outcomes
201	Primary outcome: death within 30 days of admission
201	i mary outcome. death within so days of dumission.
202	Statistical analysis
202	Statistical allarysis
202	We see the second
203	We report the number and percentage of patients for categorical variables and the
204	median and interquartile range for continuous variables (non-normal distribution
205	confirmed by the Kolmogorov-Smirnov test). Categorical variables were compared
206	using the χ^2 test. Continuous variables were compared using the nonparametric Mann-
207	Whitney test.
208	Derivation cohort
200	
200	We used an undated version of the nublished SensisEinder (SeE) model (see
209	we used an updated version of the published sepsisrinder (see
210	a selection of the first state of the second difference defines and the ball
210	supplementary material for a description of the modelling techniques used including
211	variable selection and parameterization). The main differences in input variables
212	between SeF and the adapted model (SeF-ML) are shown in Table 1.
213	We adapted the respiratory component of SeF, which comprises acid-base balance
214	(pH, HCO3-), respiratory rate, and oxygen perfusion (measured through PaO2, SaO2)
215	and FiO2 (FiO2 was discretized based on a conversion from oxygen flow rate in L/min.
216	PaO2 was discretized into 8 bins]) We retrained this portion of the model using data
210	
217	included in the study. The remainder of the model was deemed invariant. To evaluate
21/	included in the study. The remainder of the model was deemed invariant. To evaluate
210	
218	the degree of overfitting to the derivation data, we performed a 10-fold cross-
• 1 0	
219	validation, shuffling the data and stratifying to ensure a similar proportion of outcomes
220	in each cross-validation fold. We then retrained the model using the full derivation
221	dataset.
222	No explicit steps to handle missing data were required. CPNs are inherently tolerant of
	,
223	missing information and are able to perform inference with partial evidence:
	200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 220 221 222

combining the model's structure and conditional probability tables with the axioms of probability theory allows the marginal probabilities of all nodes in a CPN to be calculated when only some of the nodes are observed (supplemental material and e-Table 1)^{22,23}. We then used the SeF-ML model to calculate the probability of death within 30 days (Pmort) for all patients with >2 input variables recorded.

We assessed predictive performance of SeF and SeF-ML by using the area under the receiver operating characteristic curve (AUC-ROC). As a general rule, the relation between AUC and diagnostic accuracy is as follows: AUC between 0.90 and 1.00 has outstanding discrimination ability; 0.80 - 0.90, excellent; 0.70 - 0.80, acceptable; 0.60 -0.70, poor, and 0.50 - 0.60 fails to accurately diagnose a certain disease or condition^{24,25}. Kaplan-Meier survival curves were also constructed.

In our preliminary analysis of this data, we noted that SeF-ML performance was not
negatively impacted by the greater degree of missing data amongst patients included
from 2003 to 2006 (e-Table 2).

We compared SeF and SeF-ML performance (Pmort as a predictor of death within 30 days) with other scoring systems (PSI, SOFA, qSOFA and CURB-65). Differences between ROC curves were assessed using De Long's method for correlated ROC curves as implemented in the pROC package of R.

In addition to assessing the model's performance, we computed an example for a
 potential use-case for TREAT Lab, showing patients being stratified into groups of 40%
 low-risk, 40% medium-risk and 20% high-risk.

245 Validation cohort

We used the SeF-ML model (Table 1), where the respiratory components were a 247 adapted through learning from 4,531 patients with CAP at the Hospital Clinic of

CHEST

Barcelona. We used the adapted model to calculate the probability of death within 30days (Pmort) for all patients.

We assessed predictive performance using the AUC. We compared SeF-ML's performance (Pmort as a predictor of death within 30 days) with other scoring systems (i.e., PSI, CURB-65, SOFA and qSOFA). Differences between ROC curves were assessed using DeLong's method for correlated ROC curves as implemented in the pROC package of R. We also compared mortality for a range of risk cut-offs to assess potential operating points for clinical implementation of risk stratification. Model calibration was assessed using the Hosmer-Lemeshow goodness of fit test and by calculating the Brier-score loss.

258 More details about the selection of variables and the ML model are displayed in the 259 supplementary material (e-Figure 1, e-Figure 2 and e-Figure 3).

2		
3 4	261	Results
5	262	Patients' characteristics
6 7	202	
8 9	263	The derivation cohort comprised 4,531 (15% outpatients and 85% inpatients) patients
10 11 12	264	and the validation cohort 1,034 patients. Clinical characteristics of the studied
13 14	265	population are shown in Table 2.
15 16 17	266	Predictive performance of risk assessment
18 19	267	Derivation cohort
20 21 22	268	We used the SeF-ML model to calculate Pmort for all patients with >2 input variables –
23 24	269	4,500/4,531=99.3% of patients. The AUC for 30-day mortality prediction was 0.801 for
25 26 27	270	SeF-ML, being significantly better than CURB-65 (0.759, p<0.001), SOFA (0.671,
27 28 29	271	p<0.001), and qSOFA (0.642, p<0.001) (Table 3). The mean cross-validation
30 31	272	performance, measured by AUC, was 0.800 (range: 0.749-0.832) which did not differ
32 33 34	273	from the AUC for the full dataset. Details of the cross-validation assessment are
35 36	274	included in the supplemental material.
37 38 39	275	PSI was only available for 58% of patients and had an AUC of 0.799; it was not
40 41	276	significantly different from SeF-ML (p=0.58). CURB-65 provided a "fair" prediction of
42 43	277	mortality, while SeF-ML and PSI provided a "good" prediction of 30-day mortality
44 45 46	278	(Figure 1). The calibration of SeF-ML was measured using the Hosmer-Lemeshow
47 48	279	statistic and the Brier-score loss. The Hosmer-Lemeshow statistic was 15.62 (p=0.048)
49 50 51	280	which suggests the model may not be well calibrated. However, the Hosmer-
52 53	281	Lemeshow statistic is known to be very sensitive to sample size. The model appears
54 55 56	282	visually well-calibrated. The Brier-score loss for the model was 0.056.
57 58	283	Survival curves are shown in Figure 2 for patients stratified according to Pmort; the
59	284	low-risk group comprises 40% of patients with the lowest Pmort; the medium-risk

CHEST

2	
3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
11	
12	
13	
14	
15	
15	
16	
17	
18	
10	
20	
20	
21	
22	
23	
23	
24	
25	
26	
27	
20	
28	
29	
30	
31	
27	
32	
33	
34	
35	
20	
30	
37	
38	
39	
40	
40	
41	
42	
43	
11	
44	
45	
46	
47	
18	
40	
49	
50	
51	
52	
52	
53	
54	
55	
56	
57	
58	
59	
60	
00	

group represents the next highest 40%; and the high-risk group accounts for the top
20% according to Pmort. The additional value provided by SeF-ML with respect to
CURB-65 and PSI scores is shown in Figure 3: SeF-ML provides additional discrimination
for mortality, e.g., for patients with PSI ≥4.

289 <u>Validation cohort</u>

For the validation cohort, the AUC for SeF-ML was 0.826. It was not significantly different from performance in the derivation cohort (p=0.51) (Figure 4). According to the Hosmer-Lemeshow statistic, SeF-ML was well-calibrated (HL statistic = 11.93, p=0.15). The Brier-score loss for SeF-ML was 0.036. Notably, only 23.1% patients had complete information to calculate the SOFA score, being not significantly different from SeF-ML (p=0.85). On the contrary, SeF-ML had a significantly higher AUC than both CURB-65 and qSOFA (p=0.03 and p=0.005, respectively) (Table 4).

297 When analysing the imputed SOFA score (assuming missing components were normal), 298 the AUC of SOFA improved from 0.728 to 0.771; the difference from SeF-ML remained 299 non-significant.

300 Supplementary e-Table 3 shows the number of patients in each PSI risk class and 301 associated 30-day mortality. Also, 30-day mortality within quantile-matched risk 302 classes for SeF-ML is also shown. SeF-ML risk groups were defined by choosing cut-offs 303 that resulted in the same number of patients assigned to each risk group like in the 304 corresponding PSI risk class.

2 305 Poten

Potential use of risk assessment

306 Patients can be stratified by their probability of death within 30 days. An odds ratio 307 (OR) for death can also be calculated for high-risk patients vs. others. The OR 308 represents the degree of separation between high-risk and low-risk patients. For this

Page 14 of 81

CHEST

analysis, we considered two strategies for stratification: high-risk patients receive rapid diagnostics and low-risk patients receive minimum standard of care. For example, the top 20% had a probability of death within 30 days of 18.4%, whilst OR for the high-risk group vs. remaining patients was 7.6 (e-Figure 4). One advantage of SeF-ML is the continuous probability output that allows custom thresholds to be set depending on the care environment versus the five potential operating points for CURB-65, for example. For comparison, three potential scenarios are shown below: 1) PSI=5 (N=98) 30-day mortality = 21.9% (PSI), 24% (SeF-ML); OR for high-risk vs. low risk = 11.5 (PSI), 14.3 (SeF-ML); 2) PSI ≥4 (N=428) 30-day mortality = 8.6% (PSI), 8.4% (SeF-ML); OR for high-risk vs. low risk = 9.3 (PSI), 7.7 (SeF-ML); 3) CURB-65 ≥3 (N=153) 30-day mortality = 14.4% (CURB-65), 18.3% (SeF-ML); OR for high-risk vs. low risk = 6.8 (CURB-65), 12.8 (SeF-ML). A fuller picture of the effect of choosing different operating points is shown in Supplementary e-Figure 4 and e-Figure 5. It shows the 30-day mortality as functions of the size of the high-risk group along with operating points for PSI (state) and CURB-65. The smaller the high-risk group, the higher the 30-day mortality.

CHEST

325 Discussion

In our study, we applied a ML approach to develop and validate a 30-day mortality prediction clinical model in patients with CAP. Although not pneumonia-specific, the SeF-ML model shows potential for improving mortality prediction amongst patients with CAP. Remarkably, SeF-ML's performance in the validation set matched that of the training data in the derivation cohort, as did those of PSI and CURB-65, whereas the performance of SOFA and qSOFA scores did not match. SeF-ML not only enhances the mortality prediction ability of currently available tools but optimizes the use and quality of available electronic health records (EHR) data. Hence, although the advantages conferred by SeF-ML need further refining and interpretation, this ML model might be applied in clinical practice in the near future, i.e., patients can be stratified by their probability of death and such stratification might then be used to determine a patient's care trajectory and diagnostic workup.

This study is one of the first to use a large cohort of hospitalised patients with CAP to generate a ML model validated through an external cohort of patients with CAP. A barrier for physicians in using ML is its potential "black box" opacity. However, studies like ours show that results obtained with ML predictions are consistent with other severity scores that we have used so far^{26,27}. In the near future, ML techniques will allow us to analyse a large volume of data that current techniques cannot do, facilitating the possibility amongst investigators to directly collect data from EHR ^{7,26–28}. The potential of SeF-ML to improve the current ability of available clinical scores for CAP primarily relies on the findings suggesting that SeF-ML better predicts 30-day mortality than qSOFA and CURB-65 according to our data, which nonetheless require further clinical validation. Compounding this is also the fact that SeF-ML had a higher

Page 16 of 81

CHEST

AUC than SOFA and PSI, albeit non-significant. Furthermore, the AUC value of SeF-ML was consistent between the predictive and validation model. There is strong evidence supporting the importance of accurately assessing the severity of CAP and stratifying patients based on their mortality risk to improve clinical outcomes^{1,3,29}. Despite that, PSI and CURB-65 remain the most widely used CAP severity scores and recommended by international guidelines^{3,29,30}. The ability of these scores to predict mortality has some limitations³¹. Moreover, although qSOFA score ≥ 2 is strongly associated with mortality in patients with pneumonia, its use in early identification of patients with CAP and mortality risk is hindered^{32,33}. SeF-ML seems to provide better discriminative capacity to discern between high- and low-risk patients, which is key to adapting the intensity of care and resources per foreseen prognosis. However, this still needs further validation to prove its actual clinical validity. In addition, SeF-ML increases efficiency in exploiting available data. The lower requirements set for minimum data and the ability to handle missing data mean that 99% of patients, on average, would have sufficient data for predictions to be established. This would facilitate optimization of EHR use, with fewer investments on data collection and curation when compared to other scores. Also, the continuous output achieved with SeF-ML allows for adaptive fine-tuning of patient classification. Cut-offs for defining risk can be smoothly and accurately adjusted. In particular, the ability of SeF-ML to identify high-risk patients with low CURB65 and PSI scores, as well as low-risk patients with high CURB65 and PSI scores paramount as these CAP-specific clinical scores are the two most widely used. It is worth noting, however, that enhanced mortality prediction through continuous appraisals is not unique to ML but is also found in other models with continuous

CHEST

output, e.g. the eCURB model, which achieved greater predictive accuracy using CURB-65 variables using regression splines³⁴.

