Original research

Early prognostication of COVID-19 to guide hospitalisation versus outpatient monitoring using a point-of-test risk prediction score

Felix Chua ,[1,2] Rama Vancheeswaran,[3] Adrian Draper,[4] Tejal Vaghela,[5] Matthew Knight,[3] Rahul Mogal,[3] Jaswinder Singh,[6] Lisa G Spencer ,[7] Erica Thwaite,[8] Harry Mitchell,[3] Sam Calmonson,[3] Noor Mahdi,[3] Shershah Assadullah,[3] Matthew Leung,[3] Aisling O'Neill,[3] Chhaya Popat,[3] Radhika Kumar,[3] Thomas Humphries,[7] Rebecca Talbutt,[7] Sarika Raghunath,[7] Philip L Molyneaux ,[1,2] Miriam Schechter,[5] Jeremy Lowe,[5] Andrew Barlow[3]

► Prepublication history and additional material is published online only. To view please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2020-216425).

For numbered affiliations see end of article.

Correspondence to
Dr Felix Chua, Interstitial Lung Disease Unit, Department of Respiratory Medicine, Royal Brompton and Harefield NHS Foundation Trust, London SW3 6NP, UK; f.chua@rbht.nhs.uk

FC, RV and AD contributed equally.

Received 19 October 2020
Revised 17 January 2021
Accepted 18 January 2021

To cite: Chua F, Vancheeswaran R, Draper A, *et al*. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thoraxjnl-2020-216425

ABSTRACT

Introduction Risk factors of adverse outcomes in COVID-19 are defined but stratification of mortality using non-laboratory measured scores, particularly at the time of prehospital SARS-CoV-2 testing, is lacking.

Methods Multivariate regression with bootstrapping was used to identify independent mortality predictors in patients admitted to an acute hospital with a confirmed diagnosis of COVID-19. Predictions were externally validated in a large random sample of the ISARIC cohort (N=14 231) and a smaller cohort from Aintree (N=290).

Results 983 patients (median age 70, IQR 53–83; in-hospital mortality 29.9%) were recruited over an 11-week study period. Through sequential modelling, a five-predictor score termed SOARS (SpO2, Obesity, Age, Respiratory rate, Stroke history) was developed to correlate COVID-19 severity across low, moderate and high strata of mortality risk. The score discriminated well for in-hospital death, with area under the receiver operating characteristic values of 0.82, 0.80 and 0.74 in the derivation, Aintree and ISARIC validation cohorts, respectively. Its predictive accuracy (calibration) in both external cohorts was consistently higher in patients with milder disease (SOARS 0–1), the same individuals who could be identified for safe outpatient monitoring. Prediction of a non-fatal outcome in this group was accompanied by high score sensitivity (99.2%) and negative predictive value (95.9%).

Conclusion The SOARS score uses constitutive and readily assessed individual characteristics to predict the risk of COVID-19 death. Deployment of the score could potentially inform clinical triage in preadmission settings where expedient and reliable decision-making is key. The resurgence of SARS-CoV-2 transmission provides an opportunity to further validate and update its performance.

Key messages

What is the key question?
► Can patients with COVID-19 be risk stratified in the prehospital setting without laboratory-measured data?

What is the bottom line?
► A five-predictor risk prediction score (SOARS) based on demographic and clinical characteristics can quickly and reliably identify COVID-19-positive patients who have a low probability of mortality for outpatient monitoring and management.

Why read on?
► Information from the prognostication of SARS-CoV-2-infected individuals early in their illness can be used to guide clinical decision-making with respect to the level of subsequent care.

INTRODUCTION

Rapid and accurate prediction of the probability of adverse clinical outcomes is central to the management of global outbreaks of infection.[1–3] Stratification by predicted risk, most commonly for death, can support clinical judgement and potentially assist clinicians in community settings to decide how urgently to refer patients to hospital. Used appropriately, predictive scores can also help inform treatment-related decision-making. The pandemic caused by SARS-CoV-2 lends itself to predictive modelling by having a large at-risk population and a high adverse event rate including death.[4]

Although the recent incidence of COVID-19 has decreased in some parts of the world at the time of writing, many countries are already experiencing a 'second wave' of new cases.[5–8] An increase in incident cases in many localities is already evident in the UK. It is widely anticipated that viral transmission will continue to surge in the months ahead, particularly with the onset of winter in the northern hemisphere. Not all patients infected with SARS-CoV-2 will require hospitalisation but even among those who initially experience mild symptoms, a sizeable proportion remain at risk of subsequent life-threatening clinical decline. The availability of a practical prehospital predictive tool to triage patients for safe discharge to an outpatient (virtual) monitoring system versus direct

admission to hospital for observation or treatment would be highly advantageous.

Reliable prediction tools to differentiate between levels and sites of clinical care already exist and have been successfully implemented in prehospital practice. For example, both the CURB-65 and the CRB-65 scoring systems for the assessment of community-acquired pneumonia include recommendations for out-of-hospital care.[9 10] Recent research by the ISARIC-4C Consortium has provided an accurate tool to similarly prognosticate for COVID-19-attributed death in hospitalised patients but its reliance on laboratory-measured indices limits its applicability outside the institutional environment.[11] Prognostic evaluation of individuals with suspected SARS-CoV-2 infection at the time of diagnostic testing is potentially achievable but has yet to be examined.

Our objective was to develop and evaluate an easy-to-apply and accurate prognostic score to predict mortality and aid early clinical decision-making by identifying patients infected with SARS-CoV-2 who might benefit from an urgent hospital assessment. To develop the initial risk score, we used multivariate logistic regression to explore the relationships between a large panel of candidate predictors and COVID-19 death. Iterative modelling resulted in a pragmatic predictive score based on five widely available patient variables. The scoring of patients against these selected predictors permitted three distinct risk classes to be defined. The performance of the score was then assessed against two validation cohorts—a large subgroup of the ISARIC study patients and a smaller single-hospital cohort, the latter to better reflect local population characteristics and practice.

