Title page

Title

Development and validation of the 4C Deterioration model for adults hospitalised with COVID-19.

Authors

Rishi K Gupta[1], Ewen M Harrison[2,3], Antonia Ho[4,5], Annemarie B Docherty[2,6], Stephen R Knight[2], Maarten van Smeden[7], Ibrahim Abubakar[1], Marc Lipman[8,9], Matteo Quartagno[10], Riinu Pius[2], Iain Buchan[11], Gail Carson[12], Thomas M Drake[2], Jake Dunning[13,14], Cameron J Fairfield[2], Carrol Gamble[15], Christopher A Green[16], Sophie Halpin[17], Hayley E Hardwick[18], Karl A Holden[18], Peter W Horby[12], Clare Jackson[15], Kenneth A Mclean[2], Laura Merson[12], Jonathan S Nguyen-Van-Tam[19], Lisa Norman[2], Piero L Olliaro[20], Mark G Pritchard[20], Clark D Russell[21], James Scott-Brown[22], Catherine A Shaw[2], Aziz Sheikh[2], Tom Solomon[18,23], Cathie Sudlow[24], Olivia V Swann[25], Lance Turtle[18,26], Peter JM Openshaw*[14], J Kenneth Baillie*[6,27], Malcolm G Semple*[18,28], Mahdad Noursadeghi*[29] on behalf of the ISARIC4C Investigators.

**Contributed equally to this work*

Affiliations

1. Institute for Global Health, University College London, London, UK
2. Centre for Medical Informatics, The Usher Institute, University of Edinburgh, Edinburgh, UK
3. Department of Clinical Surgery, University of Edinburgh, Edinburgh, UK
4. Medical Research Council, University of Glasgow Centre for Virus Research, Glasgow, UK
5. Department of Infectious Diseases, Queen Elizabeth University Hospital, Glasgow, UK
6. Intensive Care Unit, Royal Infirmary Edinburgh, Edinburgh, UK
7. Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands
8. UCL Respiratory, Division of Medicine, University College London, London, UK
9. Royal Free Hospitals NHS Trust, London, UK
10. MRC Clinical Trials Unit, Institute of Clinical Trials and Methodology, University College London, London, UK
11. Institute of Population Health Sciences, University of Liverpool, Liverpool, UK
12. ISARIC Global Support Centre, Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK
13. National Infection Service, Public Health England, London, UK
14. National Heart and Lung Institute, Imperial College London, London, UK
15. Liverpool Clinical Trials Centre, University of Liverpool, Liverpool, UK

16. Institute of Microbiology & Infection, University of Birmingham, Birmingham, UK
17. Institute of Global Health, University College London, London, UK
18. NIHR Health Protection Research Unit, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK
19. Division of Epidemiology and Public Health, University of Nottingham School of Medicine, Nottingham, UK
20. Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK
21. Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK
22. School of Informatics, University of Edinburgh, Edinburgh, UK
23. Walton Centre NHS Foundation Trust, Liverpool, UK
24. Health Data Research UK, London, UK
25. Department of Child Life and Health, University of Edinburgh, Edinburgh, UK
26. Tropical & Infectious Disease Unit, Royal Liverpool University Hospital, Liverpool, UK
27. Roslin Institute, University of Edinburgh, Edinburgh, UK
28. Respiratory Medicine, Alder Hey Children's Hospital, Institute in The Park, University of Liverpool, Alder Hey Children's Hospital, Liverpool L12 2AP, UK
29. Division of Infection and Immunity, University College London, UK

Correspondence

For ISARIC4C study processes:

Malcolm G Semple (*http://orcid.org/0000-0001-9700-0418)* NIHR Health Protection Research Unit, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, L69 7BE, UK. Email: m.g.semple@liverpool.ac.uk (Twitter @ProfCalumSemple)

For model derivation:

Prof Mahdad Noursadeghi (*https://orcid.org/0000-0002-4774-0853*), Division of Infection & Immunity, Cruciform Building, University College London, London WC1E 6BT, United Kingdom. Email: m.noursadeghi@ucl.ac.uk

Abstract

Prognostic models to predict the risk of clinical deterioration in acute COVID-19 are required to inform clinical management decisions. Among 75,016 consecutive adults across England, Scotland and Wales prospectively recruited to the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) study, we developed and validated a multivariable logistic regression model for in-hospital clinical deterioration (defined as any requirement of ventilatory support or critical care, or death) using 11 routinely measured variables. We used internal-external cross-validation to show consistent measures of discrimination, calibration and clinical utility across eight geographical regions. We further validated the final model in held-out data from 8,252 individuals in London, with similarly consistent performance (C-statistic 0.77 (95% CI 0.75 to 0.78); calibration-in-the-large 0.01 (-0.04 to 0.06); calibration slope 0.96 (0.90 to 1.02)). Importantly, this model demonstrated higher net benefit than using other candidate scores to inform decision-making. Our 4C Deterioration model thus demonstrates unprecedented clinical utility and generalisability to predict clinical deterioration among adults hospitalised with COVID-19.