The ability of SeF-ML to reclassify risk of patients with CAP might be more beneficial for certain patient subgroups, e.g., those with low CURB-65/PSI yet high SeF and mortality, or those with high CURB-65/PSI yet low SeF and mortality.

A foremost strength of our study is the use of a large set of clinical data that are representative of the population in a defined area of influence. This study feature enabled the integration of surveillance data into direct clinical care of individual patients and can be helpful in making decisions by applying ML models like SeF-ML. When developing and validating the algorithm, we used two large patient slices to ensure that our model can be implemented using real-time patient data.

However, some limitations need to be highlighted. This score validation against mortality is only the first step toward clinical utility. Remarkably, the ability of SeF-ML to predict ICU admission remains unproven. In particular, our study lacked information on post-admission disposition, and therefore we were not able to test SeF-ML against potentially different clinical pathways other than direct admission to the ICU. Furthermore, the use of closed databases does not incorporate new information and therefore does not allow "learning" of ML models. Besides, despite the large sample sizes of both cohorts, some baseline variables and clinical features, including CAP severity and outcomes, differed. However, this does not hamper the training and validation of the CPN model. There are components of the current SeF that were not available in the dataset due to them not generally being measured in the ED for CAP. However, SeF performs well despite not including these variables, which helps to demonstrate its robustness to missing values.

396 Interpretation

SeF-ML performance at predicting 30-day mortality appears to be overall superior than that of existing CAP-specific scores, with the exception of PSI in the validation model. SeF-ML offers some advantages over current scoring methods, eg., calculations easily made on routinely collected data and based upon; structured EHR data (vs. subjective criteria and arterial blood gas sampling needed for PSI for instance); and tunable performance so as to allow risk cut-offs to be tailored to workflow requirements and capacities of the individual institution (compared with fewer states in CURB-65). In addition, SeF-ML performance seems to not be dependent upon data availability, therefore allowing for more effective calculation of risk scores for CAP based upon data sources with limited access to or completeness of certain variables. Our findings need further validation in other cohorts from different settings to assess the actual clinical utility of SeF-ML in predicting CAP prognosis.

CHEST

1		
2 3 4	410	Acknowledgements
5 6 7	411	Guarantor:
7 8 9	412	Dr. Antoni Torres is the guarantor of the article
10 11	413	Author contributions:
12 13	414	Conception and design of the study: CC, LW, ML, MF, JP, RM, RM, AT. Data acquisition:
14 15	415	all authors. Data analysis and interpretation: LW, ML, CC, AG. Manuscript drafting or
16	416	critical revision for important intellectual content: all authors. Final approval of the
17 18	417	submitted version: all authors.
19 20	418	Financial/nonfinancial disclosures:
21 22	419	The authors have reported to Chest the following: This study was supported by CIBER
23 24	420	de Enfermedades Respiratorias (CIBERES CB06/06/0028), and by 2009 Support to
25	421	Research Groups of Catalonia 911, IDIBAPS. The founders of the study had no role in
20	422	the study design, data collection, analysis, or interpretation, writing of the report, or
28 29	423	decision to submit for publication. SEPAR integrated respiratory infections program.
30 31	424	Role of sponsors:
32 33	425	The funders did not have a role in the conduct of the study; in the collection,
34 35	426	management, analysis, or interpretation of data; or in the preparation of the
36	427	manuscript.
37 38	428	
39 40		
41		
42 43		
44		
45		
46 47		
48		
49		
50		
51		
52		

3 4	429	Refe	erences
5 6	430	1.	Torres A, Cilloniz C, Niederman MS, et al. Pneumonia. Nature Reviews Disease
7 8	431		Primers 2021;7(1):1–28.
9	432	2.	GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and
10	433		injuries in 204 countries and territories, 1990-2019: a systematic analysis for the
11 12 13	434		Global Burden of Disease Study 2019. Lancet 2020;396(10258):1204–1222.
14	435	3.	Metlay JP, Waterer GW, Long AC, et al. Diagnosis and Treatment of Adults with
15	436		Community-acquired Pneumonia. An Official Clinical Practice Guideline of the
16	437		American Thoracic Society and Infectious Diseases Society of America. Am J Respir
17	438		Crit Care Med 2019:200(7):e45–e67.
18	150		
19	439	4	Fine ML Auble TF. Yealy DM, et al. A prediction rule to identify low-risk patients
20	440		with community-acquired pneumonia N Engl Med 1997:336(4):243-250
21	440		
23	441	5	Lim WS Lewis S Macfarlane IT Severity prediction rules in community acquired
24	112	5.	nnoumonia: a validation study. Thorax 2000:55(2):210-222
25	442		preumonia. a valuation study. morux 2000,55(5).215–225.
26	443	6	Peiffer-Smadia N. Rawson TM. Ahmad R. et al. Machine learning for clinical
27	111	0.	decision support in infectious diseases: a parrative review of surrent applications
28	444		Clin Migrahial Infant 2020.26(5):504 505
29	445		CIIN MICrobiol Infect 2020;26(5):584–595.
31	116	7	Carcia Vidal C. Saniyan C. Buerta Alcaldo B. Morono Carcía E. Soriano A. Artificial
32	440	7.	Garcia-Viuar C, Sanjuari G, Puerta-Alcaide P, Moreno-Garcia E, Sonano A. Artificiar
33	44 /		intelligence to support clinical decision-making processes. <i>EBioMedicine</i>
34	448		2019;46:27–29.
35	1.10		
36	449	8.	Gonem S, Janssens W, Das N, Topalovic M. Applications of artificial intelligence
37	450		and machine learning in respiratory medicine. <i>Thorax</i> 2020;75(8):695–701.
38 30	4.5.1		
40	451	9.	Kang SY, Cha WC, Yoo J, et al. Predicting 30-day mortality of patients with
41	452		pneumonia in an emergency department setting using machine-learning models.
42	453		Clin Exp Emerg Med 2020;7(3):197–205.
43			
44	454	10.	Chumbita M, Cillóniz C, Puerta-Alcalde P, et al. Can Artificial Intelligence Improve
45	455		the Management of Pneumonia. <i>J Clin Med</i> 2020;9(1):248.
46			
4/ 10	456	11.	Ward L, Andreassen S, Astrup JJ, Rahmani Z, Fantini M, Sambri V. Clinical- vs.
40 10	457		model-based selection of patients suspected of sepsis for direct-from-blood rapid
42 50	458		diagnostics in the emergency department: a retrospective study. Eur J Clin
51	459		Microhiol Infect Dis 2019:38(8):1515–1522
52	.07		
53	460	12.	Ward L. Møller JK. Eliakim-Raz N. Andreassen S. Prediction of Bacteraemia and of
54	461		30-day Mortality Among Patients with Suspected Infection using a CPN Model of
55	-01 /67		Systemic Inflammation JEAC_DanarsOnLine 2019.116 121
56	402		System t manimation. nAC-rupersonline 2010,110-121.
5/			
50 50			
60			

CHEST

2			
3	463	13.	Andreassen S, Møller JK, Eliakim-Raz N, Lisby G, Ward L. A comparison of
4	464		predictors for mortality and bacteraemia in patients suspected of infection. BMC
5	465		Infect Dis 2021:21(1):864
0 7	100		
, 8	466	14.	Leisman DF. Harhay MO. Lederer DJ. et al. Development and Reporting of
9	467		Prediction Models: Guidance for Authors From Editors of Respiratory Sleen and
10	468		Critical Care Journals Crit Care Med 2020:48(5):623–633
11	400		
12	469	15	Torres A Barberán I Falguera M et al [Multidisciplinary guidelines for the
13	470	10.	management of community-acquired nneumonia] Med Clin (Barc)
14 15	470		
15	4/1		2013,140(3).223.01-223.013.
17	472	16	Singer M. Deutschman CS. Seymour CW. et al. The Third International Consensus
18	472	10.	Definitions for Sensis and Sentic Shock (Sensis 3) //M/A 2016:315(8):801–810
19	475		
20	474	17	Logan Morgan Ward, Gradation of the Severity of Sensis: Learning in a Causal
21	475	17.	Probabilistic Network 2016:55
22	475		Probabilistic Network. 2010,55.
23	476	18	Vincent II Moreno R. Takala L et al. The SOEA (Sensis-related Organ Eailure
25	470	10.	Assossment) score to describe organ dysfunction/failure. On behalf of the
26	4//		Assessment) score to describe organ dystunction/failure. On benan of the
27	4/8		Construction of Sepsis-Related Problems of the European Society of Intensive
28	4/9		Care Medicine. Intensive Care Med 1996;22(7):707–710.
29	100	10	ADDS Definition Task Force, Denieri VIA Dubenfeld CD, et al. Aquite respiratory
30	400	19.	ARDS Deminition Task Force, Ramen VIVI, Ruberneid GD, et al. Acute respiratory
32	481		distress syndrome: the Berlin Definition. JAIVIA 2012;307(23):2526–2533.
33	107	20	Bhodos A. Evans J.E. Albazzani W. et al. Surviving Sonsis Compaign: International
34	402	20.	Cuidelines for Management of Sensis and Sentia Shaely 2016. International
35	483		Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Nied
36	484		2017;43(3):304-377.
3/	105	21	Pollomo P. Bonco C. Kollum IA. Mohto Pl. Dolovaly, P. Aquito Dialysis Quality
30	483	21.	Bellomo R, Ronco C, Relium JA, Menta RL, Palevsky P, Acute Dialysis Quality
40	486		Initiative workgroup. Acute renal failure - definition, outcome measures, animal
41	48/		models, fluid therapy and information technology needs: the Second
42	488		International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI)
43	489		Group. <i>Crit Care</i> 2004;8(4):R204-212.
44 45	100		
45 46	490	22.	Steen Andreassen, Leonard Leibovici, Mical Paul, Anders D. Nielsen, Alina
47	491		Zalounina, Leif E. Kristensen, Karsten Falborg, Brian Kristensen, Uwe Frank, Henrik
48	492		C. Schønheyder. A Probabilistic Network for Fusion of Data and Knowledge in
49	493		Clinical Microbiology [Internet]. springerprofessional.de. 2005 [cited 2022 Mar
50	494		22];Available from: https://www.springerprofessional.de/en/a-probabilistic-
51	495		network-for-fusion-of-data-and-knowledge-in-clin/1020738
52 53			
53 54	496	23.	Lauritzen SL, Spiegelhalter DJ. Local Computations with Probabilities on Graphical
55	497		Structures and Their Application to Expert Systems. Journal of the Royal Statistical
56	498		Society: Series B (Methodological) 1988;50(2):157–194.
57			
58	499	24.	Cole TJ. Applied logistic regression. D. W. Hosmer and S. Lemeshow, Wiley, New
59 60	500		York, 1989. Statistics in Medicine 1989;10(7):1162–1163.
00			