METHODS

Study design and characteristics of the derivation cohort

All individuals aged 18 or older who tested positive for SARS-CoV-2 nucleic acid by real-time reverse transcriptase PCR between 1 March and 16 May 2020 after presenting to the emergency department (ED) at Watford Hospital, West Hertfordshire NHS Hospitals Trust were prospectively recruited. Baseline clinical characteristics and investigation results were collected according to a prespecified protocol. Patients were either referred to the virtual hospital (VH) for outpatient monitoring or admitted to a medical ward.

Laboratory, physiologic and radiographic data

All laboratory tests were performed as part of routine clinical care. Nasopharyngeal mucosal swabs for rRT-PCR were couriered to the regional UK Public Health England laboratory. Baseline vital observations included all the parameters recommended by the National Early Warning Score.[12] Chest radiographs acquired in ED were collated and scored at the end of the recruitment period.

Location and level of care

After presentation, patients who were clinically judged to have mild illness were referred to the VH for subsequent monitoring. To avoid missing early clinical deterioration in the post-assessment period, they were observed for up to 24 hours in hospital. Patients who remained admitted after the first 24 hours but who did not require additional respiratory support beyond wall-based oxygen were managed on designated medical wards. Where clinically indicated, continuous positive airway pressure (CPAP) was provided on such wards or on the intensive care unit (ICU); intubation and mechanical ventilation were undertaken on the ICU.

Identifying predictors of death in the derivation cohort

The primary outcome of the study was in-hospital death. TRIPOD (Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis) recommendations were followed for multivariate model evaluation and reporting.[13] Seventy-five baseline clinical and non-clinical variables were initially collected based on their reported association with COVID-19 and analysed by univariate and multivariate logistic regression with bootstrap resampling.[14 15] Of these, variables with numerically small ORs or a p value of >0.05 were not included in the final analysis. Candidate predictors of death were assessed for potential clustering effects and missing at random values were addressed by multiple imputation with chained equations (MICE),[16] with 10–20 random draws to account for data variability.

Development and external validation of the clinical risk score

The large external cohort comprised a randomly selected subpopulation of the ISARIC 4C derivation population (N=20 000 provided; 14 231 with complete data for scoring). The primary data of these individuals were submitted by 260 hospitals across England, Scotland and Wales to the prospective ISARIC WHO Clinical Characterization Protocol UK (CCP-UK) study.[11] We also tested our score against a smaller population of SARS-CoV-2-positive cases from Aintree Hospital, Liverpool (N=303 provided; N=290 with complete data for scoring) as a single-setting validation control.

In the initial stages, a preliminary score comprising 12 independent predictors of death, including care home residency, was developed; to enable external validation against the ISARIC cohort (which did not include residential data), care home status was excluded as a variable to yield an 11-predictor score. Its ability to discriminate for in-hospital mortality was assessed by the area under the receiver operating characteristic (AUROC). From this score, a condensed version comprising five clinical predictors was developed for prehospital application. Mortality cut-points at each risk level were assessed to define mild, moderate and high risk classes, followed by determination of positive and negative predictive values, as well as sensitivity and specificity thresholds. Model performance was further assessed by calibration using a graphical representation of the Hosmer-Lemeshow 'goodness-of-fit' test to depict agreement between the expected (predicted) and observed (actual) outcome across the entire COVID-19 severity range in both validation cohorts.[17] The summary relationship between the dependent variable (death) and different levels of disease severity in the external ISARIC population was expressed as McFadden's R^2.[18]

Statistical analysis

Categorical variables were expressed as frequency (%), with significance determined by the χ^2 test. Continuous variables were expressed as median (IQR) or mean (SD) and analysed by the t-test, Kruskal-Wallis or Mann-Whitney U test, as appropriate. ORs were assessed as unadjusted and adjusted values with respect to in-hospital death, the latter determined by multivariate regression with bootstrapping of 1000 resamples. We used this method as internal validation to improve statistical inference by deriving a better estimate of the sampling distribution. Bootstrapping involves randomly drawing repeat samples from the core dataset to calculate SEs and CIs for the final regression analysis. MICE was used to generate valid estimates of randomly missing values in the derivation model. A p value of <0.05 was considered statistically significant. All statistical

analyses including risk modelling calculations were performed using STATA, V.16 (Stata, Texas, USA).

RESULTS

Baseline demographic and clinical characteristics of the derivation cohort

Nine hundred eighty-three patients (52.5% male) confirmed as SARS-CoV-2 rRT-PCR positive were recruited over the 11-week study period. Five patients remained in hospital at the time of data cut-off on 31 May 2020. The median age of the cohort was 70 (IQR 53–83; range 23–99); median age was lowest in the virtual hospital pathway (53; IQR 43–67) and highest among hospitalised patients who did not receive CPAP (77; IQR 61–86) (p<0.001) (online supplemental appendix 1).

The most common comorbidities were hypertension (48.4%), pulmonary disease (30.0%), cardiac disease (26.6%), diabetes mellitus (23.6%), chronic kidney disease (CKD; 20.0%) and dementia (15.4%). Obesity, defined as body mass index (BMI)>30, was present in 24.7% (243) of the cohort and is associated with an unadjusted OR for death of 1.40 (95% CI 1.18 to 2.68, p<0.05).

Overall, 294 out of 983 (29.9%) patients died in hospital, the vast majority (97.3%) aged 50 or older. The mortality rates of different age brackets in the cohort (compared with the ISARIC and Aintree validation cohorts) are shown in online supplemental appendix 2. The univariate OR for death increased with rising age, and was 14.86 (95% CI 6.89 to 32.04) for those aged 70–79 and 20.87 (95% CI 9.93 to 43.86) for those aged 80 or older (table 1). When stratified by maximal levels of care, mortality rate was lowest in the VH (1.8%) and highest in the ICU group (62.1%) (p<0.001) (online supplemental appendix 3).

