Prospective validation of the 4C prognostic models for adults hospitalised with COVID-19 using the ISARIC WHO Clinical Characterisation Protocol

Stephen R Knight,1 Rishi K Gupta2, Antonia Ho,3 Riinu Pius,1 Iain Buchan,4,5 Gail Carson,6 Thomas M Drake,1 Jake Dunning,7,8 Cameron J Fairfield,1 Carrol Gamble,9 Christopher A Green,10 Sophie Halpin,9 Hayley E Hardwick,11 Karl A Holden,11 Peter W Horby,6 Clare Jackson,9 Kenneth A Mclean,1 Laura Merson,6 Jonathan S Nguyen-Van-Tam,12 Lisa Norman,1 Piero L Olliaro,13 Mark G Pritchard,13 Clark D Russell,14 Catherine A Shaw,1 Aziz Sheikh,15 Tom Solomon,11 Cathie Sudlow,16 Olivia V Swann,17 Lance C W Turtle18,19 Peter J M Openshaw,20 J Kenneth Baillie21,22 Annemarie Docherty,1 Malcolm G Semple11,22 Mahdad Noursadeghi23, Ewen M Harrison,1 ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) Investigators, On behalf of the ISARIC4C investigators

ABSTRACT

Purpose To prospectively validate two risk scores to predict mortality (4C Mortality) and in-hospital deterioration (4C Deterioration) among adults hospitalised with COVID-19.

Methods Prospective observational cohort study of adults (age ≥18 years) with confirmed or highly suspected COVID-19 recruited into the International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) WHO Clinical Characterisation Protocol UK (CCP-UK) study in 306 hospitals across England, Scotland and Wales. Patients were recruited between 27 August 2020 and 17 February 2021, with at least 4 weeks follow-up before final data extraction. The main outcome measures were discrimination and calibration of models for in-hospital deterioration (defined as any requirement of ventilatory support or critical care, or death) and mortality, incorporating predefined subgroups.

Results 76 588 participants were included, of whom 27 352 (37.4%) deteriorated and 12 581 (17.4%) died. Both the 4C Mortality (0.78 (0.77 to 0.78)) and 4C Deterioration scores (pooled C-statistic 0.76 (95% CI 0.75 to 0.77)) demonstrated consistent discrimination across all nine National Health Service regions, with similar performance metrics to the original validation cohorts. Calibration remained stable (4C Mortality: pooled slope 1.09, pooled calibration-in-the-large 0.12; 4C Deterioration: 1.00, –0.04), with no need for temporal recalibration during the second UK pandemic wave of hospital admissions.

Conclusion Both 4C risk stratification models demonstrate consistent performance to predict clinical deterioration and mortality in a large prospective second wave validation cohort of UK patients. Despite recent advances in the treatment and management of adults hospitalised with COVID-19, both scores can continue to inform clinical decision making.

Key messages

What is the key question?

► The clinical characteristics and management of hospitalised patients with confirmed or highly suspected COVID-19 have changed over time; ongoing prospective validation of risk prediction scores is therefore required.

What is the bottom line?

► Both 4C prediction scores performed well in a large prospective UK cohort, with stable validation metrics across National Health Service regions and ethnicity, despite reduced overall deterioration and mortality risk compared with original derivation and validation cohorts.

Why read on?

► This is the first large prospective revalidation of these risk stratification tools, which demonstrated stable performance in over 75 000 hospitalised UK patients.

Trial registration number ISRCTN66726260.

INTRODUCTION

Disease resulting from infection with SARS-CoV-2 has a high mortality rate with deaths predominantly caused by respiratory failure.1 We previously reported two prognostic scores for in-hospital mortality2 and deterioration3 from the International Severe Acute Respiratory and Emerging Infections Consortium Coronavirus Clinical Characterisation...
Consortium (ISARIC4C) study derived and validated in large UK cohorts during the first pandemic wave.

As hospitals around the world faced a continued influx of patients with COVID-19, these easy-to-use risk stratification tools have facilitated early identification of patients infected with SARS-CoV-2 who are at the highest risk of deterioration and death to guide management and optimise resource allocation. Both scores use readily available clinical and biochemical parameters at the time of assessment, with predicted risk used to guide antiviral treatment across the UK. Furthermore, independent external validation has demonstrated consistent performance worldwide.

Management and treatment of patients admitted or diagnosed in hospital with COVID-19 have changed markedly over the past year, notably with the introduction of corticosteroids for people requiring supplemental oxygen or ventilatory support. Ongoing prospective validation is therefore necessary to ensure adequate performance and inform the need for temporal recalibration.

In this article, we extend our previous work by prospectively validating these scores in a large study cohort using the wide geographical coverage of the ISARIC4C study cohort in England, Wales, and Scotland, among adults recruited during the second pandemic wave.

**METHODS**

**Study population and data collection**

The International Severe Acute Respiratory and emerging Infections Consortium (ISARIC)–WHO Clinical Characterisation Protocol UK (CCP-UK) study is an ongoing prospective multi-centre cohort study being conducted by ISARIC4C in 306 hospitals across England, Scotland and Wales (National Institute for Health Research (NIHR) Clinical Research Network Central Portfolio Management System ID 14152). The study was part of a suite of ‘sleeping’ protocols established and approved in the UK prior to the pandemic and was activated on 17 January 2020. CCP-UK is a version of a standardised open-source international research protocol (the ISARIC/WHO CCP) created in 2012, which has been employed worldwide for harmonised observational studies of COVID-19.