3 4 5	501 502	25.	Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. <i>BMJ</i> 1994;309(6948):188.
6 7 8	503 504	26.	Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. Lancet 2019;393(10181):1577–1579.
9 10 11 12	505 506	27.	Mekov E, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. <i>Expert Rev Respir Med</i> 2020;14(6):559–564.
13 14 15	507 508	28.	He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. <i>Nat Med</i> 2019;25(1):30–36.
16 17 18 19 20	509 510 511	29.	Menéndez R, Cilloniz C, España PP, et al. Community-Acquired Pneumonia. Spanish Society of Pulmonology and Thoracic Surgery (SEPAR) Guidelines. 2020 Update. <i>Arch Bronconeumol</i> 2020;56 Suppl 1:1–10.
21 22 23 24 25	512 513 514	30.	Lim WS, Baudouin SV, George RC, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. <i>Thorax</i> 2009;64 Suppl 3:iii1-55.
26 27 28 29	515 516 517	31.	Loke YK, Kwok CS, Niruban A, Myint PK. Value of severity scales in predicting mortality from community-acquired pneumonia: systematic review and meta-analysis. <i>Thorax</i> 2010;65(10):884–890.
30 31 32 33 34	518 519 520	32.	Jiang J, Yang J, Jin Y, Cao J, Lu Y. Role of qSOFA in predicting mortality of pneumonia: A systematic review and meta-analysis. <i>Medicine (Baltimore)</i> 2018;97(40):e12634.
35 36 37 38 39	521 522 523	33.	Ranzani OT, Prina E, Menéndez R, et al. New Sepsis Definition (Sepsis-3) and Community-acquired Pneumonia Mortality. A Validation and Clinical Decision-Making Study. <i>Am J Respir Crit Care Med</i> 2017;196(10):1287–1297.
40 41 42	524 525	34.	Jones BE, Jones J, Bewick T, et al. CURB-65 pneumonia severity assessment adapted for electronic decision support. <i>Chest</i> 2011;140(1):156–163.
43 44 45	526		
46 47 48 49 50 51 52 53 54	527		
55 56 57			

Variable	SeF	SeF-ML	PSI	SOFA	qSOFA	CURB-6
Demography						
Age	Yes	Yes	Yes			Yes
Sex			Yes			
Nursing home resident			Yes			
Comorbidities						
Neoplastic disease			Yes			
Liver disease			Yes			
Congestive heart failure	2		Yes			
Cerebrovascular diseas	e		Yes			
Renal disease			Yes			
Vitals						
Temperature	Yes	Yes	Yes			
Chills	Yes	Yes				
Systolic Blood Pressure			Yes		Yes	Yes
Diastolic Blood Pressure	9					Yes
Mean Arterial Pressure	Yes	Yes		Yes		
Heart Rate	Yes	Yes	Yes			
Respiratory Rate		Yes	Yes		Yes	Yes
Mental Status	Yes	Yes	Yes	Yes, GCS	Yes	Yes
SpO2		Yes				
Labs						
CRP	Yes	Yes				
Creatinine	Yes	Yes		Yes		
Albumin	Yes*	Yes*				
Lactate	Yes*	Yes*				
Platelets	Yes	Yes		Yes		
Neutrophils%	Yes	Yes				
Bilirubin		Yes		Yes		
BUN or Urea		Yes	Yes			Yes
Sodium			Yes			
Glucose			Yes			
рН		Yes	Yes			

	Variable	SeF	SeF-ML	PSI	SOFA	qSOFA	CURB-6
	Hematocrit		Yes	Yes			
	PaO2		Yes	Yes	Yes		
	FiO2		Yes		Yes		
	Bicarbonate		Yes				
	Leukocytes		Yes				
	Radiology						
	Chest x-ray			Yes			
	Treatment/devices						
	Mechanical ventilation				Yes		
	Pressors				Yes		
529	Abbreviations: Glasgow co	oma scale (GCS); * not p	art of st	andard tes	ting for CAP	patients
530	at the emergency denar	tment (FD))			0	
521	at the energency deput						
551							
532							
522							
533							

	Derivation	Validation	
	cohort	cohort	
Variable	(N = 4,531)	(N = 1,034)	P-value
Age, median (Q1; Q3), years	73 (55; 82)	72 (59; 81)	0.650
Male sex, n (%)	2,708 (60)	642 (62)	0.176
Previous antibiotic, n (%)	1,057 (25)	338 (33)	<0.001
Influenza vaccine, n (%)	1,770 (44)	438 (42)	0.350
Pneumococcal vaccine, n (%)	695 (17)	79 (8)	<0.001
Previous inhaled corticosteroids, n (%)	818 (18)	-	-
Previous systemic corticosteroids, n (%)	180 (4)	-	-
Previous episode of pneumonia (last year),			
n (%)	630 (14)	-	-
Comorbidities, n (%) ^a	2,982 (66)	731 (71)	0.005
Chronic respiratory disease	1,729 (39)	325 (31)	<0.001
Chronic cardiovascular disease	635 (14)	347 (34)	<0.001
Diabetes mellitus	914 (20)	288 (28)	<0.001
Neurologic disease	836 (19)	176 (17)	0.137
Chronic renal disease	318 (7)	139 (13)	<0.001
Chronic liver disease	212 (5)	38 (4)	0.146
Nursing home, n (%)	352 (8)	43 (4)	<0.001
Confusion, n (%)	820 (18)	77 (7)	<0.001
Respiratory rate, median (Q1; Q3),			
breaths/min	24 (20; 30)	18 (16; 24)	<0.001
Heart rate, median (Q1; Q3), beats/min	97 (84; 110)	97 (85; 110)	0.685
Systolic blood pressure, median (Q1; Q3),			
mmHg	129 (112; 148)	134 (118; 152)	<0.001
Diastolic blood pressure, median (Q1; Q3),			
mmHg	72 (64; 80)	70 (61; 80)	0.007
Creatinine, median (Q1; Q3), mg/dL	1.0 (0.8; 1.4)	1.0 (0.8; 1.4)	0.003
Glucose level, median (Q1; Q3), mg/dL	124 (105; 157)	-	-

	Derivation	Validation	
	cohort	cohort	
Variable	(N = 4,531)	(N = 1,034)	P-value
PaO ₂ /FiO ₂ , median (Q1; Q3)	281 (238; 327)	271 (238; 311)	0.012
PSI score, median (Q1; Q3)	98 (74; 123)	86 (66; 105)	<0.001
Severe CAP, n (%)	868 (26)	144 (14)	<0.001
Bacteraemia, n (%) ^b	390 (12)	54 (8)	0.001
Appropriate empiric treatment, n (%)	2,844 (96)	-	-
Length of hospital stay, median (Q1; Q3)),		
days	7 (4; 11)	6 (5; 9)	0.261
Mechanical ventilation, n (%)	375 (10)	26 (3)	<0.001
Non-invasive	169 (4)	-	-
Invasive	206 (5) ^c	-	-
In-hospital mortality, n (%)	272 (6)	35 (3)	<0.001
30-day mortality, n (%)	293 (7)	43 (4)	0.004

Abbreviations: CAP indicates community-acquired pneumonia; PSI, pneumonia severity index; Q1, first quartile; Q3, third quartile. Percentages calculated on non-missing data. ^a May have >1 comorbid condition. ^b Calculated only for patients with blood samples (3,206 in the derivation cohort and 696 in the validation cohort). ^c Patients who initially received non-invasive ventilation yet needed intubation subsequently were included in the invasive mechanical ventilation group.

Table 3. Discriminatory ability of SeF-ML vs. other scoring systems for 30-day

mortality in the derivation cohort

			P-value (vs.
Model/Score	Availability	AUC (95% CI)	SeF-ML)
SeF-ML	99.3%	0.801 (0.777-0.825)	-
(limited to those where PSI			-
was available)		0.808 (0.776-0.839)	
(limited to those where			-
SOFA was available)		0.804 (0.780-0.829)	
(limited to those where			-
CURB-65 was available)		0.814 (0.788-0.839)	
PSI (raw score)	57.7%	0.799 (0.768-0.830)	0.58
SOFA	91.4%	0.671 (0.638-0.704)	<0.001
qSOFA ^a	100%	0.642 (0.611-0.673)	<0.001
CURB-65	82.5%	0.759 (0.732-0.786)	<0.001
Abbreviations: AUC indicate	es area under the	e receiver operating charac	cteristic curve; Cl
confidence interval; PSI, p	neumonia seve	rity index. ^a Calculated a	assuming missing
values were normal.			

3 1	548
5	549
6 7	
8 9	
10 11	
12	
13 14	
15 16	
17 18	
19	550
20	550
22 23	552
24 25	553
26 27	
28	
29 30	
31 32	
33 34	
35 36	
37	
38 39	
40 41	
42 43	
44	
46	
47 48	
49 50	
51 52	
53 54	
55	
56 57	
58 59	

1 2

548 Table 4. Discriminatory ability of SeF-ML vs. other scoring systems for 30-day

549 mortality in the validation cohort

Model/Score	Availability	AUC (95% CI)	P-value (vs. SeF-ML)
SeF-ML	100%	0.826 (0.753-0.899)	-
PSI (raw score)	100%	0.830 (0.753-0.90)	0.92
CURB-65	100%	0.764 (0.694-0.834)	0.03
SOFA	23.1%	0.728 (0.588-0.869)	0.85 ª
SOFA-imputed ^b	100%	0.771 (0.706-0.836)	0.14
qSOFA	98.3%	0.729 (0.653-0.804)	0.005 ª

Abbreviations: AUC indicates area under the receiver operating characteristic curve; CI,

confidence interval; PSI, pneumonia severity index. ^a Compared only for patients with

complete SOFA/qSOFA score. ^b Calculated assuming missing values were normal.

Figure Legends

cohort

1 С

Figure 1. Model performance for 30-day mortality prediction in the derivation cohort.

Figure 2. Survival curves for patients stratified according to Pmort in the derivation

Figure 3. Survival curves stratified by SF risk group, set together according to CURB-

Figure 4. Model performance for 30-day mortality prediction in the validation cohort.

J, LURVES I.

A: ROC curves B: Calibration curves for SeF models

A: ROC curves B: Calibration curves for SeF models

65/PSI score in the derivation cohort

2 3 4	554
5 6	555
7 8	556
9 10	557
10 11 12	558
12	559
14 15	560
16 17	561
18 19	562
20 21	563
22 23	
24 25	
25 26	
27 28	
29 30	
31 32	
33 34	
35 36	
37	
39	
40 41	
42 43	
44 45	
46 47	
48 49	
50 51	
52	
53 54	
55 56	
57 58	
59 60	

Take-home Points:

- 565 Study Questions: Did a CPN model could predict mortality in patients with CAP better
- 566 than the frequent severity scores?
 - **Results:** SeF-ML performance at predicting 30-day mortality was overall significantly
- 568 better than that of existing CAP-specific scores.
- 569 Interpretation: Our results showed that SeF-ML shows potential for improving
- 570 mortality prediction amongst patients with CAP using structured health data.







Figure 2. Survival curves for patients stratified according to Pmort in the derivation cohort

152x101mm (72 x 72 DPI)



Figure 3. Survival curves stratified by SF risk group, set together according to CURB-65/PSI score in the derivation cohort

381x254mm (72 x 72 DPI)

ScholarOne - http://mchelp.manuscriptcentral.com/gethelpnow/index.html - (434) 964-4100



60





Figure 4. Model performance for 30-day mortality prediction in the validation cohort. A: ROC curves B: Calibration curves for SeF models

406x177mm (72 x 72 DPI)

Page 35 of 81

60

CHEST




e-Figure 2. Example of composite Gaussian distributions created by mapping across severity states (conditional probabilities learned in the CPN). A: learned distributions B: univariate outcome prediction across the predictor domain

254x190mm (96 x 96 DPI)





60



e-Figure 4. 30-day mortality as a function of the size of the high-risk group, compared with PSI and CURB-65 operating points in the derivation cohort

203x101mm (72 x 72 DPI)



50 SeF-ML Pmort • PSI operating points 40 CURB-65 operating points ٠ 30-day mortality 30 20 10 0 ó 20 40 60 80 100 High-risk group size, % of all patients





Supplementary Material

Machine Learning Model for Mortality Prediction in Patients with Community-

acquired Pneumonia: Development and validation study

Catia Cilloniz, PhD^{1*}, Logan Ward, PhD^{2*}, Mads Lause Mogensen, PhD^{2*}, Juan M Pericàs, PhD^{3,4*}, Raúl Méndez, PhD⁵, Albert Gabarrús, MSc¹, Miquel Ferrer, PhD¹, Carolina Garcia-Vidal, PhD⁶, Rosario Menendez, PhD⁵, Antoni Torres, PhD^{1.6}

Changes made to the published SepsisFinder model

Addition of input variables

The published version of the SepsisFinder model [1] was updated to include a wider set of input variables. Variables which are to include or extend modelling of dysfunction associated with the hepatic system (bilirubin), renal system (urea), haematology (WBC, hematocrit) and respiratory system (respiratory rate, SpO2, PaO2, FiO2, pH and HCO3). Background models were added for both bacteraemia (pacemaker, urinary catheter, central IV line, diabetes, recent surgery, heart failure) and mortality (malignancy, functional capacity, nasogastric/endotracheal tubes, heart failure). However, background variables were not used as inputs for analysis in this manuscript. Placement of the neutrophil fraction variable in the model was changed to make it an independent predictor of bacteraemia unrelated to severity.