White Caucasian ethnicity constituted 77.3% (760/983) of the whole cohort, while Asian, Black and other minor ethnicities (BAME) represented 16.5%, 4.5% and 1.7%, respectively. Overall, white ethnicity was associated with the highest proportion of non-survivors (85.0%); in comparison, patients of Asian (OR 0.57, 95% CI 0.38 to 0.85, p<0.01) or black (OR 0.39, 95% CI 0.17 to 0.88, p<0.05) background in this cohort had lower ORs for death from COVID-19. The proportion of non-survivors within each ethnic group was also highest in white (32.9%), followed by Asian (21.6%), Black (15.9%) and other minority groups (17.6%) (p<0.01). Of note, white patients were significantly older by median age (74, IQR 58–85) compared with Asian (57, IQR 46–71; p<0.0001) or black (58, IQR 50–72; p<0.001).

Care home residency (204/983; 20.8%) was more common among non-survivors (p<0.001) and was associated with an unadjusted OR for death of 3.14 (95% CI 2.28 to 4.32, p<0.001). Based on data from 644 patients aged 65 or older, the univariate OR of frailty for death was 2.52 (95% CI 1.73 to 3.69; p<0.001) in those with a group 1 frailty score and 2.56 (95% CI 1.62 to 4.06; p<0.001) in those with a group 2 frailty score (table 1).

The median time from symptom onset to presentation for the derivation cohort was 6 days (IQR 2.0–11.0), with no difference between survivors and non-survivors. The four most common reported symptoms were fever (61%), breathlessness (57.9%), cough (52.8%) and myalgia (21.7%). Tachypnoea (respiratory rate>24/minute) and hypoxia (SpO_2≤92% on ambient air) were evident in 35.9% and 31.4% of patients, respectively, and is associated with crude OR for death of 2.15 (95% CI 1.63 to 2.84, p<0.001) and 3.74 (95% CI 2.73 to 5.12, p<0.001), respectively. C reactive protein>50 mg/L was more frequently documented in non-survivors (76.7% vs 58.5%; p<0.001) and associated with an unadjusted OR for mortality of 2.40 (95% CI 1.73 to 3.32, p<0.001). Lymphopenia was similarly more common in non-survivors (44.7% vs 29.1%; p<0.001), with an unadjusted OR for death of 1.97 (95% CI 1.47 to 2.64; p<0.001). A baseline chest radiograph (CXR) was available in 91% (895/983) patients; abnormalities in ≥4 radiographic zones were evident in 338 (37.8%) of cases and was associated with increased mortality on univariate analysis (OR 1.89, 95% CI 1.42 to 2.52, p<0.001).

Multivariate regression for independent risk factors of mortality

The bootstrapped multivariate regression analysis included the whole derivation cohort of 983 patients comprising 689 (70.1%) survivors and 294 (29.9%) non-survivors with complete or multiply imputed values for data. Older age, CKD stage 5, baseline hypoxia, elevated BMI, tachypnoea, leucocytosis and a history of stroke were identified as the strongest independent predictors of mortality (table 1). In particular, age ≥70 had the highest OR for death over the other individual variables. Following multivariate regression, care home residency no longer independently predicted mortality.

Iterative modelling to construct an 11-predictor and 5-predictor risk prediction scores

ORs and their respective p values from the multivariate logistic regression model were used to identify constituent variables for developing risk prediction scores. We began by constructing an initial 11-predictor score that ranged from 1 to 18 points, using data from 770 patients in the derivation cohort with a complete dataset (online supplemental appendix 4). The lowest score of 1 point reflected the KDIGO (kidney disease improving global outcomes) categorisation of CKD.[19] In this score, the correlation between increasing COVID-19 severity and in-hospital mortality followed a linear dose–response relationship, particularly between a score of 3 (below which no deaths occurred) and 12 (above which no patient survived), and an accuracy (AUROC) of 0.84 (figure 1). A shorter five-predictor score based solely on clinical parameters and scaled from 0 to 8 points was also developed (table 2; figure 1). This short score, abbreviated as **SOARS** (SpO2, Obesity, Age, Respiratory rate, Stroke history), demonstrated an AUROC of 0.82 and was retained for further evaluation as a practical prehospital risk stratification tool.

External validation and performance of the SOARS score

The performance metrics of the long 11-predictor and 5-predictor (SOARS) scores were assessed by their ability to discriminate for in-hospital mortality against both the ISARIC and Aintree validation cohorts (table 3). The longer score showed higher discrimination in the Aintree (AUROC 0.87) than the ISARIC cohort (0.77). The SOARS score had a slightly lower AUROC against both cohorts, namely, 0.80 (Aintree) and 0.74 (ISARIC). In comparison, the performance of other scores based solely on different cut-offs of age were associated with inferior discriminatory ability. Comparison of some of the main population parameters between the derivation and both external cohorts is shown in online supplemental appendix 5.

The mortality rate at each level of the SOARS score in the derivation and both validation cohorts are shown in table 4. For increased applicability, the SOARS score results were further categorised into three risk classes: low (SOARS 0–1), moderate (SOARS 2) and high (SOARS≥3) (table 5). Between 2.3% and

Table 1 Risk factors of mortality in the derivation cohort (N=983)