The CCP-UK protocol and further study details are available online. In this analysis, we included consecutive adults (aged ≥18 years) who had highly suspected or PCR-confirmed COVID-19. As before, we included patients with suspected COVID-19 in the analysis because both models were intended for use in participants at the point of initial evaluation for COVID-19, when virological confirmation might not be available. We also included nosocomial COVID-19 acquisition to test the hypothesis that acquisition of infection in hospital might be associated with differential risk. Community-acquired infection was defined as symptom onset or first positive SARS-CoV-2 PCR result within 7 days from admission; participants who did not meet these criteria and had either symptom onset or first positive SARS-CoV-2 PCR result more than 7 days from admission were classified as nosocomial cases. Among nosocomial cases, patients who met the deterioration outcome before the onset of COVID-19 were excluded. Northern Ireland was excluded from model validation due to the small numbers of patients recruited.

For both scores, we included eligible hospitalised participants with confirmed or highly suspected COVID-19 between 27 August 2020 and 17 February 2021. All patients had at least four weeks follow-up to reduce selection bias (final data extraction date: 23 April 2021). Patients included within original derivation or validation cohorts for either score were excluded from this analysis. Participants who had ongoing hospital care at the end of follow-up (the point at which a final outcome was recorded in the case record form) were classified as not meeting the endpoint because the risk of deterioration declines with time since admission. Hospitals that recruited participants admitted to intensive care unit (ICU) exclusively were not included since these participants had already met the 4C Deterioration outcome by definition. Both scores are summarised in table 1, with additional information on how coefficients were transformed into points systems contained in online supplemental appendix 1.

The study is reported in accordance with Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD) guidance. Demographic, clinical and outcome data were collected by research nurses and medical student volunteers using a publicly available standardised case record form and uploaded to a Research Electronic Data Capture Database (REDCap, Vanderbilt University, US, hosted by University of Oxford, UK). Co-morbidities were defined by a modified Charlson comorbidity index and obesity was clinician-defined. Consent was not required for the use of routinely collected clinical data from medical records. The Control of Patient Information notice 2020 for urgent public health research makes provision for this in England and Wales. A waiver for consent was obtained from the Public Benefit and Privacy Panel for Health and Social Care in Scotland.

**Outcomes**

For the 4C Deterioration model, we used a composite primary outcome comprising any of the following during hospital admission and were equally weighted: initiation of ventilatory support (non-invasive ventilation, invasive mechanical ventilation or extracorporeal membrane oxygenation); admission to a high-dependency unit or ICU; or death (all-cause), as reported previously.

**Model validation**

All validation analyses were conducted as described in our previous publication. Briefly, we assessed model discrimination (how well predictions differentiated participants who experienced the outcomes from those who did not, quantified as the C-statistic) and calibration (agreement between predicted and observed risk, assessed using calibration slopes, calibration-in-the-large and calibration plots). To assess calibration for the 4C Mortality score, we transformed points scores to the probability scale using the observed mortality proportion for each distinct total point score in the original reported validation cohort.

We used multiple imputation with chained equations to account for missing data using the mice package in R, as previously described. We included all predictors (including restricted cubic spline transformations) and the outcome in the imputation models. Analyses were done in each of the 10 multiply imputed datasets and pooled using Rubin’s rules. Our primary analyses were performed stratified by National Health Service (NHS) region in order to examine for evidence of between-region heterogeneity in model performance. We visualised C-statistics, calibration-in-the-large and slope estimates across regions in forest plots and calculated pooled estimates using random effects meta-analysis, as previously recommended.

Decision curve analysis allows assessment of clinical utility by quantifying the trade-off between correctly identifying true positives and incorrectly identifying false positives weighted
Table 1  Model parameters for 4C Mortality and 4C Deterioration prognostic model