Introduction

The coronavirus disease 2019 (COVID-19) pandemic continues to overwhelm healthcare systems worldwide[1]. Effective triage of patients presenting to hospital for risk of progressive deterioration is critical to inform clinical decision-making and facilitate effective resource allocation. Moreover, early identification of higher risk subgroups enables targeted recruitment for randomised controlled trials of therapies with equipoise[2], and more precise delivery of treatments for which effectiveness is known to vary according to disease severity[3–5].

A large number of multivariable clinical prognostic models for patients with COVID-19 have rapidly accrued to predict adverse outcomes of mortality or clinical deterioration[6]. The vast majority of candidate models subjected to comprehensive quality assessment have been classified as being at high risk of bias, and therefore may not be generalisable[6,7]. Moreover, none of the multivariable prognostic models included in a systematic head-to-head external validation study outperformed univariable predictors[8], highlighting a critical need to adhere to rigorous model development methodology using large scale multi-site data, in order to facilitate generalisability.

We have previously reported a pragmatic prognostic index for in-hospital mortality from the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) study[9]. Here, we extend this work through a larger study cohort to develop and validate a prognostic model for in-hospital clinical deterioration. We use the wide geographic coverage of the ISARIC4C study cohort in England, Wales and Scotland to explore between-region heterogeneity, and to comprehensively assess model generalisability with respect to discrimination, calibration and clinical utility. We have called this the 4C Deterioration model.

Methods

Study population and data collection

The International Severe Acute Respiratory and emerging Infections Consortium (ISARIC) World Health Organization (WHO) Clinical Characterisation Protocol UK (CCP-UK) study is being conducted by the ISARIC Coronavirus Clinical Characterisation Consortium (ISARIC4C) in 260 hospitals across England, Scotland, and Wales (National Institute for Health Research Clinical Research Network Central Portfolio Management System ID 14152). Further details of this prospective cohort study have been reported previously[10]. In this analysis, we included consecutive adults (≥18 years) with highly suspected or confirmed COVID-19 in whom eligibility criteria were confirmed. The study is reported in accordance with transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD) guidance[11]. Demographic, clinical, and outcome data were collected through a standardised case record form as reported previously[9,10,12]. In this analysis, hospitals were classified and analysed by UK National Health Service (NHS) region[13].

Outcomes

We defined a composite primary outcome of in-hospital 'clinical deterioration' that includes: (a) initiation of ventilatory support (non-invasive ventilation, invasive mechanical ventilation or extra-corporeal membrane oxygenation); (b) high-dependency or intensive care unit admission; or (c) death. This composite outcome aligns closely with World Health Organization Clinical Progression Scale ≥6 and ensures that the outcome is generalisable between hospitals, since respiratory support practices may vary considerably[14]. We only included participants admitted or first assessed for COVID-19 until 26/08/2020 in order to allow a minimum four-week interval for registration of outcome events (until final data extraction date of 24/09/2020). Participants who had ongoing hospital care at the end of follow-up were classified as not meeting the endpoint, since the risk of deterioration declines with time since admission[8]. Missing outcomes were imputed, as described below.

Candidate predictors

We specified candidate predictor variables *a priori* based on previous prognostic scores, and emerging literature describing routinely measured biomarkers associated with COVID-19 prognosis, as described previously[9]. We only considered predictors available in at least 60% of the study population for further analysis. In addition to the variables included in our previous analysis, we

included nosocomial COVID-19 acquisition as an additional candidate predictor, since we hypothesised that acquisition of infection in hospital may be associated with differential outcomes. Community-acquired infection was defined as symptom onset or first positive SARS-CoV-2 PCR within 7 days from admission; participants who did not meet these criteria and had either symptom onset or first positive SARS-CoV-2 PCR >7 days from admission were classified as nosocomial cases[15]. Nosocomial cases who met the deterioration outcome prior to onset of COVID-19 were excluded.