	Univariate OR (95% CI)	P value	Likelihood ratio χ^2 of five final predictors	Multivariate OR (95% CI)	P value
Age (years)			2177.27		
<50	1.00			1.00	
50–59	1.30 (1.28 to 7.02)	0.01		2.39 (1.49 to 3.84)	<0.001
60–69	6.10 (2.73 to 13.66)	<0.001		2.15 (1.32 to 3.50)	<0.01
70–79	14.86 (6.89 to 32.04)	<0.001		7.40 (4.67 to 11.74)	<0.001
≥80	20.87 (9.93 to 43.86)	<0.001		10.73 (6.82 to 16.90)	<0.001
Male sex (vs female)	1.23 (0.94 to 1.62)	0.137		–	
Ever smoked (vs never smoked)	2.14 (1.52 to 3.01)	<0.001		2.51 (1.91 to 3.29)	<0.001
Ethnicity (vs White)					
Asian	0.57 (0.38 to 0.85)	0.006		1.44 (0.71 to 2.94)	0.301
Black	0.39 (0.17 to 0.88)	0.024		0.51 (0.16 to 1.64)	0.257
Symptoms (vs none)					
Breathlessness	0.80 (0.59 to 1.07)	0.135		–	
Fever	0.84 (0.62 to 1.14)	0.26		–	
Cough	0.61 (0.45 to 0.82)	0.001		0.67 (0.41 to 1.09)	0.108
Myalgia	0.41 (0.28 to 0.63)	<0.001		0.79 (0.44 to 1.41)	0.423
Headache	0.51 (0.26 to 0.97)	0.039		1.70 (0.64 to 4.51)	0.286
Clinical parameters					
SpO_2 (≤92% on air)	3.74 (2.73 to 5.12)	<0.001	31.87	2.69 (1.80 to 4.01)	<0.001
Respiratory rate (>24/min)	2.15 (1.63 to 2.84)	<0.001	158.21	2.12 (1.35 to 3.32)	0.001
Systolic BP (≤90 mm Hg)	2.19 (0.96 to 5.03)	0.064		–	
BMI (>30)	1.40 (1.03 to 1.90)	0.033	11.13	2.18 (1.46 to 3.20)	<0.001
Frailty (vs not frail, CFS 0–4)					
1 (CFS 5–6)	2.53 (1.73 to 3.70)	<0.001		1.26 (0.82 to 1.96)	0.294
2 (CFS 7–9)	2.56 (1.62 to 4.06)	<0.001		1.28 (0.77 to 2.13)	0.341
Residency in care home (vs own home)	3.14 (2.28 to 4.32)	<0.001		1.38 (0.90 to 2.11)	0.137
Peripheral blood markers					
CRP (>50 mmol/L)	2.40 (1.73 to 3.32)	<0.001		1.42 (0.87 to 2.31)	0.16
Total white cell count					
≤4 × 109/L	0.93 (0.56 to 1.55)	0.79		–	
>11 × 109/L	2.04 (2.09 to 4.15)	<0.001		1.76 (1.18 to 2.61)	<0.01
Lymphocytes (<0.7 × 109/L)	1.97 (1.47 to 2.64)	<0.001		1.67 (1.17 to 2.37)	<0.01
Chronic kidney disease stage (vs 1, eGFR≥90 mL/min/1.73 m^2)	1.00				
2 eGFR 60–89	2.35 (1.53 to 3.62)	<0.001		0.91 (0.49 to 1.69)	0.769
3 eGFR 30–44 and 45–59	6.09 (3.90 to 9.50)	<0.001		1.46 (0.78 to 2.71)	0.234
4 eGFR 15–29	8.88 (4.82 to 16.34)	<0.001		1.78 (0.82 to 3.85)	0.145
5 eGFR<15	15.73 (6.85 to 36.14)	<0.001		3.82 (1.43 to 10.26)	<0.01
CXR (≥4 zones affected, vs no abnormal zones)	1.89 (1.42 to 2.52)	<0.001		1.73 (1.22 to 2.46)	<0.01
Medications (≥5 different)	2.59 (1.93 to 3.48)	<0.001		0.90 (0.53 to 1.52)	0.695
Co-morbidities (vs none)					
Dementia	3.53 (2.46 to 5.08)	<0.001		1.54 (0.97 to 2.44)	0.066
CVA/stroke	3.50 (2.32 to 5.28)	<0.001	44.98	1.92 (1.18 to 3.11)	<0.01
Cardiac disease	2.84 (2.13 to 3.80)	<0.001		1.15 (0.70 to 1.91)	0.579
Cancer	2.60 (1.77 to 3.83)	<0.001		1.85 (0.91 to 3.75)	0.091
Hypertension	2.46 (1.85 to 3.26)	<0.001		1.10 (0.65 to 1.88)	0.717
Diabetes mellitus	1.40 (1.02 to 1.91)	0.036		1.00 (0.58 to 1.75)	0.981
Anxiety/psychosis	1.20 (0.84 to 1.71)	0.318		–	
Lung disease	1.15 (0.86 to 1.55)	0.35		–	

Out of the starting 75 variables, those with numerically small ORs or a p value of >0.05 following univariate regression were not included in table 1.
BMI, body mass index; BP, blood pressure; CFS, Clinical Frailty Scale; CRP, c-reactive protein; CVA, cerebrovascular accident; CXR, chest radiograph; eGFR, estimated glomerular filtration rate.

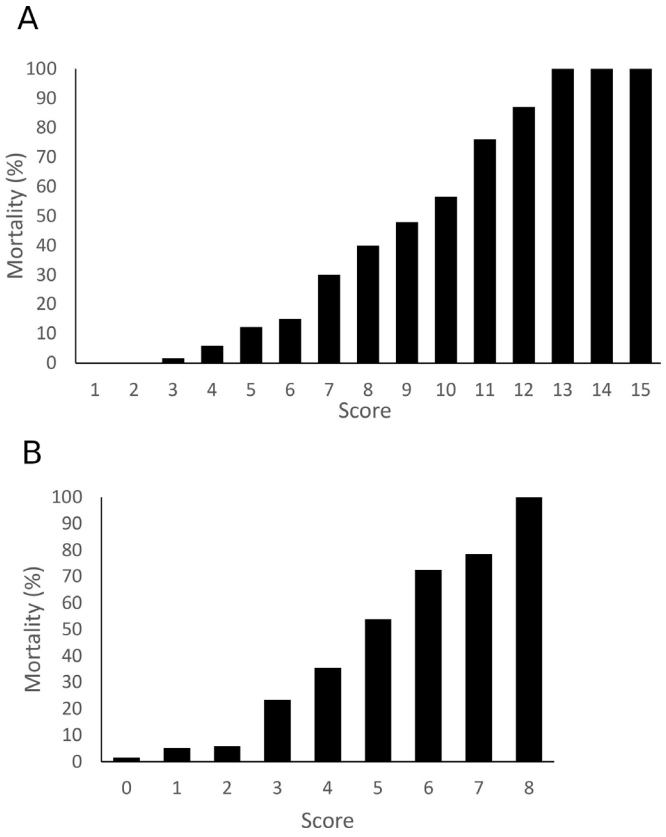

Figure 1 (A) In-patient mortality stratified according to the 11-predictor (SOARS) scores; (B) In-patient mortality stratified according to the 5-predictor (SOARS) scores.