<table>
<thead>
<tr>
<th>Variable</th>
<th>4C Mortality score</th>
<th>4C Deterioration</th>
<th>OR (OR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;50</td>
<td>–</td>
<td>Intercept</td>
<td>4.033</td>
</tr>
<tr>
<td>50–59</td>
<td>+2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60–69</td>
<td>+4</td>
<td>Age (years)</td>
<td>0.0159</td>
</tr>
<tr>
<td>70–79</td>
<td>+6</td>
<td>Age (spline 1)</td>
<td>–0.1219</td>
</tr>
<tr>
<td>≥80</td>
<td>+7</td>
<td>Age (spline 2)</td>
<td>0.1265</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>–</td>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>+1</td>
<td>Female</td>
<td>–</td>
</tr>
<tr>
<td>Number of comorbidities</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>–</td>
<td>Nosocomial</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>+1</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>≥2</td>
<td>+2</td>
<td>Yes</td>
<td>0.2439</td>
</tr>
<tr>
<td>Respiratory rate (breaths/minute)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>–</td>
<td>Radiographic infiltrates</td>
<td></td>
</tr>
<tr>
<td>20–29</td>
<td>+1</td>
<td>No</td>
<td>–</td>
</tr>
<tr>
<td>≥30</td>
<td>+2</td>
<td>Yes</td>
<td>0.3252</td>
</tr>
<tr>
<td>Peripheral oxygen saturation on room air (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≥92</td>
<td>–</td>
<td>Respiratory rate (breaths/minute)</td>
<td>–0.0145</td>
</tr>
<tr>
<td>&lt;92</td>
<td>+2</td>
<td>Respiratory rate (spline 1)</td>
<td>0.5992</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Respiratory rate (spline 2)</td>
<td>–1.078</td>
</tr>
<tr>
<td>Glasgow Coma Scale score</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>–</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>&lt;15</td>
<td>+2</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤7</td>
<td>–</td>
<td>Peripheral oxygen saturation (SpO₂)</td>
<td>–0.07078</td>
</tr>
<tr>
<td>7–14</td>
<td>+1</td>
<td>SpO₂ (spline 1)</td>
<td>–0.0248</td>
</tr>
<tr>
<td>&gt;14</td>
<td>+3</td>
<td>SpO₂ (spline 2)</td>
<td>1.024</td>
</tr>
<tr>
<td>C reactive protein (mg/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤50</td>
<td>–</td>
<td>Room air or oxygen</td>
<td></td>
</tr>
<tr>
<td>50–99</td>
<td>+1</td>
<td>Room air</td>
<td>–</td>
</tr>
<tr>
<td>≥100</td>
<td>+2</td>
<td>Oxygen therapy</td>
<td>0.7450</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lymphocyte count (x10⁹/L)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤4564</td>
<td>–</td>
<td>Lymphocytes (spline 1)</td>
<td>0.7309</td>
</tr>
<tr>
<td>≥8113</td>
<td>+1</td>
<td>Lymphocytes (spline 2)</td>
<td>–0.8113</td>
</tr>
</tbody>
</table>

Restricted cubic spline knot positions are:

- Age = 38.5, 67.7, 81.1, 92.9;
- Respiratory rate=16, 19, 24, 37;
- SpO₂=84, 94, 96, 100;
- Urea = 2.9, 5.7, 9.3, 25.5;
- C reactive protein=5, 45, 113, 297;
- Lymphocytes = 0.3, 0.7, 1.1, 2.4.

The threshold probability represents the risk cut-off above which any given treatment or intervention might be considered and reflects the perceived risk:benefit ratio for the intervention. Decision curve analysis was used to quantify the net benefit of implementing the model in clinical practice compared with a treat-all approach; a treat-none approach. All decision curves were smoothed by locally weighted smoothing (LOESS) from stacked multiply imputed datasets.

A sensitivity analysis was also performed, with stratification of the validation cohort by ethnic group and month of admission, in view of previously reported differences in COVID-19 outcomes by ethnicity and over time. All analyses were done in R (V3.6.3).

Patient and public involvement

This was an urgent public health research study in response to a Public Health Emergency of International Concern. Patients or the public were not involved in the design, conduct or reporting of this rapid response research.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation or writing of the report. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

RESULTS

Between 27 August 2020 and 17 February 2021, 76 588 eligible adults were recruited to the ISARIC4C study and included in the current analysis, of whom 69 260 (90.4%) were known to have PCR-confirmed COVID-19. Baseline demographic, physiological and laboratory characteristics are shown stratified by outcome (table 2). The median age of patients in the cohort was 72 years (IQR 57–83); 35 231 (46.0%) were female and 52 704 (72.6%) had at least one comorbidity. The temporal distribution of participant admissions, stratified by NHS region, is shown in figure 1. For patients with nosocomial infections, the median time from admission to recruitment was 15 days (IQR 10–29). A summary of missingness for predictor variables included in both scores is shown in online supplemental appendix 2.

For the deterioration outcome, 73 078 (95.4%) participants had an outcome available, and in-hospital clinical deterioration occurred in 27 352 (37.4%), with a median time to deterioration of 5 days (IQR 1–12). For mortality, 72 481 (94.6%) participants had an outcome available, with in-hospital death occurring among 12 581 (17.4%). The median time to death was 11 days (IQR 6–18).

For the deterioration outcome, 73 078 (95.4%) participants had an outcome available, and in-hospital clinical deterioration occurred in 27 352 (37.4%), with a median time to deterioration of 5 days (IQR 1–12). For mortality, 72 481 (94.6%) participants had an outcome available, with in-hospital death occurring among 12 581 (17.4%). The median time to death was 11 days (IQR 6–18).

Forest plots showing model discrimination (C-statistic) and calibration metrics (slope and calibration-in-the-large) for both 4C Mortality and 4C Deterioration scores are shown in figure 2. C-statistics were consistent across NHS regions for 4C Mortality scores (point estimates 0.77–0.81; pooled random-effects meta-analysis estimate 0.78 (95% CI 0.77 to 0.78), I²=35%) and 4C Deterioration (point estimates 0.74–0.78; pooled random-effects meta-analysis estimate 0.76 (0.75 to 0.77), I²=57%). Calibration slopes were also consistent across regions, for the 4C Mortality score (0.97–1.31; pooled estimate 1.09 (1.03 to 1.16), I²=63%) and 4C Deterioration (point estimates 0.94–1.04; pooled estimate 1.00 (0.98 to 1.03), I²=30%; figure 3).