Relearning conditional probability distributions

Conditional probabilities in the CPN model were learned using the same derivation data and learning technique described in [1–4].

Variable selection

Variable selection for the published version of the SepsisFinder model is described in the literature. Variables were selected if they showed utility in predicting bacteraemia, 30-day mortality or both, as well as a documented link to some underlying physiological process which suggests clinical relevance regarding linkages to infection/organ dysfunction/critical illness. Stepwise selection of variables was not performed.

Variable parameterization

The measured lab/vital variables are input as continuous variables. The SepsisFinder model models a set of severity states, for each input variable this is modeled as a Gaussian mixture. Design of the individual Gaussian components is described in the literature for earlier versions of the SepsisFinder model[3,4], while components for variables added to the latest version were generated by fitting a Gaussian Mixture Model (GMM). GMM modelling was performed in python (scikit-learn).

To determine the optimal number of Gaussian components and their mean/variance, a grid search across the number of components (min=4, max=10) was performed to assess the 10-fold cross-validation performance. The number of components with the highest likelihood (best overall fit of the input data) was selected and the GMM was then relearned with the complete training set. Manual adjustments were then made to avoid distributions with very high variance as this can create undesired behavior near the edges of other distributions. An example is shown in e-Figure 1 for Urea.







1 2 3

Models

The mapping of severity states across the individual Gaussian components occurs when learning the conditional probability distributions in the model. Observation of a certain value propagates an odds ratio equivalent to the relative probability of each state given the observed value, which is used to adjust the model's *a priori* beliefs. An example of such a mapping is shown for urea in e-Figure 2. This concept is also described in the literature [1,3].

e-Figure 2. Example of composite Gaussian distributions created by mapping across severity states (conditional probabilities learned in the CPN). A: learned distributions B: univariate outcome prediction across the predictor domain







Evaluation of overfitting to the derivation dataset

To ensure the model did not significantly overfit to the derivation data, a ten-fold cross-validation was performed to assess the internal performance. The derivation data were shuffled and split into ten cross-validation folds, stratifying by outcome to ensure a similar proportion of outcomes in each fold. Ten models were trained and evaluated, in each case holding out one of the cross-validation folds as a validation set and training using the remaining nine. The ROC AUC was then computed for each model on its respective held-out validation set. The mean cross validation AUC was 0.800 (range: 0.749-0.832). The final model was then learned using the entire derivation cohort. The internal performance of the final model on the derivation cohort was AUC = 0.801 (95% CI 0.777-0.825) which did not differ significantly from the cross-validation performance.

Handling of missing data

The CPN model does not require any explicit handling of missing data. Ability to handle missing data is an inherent property of CPNs. e-Table 1 shows the measurement rates of the CPN input variables for the derivation and validation datasets.

For three of the variables (CRP, platelets, neutrophil fraction) where there was a discrepancy in measurement rates between the cohorts, this was a temporal feature of the derivation data. Prior to 2007, measurement of these parameters was less common. However, the lower availability of these parameters did not adversely affect model performance, as shown in e-Table 2.

e-Table 1. Measurement rates

Measurement rates	Derivation	Validation
	n (%)	n (%)
Temperature	4235 (93.5)	1012 (97.9)
Chills	4388 (96.8)	na
MAP	4241 (93.6)	1000 (96.7)
HR	4184 (92.3)	1032 (99.8)
RR	3907 (86.2)	1016 (98.3)
Mental status	4531 (100.0)	1031 (99.7)
SpO2	3098 (68.4)	992 (95.9)
CRP	3971 (87.6)	1017 (98.4)
Creatinine	4457 (98.4)	1031 (99.7)
Platelets	3388 (74.8)	1027 (99.3)
Neutrophil fraction	3884 (85.7)	1032 (99.8)
Urea	744 (16.4)	1021 (98.7)
рН	3317 (73.2)	668 (64.6)
PaO2	3334 (73.6)	645 (62.4)
FiO2	3167 (69.9)	983 (95.1)
HCO3	2998 (66.2)	653 (63.2)
Leukocytes	4452 (98.3)	1031 (99.7)
Lactate	na	181 (17.5)
Bilirubin	na	253 (24.5)
Hematocrit	na	1031 (99.7)

na: not available

3
4
5
6
7
~
8
9
10
11
12
12
13
14
15
16
17
10
10
19
20
21
22
23
23
24
25
26
27
28
20
29
30
31
32
33
34
25
35
36
37
38
39
10
40
41
42
43
44
45
16
40
4/
48
49
50
51
51
ےد ح
53
54
55
56
57
50
10

1 2

Years	Patients,	Mortality,	Availability of selected model inputs		AUC 30-day mortality	
	n (%)	n/N (%)	CRP	Platelets	Neut.Frac	
2003-2006	1687	90/1679	76.5%	41.6%	67.2%	0.856 [0.824-0.888]
		(5.4)				
2007-2016	2844	202/2744	94.6%	94.5%	97.0%	0.771 [0.739-0.804]
		(7.4)				
All	4531	292/4423	87.6%	74.8%	85.7%	0.801 [0.777-0.825]
		(6.6)				

e-Table 2. Impact of missing data

Abbreviations: CRP indicates C-reactive protein

e-Table 3. Mortality in PSI risk classes and quantile-matched SF risk classes in the validation cohort

PSI Risk Class	Patients, n (%)	30-day Mortality	SeF-ML – matched, 30-day mortality
1	112 (10.9)	1 (0.9)	1 (0.9)
2	185 (18.1)	0 (0.0)	2 (1.1)
3	299 (29.2)	5 (1.7)	4 (1.3)
4	332 (32.4)	16 (4.8)	13 (3.9)
5	96 (9.4)	21 (21.9)	23 (24.0)

Abbreviations: PSI indicates pneumonia severity index; SeF-ML, SepsisFinder-Machine

Learning.

CURB65 Risk	Patients, n	30-day	SeF-ML – matched, 30-day mortality
Class	(%)	Mortality	
0	207 (20.0)	2 (1.0)	2 (1.0)
1	282 (27.3)	3 (1.1)	2 (0.7)
2	390 (37.7)	16 (4.1)	12 (3.1)
3	124 (12.0)	15 (12.1)	16 (12.9)
4	29 (2.8)	6 (20.7)	12 (41.4)
5	2 (0.2)	1 (50.0)	0 (0.0)

e-Table 4. Mortality in CURB65 risk classes and quantile-matched SF risk classes in the validation cohort

Abbreviations: CURB65 indicates Confusion, Urea, Respiratory Rate, Blood pressure

and Age 65; SeF-ML, SepsisFinder-Machine Learning.

e-Figure 4. 30-day mortality as a function of the size of the high-risk group, compared with PSI and CURB-65 operating points in the derivation cohort



The coloured circles show possible operating points for the PSI and CURB-65 scores.

e-Figure 5. 30-day mortality as a function of the size of the high-risk group in the validation cohort



The coloured circles show possible operating points for the PSI and CURB-65 scores.

References.

 Ward L, Andreassen S. A Bayesian Approach to Model-Development: Automatic Learning for Tuning Predictive Performance. IFAC-PapersOnLine [Internet] 2015;48:481–6. doi:

http://linkinghub.elsevier.com/retrieve/pii/S2405896315020789doi: 10.1016/j.ifacol.2015.10.187

- 2. Ward LM. Gradation of the severity of sepsis Learning in a causal probabilistic network. 2016;doi: 10.5278/VBN.PHD.MED.00064
- 3. Ward L, Paul M, Andreassen S. Automatic Learning of mortality in a CPN model of the Systemic Inflammatory Response Syndrome. Math. Biosci. 2017;284:12–20. doi: http://dx.doi.org/10.1016/j.mbs.2016.11.004
- Ward L, Møller JK, Eliakim-Raz N, Andreassen S. Prediction of Bacteraemia and of 30-day Mortality Among Patients with Suspected Infection using a CPN Model of Systemic Inflammation. IFAC-PapersOnLine [Internet] 2018;51:116–21. doi: https://linkinghub.elsevier.com/retrieve/pii/S2405896318333731doi: 10.1016/j.ifacol.2018.11.657

STROBE Statement—Checklist of items that should be included in reports of cohort studies

	Item No	Recommendation	Page No
Title and abstract	1	(a) Indicate the study's design with a commonly used term in the title or the	
		abstract	
		(b) Provide in the abstract an informative and balanced summary of what was	1-3
		done and what was found	
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	4
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting locations and relevant dates including periods of	5-7
Setting	U	recruitment, exposure, follow-up, and data collection	
Participants	6	(a) Give the eligibility criteria, and the sources and methods of selection of	
	-	participants. Describe methods of follow-up	
		(b) For matched studies, give matching criteria and number of exposed and	5-7
		unexposed	
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and	5-7
		effect modifiers. Give diagnostic criteria, if applicable	
Data sources/	8*	For each variable of interest, give sources of data and details of methods of	5-7
measurement		assessment (measurement). Describe comparability of assessment methods if	
		there is more than one group	
Bias	9	Describe any efforts to address potential sources of bias	5-7
Study size	10	Explain how the study size was arrived at	5-7
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable,	7
		describe which groupings were chosen and why	
Statistical methods	12	(a) Describe all statistical methods, including those used to control for	
		confounding	
		(b) Describe any methods used to examine subgroups and interactions	7-11
		(c) Explain how missing data were addressed	
		(d) If applicable, explain how loss to follow-up was addressed	
		(<i>e</i>) Describe any sensitivity analyses	
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially	12
-		eligible, examined for eligibility, confirmed eligible, included in the study,	
		completing follow-up, and analysed	
		(b) Give reasons for non-participation at each stage	
		(c) Consider use of a flow diagram	
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social)	12-
		and information on exposures and potential confounders	13
		(b) Indicate number of participants with missing data for each variable of interest	
		(c) Summarise follow-up time (eg, average and total amount)	
Outcome data	15*	Report numbers of outcome events or summary measures over time	14

CHEST

16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and	7-17
	their precision (eg, 95% confidence interval). Make clear which confounders were	
	adjusted for and why they were included	
	(b) Report category boundaries when continuous variables were categorized	
	(c) If relevant, consider translating estimates of relative risk into absolute risk for a	
	meaningful time period	
17	Report other analyses done-eg analyses of subgroups and interactions, and sensitivity	Not
	analyses	performe
18	Summarise key results with reference to study objectives	18
19	Discuss limitations of the study, taking into account sources of potential bias or	20
	imprecision. Discuss both direction and magnitude of any potential bias	
20	Give a cautious overall interpretation of results considering objectives, limitations,	18-21
	multiplicity of analyses, results from similar studies, and other relevant evidence	
21	Discuss the generalisability (external validity) of the study results	18-21
on		
22	Give the source of funding and the role of the funders for the present study and, if	22
	applicable, for the original study on which the present article is based	
	16 17 17 18 19 20 21 0n 22	 16 (a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included (b) Report category boundaries when continuous variables were categorized (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period 17 Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses 18 Summarise key results with reference to study objectives 19 Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias 20 Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence 21 Discuss the generalisability (external validity) of the study results on 22 Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the arceant article is based.

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at http://www.strobe-statement.org.