Table 2 SOARS score (five predictors; range: 0–8 points) for predicting in-hospital COVID-19 death

Predictor	Points
SpO_2	
>92% on air	0
≤92% on air	1
Obesity (BMI>30)	
Absent	0
Present	1
Age (years)	
<50	0
50–59	1
60–69	2
70–79	3
≥80	4
Respiratory rate	
≤24/min	0
>24/min	1
Stroke/CVA	
Absent	0
Present	1

CVA, cerebrovascular accident; SOARS, SpO_2, Obesity, Age, Respiratory rate, Stroke history.

Table 3 Discriminatory performance (area under the receiver operating characteristic; AUROC) of different risk stratification models for predicting COVID-19 in-hospital mortality

Cohort	11-predictor	5-predictor (SOARS)	Age ≥80	Age 70–79	Age ≥50
Derivation	0.84	0.82	0.73	0.76	0.73
Aintree (validation)	0.87	0.80	0.78	0.69	0.57
ISARIC (validation)	0.77	0.74	0.68	0.70	0.63

SOARS, SpO_2, Obesity, Age, Respiratory rate, Stroke history.

3.2% of patients scoring 0 or 1 (low risk) in each of the three cohorts died due to COVID-19, whereas 2.3%–6.3% of those scoring 2 (moderate risk) failed to survive to discharge. Overall, 9 out of every 10 deaths in each of the derivation and validation cohorts had a SOARS score of 3 or greater. Within this broad range of increasing scores, the highest proportion of deaths was encountered at SOARS score 4 in both the derivation cohort and the ISARIC validation cohort (28.3% and 34.5%, respectively) and at SOARS score 5 in the Aintree validation cohort (30.9%).

Sensitivity thresholds calculated for the larger validation (ISARIC) cohort showed that the low risk class (SOARS 0–1) comprised 16.6% of patients with an in-hospital mortality of 5.4%, sensitivity of 97% and negative predictive value (NPV) of 94.6% (table 6). By comparison, 13.2% of patients were classified as moderate risk; their mortality rate was 14.5%. When combining the low and moderate risk groups (SOARS 0–2), the sensitivity reduces to 90.7% and the NPV to 90.5%. The high-risk group comprised 70.2% of the validation cohort who scored across a wide range of SOARS (scores 3–8). The specificity for a prediction of death was thus more variable, from 58.2% for a SOARS score of 3 to 99.8% at the other end of the scale when the SOARS score was 7. Only one patient scored 8; they survived to discharge.

Reliability of the risk estimates in the ISARIC cohort, modelled as calibration or goodness-of-fit between expected (predicted) and observed outcomes using SOARS, showed a calibration slope of 0.70, calibration-in-the-large (CiTL) 0.02 and an expected-to-observed (E:O) ratio of 0.990 (figure 2). Calibration was slightly improved in the Aintree cohort with a slope of 0.80, CiTL −0.16 and E:O ratio of 1.06. A LOWESS (locally weighted scatterplot smoothing algorithm) curve was generated to show differences between these outcomes in both cohorts. The plot characteristics suggested that the model, while demonstrating good concordance, had greater predictive accuracy in low-risk to moderate-risk patients whose predicted probability of mortality was under 40%. Conversely, overestimation of mortality risk was evident in patients in the high-risk group.

DISCUSSION

We show that prognostic evaluation of a small panel of baseline clinical and demographic characteristics of patients with COVID-19 enables their subsequent risk of in-hospital death to be quantified across three strata of risk. Findings were obtained by applying the five-predictor SOARS (SpO_2, Obesity, Age, Respiratory rate, Stroke history) score to a large random sample of the ISARIC cohort and a smaller single-hospital cohort from Aintree, both with individual-level data. Our objective was to enable risk stratification to be undertaken early, ideally prehospitalisation (eg, in the community), during the encounter

Table 4 Mortality in the derivation and validation cohorts at different levels of SOARS

	Derivation cohort (N=821; deaths=258)			Aintree validation cohort (N=290; deaths=94)			ISARIC validation cohort (N=14 231; deaths=4319)		
Score	*No of deaths (n/N)*	*Mortality rate (%)*	*Proportion of deaths (%)*	*No of deaths (n/N)*	*Mortality rate (%)*	*Proportion of deaths (%)*	*No of deaths (n/N)*	*Mortality rate (%)*	*Proportion of deaths (%)*
0	1/69	1.4%	0.4%	1/12	8.3%	1.1%	34/833	4.1%	0.8%
1	5/94	5.3%	1.9%	2/23	8.7%	2.1%	94/1529	6.1%	2.2%
2	6/102	5.9%	2.3%	3/37	8.1%	3.2%	273/1879	14.5%	6.3%
3	29/124	23.4%	11.2%	11/60	18.3%	11.7%	650/2577	25.2%	15.1%
4	73/206	35.4%	28.3%	24/72	33.3%	25.5%	1490/4013	37.1%	34.5%
5	63/117	53.9%	24.4%	29/48	60.4%	30.9%	1170/2402	48.7%	27.1%
6	58/80	72.5%	22.6%	16/30	53.3%	17.0%	571/939	60.8%	13.2%
7	22/28	78.6%	8.5%	6/6	100.0%	6.4%	36/58	62.1%	0.8%
8	1/1	100.0%	0.4%	2/2	100.0%	2.1%	0/1	0.0%	0.0%

The proportion of deaths at each score is defined as the number of deaths at that score divided by the total number of deaths in that particular cohort.
SOARS, SpO_2, Obesity, Age, Respiratory rate, Stroke history.

with COVID-19-suspected individuals. This role is not met by currently available prediction tools that rely on laboratory measurements.

The SOARS score discriminated well for COVID-19 mortality and its simplicity obviated the need for complex calculations. It also retained good predictive accuracy in two external validation cohorts, with performance metrics that were primarily reflected in its high negative predictive values for mortality among patients with the lowest risk scores (0 or 1). This characteristic is consistent with a high accuracy for predicting a non-fatal outcome in its key target group, namely, individuals with milder COVID-19.