Comorbidities were defined according to a modified Charlson comorbidity index[16], with the addition of clinician-defined obesity, as reported previously[9]. We considered a composite 'number of comorbidities' variable for inclusion in the model, which included the following co-morbidities: chronic cardiac disease, chronic respiratory disease (excluding asthma), chronic renal disease, mild to severe liver disease, dementia, chronic neurological disease, connective tissue disease, diabetes mellitus, HIV or AIDS, malignancy and clinician-defined obesity.

All predictors were taken from the day of hospital admission, or the day of first clinical suspicion of COVID-19 for patients with nosocomial infection. Continuous predictors were modelled with restricted cubic splines using a default of four knots, placed at recommended locations based on percentiles[17]. Glasgow coma scale was categorised as 15 *vs*. <15 since there were insufficient data points below 15 to fit spline functions.

Sample size

We assumed *a priori* that at least 30% of hospital admission would reach the primary outcome. The most comparable existing tools for risk stratification are the qSOFA, NEWS2, and CURB-65 scores[18–20]. qSOFA achieved C-statistics ranging from 0.71-0.78 during external validation[18]. Based on a conservative C-statistic of 0.71, we assumed a Cox-Snell R^2 of 0.24. With 40 candidate predictor parameters (including transformations), the minimum sample size required for new model development was estimated to be 1,290, with 387 events[21,22].

Missing data

We handled missing data (including predictors and outcomes) using multiple imputation with chained equations, assuming missingness at random[23], using the mice package in R[24]. We included all

predictors (including restricted cubic spline transformations) and the outcome in the imputation models to ensure compatibility. Imputation was done separately for each NHS region to preserve potential inter-region heterogeneity. We generated 10 multiply imputed datasets; all primary analyses were performed in each imputed dataset and model and validation parameters were pooled using Rubin's rules[25].

Model development

We hypothesised that heterogeneity among populations, healthcare services and clinical management may be present between NHS regions, which may contribute to differences in model performance; we therefore split the dataset according to the region of the contributing hospital. We included eight regions in model development and internal-external cross validation (East of England, the Midlands, North East England and Yorkshire, North West England, Scotland, South East England, South West England and Wales), with one region held out for further validation (London).

We used a logistic regression modelling approach and performed backward elimination of the *a priori* candidate variables using Akaike information criterion (AIC). This process was done separately in each multiply imputed dataset and in each NHS region in the development set. Predictors were required to be retained in >50% of multiply imputed datasets in >50% of development NHS regions in order to enter the final model. We specified this in order to retain a parsimonious set of predictors that had consistent prognostic value across the development NHS regions.

Internal-external cross-validation

The model including the selected variables was then validated in the development dataset using the internal-external cross-validation framework, in order to concurrently examine between-region heterogeneity and assess generalisability[26]. During each internal-external cross-validation cycle, one of the contributing NHS regions within the development set was iteratively discarded from model training. Validation was then evaluated in the omitted NHS region by quantifying the model C-statistic, calibration slope and calibration-in-the-large[27]. We used random-effects meta-analysis to calculate pooled C-statistic, calibration slope and calibration-in-the-large statistics from internal-external cross-validation, and forest plots were examined to assess heterogeneity between regions. Calibration plots were also generated for each internal-external cross-validation cycle by fitting a loess smoother between the model predictions and the outcome in the stacked multiply imputed datasets.

Recalibration to each region was performed by re-estimating the model intercept in the validation sets during each internal-external cross-validation cycle.

Decision curve analysis was also done in the internal-external cross-validation validation sets to quantify the net benefit of implementing the model in clinical practice[28], compared to: (a) a 'treat all' approach; (b) a 'treat none' approach; and (c) using other candidate generic and COVID-specific clinical prognostic models to stratify treatment, identified by recent systematic reviews[6,8,9]. Only candidate models where constituent variables were available among >60% of the cohort were considered. Candidate models using points scores were calibrated to the validation data during decision curve analysis, resulting in optimistic estimates of their net benefit. All decision curves were loess-smoothed, from stacked multiply imputed datasets.

Model validation in held-out NHS region

The final model was then trained using the full development dataset and validation was further evaluated in the held-out NHS region (London) by quantifying the C-statistic, calibration slope and calibration-in-the-large, and by visualisation of calibration plots[27]. Decision curve analysis was also performed, with comparison to other candidate models.

All analyses were conducted in R (version 3.6.3) using tidyverse (version 1.3.0)[29], rms (version 6.0-1)[30] for logistic regression modelling, mice (version 3.11.0)[24] for multiple imputation and rmda (version 1.6)[31] for decision curve analysis.