1	Word counts abstract: 275
2	Word counts text: 3006
3	
4	Machine-Learning Model for Mortality Prediction in Patients with Community-
5	acquired Pneumonia: Development and validation study
6	Running title: Machine-Learning for Mortality Prediction in CAP
8	Catia Cilloniz, PhD ^{1*} , Logan Ward, PhD ^{2*} , Mads Lause Mogensen, PhD ^{2*} , Juan M
9	Pericàs, MD, PhD ^{3,4*} , Raúl Méndez, MD,PhD ⁵ , Albert Gabarrús, MSc ¹ , Miquel Ferrer,
10	MD, PhD ¹ , Carolina Garcia-Vidal, MD, PhD ⁶ , Rosario Menendez, MD, PhD ⁵ , Antoni
11	Torres, MD, PhD ^{1.6}
12	¹ Department of Pneumology, Hospital Clinic of Barcelona; August Pi i Sunyer
13	Biomedical Research Institute–IDIBAPS, University of Barcelona; Biomedical Research
14	Networking Centres in Respiratory Diseases (Ciberes) Barcelona, Spain.
15	² Treat Systems ApS, Aalborg, Denmark
16	³ Department of Infectious Diseases, Hospital Clinic of Barcelona, Barcelona, Spain.
17	⁴ Liver Unit, Vall d'Hebron University Hospital, Vall d'Hebron Institute for Research
18	(VHIR), Barcelona, Spain.
19	⁵ Department of Pneumology, Hospital La Fe de Valencia, Valencia, Spain
20	⁶ ICREA, Barcelona, Spain
21	*Equal contribution
22	
23	Corresponding author: Prof. Antoni Torres (primary) or Dr. Catia Cillóniz (alternative)
24	Department of Pulmonary Medicine, Hospital Clinic of Barcelona
25	C/ Villarroel 170, 08036 Barcelona, Spain
26	Tel: (+34) 93-227-5779, fax: (+ 34) 93-227-9813
27	Email: <u>atorres@clinic.cat</u> (primary) or <u>catiacilloniz@yahoo.com</u> (alternative)

CHEST

3	
4	
5	
6	
7	
/	
8	
9	
10	
11	
12	
13	
11	
15	
10	
16	
17	
18	
19	
20	
21	
22	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
27	
32	
33	
34	
35	
36	
37	
38	
30	
40	
40	
41	
42	
43	
44	
45	
46	
47	
10	
40	
49	
50	
51	
52	
53	
54	
55	
55	
50	
5/	
58	
59	

60

28 Summary conflict of interest statements:

29 The authors declare that they have no conflicts of interest with the study. Dr Cillóniz is 30 the recipient of the 2018 SEPAR fellowship and a grant from the Fondo de 31 Investigación Sanitaria (PI19/00207). RM has received honoraria for lectures from 32 Pfizer.

33 **Funding information:**

34 This study was supported by CIBER de Enfermedades Respiratorias (CIBERES 35 CB06/06/0028), and by 2009 Support to Research Groups of Catalonia 911, IDIBAPS. 36 The founders of the study had no role in the study design, data collection, analysis, or 37 interpretation, writing of the report, or decision to submit for publication. SEPAR 38 integrated respiratory infections program.

40 Notation of prior abstract publication/presentation:

- 41 None
- 42

1 2		
3	43	Key words list:
4 5	44	Community-acquired pneumonia
6 7	45	Machine learning
8 9	46	Pneumonia
10 11 12 13	47	Artificial intelligence
	48	Mortality prediction
14	49	Abbreviation list:
16	50	ML: machine learning
17 18	51	AI: artificial intelligence
19 20	52	CAP: community-acquired pneumonia
21 22	53	CPN: causal probabilistic network
23 24	54	SeF: SepsisFinder
25	55	SeF-ML: SeF adapted model
26 27 28 29 30 31 32 33 34 35 36 37 38 20	56	ATS/IDSA: American Thoracic Society/Infectious Diseases Society of America
	57	SOFA: Sequential Organ Failure Assessment
	58	qSOFA: quick Sequential Organ Failure Assessment
	59	ED: emergency department
	60	MAP: median arterial pressure
	61	SBP: systolic blood pressure
	62	DBP: diastolic blood pressure
39 40	63	CRP: C-reactive protein
41 42	64	ARDS indicates acute respiratory distress syndrome
43 44	65	AUC-ROC: area under the receiver operating characteristic curve
45 46	66	CI: confidence intervals
47	67	Q1: first quartile
48 49 50 51 52 53 54 55 56 57 58 58	68	Q3: third quartile
	69	FiO ₂ , fraction of inspired oxygen
	70	IDSA/ATS: Infectious Diseases Society of America/American Thoracic Society
	71	PaO ₂ : partial pressure of arterial oxygen
	72	PSI, pneumonia severity index
	73	Q1: first quartile
60	74	Q3: third quartile

75 Abstract

Background: Artificial intelligence tools and techniques such as machine learning (ML) are increasingly seen as a suitable manner to increase the prediction capacity of currently available clinical tools, including prognostic scores. However, studies evaluating the efficacy of ML methods in enhancing the predictive capacity of existing scores for community-acquired pneumonia (CAP) are limited. We aimed to apply and validate a causal probabilistic network (CPN) model to predict mortality in patients with CAP.

Research question: Is a CPN model able to predict mortality in patients with CAP
 better than the commonly-used severity scores?

Study design and methods: Derivation-validation retrospective study conducted in two Spanish University hospitals. The ability to predict 30-day mortality of a CPN designed to predict mortality in sepsis (SeF) and adapted for CAP (SeF-ML) was assessed and compared to other scoring systems (PSI, SOFA, qSOFA and CURB-65). The SeF models are proprietary software. Differences between ROC curves were assessed using De Long's method for correlated ROC curves.

- **Results:** The derivation cohort comprised 4,531 patients whilst the validation cohort had 1,034 patients. In the derivation cohort, the AUC of SeF-ML, CURB-65, SOFA, PSI and qSOFA were 0.801, 0.759, 0.671, 0.799 and 0.642, respectively, for 30-day mortality prediction. In the validation study, the AUC of SeF-ML was 0.826, concordantly with the AUC (0.801) in the derivation data (p=0.51). The AUC of SeF-ML was significantly higher than those of CURB-65 (0.764, p=0.03) and qSOFA (0.729, p=0.005). However, it did not differ significantly from PSI (0.830, p=0.92) and SOFA (0.771, p=0.14).
- Interpretation: SeF-ML shows potential for improving mortality prediction amongst
 patients with CAP using structured health data. Additional external validation studies
 should be conducted to support generalisability.

3	
4	
5	
6	
7	
8	
9 10	
10	
12	
13	
14	
15	
16	
17	
18	
19 20	
20	
22	
23	
24	
25	
26	
27	
20 20	
30	
31	
32	
33	
34	
35	
30 27	
38	
39	
40	
41	
42	
43	
44 45	
45 46	
47	
48	
49	
50	
51	
52	
53 5∕I	
55	
56	
57	
58	
59	
60	

1 2

104 Introduction

105

106 Community-acquired pneumonia (CAP) remains a major cause of morbidity and 107 mortality worldwide, with children and elderly being the most affected population^{1,2}. 108 Early-risk stratification of severity and adequate antimicrobial therapy are critical to 109 improving CAP outcomes^{1,3}. Scoring systems such as Pneumonia Severity Index (PSI)⁴ 110 and CURB-65 score⁵ have been widely used to stratify patients with CAP and support 111 clinical decision-making processes³. However, these scores have several limitations 112 that hinder their capacity as clinical decision-making supporting tools, e.g., low positive 113 and negative predictive values in predicting the need of intensive care, or 114 underestimated severity in certain age groups.

115 Artificial intelligence (AI) tools and techniques such as machine learning (ML) are 116 increasingly seen as a suitable manner to increase the prediction capacity of currently 117 available tools in infectious diseases, e.g., sepsis, antimicrobial resistance and COVID-118 19^{6,7}. Within respiratory medicine, the main applications of AI and ML have included 119 the interpretation of thoracic imaging, lung pathology slides and physiologic data such 120 as pulmonary function tests⁸. Nonetheless, studies evaluating the efficacy of ML 121 methods in enhancing the predictive capacity of existing scores for CAP are limited^{9,10}. 122 We aimed to apply a causal probabilistic network (CPN) model previously used in 123 sepsis (SeF) ^{11–13} to predict 30-day mortality in patients with CAP, comparing the 124 accuracy of this model to that of the established clinical scores. In addition, we 125 pursued validating the ML model in CAP using a large cohort of patients with CAP. 126

128

127

1 2		
2 3 4	129	Methods
5 6 7	130	Study design and participants
7 8 9	131	A derivation-validation retrospective study was performed when using an innovative
10 11	132	CPN model ^{12,13} to predict mortality in adult patients hospitalised with CAP. The
12 13 14	133	rationale for using this new statistical approach is trying to diminish the effect of
15 16	134	correlations and make better use of the variables, thus avoiding losing patients with
17 18 19	135	missing information. We applied the guidelines provided by Leisman et al 14 for
20 21	136	reporting of prediction models.
22 23 24	137	Institutional approval was provided by the IRBs of both University Hospital La Fe of
24 25 26	138	Valencia (EC2011/2019) and Hospital Clínic of Barcelona (HCP2009/5451), which
27 28	139	waived the need for informed consent due to the retrospective nature of the study.
29 30 31	140	Data were collected within the first 24 hours of admission. Collected data had similar
32 33	141	definitions for both cohorts, and harmonization between cohorts was elementary. All
34 35 36	142	data set were anonymously analyzed, and the study was performed following current
37 38	143	recommendation of the Declaration of Helsinki.
39 40	144	Definitions
41 42 43	145	CAP was diagnosed if new acute respiratory symptoms, signs, and compatible
44 45	146	infiltrate(s) on chest x-ray were present. Severe CAP was defined according to the
46 47 48	147	ATS/IDSA guidelines ³ . Prior antibiotic treatment was defined as the intake of
49 50	148	antibiotics during the week before hospital admission. The appropriateness of empiric
51 52 53	149	antibiotic treatment was determined according to multidisciplinary guidelines for the
54 55	150	management of CAP ¹⁵ . Sepsis was defined as the presence of pneumonia and an
56 57 58	151	increase of ≥ 2 points in the Sequential Organ Failure Assessment (SOFA) score ¹⁶ . We
59 60	152	also calculated median arterial pressure (MAP) from systolic blood pressure (SBP) and

diastolic blood pressure (DBP) as 1/3*SBP + 2/3*DBP. The ML technique used is based on CPN ^{11,12,17}. TREAT-Lab (Treat Systems ApS, Aalborg, Denmark) is a medical device software program that provides a risk assessment of patients with suspected infection (proprietary software). The aim of such software is to inform the use of additional or adjunct diagnostics. For example, the clinician (or clinical microbiologist) can use the risk score to identify high-risk patients, i.e., classified as those with high predicted probability of mortality, who may benefit from rapid diagnostics. Conversely, it may be used to identify low-risk patients who only receive standard of care. Customisable risk thresholds can be set for individual clinical installations depending on resources available and local practice requirements. The risk assessment model used within TREAT-Lab is the SeF CPN.

164 Patient selection, inclusion and exclusion criteria

165 Derivation cohort

We enrolled all consecutive adult patients with a CAP diagnosis in the Emergency Department of Hospital Clinic of Barcelona between January 2003 and December 2016. We excluded patients with severe immunosuppression due but not limited to human immunodeficiency virus (HIV) infection, active solid or haematologic malignancy treated with chemotherapy, oral corticosteroid treatment with at least 20 mg of prednisone (or equivalent) per day for at least two weeks, and treatment with other immunosuppressive drugs. We also excluded those with active tuberculosis or a confirmed alternative diagnosis.

174 Validation cohort

All consecutive patients admitted with CAP to Hospital Universitario y Politécnico La Fe(Valencia, Spain) between January 2012 and December 2018 were included. The

CHEST

inclusion criteria were CAP diagnosis based on a new radiologic infiltrate with at least
two compatible clinical symptoms and age ≥18 years. Exclusion criteria were hospital
admission within the previous 15 days, immunosuppressive treatments and HIV
infection.

181 Data collection and evaluation

182 Derivation cohort

Demographic variables, comorbidities, and physiologic parameters were collected at the emergency department within 24 hours of admission. The PSI, CURB-65, qSOFA and SOFA score were calculated at admission^{4,5,18}. We recorded whether patients had specific complications, including multilobar infiltration, pleural effusions, acute respiratory distress syndrome (ARDS)¹⁹, septic shock²⁰ and acute renal failure²¹ during hospitalisation. All surviving patients were visited or contacted by telephone within 30 days of discharge; hospital records and the Catalunya Health Department database were reviewed at the 1-year mark. We included all available patients in this analysis. We also calculated MAP from SBP and DBP as 1/3*SBP + 2/3*DBP. We discretised PO2 and FiO2 as required by the model. Finally, we transformed creatinine, C-reactive protein (CRP), lactate, bilirubin and platelets through the natural logarithm.