Patients who score SOARS 0 or 1 could be discharged home with advice to re-establish urgent contact if their symptoms worsened. Patients stratified as moderate risk (SOARS 2) could be virtually monitored with a predefined plan for care escalation if specific thresholds relating to deteriorating symptoms or self-recorded SpO_2 were triggered. Patients in the high-risk class (score ≥3) are highly likely to be symptomatic and would, in all probability, be referred directly to the ED for hospital-based management. Thus, the target individuals for the SOARS score are those with a low or moderate risk of COVID-19 mortality.

Our data concur with other reports that advanced age is the strongest predictor of death from COVID-19.[20–23] The increase in mortality in patients within the derivation cohort who were in or beyond their seventh decade of life was reflected in the magnitude of their respective adjusted ORs for in-hospital death, namely, 7.4 (aged 70–79) and 10.7 (aged ≥80). Even so, predictive models generated with age as the lone variable showed poorer discriminatory ability than the SOARS score.

The early development of physiological abnormalities in COVID-19 does not always result in timely clinical presentation. In our study, two measures of acute physiological perturbation proved to be important predictors of COVID-19 mortality: hypoxia and tachypnoea. Although persistent hypoxia is more common in non-survivors of COVID-19, its relationship with tachypnoea remains incompletely understood.[24 25] Fewer than half of patients with COVID-19 who present to hospital with decreased oxygen saturation report experiencing subjective breathlessness.[26–29] One reason for this observation might be the so-called 'silent hypoxia' where a blunted symptomatic perception of the effects of hypoxaemia is apparent even when low arterial oxygen tension is evident.[30] This phenomenon may be responsible for delays in seeking clinical attention. Such danger could be mitigated by accurate risk assessment including the measurement and tracking of SpO_2 in patients who are deemed to not require immediate hospitalisation. The absence of oxygen determination in the CURB-65 score has been cited as limiting its utility in stratifying patients with COVID-19 for outpatient management.[31]

The SOARS score was constructed with data from hospitalised patients as the very low adverse event rate among non-admitted cases (eg, in the VH pathway) curtailed the development of a prognostic tool. This issue has previously been highlighted in the context of CRB-65 where low event rates in community studies of pneumonia made predictive inferences difficult to conclude.[9] Other scores that have been used in COVID-19 studies have either not been designed for this disease or have relied heavily on laboratory-measured data.[32–35]

Table 5 COVID-19 mortality risk stratification based on SOARS score

	Derivation cohort (N=821 with full data; deaths=258)		Aintree validation cohort (N=290 with full data; deaths=94)		ISARIC validation cohort (N=14 231; deaths=4319)	
Risk class (score level)	Mortality by risk class (%)	Proportion of deaths by risk class (%)	Mortality by risk class (%)	Proportion of deaths by risk class (%)	Mortality by risk class (%)	Proportion of deaths by risk class (%)
Low (0–1)	6/163 (3.7%)	6/258 (2.3%)	3/35 (8.6%)	3/94 (3.2%)	128/2362 (5.4%)	128/4319 (3.0%)
Moderate (2)	6/102 (5.9%)	6/258 (2.3%)	3/37 (8.1%)	3/94 (3.2%)	273/1879 (14.5%)	273/4319 (6.3%)
High (≥3)	246/556 (44.2%)	246/258 (95.3%)	88/218 (40.4%)	88/94 (93.6%)	3917/9990 (39.2%)	3917/4319 (90.7%)

SOARS, SpO_2, Obesity, Age, Respiratory rate, Stroke history.

Table 6 Sensitivity analysis of the SOARS score for predicting mortality in the ISARIC validation cohort

Score cut-off	Patients N (% of total)	True positive	True negative	False positive	False negative	Sensitivity	Specificity	PPV	NPV	Cumulative mortality (%)
>0	833 (5.9%)	4284	799	9114	34	99.2%	8.1%	32.0%	95.9%	34/833 (4.1%)
>1	2362 (16.6%)	4190	2234	7679	128	97.0%	22.5%	35.3%	94.6%	128/2362 (5.4%)
>2	4241 (29.8%)	3917	3840	6073	401	90.7%	38.7%	39.2%	90.5%	401/4241 (9.5%)
>3	6818 (47.9%)	3267	5767	4146	1051	75.7%	58.2%	44.1%	84.6%	1051/6818 (15.4%)
>4	10831 (76.1%)	1777	8290	1623	2541	41.2%	83.6%	52.3%	76.5%	2541/10831 (23.5%)
>5	13233 (93.0%)	607	9522	391	3711	14.1%	96.1%	60.8%	72.0%	3711/13233 (28.0%)
>6	14172 (99.6%)	36	9890	23	4282	0.8%	99.8%	61.0%	69.8%	4282/14172 (30.2%)
>7	14231 (100.0%)	0	9912	1	4318	0.0%	100.0%	0.0%	69.7%	4318/14231 (30.3%)

SOARS, SpO_2, Obesity, Age, Respiratory rate, Stroke history.

Our multivariate regression model was bootstrapped to reduce overfitting but not penalised prior to external validation. In common with other severity scores for COVID-19, we dichotomised several continuous data parameters which may have potentially obscured non-linear effects between predictors and outcome, contributing to the difference in AUROC values between our derivation and validation cohorts and between both validation cohorts.[11 32 34] Other prediction systems, notably CURB-65 and the Pneumonia Severity Index for pneumonia similarly categorise some of their score parameters.[10 36] We also used in-hospital mortality as an unambiguous disease-related primary outcome rather than 30-day or 60-day mortality. The better performance of the SOARS score in the smaller Aintree validation cohort compared with the much larger ISARIC cohort may have been due to its more homogeneous case-mix. This comparison suggested that, on balance, the simplicity of a prehospital risk prediction tool, provided it retained acceptable accuracy, may outweigh any minor diminution of its performance arising from improved practicality.