Sensitivity analyses

We assessed validation of the final model using complete case data only in the held-out NHS region. We also recalculated validation metrics when: (a) excluding participants who experienced the outcome on the day of admission, in order to assess discrimination of the final model without these early events; (b) excluding participants in the validation cohort who had ongoing hospital care at the end of follow-up; (c) stratifying the validation cohort by community *vs.* nosocomial infection; and (d) excluding community-acquired cases who developed symptoms in the interval between admission and the temporal threshold for nosocomial infection, in order to assess any effect of incorrect inclusion of nosocomial infections within the community acquired cases. We also repeated the analysis using an alternative multiple imputation approach, using the aregImpute function from the

rms package in R[30], and recalculated model parameters using alternative temporal definitions of nosocomial SARS-CoV-2 infection (>5 days and >10 days after admission, compared to >7 days in the primary analysis).

Ethical approval

Ethical approval was given by the South Central-Oxford C Research Ethics Committee in England (reference 13/SC/0149), and by the Scotland A Research Ethics Committee (reference 20/SS/0028). The study was registered at https://www.isrctn.com/ISRCTN66726260.

Results

Overview of study cohort

A total of 75,016 adults were recruited to the ISARIC4C study during the study period. Baseline demographic, physiological and laboratory characteristics of the cohort are shown stratified by outcome in Table 1 and by community *vs.* nosocomial infection in Supplementary Table 1. A total of 31,993/74,018 (43.2%) participants with outcomes available met the composite primary outcome of clinical deterioration during hospital admission. The interval between hospital admission and deterioration events, stratified by deterioration category, is shown in Supplementary Figure 1 (median time to deterioration 4 days; interquartile range 1-9). The overall risk of deterioration generally declined with increasing time from admission, supporting our approach to classify patients requiring ongoing hospital care at the end of follow-up as not meeting an endpoint in the primary analysis. Outcomes were missing for 998/75,016 (1.3%) participants.

Development of the 4C Deterioration model

Since we hypothesised that geographic heterogeneity may contribute to model performance, we analysed the dataset by UK National Health Service (NHS) region. We included eight regions in model development (sample sizes range 3,068 to 15,599; total n=66,764), with one region (London; n=8,252) held out for additional validation. Candidate predictors and their proportions of missingness, stratified by NHS region, are shown in Supplementary Figure 2. Proportions of missingness appeared similar across regions for each variable.

Following our backward elimination procedure, 11 predictors were retained in >50% of multiply imputed datasets in >50% of NHS regions in the development cohort. These were: age, sex, nosocomial infection, Glasgow coma scale, admission oxygen saturation, breathing room air or oxygen therapy, respiratory rate, urea, C-reactive protein, lymphocyte count and presence of radiographic chest infiltrates. Associations (including non-linearities) between these predictors and the outcome from the model trained on the full development cohort are shown visually in Figure 1. Full model coefficients are presented to enable independent model reconstruction in Supplementary Table 2.

Internal-external cross-validation

In order to examine potential heterogeneity between NHS regions and evaluate generalisability, we conducted internal-external cross-validation[26] of the prognostic model in the development cohort. In this process, we iteratively excluded one of the eight contributing NHS regions from the development set; the model was then trained using the 11 selected predictors in the remaining 7 development regions before being validated in the omitted region.

Forest plots showing model discrimination (C-statistic) and calibration metrics (slope and calibration-in-the-large (CITL) from internal-external cross-validation are shown in Figure 2. An ideal calibration slope is 1, while CITL should be 0 if the number of observed outcome events matches the number predicted. C-statistics were consistent across development NHS regions (point estimates 0.75-0.77; pooled random effects meta-analysis estimate 0.76; 95% confidence interval (95% CI) 0.75 to 0.77). Calibration slopes were also consistent across regions, with little evidence of heterogeneity (point estimates 0.95-1.06; pooled estimate 0.99; 95% CI 0.97 to 1.02). There was minor heterogeneity across NHS regions in CITL, likely reflecting some variation in baseline risk between regions (point estimates -0.20 to 0.13; pooled estimate -0.01; 95% CI -0.11 to 0.09). Overall risk was slightly underestimated in South East England (CITL 0.09; 95% CI 0.04 to 0.14) and Wales (CITL 0.13; 95% CI 0.06 to 0.21), and overestimated in Scotland (CITL -0.20; 95% CI -0.28 to -0.11) and South West England (CITL -0.19; 95% CI -0.26 to -0.11). Pooled calibration plots by NHS region are shown in Supplementary Figure 3a.