194 Validation cohort

Demographic characteristics and comorbidities (diabetes mellitus, respiratory, heart, liver, neurological, and renal diseases) were collected at time of admission. The severity of disease at presentation was assessed with the PSI. Antibiotic treatment before CAP diagnosis in the current episode was recorded. We included all available patients in this analysis.

200	C
201	F
202	S
203	١
204	r
205	С
206	ι
207	٧
208	C
209	٧
210	s
211	v
212	k
213	٧
214	(
215	a
216	F
217	i
218	t
219	v
220	i
221	С
222	٢
223	r
	200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 217 218 219 220 221 221 222

200 Outcomes

1

201 Primary outcome: death within 30 days of admission.

202 Statistical analysis

We report the number and percentage of patients for categorical variables and the median and interquartile range for continuous variables (non-normal distribution confirmed by the Kolmogorov-Smirnov test). Categorical variables were compared using the χ^2 test. Continuous variables were compared using the nonparametric Mann-

207 Whitney test.

208 Derivation cohort

We used an updated version of the published SepsisFinder (SeF) model (see supplementary material for a description of the modelling techniques used including variable selection and parameterization). The main differences in input variables between SeF and the adapted model (SeF-ML) are shown in Table 1.

We adapted the respiratory component of SeF, which comprises acid-base balance (pH, HCO3-), respiratory rate, and oxygen perfusion (measured through PaO2, SaO2 and FiO2 [FiO2 was discretized based on a conversion from oxygen flow rate in L/min. PaO2 was discretized into 8 bins]). We retrained this portion of the model using data included in the study. The remainder of the model was deemed invariant. To evaluate the degree of overfitting to the derivation data, we performed a 10-fold crossvalidation, shuffling the data and stratifying to ensure a similar proportion of outcomes in each cross-validation fold. We then retrained the model using the full derivation dataset.

222 No explicit steps to handle missing data were required. CPNs are inherently tolerant of 223 missing information and are able to perform inference with partial evidence: Page 61 of 81

 CHEST

combining the model's structure and conditional probability tables with the axioms of probability theory allows the marginal probabilities of all nodes in a CPN to be calculated when only some of the nodes are observed (supplemental material and e-Table 1)^{22,23}. We then used the SeF-ML model to calculate the probability of death within 30 days (Pmort) for all patients with >2 input variables recorded.

We assessed predictive performance of SeF and SeF-ML by using the area under the receiver operating characteristic curve (AUC-ROC). As a general rule, the relation between AUC and diagnostic accuracy is as follows: AUC between 0.90 and 1.00 has outstanding discrimination ability; 0.80 - 0.90, excellent; 0.70 - 0.80, acceptable; 0.60 -0.70, poor, and 0.50 - 0.60 fails to accurately diagnose a certain disease or condition^{24,25}. Kaplan-Meier survival curves were also constructed.

235 In our preliminary analysis of this data, we noted that SeF-ML performance was not
236 negatively impacted by the greater degree of missing data amongst patients included
237 from 2003 to 2006 (e-Table 2).

We compared SeF and SeF-ML performance (Pmort as a predictor of death within 30
days) with other scoring systems (PSI, SOFA, qSOFA and CURB-65). Differences
between ROC curves were assessed using De Long's method for correlated ROC curves
as implemented in the pROC package of R.

In addition to assessing the model's performance, we computed an example for a
 potential use-case for TREAT Lab, showing patients being stratified into groups of 40%
 low-risk, 40% medium-risk and 20% high-risk.

245 Validation cohort

57 246 We used the SeF-ML model (Table 1), where the respiratory components were 58 59 247 adapted through learning from 4,531 patients with CAP at the Hospital Clinic of

Page 62 of 81

CHEST

Barcelona. We used the adapted model to calculate the probability of death within 30days (Pmort) for all patients.

We assessed predictive performance using the AUC. We compared SeF-ML's performance (Pmort as a predictor of death within 30 days) with other scoring systems (i.e., PSI, CURB-65, SOFA and qSOFA). Differences between ROC curves were assessed using DeLong's method for correlated ROC curves as implemented in the pROC package of R. We also compared mortality for a range of risk cut-offs to assess potential operating points for clinical implementation of risk stratification. Model calibration was assessed using the Hosmer-Lemeshow goodness of fit test and by calculating the Brier-score loss.

258 More details about the selection of variables and the ML model are displayed in the 259 supplementary material (e-Figure 1, e-Figure 2 and e-Figure 3).

CHEST

2 3	261	Results
4 5	-	
6 7	262	Patients' characteristics
, 8 9	263	The derivation cohort comprised 4,531 (15% outpatients and 85% inpatients) patients
10 11 12	264	and the validation cohort 1,034 patients. Clinical characteristics of the studied
12 13 14	265	population are shown in Table 2.
15 16	266	Predictive performance of risk assessment
17 18 19	267	Derivation cohort
20 21	268	We used the SeF-ML model to calculate Pmort for all patients with >2 input variables –
22 23	269	4,500/4,531=99.3% of patients. The AUC for 30-day mortality prediction was 0.801 for
24 25 26 27 28 29 30 31	270	SeF-ML, being significantly better than CURB-65 (0.759, p<0.001), SOFA (0.671,
	271	p<0.001), and qSOFA (0.642, p<0.001) (Table 3). The mean cross-validation
	272	performance, measured by AUC, was 0.800 (range: 0.749-0.832) which did not differ
32 33	273	from the AUC for the full dataset. Details of the cross-validation assessment are
34 35 26	274	included in the supplemental material.
37 38	275	PSI was only available for 58% of patients and had an AUC of 0.799; it was not
39 40	276	significantly different from SeF-ML (p=0.58). CURB-65 provided a "fair" prediction of
41 42 43	277	mortality, while SeF-ML and PSI provided a "good" prediction of 30-day mortality
43 44 45	278	(Figure 1). The calibration of SeF-ML was measured using the Hosmer-Lemeshow
46 47 48	279	statistic and the Brier-score loss. The Hosmer-Lemeshow statistic was 15.62 (p=0.048)
49 50	280	which suggests the model may not be well calibrated. However, the Hosmer-
51 52	281	Lemeshow statistic is known to be very sensitive to sample size. The model appears
53 54 55	282	visually well-calibrated. The Brier-score loss for the model was 0.056.
56 57	283	Survival curves are shown in Figure 2 for patients stratified according to Pmort; the
58 59 60	284	low-risk group comprises 40% of patients with the lowest Pmort; the medium-risk

> group represents the next highest 40%; and the high-risk group accounts for the top 20% according to Pmort. The additional value provided by SeF-ML with respect to CURB-65 and PSI scores is shown in Figure 3: SeF-ML provides additional discrimination for mortality, e.g., for patients with $PSI \ge 4$.

Validation cohort

For the validation cohort, the AUC for SeF-ML was 0.826. It was not significantly different from performance in the derivation cohort (p=0.51) (Figure 4). According to the Hosmer-Lemeshow statistic, SeF-ML was well-calibrated (HL statistic = 11.93, p=0.15). The Brier-score loss for SeF-ML was 0.036. Notably, only 23.1% patients had complete information to calculate the SOFA score, being not significantly different from SeF-ML (p=0.85). On the contrary, SeF-ML had a significantly higher AUC than both CURB-65 and qSOFA (p=0.03 and p=0.005, respectively) (Table 4).

When analysing the imputed SOFA score (assuming missing components were normal), the AUC of SOFA improved from 0.728 to 0.771; the difference from SeF-ML remained non-significant.

Supplementary e-Table 3 shows the number of patients in each PSI risk class and associated 30-day mortality. Also, 30-day mortality within quantile-matched risk classes for SeF-ML is also shown. SeF-ML risk groups were defined by choosing cut-offs that resulted in the same number of patients assigned to each risk group like in the corresponding PSI risk class.

Potential use of risk assessment

Patients can be stratified by their probability of death within 30 days. An odds ratio (OR) for death can also be calculated for high-risk patients vs. others. The OR represents the degree of separation between high-risk and low-risk patients. For this

Page 65 of 81

CHEST

analysis, we considered two strategies for stratification: high-risk patients receive rapid diagnostics and low-risk patients receive minimum standard of care. For example, the top 20% had a probability of death within 30 days of 18.4%, whilst OR for the high-risk group vs. remaining patients was 7.6 (e-Figure 4). One advantage of SeF-ML is the continuous probability output that allows custom thresholds to be set depending on the care environment versus the five potential operating points for CURB-65, for example. For comparison, three potential scenarios are shown below: 1) PSI=5 (N=98) 30-day mortality = 21.9% (PSI), 24% (SeF-ML); OR for high-risk vs. low risk = 11.5 (PSI), 14.3 (SeF-ML); 2) PSI ≥4 (N=428) 30-day mortality = 8.6% (PSI), 8.4% (SeF-ML); OR for high-risk vs. low risk = 9.3 (PSI), 7.7 (SeF-ML); 3) CURB-65 ≥3 (N=153) 30-day mortality = 14.4% (CURB-65), 18.3% (SeF-ML); OR for high-risk vs. low risk = 6.8 (CURB-65), 12.8 (SeF-ML). A fuller picture of the effect of choosing different operating points is shown in Supplementary e-Figure 4 and e-Figure 5. It shows the 30-day mortality as functions of the size of the high-risk group along with operating points for PSI (state) and CURB-65. The smaller the high-risk group, the higher the 30-day mortality.

325 Discussion

In our study, we applied a ML approach to develop and validate a 30-day mortality prediction clinical model in patients with CAP. Although not pneumonia-specific, the SeF-ML model shows potential for improving mortality prediction amongst patients with CAP. Remarkably, SeF-ML's performance in the validation set matched that of the training data in the derivation cohort, as did those of PSI and CURB-65, whereas the performance of SOFA and qSOFA scores did not match. SeF-ML not only enhances the mortality prediction ability of currently available tools but optimizes the use and quality of available electronic health records (EHR) data. Hence, although the advantages conferred by SeF-ML need further refining and interpretation, this ML model might be applied in clinical practice in the near future, i.e., patients can be stratified by their probability of death and such stratification might then be used to determine a patient's care trajectory and diagnostic workup.

This study is one of the first to use a large cohort of hospitalised patients with CAP to generate a ML model validated through an external cohort of patients with CAP. A barrier for physicians in using ML is its potential "black box" opacity. However, studies like ours show that results obtained with ML predictions are consistent with other severity scores that we have used so far^{26,27}. In the near future, ML techniques will allow us to analyse a large volume of data that current techniques cannot do, facilitating the possibility amongst investigators to directly collect data from EHR ^{7,26–28}. The potential of SeF-ML to improve the current ability of available clinical scores for CAP primarily relies on the findings suggesting that SeF-ML better predicts 30-day mortality than qSOFA and CURB-65 according to our data, which nonetheless require further clinical validation. Compounding this is also the fact that SeF-ML had a higher

Page 67 of 81

CHEST

AUC than SOFA and PSI, albeit non-significant. Furthermore, the AUC value of SeF-ML was consistent between the predictive and validation model. There is strong evidence supporting the importance of accurately assessing the severity of CAP and stratifying patients based on their mortality risk to improve clinical outcomes^{1,3,29}. Despite that, PSI and CURB-65 remain the most widely used CAP severity scores and recommended by international guidelines^{3,29,30}. The ability of these scores to predict mortality has some limitations³¹. Moreover, although qSOFA score ≥ 2 is strongly associated with mortality in patients with pneumonia, its use in early identification of patients with CAP and mortality risk is hindered^{32,33}. SeF-ML seems to provide better discriminative capacity to discern between high- and low-risk patients, which is key to adapting the intensity of care and resources per foreseen prognosis. However, this still needs further validation to prove its actual clinical validity. In addition, SeF-ML increases efficiency in exploiting available data. The lower requirements set for minimum data and the ability to handle missing data mean that 99% of patients, on average, would have sufficient data for predictions to be established. This would facilitate optimization of EHR use, with fewer investments on data collection and curation when compared to other scores. Also, the continuous output achieved with SeF-ML allows for adaptive fine-tuning of patient classification. Cut-offs for defining risk can be smoothly and accurately adjusted. In particular, the ability of SeF-ML to identify high-risk patients with low CURB65 and PSI scores, as well as low-risk patients with high CURB65 and PSI scores paramount as these CAP-specific clinical scores are the two most widely used. It is worth noting, however, that enhanced mortality prediction through continuous appraisals is not unique to ML but is also found in other models with continuous

Page 68 of 81

CHEST

372 output, e.g. the eCURB model, which achieved greater predictive accuracy using CURB-

373 65 variables using regression splines³⁴.