Other limitations in the study include the occurrence of missing information despite prospective data collection. The use of multiple imputation to estimate missing values for multivariate regression and the availability of nearly 85% of observations for constructing the risk stratification rule helped to mitigate against underestimating their role. The modest sample size of our derivation cohort was dictated by the incident caseload during the pandemic. However, selective sampling of the pandemic timeline was avoided by including all COVID-19 cases from the initial rise to the subsequent decline in new case numbers over the 11-week study period. Finally, reduced score calibration at the high-risk end suggests that SOARS may overestimate the probability of death in the highest risk cases. However, the principal objective of this score was to enhance frontline decision-making in patients with a low predicted risk of mortality at a time when demand for in-patient resources is likely to be high.

In summary, prognostication using the SOARS score can be undertaken concomitantly with SARS-CoV-2 diagnostic testing to inform clinical triaging, including decisions about the placement of the patient for ongoing care. Analysis of the ISARIC validation cohort in this study showed that between 16.6% and 29.8% (those scoring up to SOARS 1 or 2, respectively) could potentially have avoided admission provided a safe alternative to hospitalisation was in place. Prospective studies of SOARS implementation will enable the score to be calibrated against other independent cohorts of patients with COVID-19 to examine its performance under conditions that may be unique to different localities. Such an opportunity may soon present itself if SARS-CoV-2 transmission continues to increase in the UK and beyond.

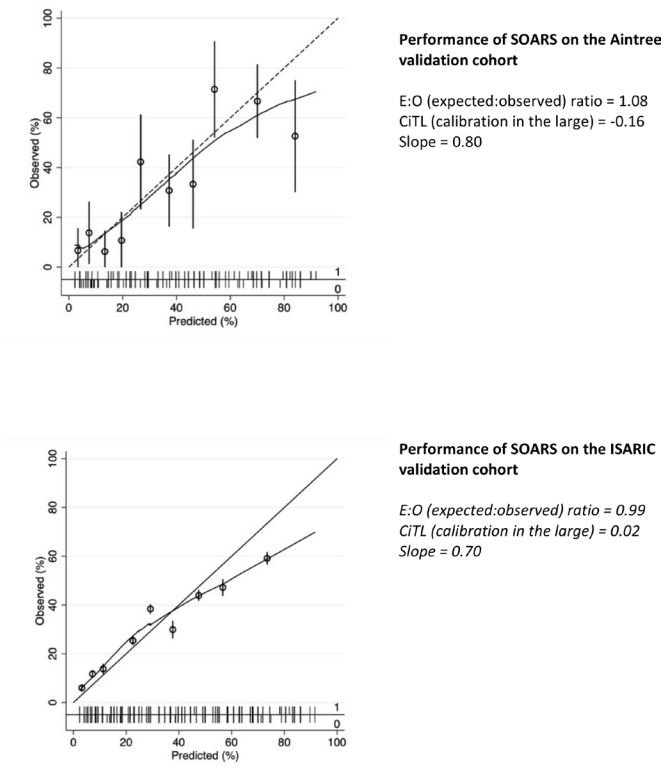

Figure 2 Calibration accuracy of the SOARS (SpO_2, Obesity, Age, Respiratory rate, Stroke history) score on external validation cohorts.

Author affiliations
[1]Interstitial Lung Disease Unit, Department of Respiratory Medicine, Royal Brompton and Harefield NHS Foundation Trust, London, UK
[2]National Heart and Lung Institute, Imperial College London, London, UK
[3]Respiratory Medicine, West Hertfordshire Hospitals NHS Trust, Watford, UK
[4]Respiratory Medicine, St. George's Hospital, London, UK
[5]Information Governance, West Hertfordshire Hospitals NHS Trust, Watford, UK
[6]Radiology, West Hertfordshire Hospitals NHS Trust, Watford, UK
[7]Respiratory Medicine, Aintree site, Liverpool Hospitals NHS Foundation Trust, Liverpool, UK
[8]Radiology, Aintree site, Liverpool Hospitals NHS Foundation Trust, UK, Liverpool, UK

Twitter Matthew Knight @mjknight0380 and Andrew Barlow @Andyatfrogmore

Acknowledgements We are grateful to the healthcare teams whose efforts in the clinical field were fundamental to this work. We would like to thank all the patients involved for their vital contributions. We are very grateful to the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC-4C) Investigators, in particular, J

Kenneth Baillie (lead investigator) and Malcolm G Semple (chief investigator) for providing access to data for the external validation of our model.

Contributors RV, AB, MK and RM developed the clinical algorithm and supervised patient management. FC, RV and AD conceived and designed the investigational plan. FC drafted the manuscript with contributions of intellectual content from RV, AD, LGS, PLM and AB. RV, TV, MK, RM, JS, LS, ET, HM, SC, NM, SA, ML, AO, CP, RK, TH, RT, SR, MS and JL collected the data at respective sites. AD, FC and RV examined the data and undertook statistical analyses. All authors approved the final version of the manuscript for submission. RV is guarantor and attests that all named authors and contributors meet authorship criteria and that no others meeting such criteria have been omitted.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval Ethical approval was provided by Stanmore Research Ethics Committee, London, England (IRAS ID: 283888). The study is registered by the National Health Service Health Research Authority under the reference 20/HRA/2344.

Provenance and peer review Not commissioned, externally peer reviewed. No part of this work has been written by a medical writer or published in printed or electronic form. Some of the findings of this study have been accepted for presentation at the British Thoracic Society Winter Meeting 2020 to be held in February 2021, at which point the abstract will be published in printed format in a supplement of Thorax. A copy of the originally submitted manuscript was uploaded to the medRixv preprint website; https://doi.org/10.1101/2020.10.19.20215426.