In view of the minor variation in CITL between NHS regions, we also repeated internal-external cross validation with recalibration to each NHS region by re-estimation of the model intercept; calibration plots with recalibrated intercepts confirmed small improvements in model calibration (Supplementary Figure 3b).

Decision curve analyses in the validation sets in internal-external cross-validation, without recalibration of the new model, are shown in Supplementary Figure 4, with benchmarking to 11 existing candidate prognostic models for which the constituent variables were available in >60% of participants in our data. In decision curve analysis, net benefit allows assessment of clinical utility by quantifying the trade-off between correctly identifying true positives and incorrectly identifying false positives weighted according to the threshold probability[28]. The threshold probability represents the

risk cut-off above which any given treatment or intervention might be considered, and reflects the underlying risk:benefit ratio for the intervention. The new model for clinical deterioration had higher net benefit than any of the existing models as well as 'treat all' or 'treat none' strategies, across a broad range of threshold probabilities, in all development NHS regions (without local recalibration).

Validation in held-out NHS region

Next, we validated the final prognostic model, trained on the full development cohort, in the held-out NHS region (London; n=8,252). Discrimination and calibration metrics for the 4C Deterioration model were similar to the estimates from internal-external cross validation (Table 2), with C-statistic 0.77 (95% CI 0.75 to 0.78), CITL 0.01 (-0.04 to 0.06) and slope 0.96 (0.90 to 1.02). Discrimination was higher for the 4C Deterioration model than for the other existing candidates. A loess-smoothed calibration curve for the held-out London region is shown in Figure 3a.

We then conducted decision curve analysis in the held-out NHS region to further examine clinical utility for the 4C Deterioration model. Importantly, this demonstrated higher net benefit than all other candidates that we were able to recreate, as well as the 'treat all' and 'treat none' approaches, across a range of threshold probabilities (Figure 3b).

We anticipate that clinicians may wish to evaluate risk of deterioration or death separately. Therefore, for illustration, we compared predictions from the 4C Deterioration model to our previously reported 4C Mortality Score[9] in the London validation cohort, stratified by age (Figure 4a). In addition, 10 example participants selected at random from each decile of 4C Deterioration predictions in the London cohort are shown in Figure 4b, with their clinical characteristics summarised in Figure 4c. Overall, deterioration predictions appeared appropriately higher than those for mortality, but these differences were exaggerated among younger age groups.

Sensitivity analyses

Recalculation of model validation metrics in complete case data from the held-out London region showed similar results to the primary analyses (Supplementary Table 3). Exclusion of deterioration events on the day of admission in the London validation cohort resulted in slightly lower C-statistics for most models; discrimination remained higher for the 4C Deterioration model, compared to other candidates (Supplementary Table 4). Validation metrics in the London cohort appeared similar to the

primary analysis when excluding participants who had ongoing hospital care at the end of follow-up (Supplementary Table 5), when restricted to community-acquired infections (Supplementary Table 6a), and when community-acquired infections with symptom onset after admission were excluded (Supplementary Table 7). Among nosocomial infections, the C-statistic appeared slightly lower for the 4C Deterioration model than in the primary analysis (0.72; 95% CI 0.67 to 0.77), though discrimination remained higher than other candidate models, and CITL was 0.39 (95% CI 0.2 to 0.59), suggesting some underestimation of risk (Supplementary Table 6b). Repeating the analyses using an alternative multiple imputation approach and with shorter and longer temporal definitions of nosocomial infection led to similar results to the primary analysis (Supplementary Figures 5-8; Supplementary Table 8).

Discussion

In this study, we developed and validated a prognostic model for clinical deterioration in a population of 75,016 consecutive adults hospitalised with COVID-19 and recruited to the ISARIC4C study across 260 hospitals in England, Scotland and Wales. The final model integrates 11 predictors that are routinely measured in clinical practice. The model was initially validated through internal-external cross-validation, demonstrating consistent discrimination, calibration and net benefit across NHS regions. The final model was then further validated in the held-out London region and demonstrated consistent performance. Importantly, the 4C Deterioration model achieved higher net benefit than other candidate risk stratification tools across a broad range of risk thresholds - in all NHS regions. Thus, the 4C Deterioration model demonstrates strong potential for clinical utility and generalisability.