Heterogeneity across NHS regions in calibration-in-the-large was seen. For 4C Mortality, points estimates were –0.15 to 0.41 (pooled estimate 0.12 (–0.01 to 0.25), I²=97%); for 4C Deterioration, point estimates ranged from –0.15 to 0.19 (pooled estimate –0.04 (–0.13 to 0.04), I²=93%). The sensitivity, specificity, positive-predictive value and negative-predictive value across the full range of probability thresholds from the model are shown in online supplemental appendix S3 and S4. Decision curve analysis to further examine clinical utility showed higher
Table 2  Baseline characteristics, stratified by first chronological deterioration category through which they met the composite 4C Deterioration primary outcome (HDU admission, ICU admission or ventilatory support, or death)

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall, N=76 588*</th>
<th>Ventilatory support or HDU/ICU, N=14 771*</th>
<th>Died, N=12 581*</th>
<th>No deterioration, N=45 726*</th>
<th>(Missing), N=3510*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (years)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>72 (57–83)</td>
<td>65 (54–75)</td>
<td>84 (77–89)</td>
<td>70 (54–82)</td>
<td>75 (60–84)</td>
</tr>
<tr>
<td><strong>Sex</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>42 731 (46)</td>
<td>5307 (36)</td>
<td>5533 (44)</td>
<td>22 717 (50)</td>
<td>1674 (48)</td>
</tr>
<tr>
<td>Male</td>
<td>41 250 (54)</td>
<td>9447 (64)</td>
<td>7028 (56)</td>
<td>22 955 (50)</td>
<td>1820 (52)</td>
</tr>
<tr>
<td>Missing</td>
<td>107 (0.1)</td>
<td>17 (0.1)</td>
<td>20 (0.2)</td>
<td>54 (0.1)</td>
<td>16 (0.5)</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>56 578 (85)</td>
<td>10 063 (78)</td>
<td>10 291 (92)</td>
<td>33 567 (84)</td>
<td>2657 (90)</td>
</tr>
<tr>
<td>South Asian</td>
<td>4098 (6.1)</td>
<td>1179 (9.2)</td>
<td>294 (2.6)</td>
<td>2565 (6.4)</td>
<td>60 (2.0)</td>
</tr>
<tr>
<td>Black</td>
<td>1536 (2.3)</td>
<td>389 (3.0)</td>
<td>95 (0.9)</td>
<td>1005 (2.5)</td>
<td>47 (1.6)</td>
</tr>
<tr>
<td>East Asian</td>
<td>337 (0.5)</td>
<td>105 (0.8)</td>
<td>32 (0.3)</td>
<td>188 (0.5)</td>
<td>12 (0.4)</td>
</tr>
<tr>
<td>Other</td>
<td>4257 (6.4)</td>
<td>1108 (8.6)</td>
<td>423 (3.8)</td>
<td>2565 (6.4)</td>
<td>161 (5.5)</td>
</tr>
<tr>
<td>Missing</td>
<td>9782 (13)</td>
<td>1927 (13)</td>
<td>1446 (11)</td>
<td>5836 (13)</td>
<td>573 (16)</td>
</tr>
<tr>
<td>SARS-CoV-2 PCR positive</td>
<td>69 260 (99)</td>
<td>13 226 (99)</td>
<td>11 642 (99)</td>
<td>41 935 (99)</td>
<td>2457 (99)</td>
</tr>
<tr>
<td>Missing</td>
<td>6841 (8.9)</td>
<td>1432 (9.7)</td>
<td>873 (6.9)</td>
<td>3502 (7.7)</td>
<td>1034 (29)</td>
</tr>
<tr>
<td><strong>Number of comorbidities</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 (0–2)</td>
<td>52 704 (73)</td>
<td>10 118 (71)</td>
<td>10 865 (91)</td>
<td>30 102 (68)</td>
<td>1619 (46)</td>
</tr>
<tr>
<td>Missing</td>
<td>3985 (5.2)</td>
<td>478 (3.2)</td>
<td>621 (4.9)</td>
<td>1674 (3.7)</td>
<td>1212 (35)</td>
</tr>
<tr>
<td>Nosocomial infection</td>
<td>9603 (13)</td>
<td>553 (3.8)</td>
<td>2489 (20)</td>
<td>6088 (14)</td>
<td>473 (16)</td>
</tr>
<tr>
<td>Missing</td>
<td>1973 (2.6)</td>
<td>132 (0.9)</td>
<td>329 (2.6)</td>
<td>1027 (2.2)</td>
<td>485 (14)</td>
</tr>
<tr>
<td>Radiographic infiltrates</td>
<td>26 079 (64)</td>
<td>8320 (81)</td>
<td>3563 (59)</td>
<td>13 485 (58)</td>
<td>711 (54)</td>
</tr>
<tr>
<td>Missing</td>
<td>35 735 (46)</td>
<td>4533 (31)</td>
<td>6569 (52)</td>
<td>22 450 (49)</td>
<td>2183 (62)</td>
</tr>
<tr>
<td>Respiratory rate (per min)</td>
<td>20 (18–24)</td>
<td>24 (20–30)</td>
<td>20 (18–24)</td>
<td>20 (18–23)</td>
<td>20 (18–23)</td>
</tr>
<tr>
<td>Missing</td>
<td>6535 (8.5)</td>
<td>760 (5.1)</td>
<td>962 (7.6)</td>
<td>3431</td>
<td>1382 (39)</td>
</tr>
<tr>
<td>SpO₂ (%)</td>
<td>96 (93–97)</td>
<td>94 (89–96)</td>
<td>96 (93–97)</td>
<td>96 (94–98)</td>
<td>96 (94–97)</td>
</tr>
<tr>
<td>Missing</td>
<td>18 897 (27)</td>
<td>6672 (48)</td>
<td>3324 (24)</td>
<td>8373 (74)</td>
<td>528 (3.9)</td>
</tr>
<tr>
<td>SpO₂ &lt;94%</td>
<td>6401 (8.4)</td>
<td>763 (5.1)</td>
<td>964 (7.7)</td>
<td>3405 (7.4)</td>
<td>1269 (36)</td>
</tr>
<tr>
<td><strong>Room air or oxygen</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Room air</td>
<td>47 061 (68)</td>
<td>5885 (42)</td>
<td>7810 (68)</td>
<td>31 831 (76)</td>
<td>1535 (70)</td>
</tr>
<tr>
<td>Oxygen</td>
<td>22 384 (32)</td>
<td>8018 (58)</td>
<td>3655 (32)</td>
<td>10 038 (24)</td>
<td>673 (30)</td>
</tr>
<tr>
<td>Missing</td>
<td>7143 (9.3)</td>
<td>868 (5.9)</td>
<td>1116 (8.9)</td>
<td>3857 (8.4)</td>
<td>1302 (37)</td>
</tr>
<tr>
<td>Missing</td>
<td>6332 (8.2)</td>
<td>873 (5.9)</td>
<td>1189 (9.5)</td>
<td>2780 (6.1)</td>
<td>1490 (42)</td>
</tr>
<tr>
<td>Lymphocytes (x10⁹/L)</td>
<td>0.90 (0.60–1.30)</td>
<td>0.80 (0.56–1.10)</td>
<td>0.80 (0.50–1.18)</td>
<td>0.98 (0.70–1.40)</td>
<td>0.90 (0.60–1.27)</td>
</tr>
<tr>
<td>Missing</td>
<td>16 087 (21)</td>
<td>1448 (11)</td>
<td>3073 (24)</td>
<td>10 778 (23.6)</td>
<td>788 (22)</td>
</tr>
<tr>
<td>Urea (mmol/L)</td>
<td>6.6 (4.6–9.9)</td>
<td>6.3 (4.9–10.3)</td>
<td>9.7 (6.6–15.1)</td>
<td>5.9 (4.3–8.6)</td>
<td>6.5 (4.6–9.7)</td>
</tr>
<tr>
<td>Missing</td>
<td>20 552 (27)</td>
<td>2403 (16)</td>
<td>3701 (29)</td>
<td>13 405 (29)</td>
<td>1043</td>
</tr>
<tr>
<td>C reactive protein (mg/L)</td>
<td>69 (27–133)</td>
<td>109 (57–181)</td>
<td>76 (34–142)</td>
<td>55 (20–110)</td>
<td>57 (22–114)</td>
</tr>
<tr>
<td>Missing</td>
<td>21 543 (28)</td>
<td>2262 (15)</td>
<td>3863 (31)</td>
<td>14 320 (31)</td>
<td>1098</td>
</tr>
<tr>
<td><strong>NHS region</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>East of England</td>
<td>9507 (12)</td>
<td>1446 (9.8)</td>
<td>1819 (14)</td>
<td>5132 (11)</td>
<td>1110 (32)</td>
</tr>
<tr>
<td>London</td>
<td>3886 (5.1)</td>
<td>1272 (8.6)</td>
<td>306 (2.4)</td>
<td>2184 (4.8)</td>
<td>124 (3.5)</td>
</tr>
<tr>
<td>Midlands</td>
<td>17 988 (23)</td>
<td>3043 (21)</td>
<td>3100 (25)</td>
<td>11 333 (25)</td>
<td>512 (15)</td>
</tr>
<tr>
<td>North East and Yorkshire</td>
<td>12 081 (16)</td>
<td>2465 (17)</td>
<td>1845 (15)</td>
<td>7520 (16)</td>
<td>251 (7.2)</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>31 (&lt;0.1)</td>
<td>17 (0.1)</td>
<td>1 (&lt;0.1)</td>
<td>13 (&lt;0.1)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>North West</td>
<td>13 731 (18)</td>
<td>2847 (19)</td>
<td>2321 (18)</td>
<td>8120 (18)</td>
<td>443 (13)</td>
</tr>
<tr>
<td>Scotland</td>
<td>1558 (2.0)</td>
<td>340 (2.3)</td>
<td>205 (1.6)</td>
<td>916 (2.0)</td>
<td>97 (2.8)</td>
</tr>
</tbody>
</table>