Data availability statement Data are available upon reasonable request. Deidentified participant data may be requested from the corresponding author following publication of the study (ORCID identifier: 0000-0001-7845-0173)

ORCID iDs
Felix Chua http://orcid.org/0000-0001-7845-0173
Lisa G Spencer http://orcid.org/0000-0003-3558-992X
Philip L Molyneaux http://orcid.org/0000-0003-1301-8800

REFERENCES

1 Challen K, Goodacre SW, Wilson R, *et al*. Evaluation of triage methods used to select patients with suspected pandemic influenza for hospital admission. *Emerg Med J* 2012;29:383–8.
2 PL H, Chau PH, PSF Y. A clinical prediction rule for clinical diagnosis of severe acute respiratory syndrome. *Eur Resp J*;200:474–9.
3 Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. *BMJ* 2020;369:m1328.
4 Intensive Care National Audit and Research Centre. ICNARC reports on COVID-19 in critical care: England, Wales and Northern Ireland. Available: https://www.icnarc.org [Accessed 4 Dec 2020].
5 Dighe A, Cattarino L, Cuomo-Dannenburg G, *et al*. Response to COVID-19 in South Korea and implications for lifting stringent interventions. *BMC Med* 2020;18:321.
6 Chen B, Zhong H, Ni Y. Epidemiological trends of coronavirus disease 2019 in China. *Front Med* 2020;7:250.
7 Center for Strategic & International Studies Southeast Asia national response to Covid-19 tracker. Available: https://www.csis.org/programs/southeast-asia-program/southeast-asia-covid-19-tracker-0 [Accessed 15 Nov 2020].
8 Cacciapaglia G, Cot C, Sannino F. Second wave COVID-19 pandemics in Europe: a temporal playbook. *Sci Rep* 2020;10:15514.
9 McNally M, Curtain J, O'Brien KK, *et al*. Validity of British thoracic Society guidance (the CRB-65 rule) for predicting the severity of pneumonia in general practice: systematic review and meta-analysis. *Br J Gen Pract* 2010;60:e423–33.
10 Lim WS, van der Eerden MM, Laing R, *et al*. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* 2003;58:377–82.
11 Knight SR, Ho A, Pius R, *et al*. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC who clinical characterisation protocol: development and validation of the 4C mortality score. *BMJ* 2020;370:m3339.
12 National Early Warning Score (NEWS)2. Royal College of Physicians, London. Available: https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 [Accessed 30 May 2020].
13 Collins GS, Reitsma JB, Altman DG, *et al*. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Ann Intern Med* 2015;162:55–63.
14 Shipe ME, Deppen SA, Farjah F, *et al*. Developing prediction models for clinical use using logistic regression: an overview. *J Thorac Dis* 2019;11:S574–84.
15 Leisman DE, Harhay MO, Lederer DJ, *et al*. Development and reporting of prediction models: guidance for authors from editors of respiratory, sleep, and critical care journals. *Crit Care Med* 2020;48:623-633.
16 Resche-Rigon M, White IR. Multiple imputation by chained equations for systematically and sporadically missing multilevel data. *Stat Methods Med Res* 2018;27:1634–49.
17 Hosmer DW, Lemesbow S. Goodness of fit tests for the multiple logistic regression model. *Commun Stat Theory Methods* 1980;9:1043–69.
18 Hu B, Shao J, Palta M. Pseudo-R2 in logistic regression model. *Statistica Sinica* 2006;16:847–60.
19 Kidney Disease, Improving Clinical Outcomes. KDIGO 2012 clinical practice guideline for the evaluation and management of chronic kidney disease. Available: https://kdigo.org/wp-content/uploads/2017/02/KDIGO_2012_CKD_GL.pdf [Accessed 1 Jun 2020].
20 Williamson EJ, Walker AJ, Bhaskaran K. OpenSAFELY: factors associated with COVID-19 death in 17 million patients. *Nature* 2020;584:430–4.
21 Mikami T, Miyashita H, Yamada T, *et al*. Risk factors for mortality in patients with COVID-19 in New York City. *J Gen Intern Med* 2020. doi:10.1007/s11606-020-05983-z. [Epub ahead of print: 30 Jun 2020].
22 Feng Y, Ling Y, Bai T, *et al*. COVID-19 with different severities: a multicenter study of clinical features. *Am J Respir Crit Care Med* 2020;201:1380–8.
23 Onder G, Rezza G, Brusaferro S. Case-Fatality rate and characteristics of patients dying in relation to COVID-19 in Italy. *JAMA* 2020;323:1775–6.
24 Xie J, Covassin N, Fan Z, *et al*. Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clin Proc* 2020;95:1138–47.
25 Wu C, Chen X, Cai Y, *et al*. Risk factors associated with acute respiratory distress syndrome and death in patients with coronavirus disease 2019 pneumonia in Wuhan, China. *JAMA Intern Med* 2020;180:934–43.
26 Richardson S, Hirsch JS, Narasimhan M, *et al*. Presenting characteristics, comorbidities, and outcomes among 5700 patients hospitalized with COVID-19 in the new York City area. *JAMA* 2020;323:2052.
27 Guan W, Ni Z, Hu Y. For the China medical treatment expert group for Covid-19. clinical characteristics of coronavirus disease in China. *N Engl J Med* 2020;382:1708–20.
28 Zhou F, Yu T, Du R, *et al*. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* 2020;395:1054–62.
29 Huang C, Wang Y, Li X. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020.
30 Tobin MJ, Laghi F, Jubran A. Why COVID-19 silent hypoxemia is Baffling to physicians. *Am J Respir Crit Care Med* 2020;202:356–60.
31 Nguyen Y, Corre F, Honsel V, *et al*. Applicability of the CURB-65 pneumonia severity score for outpatient treatment of COVID-19. *J Infect* 2020;81:e96–8.
32 Fan G, Tu C, Zhou F, *et al*. Comparison of severity scores for COVID-19 patients with pneumonia: a retrospective study. *Eur Respir J* 2020;56:2002113.
33 Liang W, Liang H, Ou L. Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with COVID-19. *JAMA Intern Med*.
34 Satici C, Demirkol MA, Sargin Altunok E, *et al*. Performance of pneumonia severity index and CURB-65 in predicting 30-day mortality in patients with COVID-19. *Int J Infect Dis* 2020;98:84–9.
35 Liu S, Yao N, Qiu Y, *et al*. Predictive performance of SOFA and qSOFA for in-hospital mortality in severe novel coronavirus disease. *Am J Emerg Med* 2020;38:2074–80.
36 Fine MJ, Auble TE, Yealy DM, *et al*. A prediction rule to identify low-risk patients with community-acquired pneumonia. *N Engl J Med* 1997;336:243–50.