Our 4C Deterioration model can be implemented programmatically alongside our previously reported 4C Mortality Score[9]. The comparison of 4C Deterioration *vs.* 4C Mortality predictions among the London validation cohort showed higher predictions for deterioration overall; these differences were amplified among younger age groups. This suggests that younger people who deteriorated were more likely to have escalation of treatment through HDU/ICU admission or ventilatory support, while older people who deteriorated were more likely to die. These observations are likely to be mediated, in part, by differential treatment escalation decisions by age. Moreover, our comparison of the models for 10 randomly selected patients across the distribution of outcome risks from the held-out validation cohort illustrates examples of cases with relatively low risk of death, but moderate to high risk of deterioration. These discordances underline the need for independent prognostic models for deterioration and mortality outcomes, thus empowering clinicians to predict their desired outcome as required to inform clinical management decisions.

Our study has a number of strengths. Previous studies seeking to develop prognostic models for people with COVID-19 have been evaluated as being at high risk of bias due to suboptimal development methodology, and are often limited to single hospital sites[6], thus impeding generalisability during external validation[8]. In the current study, we adhered to TRIPOD standards[11] and retained continuous variables without arbitrary categorisation, while accounting for linearities, to avoid loss of information[32]. Moreover, we used the largest dataset to date, to our knowledge, to develop and validate the 4C Deterioration model, including data from hospitals across nine NHS

regions in England, Scotland and Wales. We exploited this wide geographic coverage to explore between-region heterogeneity in model performance using the recommended approach of internal-external cross-validation[33]. While discrimination, calibration slopes and net benefit were largely very consistent, we noted minor variation in CITL, suggesting some variation in baseline risk between regions. Our approach of recalibrating the model intercept to each NHS region demonstrated the potential to address such heterogeneity and could be used to update the model if risk is found to vary temporally (as novel therapies are implemented) and among different populations. However, it is notable that net benefit, which accounts for model discrimination and calibration in quantifying clinical utility, appeared higher for the 4C Deterioration model than all other candidates even without recalibration, across all NHS regions and in the held-out validation dataset. This was the case even when comparing to points-based models, which may achieve overly optimistic performance in during decision curve analysis since they were recalibrated to the validation data set. We also used a robust approach to missing data with multiple imputation, as widely recommended in prediction model studies[34], and performed a sensitivity analysis using an alternative approach, with similar findings.

Ongoing prospective external validation of the 4C Deterioration model will be required over time to consider the need for temporal recalibration[35], and to include diverse international settings outside of the ISARIC4C study. Another limitation is that we only included predictors that were routinely measured as part of clinical care during the study period, and specified that they had to be available among >60% of the population for inclusion in the analysis. Thus, we were unable to assess candidate models that include predictors such as lactate dehydrogenase or D-dimer, since these variables were only available in a small minority of participants. Future studies could consider standardised capture of laboratory measurements considered to have prognostic value to enable inclusion of these variables in model development and validation at scale. Moreover, we note that novel molecular biomarkers currently under investigation may also offer prognostic value[36]. Blood transcript, protein and metabolite measurements will be available from a subset of the ISARIC4C participants and could be integrated into risk-stratification tools in future studies.

In summary, we present a prognostic model for clinical deterioration among hospitalised adults with community or hospital acquired COVID-19, validated in nine NHS regions in England, Scotland and Wales. The model uses readily available clinical predictors and will be made freely available online

alongside our previously reported mortality risk score (https://isaric4c.net/outputs/4c_score/)[9] at the point of peer-reviewed publication, to inform clinical decision-making and patient stratification for therapeutic interventions. The underlying model coefficients are presented and code will be published to enable independent external validation in new datasets.