Continued
net benefit than the treat-all and treat-none approaches across a range of threshold probabilities for both scores (online supplemental appendix S5). Threshold performance and clinical utility were similar to those originally reported for both scores.2 3

Sensitivity analysis
Sensitivity analyses that used complete case data showed similar discrimination and performance metrics (online supplemental appendices S6 and S7) to analyses that used the imputed dataset. After stratification of the validation cohort by ethnicity (online supplemental appendices S8 and S9), discrimination remained similar for the two scores (4C Mortality range 0.77–0.83; 4C Deterioration 0.75–0.82). However, discrimination appeared marginally better among all non-white ethnic groups (4C Mortality range 0.82–0.83; 4C Deterioration 0.79–0.82) compared with the white group (4C Mortality range 0.77 (95% CI 0.76 to 0.77)); 4C Deterioration 0.75 (95% CI 0.75 to 0.76). Calibration-in-the-large for black ethnicity demonstrated some evidence of inconsistency for both 4C Mortality (−0.13 (95% CI −0.30 to 0.04)) and 4C Deterioration (−0.26 (95% CI −0.38 to −0.14)), particularly in complete case analysis (online supplemental appendix S9), despite excellent discrimination (4C Mortality 0.83 (95% CI 0.80 to 0.85)); 4C Deterioration 0.79 (95% CI 0.76 to 0.81).