The ability of SeF-ML to reclassify risk of patients with CAP might be more beneficial for certain patient subgroups, e.g., those with low CURB-65/PSI yet high SeF and mortality, or those with high CURB-65/PSI yet low SeF and mortality.

A foremost strength of our study is the use of a large set of clinical data that are representative of the population in a defined area of influence. This study feature enabled the integration of surveillance data into direct clinical care of individual patients and can be helpful in making decisions by applying ML models like SeF-ML. When developing and validating the algorithm, we used two large patient slices to ensure that our model can be implemented using real-time patient data.

However, some limitations need to be highlighted. This score validation against mortality is only the first step toward clinical utility. Remarkably, the ability of SeF-ML to predict ICU admission remains unproven. In particular, our study lacked information on post-admission disposition, and therefore we were not able to test SeF-ML against potentially different clinical pathways other than direct admission to the ICU. Furthermore, the use of closed databases does not incorporate new information and therefore does not allow "learning" of ML models. Besides, despite the large sample sizes of both cohorts, some baseline variables and clinical features, including CAP severity and outcomes, differed. However, this does not hamper the training and validation of the CPN model. There are components of the current SeF that were not available in the dataset due to them not generally being measured in the ED for CAP. However, SeF performs well despite not including these variables, which helps to demonstrate its robustness to missing values.

CHEST

396	Interpretation
-----	----------------

SeF-ML performance at predicting 30-day mortality appears to be overall superior than that of existing CAP-specific scores, with the exception of PSI in the validation model. SeF-ML offers some advantages over current scoring methods, eg., calculations easily made on routinely collected data and based upon; structured EHR data (vs. subjective criteria and arterial blood gas sampling needed for PSI for instance); and tunable performance so as to allow risk cut-offs to be tailored to workflow requirements and capacities of the individual institution (compared with fewer states in CURB-65). In addition, SeF-ML performance seems to not be dependent upon data availability, therefore allowing for more effective calculation of risk scores for CAP based upon data sources with limited access to or completeness of certain variables. Our findings need further validation in other cohorts from different settings to assess the actual clinical utility of SeF-ML in predicting CAP prognosis.

2 3 4	410	Acknowledgements
5	411	Guarantor:
7 8 9	412	Dr. Antoni Torres is the guarantor of the article
10 11	413	Author contributions:
12	414	Conception and design of the study: CC, LW, ML, MF, JP, RM, RM, AT. Data acquisition:
13 14	415	all authors. Data analysis and interpretation: LW, ML, CC, AG. Manuscript drafting or
15 16	416	critical revision for important intellectual content: all authors. Final approval of the
17 18	417	submitted version: all authors.
19 20	418	Financial/nonfinancial disclosures:
21	419	The authors have reported to Chest the following: This study was supported by CIBER
22 23	420	de Enfermedades Respiratorias (CIBERES CB06/06/0028), and by 2009 Support to
24 25	421	Research Groups of Catalonia 911, IDIBAPS. The founders of the study had no role in
26 27	422	the study design, data collection, analysis, or interpretation, writing of the report, or
28 29	423	decision to submit for publication. SEPAR integrated respiratory infections program
30 31	424	Role of sponsors:
32	425	The funders did not have a role in the conduct of the study: in the collection.
33 34	426	management, analysis, or interpretation of data: or in the preparation of the
35 36	42.7	manuscript.
37 38	428	
39 40	420	
40 41		
42		
43 44		
45		
46		
47 48		
49		
50		
51		
52 53		
54		
55		
56		
57		
50 59		
60		

1				
2				
3	429	Refe	References	
4 5				
6	430	1.	Torres A, Cilloniz C, Niederman MS, et al. Pneumonia. Nature Reviews Disease	
7	431		Primers 2021;7(1):1–28.	
8				
9	432	2.	GBD 2019 Diseases and Injuries Collaborators. Global burden of 369 diseases and	
10	433		injuries in 204 countries and territories, 1990-2019: a systematic analysis for the	
12	434		Global Burden of Disease Study 2019. <i>Lancet</i> 2020;396(10258):1204–1222.	
13	425	2		
14	435	3.	Metay JP, Waterer GW, Long AC, et al. Diagnosis and Treatment of Adults with	
15 16	430		Community-acquired Pneumonia. An Official Clinical Practice Guideline of the	
17	43/		American Thoracic Society and Infectious Diseases Society of America. Am J Respir	
18	438		Crit Care Mea 2019;200(7):e45–e67.	
19	130	Л	Fine ML Auble TE Yealy DM at al. A prediction rule to identify low risk nationts	
20	439	4.	with community acquired proumonia. N Engl 1 Med 1007:226(4):242, 250	
21	440		with community-acquired priedmonia. W Engl J Med 1997, 530(4).245–250.	
23	441	5	Lim WS Lewis S Macfarlane IT Severity prediction rules in community acquired	
24	442	5.	nneumonia: a validation study. <i>Thorax</i> 2000:55(3):219–223	
25	112			
26 27	443	6.	Peiffer-Smadja N, Rawson TM, Ahmad R, et al. Machine learning for clinical	
27	444		decision support in infectious diseases: a narrative review of current applications.	
29	445		<i>Clin Microbiol Infect</i> 2020;26(5):584–595.	
30				
31	446	7.	Garcia-Vidal C, Sanjuan G, Puerta-Alcalde P, Moreno-García E, Soriano A. Artificial	
32	447		intelligence to support clinical decision-making processes. EBioMedicine	
34	448		2019;46:27–29.	
35				
36	449	8.	Gonem S, Janssens W, Das N, Topalovic M. Applications of artificial intelligence	
37	450		and machine learning in respiratory medicine. <i>Thorax</i> 2020;75(8):695–701.	
38		-		
39 40	451	9.	Kang SY, Cha WC, Yoo J, et al. Predicting 30-day mortality of patients with	
41	452		pneumonia in an emergency department setting using machine-learning models.	
42	453		Clin Exp Emerg Med 2020;7(3):197–205.	
43	151	10	Churchite M. Cillégie C. Duerte Alcelde D. et al. Car. Artificial Intelligence Incorrecte	
44 45	454	10.	Chumbita IVI, Cilioniz C, Puerta-Alcaide P, et al. Can Artificial Intelligence Improve	
46	455		the Management of Pheumonia. J Clin Med 2020;9(1):248.	
47	156	11	Ward L. Andreassen S. Astrun II. Rahmani Z. Fantini M. Samhri V. Clinical, vs.	
48	450	11.	model based selection of patients suspected of sensis for direct from blood ranid	
49	457		diagnostics in the emergency department: a retrespective study. Fur I Clin	
50 51	450		Microbiol Infact Dic 2010-22(2):1515-1522	
52	439		Wilciobiol Hiject Dis 2019,38(8).1515–1522.	
53	460	12	Ward L Møller IK Fliakim-Raz N Andreassen S Prediction of Bacteraemia and of	
54	461	± - •	30-day Mortality Among Patients with Suspected Infection using a CPN Model of	
55	462		Systemic Inflammation, IFAC-PanersOnLine 2018:116–121	
20 57	102			
58				
59				
60				
CHEST

3	463	13.	Andreassen S, Møller JK, Eliakim-Raz N, Lisby G, Ward L. A comparison of
4	464		predictors for mortality and bacteraemia in patients suspected of infection. BMC
5	465		Infect Dis 2021;21(1):864.
7			
8	466	14.	Leisman DE, Harhay MO, Lederer DJ, et al. Development and Reporting of
9	467		Prediction Models: Guidance for Authors From Editors of Respiratory, Sleep, and
10	468		Critical Care Journals Crit Care Med 2020:48(5):623-633
11	-00		
12	469	15	Torres A Barberán I Falguera M et al [Multidisciplinary guidelines for the
13	402	15.	management of community acquired anoumonia] Med Clin (Parc)
14	470		
15 16	4/1		2013;140(5):223.01-223.019.
10	470	10	Cincer M. Deutechman CC. Courseur CM/ et al. The Third International Concernsus
18	472	10.	Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus
19	4/3		Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA 2016;315(8):801–810.
20	477.4	47	
21	4/4	17.	Logan Morgan Ward. Gradation of the Severity of Sepsis: Learning in a Causal
22	475		Probabilistic Network. 2016;55.
23			
24	476	18.	Vincent JL, Moreno R, Takala J, et al. The SOFA (Sepsis-related Organ Failure
25	477		Assessment) score to describe organ dysfunction/failure. On behalf of the
20	478		Working Group on Sepsis-Related Problems of the European Society of Intensive
28	479		Care Medicine. Intensive Care Med 1996;22(7):707–710.
29			
30	480	19.	ARDS Definition Task Force, Ranieri VM, Rubenfeld GD, et al. Acute respiratory
31	481		distress syndrome: the Berlin Definition. JAMA 2012;307(23):2526–2533.
32			
33	482	20.	Rhodes A, Evans LE, Alhazzani W, et al. Surviving Sepsis Campaign: International
34 25	483		Guidelines for Management of Sepsis and Septic Shock: 2016. Intensive Care Med
36	484		2017:43(3):304–377
37			
38	485	21.	Bellomo R. Ronco C. Kellum JA. Mehta RL. Palevsky P. Acute Dialysis Quality
39	486		Initiative workgroup, Acute renal failure - definition, outcome measures, animal
40	487		models fluid therapy and information technology needs: the Second
41	407		International Concensus Conference of the Acute Dialycis Quality Initiative (ADQI)
42	400		International Consensus Conference of the Acute Dialysis Quality Initiative (ADQI)
43	489		Group. Crit Care 2004;8(4):R204-212.
44 45	400	22	Chan Andresson, Leonard Leiberisi, Missi Davi, Anders D. Nielsen, Aline
46	490	22.	Steen Andreassen, Leonard Leibovici, Micai Paul, Anders D. Nielsen, Alina
47	491		Zalounina, Leif E. Kristensen, Karsten Falborg, Brian Kristensen, Uwe Frank, Henrik
48	492		C. Schønheyder. A Probabilistic Network for Fusion of Data and Knowledge in
49	493		Clinical Microbiology [Internet]. springerprofessional.de. 2005 [cited 2022 Mar
50	494		22];Available from: https://www.springerprofessional.de/en/a-probabilistic-
51	495		network-for-fusion-of-data-and-knowledge-in-clin/1020738
52			
53	496	23.	Lauritzen SL, Spiegelhalter DJ. Local Computations with Probabilities on Graphical
55	497		Structures and Their Application to Expert Systems. <i>Journal of the Royal Statistical</i>
56	498		Society: Series B (Methodological) 1988:50(2):157–194.
57			, ,
58	499	24.	Cole TJ. Applied logistic regression. D. W. Hosmer and S. Lemeshow. Wilev. New
59	500		York, 1989. Statistics in Medicine 1989:10(7):1162–1163
60	200		