Footnotes

ISARIC4C Investigators

Consortium Lead Investigator: J Kenneth Baillie, Chief Investigator: Malcolm G Semple, Co-Lead Investigator: Peter JM Openshaw. ISARIC Clinical Coordinator: Gail Carson. Co-Investigators: Beatrice Alex, Benjamin Bach, Wendy S Barclay, Debby Bogaert, Meera Chand, Graham S Cooke, Annemarie B Docherty, Jake Dunning, Ana da Silva Filipe, Tom Fletcher, Christopher A Green, Ewen M Harrison, Julian A Hiscox, Antonia Ying Wai Ho, Peter W Horby, Samreen Ijaz, Saye Khoo, Paul Klenerman, Andrew Law, Wei Shen Lim, Alexander J Mentzer, Laura Merson, Alison M Meynert, Mahdad Noursadeghi, Shona C Moore, Massimo Palmarini, William A Paxton, Georgios Pollakis, Nicholas Price, Andrew Rambaut, David L Robertson, Clark D Russell, Vanessa Sancho-Shimizu, Janet T Scott, Thushan de Silva, Louise Sigfrid, Tom Solomon, Shiranee Sriskandan, David Stuart, Charlotte Summers, Richard S Tedder, Emma C Thomson, AA Roger Thompson, Ryan S Thwaites, Lance CW Turtle, Maria Zambon. Project Managers: Hayley Hardwick, Chloe Donohue, Ruth Lyons, Fiona Griffiths, Wilna Oosthuyzen. Data Analysts: Lisa Norman, Riinu Pius, Tom M Drake, Cameron J Fairfield, Stephen Knight, Kenneth A Mclean, Derek Murphy, Catherine A Shaw. Data and Information System Managers: Jo Dalton, James Lee, Daniel Plotkin, Michelle Girvan, Egle Saviciute, Stephanie Roberts, Janet Harrison, Laura Marsh, Marie Connor, Sophie Halpin, Clare Jackson, Carrol Gamble. Data integration and presentation: Gary Leeming, Andrew Law, Murray Wham, Sara Clohisey, Ross Hendry, James Scott-Brown. Material Management: William Greenhalf, Victoria Shaw, Sarah McDonald. Patient engagement: Seán Keating. Outbreak Laboratory Staff and Volunteers: Katie A. Ahmed, Jane A Armstrong, Milton Ashworth, Innocent G Asiimwe, Siddharth Bakshi, Samantha L Barlow, Laura Booth, Benjamin Brennan, Katie Bullock, Benjamin WA Catterall, Jordan J Clark, Emily A Clarke, Sarah Cole, Louise Cooper, Helen Cox, Christopher Davis, Oslem Dincarslan, Chris Dunn, Philip Dyer, Angela Elliott, Anthony Evans, Lorna Finch, Lewis WS Fisher, Terry Foster, Isabel Garcia-Dorival, Willliam Greenhalf, Philip Gunning, Catherine Hartley, Antonia Ho, Rebecca L Jensen, Christopher B Jones, Trevor R Jones, Shadia Khandaker, Katharine King, Robyn T. Kiy, Chrysa Koukorava, Annette Lake, Suzannah Lant, Diane Latawiec, L Lavelle-Langham, Daniella Lefteri, Lauren Lett, Lucia A Livoti, Maria Mancini, Sarah McDonald, Laurence McEvoy, John McLauchlan, Soeren Metelmann, Nahida S Miah, Joanna Middleton, Joyce Mitchell, Shona C Moore, Ellen G Murphy, Rebekah Penrice-Randal, Jack Pilgrim, Tessa Prince, Will Reynolds, P. Matthew Ridley,