Despite changes in unadjusted mortality when stratified by month of admission for patients aged ≥60 years old (online supplemental appendix S10), discrimination (4C Mortality range 0.77–0.80, I²=74%; 4C Deterioration 0.75–0.79, I²=82%) and calibration slopes (4C Mortality point estimates 1.05–1.16, I²=6%; 4C Deterioration 0.97–1.09, I²=49%, online supplemental appendices 11–13) were found to remain stable for both scores. However, 4C Mortality demonstrated heterogeneity for calibration-in-the-large in December 2020 (0.27) compared with other months of admission (range −0.06 to 0.14, pooled estimate 0.10 (−0.02 to 0.22), I²=96%; online supplemental appendix S11), corresponding with an increase in unadjusted mortality for patients ≥70 years old (online supplemental appendix S10). The calibration plot for patients recruited during December demonstrated an associated underestimation of mortality risk (online supplemental appendix S12). For 4C Deterioration, calibration-in-the-large was similar across admission month (pooled estimate −0.06 (−0.11 to −0.01), I²=81%).

Across the four 4C Mortality risk groups, the corresponding mortality risks were: low risk (0–3 score, mortality rate 1.5%); intermediate risk (4–8 score, 9.5%); high risk (9–14 score, 32.8%); and very high risk (≥15 score, 63.9%). These mortality risks were similar to the original validation cohort (online supplemental appendices 14 and 15). A stepwise increase in oxygen requirement, deterioration, mortality and duration of

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Overall, N=76 588*</th>
<th>Ventilatory support or HDU/ICU, N=14 771*</th>
<th>Died, N=12 581*</th>
<th>No deterioration, N=45 726*</th>
<th>(Missing), N=3510*</th>
</tr>
</thead>
<tbody>
<tr>
<td>South East</td>
<td>9 530 (12)</td>
<td>1 903 (13)</td>
<td>1 619 (13)</td>
<td>5 672 (12)</td>
<td>336 (9.6)</td>
</tr>
<tr>
<td>South West</td>
<td>6 845 (8.5)</td>
<td>1 079 (7.3)</td>
<td>991 (7.9)</td>
<td>3 788 (8.3)</td>
<td>627 (18)</td>
</tr>
<tr>
<td>Wales</td>
<td>1 791 (2.3)</td>
<td>359 (2.4)</td>
<td>374 (3.0)</td>
<td>1 048 (2.3)</td>
<td>10 (0.3)</td>
</tr>
<tr>
<td>Oxygen received</td>
<td>4 989 (69)</td>
<td>1 427 (98)</td>
<td>9 777 (81)</td>
<td>24 859 (56)</td>
<td>244 (9)</td>
</tr>
<tr>
<td>Systemic steroids received</td>
<td>38 417 (54)</td>
<td>11 769 (84)</td>
<td>6 765 (57)</td>
<td>24 859 (44)</td>
<td>266 (47)</td>
</tr>
<tr>
<td>Mortality</td>
<td>18 211 (25)</td>
<td>5 630 (40)</td>
<td>1 2581 (100)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Length of stay (days)</td>
<td>10 (5–19)</td>
<td>12 (7–21)</td>
<td>13 (6–22)</td>
<td>8 (4–17)</td>
<td>7 (1–23)</td>
</tr>
</tbody>
</table>

Data are median (IQR) or n (%), calculated from non-missing data. Participants are shown by the first chronological deterioration category through which they met the composite primary outcome (HDU or ICU admission, ventilatory support or death).

*Statistics presented: median (IQR); n (%).

HDU, high-dependency unit; ICU, intensive care unit; NHS, National Health Service.

Figure 1  Histogram showing dates of hospital admission or first COVID-19 assessment for adults included in this analysis, stratified by National Health Service region.
hospital stay was seen for both scores as the predicted outcome risk increased (figures 4 and 5). For those identified at the lowest risk of mortality (0–3 score) or deterioration (first and second deciles), patients were the least likely to receive oxygen or deteriorate and had a shorter length of in-patient stay.

**DISCUSSION**

**Principal findings**

We prospectively validated our previously reported risk stratification tools in a prospective cohort study of 76,588 UK hospitalised patients with confirmed or highly suspected COVID-19 during the second pandemic wave. Both scores (4C Mortality and 4C Deterioration) showed consistent discrimination and calibration across NHS regions and ethnicity. Similar metrics to the original reported first wave validation cohorts were found, despite systematic changes in patient management compared with the first wave, particularly the routine use of corticosteroids.

Robust prognostic models that predict outcomes among COVID-19 patients have been urgently needed to support clinical decision-making regarding hospital admission and treatment. Our 4C scores measure patient comorbidity, abnormal physiology and inflammation using routinely measured demographics, bedside observations and laboratory tests to facilitate objective evidence-based assessments. We previously noted that our prognostic tools should be interpreted in the context of current standard treatment at the time the models were developed and validated. Hospital-based management of adults with COVID-19 has evolved with the accrual of new evidence and increasing clinical experience, including dexamethasone use and patient proning. Therefore, the temporal assessment of model performance is critical to ensure consistent performance and inform the need for temporal recalibration, by updating model intercepts, slopes or coefficients as required. Our findings suggest that recalibration is currently not required for the 4C Mortality score or 4C Deterioration model. We will continue to monitor model performance prospectively in ISARIC4C and can perform temporal recalibration to update the model if required.