60

CHEST

2 3 4	501 502	25.	Altman DG, Bland JM. Diagnostic tests 3: receiver operating characteristic plots. <i>BMJ</i> 1994;309(6948):188.
6 7 8	503 504	26.	Collins GS, Moons KGM. Reporting of artificial intelligence prediction models. <i>Lancet</i> 2019;393(10181):1577–1579.
9 10 11 12	505 506	27.	Mekov E, Miravitlles M, Petkov R. Artificial intelligence and machine learning in respiratory medicine. <i>Expert Rev Respir Med</i> 2020;14(6):559–564.
13 14 15	507 508	28.	He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. <i>Nat Med</i> 2019;25(1):30–36.
16 17 18 19 20	509 510 511	29.	Menéndez R, Cilloniz C, España PP, et al. Community-Acquired Pneumonia. Spanish Society of Pulmonology and Thoracic Surgery (SEPAR) Guidelines. 2020 Update. <i>Arch Bronconeumol</i> 2020;56 Suppl 1:1–10.
21 22 23 24 25	512 513 514	30.	Lim WS, Baudouin SV, George RC, et al. BTS guidelines for the management of community acquired pneumonia in adults: update 2009. <i>Thorax</i> 2009;64 Suppl 3:iii1-55.
26 27 28 29	515 516 517	31.	Loke YK, Kwok CS, Niruban A, Myint PK. Value of severity scales in predicting mortality from community-acquired pneumonia: systematic review and meta-analysis. <i>Thorax</i> 2010;65(10):884–890.
30 31 32 33 34	518 519 520	32.	Jiang J, Yang J, Jin Y, Cao J, Lu Y. Role of qSOFA in predicting mortality of pneumonia: A systematic review and meta-analysis. <i>Medicine (Baltimore)</i> 2018;97(40):e12634.
35 36 37 38 20	521 522 523	33.	Ranzani OT, Prina E, Menéndez R, et al. New Sepsis Definition (Sepsis-3) and Community-acquired Pneumonia Mortality. A Validation and Clinical Decision-Making Study. <i>Am J Respir Crit Care Med</i> 2017;196(10):1287–1297.
40 41 42	524 525	34.	Jones BE, Jones J, Bewick T, et al. CURB-65 pneumonia severity assessment adapted for electronic decision support. <i>Chest</i> 2011;140(1):156–163.
43 44	526		
43 46 47 48 49 50 51 52 53 54 55 56 57 58 59	527		

528 Table 1. Variables used in scoring systems

Variable	SeF	SeF-ML	PSI	SOFA	qSOFA	CURB-65
Demography						
Age	Yes	Yes	Yes			Yes
Sex			Yes			
Nursing home resident			Yes			
Comorbidities						
Neoplastic disease			Yes			
Liver disease			Yes			
Congestive heart failure			Yes			
Cerebrovascular disease			Yes			
Renal disease			Yes			
Vitals						
Temperature	Yes	Yes	Yes			
Chills	Yes	Yes				
Systolic Blood Pressure			Yes		Yes	Yes
Diastolic Blood Pressure						Yes
Mean Arterial Pressure	Yes	Yes		Yes		
Heart Rate	Yes	Yes	Yes			
Respiratory Rate		Yes	Yes		Yes	Yes
Mental Status	Yes	Yes	Yes	Yes, GCS	Yes	Yes
SpO2		Yes				
Labs						
CRP	Yes	Yes				
Creatinine	Yes	Yes		Yes		
Albumin	Yes*	Yes*				
Lactate	Yes*	Yes*				
Platelets	Yes	Yes		Yes		
Neutrophils%	Yes	Yes				
Bilirubin		Yes		Yes		
BUN or Urea		Yes	Yes			Yes
Sodium			Yes			
Glucose			Yes			
рН		Yes	Yes			

Page 75 of 81

	Variable	SeF	SeF-MI	PSI	SOFA	aSOFA	CURB-6
	Hematocrit	501	Yes	Yes		40017	
	PaO2		Yes	Yes	Yes		
	FiO2		Yes		Yes		
	Bicarbonate		Yes				
	Leukocytes		Yes				
	Radiology						
	Chest x-ray			Yes			
	Treatment/devices						
	Mechanical ventilation				Yes		
	Pressors				Yes		
529	Abbreviations: Glasgow	coma scale (GCS); * not p	art of st	andard tes	ting for CAP	patients
530	at the emergency depa	artment (ED)				
531	0 , 1		1				
532							
533							

534	Table 2. Clinical characteristics of the studied cohorts

	Derivation	Validation	
	cohort	cohort	
Variable	(N = 4,531)	(N = 1,034)	P-value
Age, median (Q1; Q3), years	73 (55; 82)	72 (59; 81)	0.650
Male sex, n (%)	2,708 (60)	642 (62)	0.176
Previous antibiotic, n (%)	1,057 (25)	338 (33)	<0.001
Influenza vaccine, n (%)	1,770 (44)	438 (42)	0.350
Pneumococcal vaccine, n (%)	695 (17)	79 (8)	<0.001
Previous inhaled corticosteroids, n (%)	818 (18)	-	-
Previous systemic corticosteroids, n (%)	180 (4)	-	-
Previous episode of pneumonia (last year),			
n (%)	630 (14)	-	-
Comorbidities, n (%) ª	2,982 (66)	731 (71)	0.005
Chronic respiratory disease	1,729 (39)	325 (31)	<0.001
Chronic cardiovascular disease	635 (14)	347 (34)	<0.001
Diabetes mellitus	914 (20)	288 (28)	<0.001
Neurologic disease	836 (19)	176 (17)	0.137
Chronic renal disease	318 (7)	139 (13)	<0.001
Chronic liver disease	212 (5)	38 (4)	0.146
Nursing home, n (%)	352 (8)	43 (4)	<0.001
Confusion, n (%)	820 (18)	77 (7)	<0.001
Respiratory rate, median (Q1; Q3),			
breaths/min	24 (20; 30)	18 (16; 24)	<0.001
Heart rate, median (Q1; Q3), beats/min	97 (84; 110)	97 (85; 110)	0.685
Systolic blood pressure, median (Q1; Q3),			
mmHg	129 (112; 148)	134 (118; 152)	<0.001
Diastolic blood pressure, median (Q1; Q3),			
mmHg	72 (64; 80)	70 (61; 80)	0.007
Creatinine, median (Q1; Q3), mg/dL	1.0 (0.8; 1.4)	1.0 (0.8; 1.4)	0.003
Glucose level, median (Q1; Q3), mg/dL	124 (105; 157)	-	-

	Derivation	Validation	
	cohort	cohort	
Variable	(N = 4,531)	(N = 1,034)	P-valu
PaO ₂ /FiO ₂ , median (Q1; Q3)	281 (238; 327)	271 (238; 311)	0.01
PSI score, median (Q1; Q3)	98 (74; 123)	86 (66; 105)	<0.00
Severe CAP, n (%)	868 (26)	144 (14)	<0.00
Bacteraemia, n (%) ^b	390 (12)	54 (8)	0.00
Appropriate empiric treatment, n (%)	2,844 (96)	-	-
Length of hospital stay, median (Q1; Q3)	,		
days	7 (4; 11)	6 (5; 9)	0.26
Mechanical ventilation, n (%)	375 (10)	26 (3)	<0.00
Non-invasive	169 (4)	-	-
Invasive	206 (5) ^c	-	-
In-hospital mortality, n (%)	272 (6)	35 (3)	<0.0
30-day mortality, n (%)	293 (7)	43 (4)	0.00

Abbreviations: CAP indicates community-acquired pneumonia; PSI, pneumonia severity index; Q1, first quartile; Q3, third quartile. Percentages calculated on nonmissing data. ^a May have >1 comorbid condition. ^b Calculated only for patients with blood samples (3,206 in the derivation cohort and 696 in the validation cohort). ^c Patients who initially received non-invasive ventilation yet needed intubation subsequently were included in the invasive mechanical ventilation group.

542 Table 3. Discriminatory ability of SeF-ML vs. other scoring systems for 30-day

543 mortality in the derivation cohort

Availability		
·····,	AUC (95% CI)	SeF-ML)
99.3%	0.801 (0.777-0.825)	-
		-
	0.808 (0.776-0.839)	
		-
	0.804 (0.780-0.829)	
		-
	0.814 (0.788-0.839)	
57.7%	0.799 (0.768-0.830)	0.58
91.4%	0.671 (0.638-0.704)	<0.001
100%	0.642 (0.611-0.673)	<0.001
82.5%	0.759 (0.732-0.786)	<0.001
	57.7% 91.4% 100% 82.5% s area under the	0.808 (0.776-0.839) 0.804 (0.780-0.829) 0.814 (0.788-0.839) 57.7% 0.799 (0.768-0.830) 91.4% 0.671 (0.638-0.704) 100% 0.642 (0.611-0.673) 82.5% 0.759 (0.732-0.786) s area under the receiver operating charac

ScholarOne - http://mchelp.manuscriptcentral.com/gethelpnow/index.html - (434) 964-4100

2 3 4 5 6 7	548 549
5 6 7 8 9 10 11 2 3 4 5 6 7 8 9 10 11 2 3 14 5 6 7 8 9 10 11 2 3 3 4 5 6 7 8 9 0 12 2 3 2 4 5 6 7 8 9 30 31 2 3 34 5 6 7 8 9 0 11 2 2 3 2 4 5 6 7 8 9 0 11 2 2 3 2 4 5 6 7 8 9 30 31 2 3 34 5 6 7 8 9 0 11 2 2 3 2 4 5 6 7 8 9 30 31 2 3 3 4 5 6 7 8 9 0 11 2 2 3 2 4 5 6 7 8 9 30 31 2 3 3 4 5 6 7 8 9 30 31 2 3 3 4 5 6 7 8 9 30 1 2 2 3 4 5 6 7 8 9 30 31 2 3 3 4 5 6 7 8 9 9 0 1 2 2 3 4 5 6 7 8 9 9 0 1 2 2 3 4 5 6 7 8 9 9 0 1 2 2 3 4 5 6 7 8 9 9 0 1 2 2 3 3 4 5 6 7 8 9 9 0 1 2 2 3 4 5 6 7 8 9 9 0 1 2 2 3 3 4 5 6 7 8 9 9 0 1 2 3 3 4 5 5 6 7 8 9 9 0 1 2 3 3 4 5 5 6 7 8 9 9 0 1 2 3 3 4 5 5 6 7 8 9 9 0 1 2 3 3 4 5 5 6 7 8 9 9 0 1 2 3 3 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	550 551 552 553
57 58 59 60	

8 Table 4. Discriminatory ability of SeF-ML vs. other scoring systems for 30-day

9 mortality in the validation cohort

Model/Score	Availability	AUC (95% CI)	P-value (vs. SeF-ML)
SeF-ML	100%	0.826 (0.753-0.899)	-
PSI (raw score)	100%	0.830 (0.753-0.90)	0.92
CURB-65	100%	0.764 (0.694-0.834)	0.03
SOFA	23.1%	0.728 (0.588-0.869)	0.85 ª
SOFA-imputed ^b	100%	0.771 (0.706-0.836)	0.14
qSOFA	98.3%	0.729 (0.653-0.804)	0.005 ^a

.ne re everity inc culated assumi. 0 Abbreviations: AUC indicates area under the receiver operating characteristic curve; CI,

1 confidence interval; PSI, pneumonia severity index. ^a Compared only for patients with

52 complete SOFA/qSOFA score. ^b Calculated assuming missing values were normal. CHEST

- 554 Figure Legends
- 555 Figure 1. Model performance for 30-day mortality prediction in the derivation cohort.
- 556 A: ROC curves B: Calibration curves for SeF models
- 557 Figure 2. Survival curves for patients stratified according to Pmort in the derivation
- 11 558 cohort
- 13 559 Figure 3. Survival curves stratified by SF risk group, set together according to CURB-
 - 560 65/PSI score in the derivation cohort
 - 561 Figure 4. Model performance for 30-day mortality prediction in the validation cohort.
 - 562 A: ROC curves B: Calibration curves for SeF models

1 2		
3	564	Take-home Points:
5	565	Study Questions: Did a CPN model could predict mortality in patients with CAP better
6 7	566	than the frequent severity scores?
8 9	567	Results: SeF-ML performance at predicting 30-day mortality was overall significantly
10 11	568	better than that of existing CAP-specific scores.
12 13	569	Interpretation: Our results showed that SeF-ML shows potential for improving
14 15	570	mortality prediction amongst patients with CAP using structured health data.
16 17		
18		
20		
21		
23 24		
25 26		
27 28		
29 30		
31 32		
33 34		
35		
37		
38 39		
40 41		
42 43		
44 45		
46 47		
48 49		
50 51		
52 53		
55 54		
55 56		
57 58		
59 60		