**Clinical management**

A key aim of risk stratification is to support clinical management decisions, as part of daily routine care and inform stratification of patients on the basis of clinical severity. The combination of
both scores could be included in the programmatic standard of care adopted by hospitals to identify clinical pathways for patients with COVID-19. As well as improving clinical management, it should also encourage better allocation of human and economic resources.30

We demonstrate that risk classes identified in original score derivation and validation cohorts continue to perform well, with a stepwise elevation in oxygen requirement, risk of deterioration and death seen as risk class increased. In addition, both scores performed well across the full range of risk. Patients identified

Figure 4  Risk of oxygen requirement, deterioration, death and length of stay stratified 4C Mortality score.

Figure 5  Risk of oxygen requirement, deterioration, death and length of stay stratified by 4C Deterioration score predicted risk deciles.
as low risk by both scores were less likely to require oxygen
and have a short hospital admission, suggesting these patients
could be managed in the community if supported by a clinician’s
overall assessment, while patients within the intermediate-risk
group (4C Mortality) or third to fifth deciles (4C Deterioration)
might be suitable for ward level monitoring initially. Meanwhile,
patients at high risk of mortality or deterioration may prompt
aggressive treatment and early escalation to critical care if appro-
priate. However, these risk stratification tools should guide, but
not replace, clinical decision making. Furthermore, care should
be taken when interpreting predicted risk, as it does not reflect
the outcome risk in the absence or presence of a particular inter-
vention.\textsuperscript{51} Unfortunately, the prediction of risk with and without
intervention is difficult to perform.\textsuperscript{51}

**Comparison with other studies**
Consistent performance of both models is perhaps surprising,
particularly as the introduction of corticosteroids is likely to
have reduced mortality among people receiving supplement-
oxygen.\textsuperscript{11,12} In this cohort, 38 417 (54\%) patients received
systemic corticosteroids compared with 14\% and 12\% in
the original 4C Deterioration and Mortality cohorts. On the
contrary, the B.1.1.7 variant was reported to be the dominant
circulating SARS-CoV-2 strain during the second wave\textsuperscript{33} and
may be associated with higher mortality.\textsuperscript{34} In our present study,
mortality (17\%) and deterioration (37\%) rates were lower
than cohorts used initially for model derivation and validation
(mortality 32\%, deterioration 43\%) despite similar patient age,
comorbidities and time to outcome. Nonetheless, both models
performed well despite these changes across the included time
period. Therefore, application of the 4C Mortality and 4C Dete-
rioration scores together still provides a validated and evidence-
based approach for clinicians to predict the appropriate outcome
as required to inform clinical management decisions.

**Strengths and limitations of this study**
This is the largest prospective validation study for prognostic risk
scores among hospitalised patients with COVID-19. As for our
original reports, we adhered to TRIPOD reporting standards,\textsuperscript{18}
used multiple imputation to deal with missing data and exam-
ined heterogeneity in detail by NHS region, ethnicity and month
of admission. We have demonstrated the ability to temporally
validate both scores during subsequent admission waves and, if
required, can temporally recalibrate both scores if performance
decreases in future. In addition, both scores were able to iden-
tify both low-risk (rule-out) and high-risk patients (rule-in)
for mortality and deterioration, which corresponded with oxygen
requirement and duration of in-hospital stay.

There are, however, some limitations. First, the patient cohort
comprised of hospitalised patients with confirmed or highly
suspected COVID-19 who were seriously ill (mortality rate of
17.4\%) and were of advanced age (median age 72 years), similar
to cohorts used to derive each score. These models are not for
use in the community and could still perform differently in
populations at lower risk of death. In addition, pooled estimates
demonstrated high heterogeneity, particularly for calibration-
in-the-large. This may represent clinical differences across each
region, including patient population and the medical manage-
ment of hospitalised patients with COVID-19. Nevertheless,
external validation in Brazil,\textsuperscript{6} Canada,\textsuperscript{7} France,\textsuperscript{3}
Netherlands\textsuperscript{9} and Pakistan\textsuperscript{10} has demonstrated consistent performance for
the 4C Mortality Score.

Second, a proportion of recruited patients had incomplete
episodes for deterioration (4.5\%) or mortality (5.4\%). We
handled missing outcome data using multiple imputation in the
primary analysis, assuming missingness at random and performed
a complete case sensitivity analysis, with consistent findings.
Furthermore, the inclusion of all-cause mortality as a primary
outcome measure, rather than covid-related mortality, may
reduce interpretability in patients with nosocomial infection.

**CONCLUSIONS AND POLICY IMPLICATIONS**
We have prospectively validated easy-to-use risk scores that
enable accurate stratification among hospitalised adults with
community-acquired or hospital-acquired COVID-19 for clini-
cal deterioration or mortality. The performance of both scores
has remained consistent despite temporal changes in manage-
ment and treatment during the second wave. Application within
the validation cohorts showed this tool could guide clinician
decisions, including treatment escalation. Although the models
showed consistent performance across England, Wales and Scot-
land, validation in other countries should be prioritised to enable
its clinical implementation internationally. The ongoing perfor-
ance of both scores will need to be assessed in the context of
increasing deployment of immunomodulatory agents\textsuperscript{15} and
COVID-19 vaccines, as well as emerging SARS-CoV-2 variants.

**Author affiliations**
1Centre for Medical Informatics, The University of Edinburgh, Usher Institute of
Population Health Sciences and Informatics, Edinburgh, UK
2University College London Institute for Global Health, London, UK
3Medical Research Council University of Glasgow Centre for Virus Research,
Glasgow, UK
4Manchester Academic Health Science Centre, Manchester, UK
5Department of Public Health and Policy, University of Liverpool, Liverpool, UK
6Nuffield Department of Clinical Medicine, JSARI Global Support Centre, Centre for
Tropical Medicine and Global Health, University of Oxford, Oxford, UK
7Public Health England National Infection Service, Salisbury, UK
8National Heart and Lung Institute, Imperial College London, London, UK
9Liverpool Clinical Trials Centre, University of Liverpool, Liverpool, UK
10Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK
11NIHR Health Protection Research Unit, Institute of Infection, Veterinary and
Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool,
Liverpool, UK
12Division of Epidemiology and Public Health, University of Nottingham, Nottingham,
UK
13Nuffield Department of Medicine, Centre for Tropical Medicine and Global Health,
University of Oxford, Oxford, UK
14Queen’s Medical Research Institute, University of Edinburgh, Edinburgh, UK
15Usher Institute of Population Health Sciences and Informatics, University of
Edinburgh, Edinburgh, UK
16Health Data Research UK, London, UK
17Department of Child Life and Health, University of Edinburgh, Edinburgh, UK
18Clinical Infection, Microbiology and Immunology, University of Liverpool Faculty of
Health and Life Sciences, Liverpool, UK
19Liverpool University Hospitals Foundation Trust, Member of Liverpool Health
Partners, Liverpool, UK
20Respiratory Department, Imperial College, London, UK
21Genetics and Genomics, Roslin Institute, University of Edinburgh, Edinburgh, UK
22Respiratory Medicine, Alder Hey Children’s Hospital, University of Liverpool,
Liverpool, UK
23Division of Infection and Immunity, University College London, London, UK

Twitter Stephen R Knight @srknight1, Antonia Ho @DrToniHo, Aziz Sheikh
@DrAzizSheikh, J Kenneth Baillie @kennethbaillie and Malcolm G Semple @
ProfCalumSemple

Acknowledgements This work uses data provided by patients and collected
by the NHS as part of their care and support #DataSavesLives. We are extremely
grateful to the 2648 frontline NHS clinical and research staff and volunteer medical
students, who collected this data in challenging circumstances; and the generosity of
the participants and their families for their individual contributions in these difficult
times. We also acknowledge the support of Jeremy J Farrar, Nahoko Shindo, Devika
Dixit, Nipunie Rajapakse, Lyndsey Castle, Martha Buckley, Debbie Malden, Katherine

Knight SR, et al. Thorax 2021;0:1–10. doi:10.1136/thoraxjnl-2021-217629
Contributors
EMH is guarantor and corresponding author for this work and attests that all listed authors meet authorship criteria and that no other meetings the criteria have been omitted. Contributor role taxonomy (CRediT): conceptualisation: SRK, RKG, AH, RP, TD, CF, KAM, MP, LT, MN, EMH; funding acquisition: PH, JSN-V-T, TS, PJ0, JKB, MGS, AD; investigation: SRK, RKG, AH, CR, OS, LT, PJ0, JKB, MGS, AD, EMH; methodology: SRK, RKG, AH, RP, IB, KAM, CS, AS, CS, AD, EMH; project administration: SH, HH, CJ, LM, LN, CR, CS, MGS, resources: CG, PH, LM, TS, MGS, EMH, software: SRK, RKG, RP, KAM, OS; supervision: SRK, RKG, CR, CG, HH, PH, PLO, PJ0, JKB. MGS, EMH; RP, MP, EMH; writing—original draft: SRK, RG, AH, EMH; writing—review and editing: SRK, RG, AH, RP, IB, GC, TMD, JD, CF, LM, JN, JN-S-V-T, PLO, MP, CR, AS, CS, OS, LT, PJ0, JKB, MGS, AD, EMH.

Funding
This work is supported by grants from: the National Institute for Health Research (NIHR) (award CO-CIN-01), the Medical Research Council (grant MC_ PC-19059) and by the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool in partnership with Amsterdam University Health and England (PHE), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford (award 200 907), NIHR HPRU in Respiratory Infections at National Institute for Health Research University College London with PHE (award 200 927), Wellcome Trust and Department for International Development (215091/2/18/B2), and the Bill and Melinda Gates Foundation (OPP1209135), and Liverpool Experimental Cancer Medicine Centre (Grant Code: CR251/251535). Our thanks also to all those who contributed to the work described in this paper.

Provenance and peer review
Not commissioned; externally peer reviewed.

This work is licensed under a Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits any legal, non-commercial use of the work on this website (https://isaric4c.net/sample_access/). Other uses to copy, redistribute, remix, transform and build upon this work are permitted only if you provide a full attribution to the work as published on this website, in accordance with the Creative Commons Attribution 4.0 Unported license (CC BY 4.0).
REFERENCES