Debby Sales, Victoria E Shaw, Rebecca K Shears, Benjamin Small, Krishanthi S Subramaniam, Agnieska Szemiel, Aislynn Taggart, Jolanta Tanianis-Hughes, Jordan Thomas, Erwan Trochu, Libby van Tonder, Eve Wilcock, J. Eunice Zhang. Local Principal Investigators: Kayode Adeniji, Daniel Agranoff, Ken Agwuh, Dhiraj Ail, Ana Alegria, Brian Angus, Abdul Ashish, Dougal Atkinson, Shahedal Bari, Gavin Barlow, Stella Barnass, Nicholas Barrett, Christopher Bassford, David Baxter, Michael Beadsworth, Jolanta Bernatoniene, John Berridge, Nicola Best, Pieter Bothma, David Brealey, Robin Brittain-Long, Naomi Bulteel, Tom Burden, Andrew Burtenshaw, Vikki Caruth, David Chadwick, Duncan Chambler, Nigel Chee, Jenny Child, Srikanth Chukkambotla, Tom Clark, Paul Collini, Catherine Cosgrove, Jason Cupitt, Maria-Teresa Cutino-Moguel, Paul Dark, Chris Dawson, Samir Dervisevic, Phil Donnison, Sam Douthwaite, Ingrid DuRand, Ahilanadan Dushianthan, Tristan Dyer, Cariad Evans, Chi Eziefula, Chrisopher Fegan, Adam Finn, Duncan Fullerton, Sanjeev Garg, Sanjeev Garg, Atul Garg, Effrossyni Gkrania-Klotsas, Jo Godden, Arthur Goldsmith, Clive Graham, Elaine Hardy, Stuart Hartshorn, Daniel Harvey, Peter Havalda, Daniel B Hawcutt, Maria Hobrok, Luke Hodgson, Anil Hormis, Michael Jacobs, Susan Jain, Paul Jennings, Agilan Kaliappan, Vidya Kasipandian, Stephen Kegg, Michael Kelsey, Jason Kendall, Caroline Kerrison, Ian Kerslake, Oliver Koch, Gouri Koduri, George Koshy, Shondipon Laha, Steven Laird, Susan Larkin, Tamas Leiner, Patrick Lillie, James Limb, Vanessa Linnett, Jeff Little, Michael MacMahon, Emily MacNaughton, Ravish Mankregod, Huw Masson, Elijah Matovu, Katherine McCullough, Ruth McEwen, Manjula Meda, Gary Mills, Jane Minton, Mariyam Mirfenderesky, Kavya Mohandas, Quen Mok, James Moon, Elinoor Moore, Patrick Morgan, Craig Morris, Katherine Mortimore, Samuel Moses, Mbiye Mpenge, Rohinton Mulla, Michael Murphy, Megan Nagel, Thapas Nagarajan, Mark Nelson, Igor Otahal, Mark Pais, Selva Panchatsharam, Hassan Paraiso, Brij Patel, Natalie Pattison, Justin Pepperell, Mark Peters, Mandeep Phull, Stefania Pintus, Jagtur Singh Pooni, Frank Post, David Price, Rachel Prout, Nikolas Rae, Henrik Reschreiter, Tim Reynolds, Neil Richardson, Mark Roberts, Devender Roberts, Alistair Rose, Guy Rousseau, Brendan Ryan, Taranprit Saluja, Aarti Shah, Prad Shanmuga, Anil Sharma, Anna Shawcross, Jeremy Sizer, Manu Shankar-Hari, Richard Smith, Catherine Snelson, Nick Spittle, Nikki Staines, Tom Stambach, Richard Stewart, Pradeep Subudhi, Tamas Szakmany, Kate Tatham, Jo Thomas, Chris Thompson, Robert Thompson, Ascanio Tridente, Darell Tupper-Carey, Mary Twagira, Andrew Ustianowski, Nick Vallotton, Lisa Vincent-Smith, Shico Visuvanathan, Alan Vuylsteke, Sam Waddy, Rachel Wake, Andrew Walden, Ingeborg Welters, Tony Whitehouse,

Paul Whittaker, Ashley Whittington, Meme Wijesinghe, Martin Williams, Lawrence Wilson, Sarah Wilson, Stephen Winchester, Martin Wiselka, Adam Wolverson, Daniel G Wooton, Andrew Workman, Bryan Yates, and Peter Young.

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 2,648 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 Newell, Kwame O'Neill, Emmanuelle Denis, Claire Petersen, Scott Mullaney, Sue MacFarlane, Nicole Maziere, Julien Martinez, Oslem Dincarslan, and Annette Lake.

Author contributions

Conceptualisation: RKG, EMH, AH, ABD, PJMO, JKB, MGS, MN;

Data curation: RP, SH, KAH, CJ, CAG, KAM, LM, LN, CDR, CAS, MGS, ABD, EMH;

Formal analysis: RKG, EMH, MN;

Funding acquisition: PWH, JSN-V-T, TS, PJMO, JKB, MGS, ABD;

Investigation: RKG, EMH, AH, ABD, PJMO, JKB, MGS, MN;

Methodology: RKG, EMH, AH, ABD, MVS, IA, ML, MQ, PJMO, JKB, MGS, MN;

Project administration: SH, HEH, CJ, LM, LN, CDR, CAS, MGS;

Resources: CG, PWH, LM, TS, MGS, EMH;

Software: N/A;

Supervision: RKG, EMH, AH, ABD, IA, ML, PJMO, JKB, MGS, MN;

Visualisation: RKG, EMH, JS-B, MN;

Writing-original draft: RKG, MN;

Writing-review and editing: RKG, EMH, AH, ABD, MVS, IA, ML, MQ, SRK, RP, IB, GC, TMD, CAG, JD, CJF, LM, JSN-V-T, PLO, MGP, CDR, JS-B, AS, CS, OVS, LCWT, PJMO, JKB, MGS, MN.

Funding

This work is supported by grants from: the National Institute for Health Research (NIHR; award CO-CIN-01), the Medical Research Council (MRC; 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 Public Health England (PHE), in collaboration with Liverpool School of Tropical Medicine and the University of Oxford (award 200907), NIHR HPRU in Respiratory Infections at Imperial College London with PHE (award 200927), Wellcome Trust and Department for International Development (DID; 215091/Z/18/Z), the Bill and Melinda Gates Foundation (OPP1209135), Liverpool Experimental Cancer Medicine Centre (grant reference C18616/A25153), NIHR Biomedical Research Centre at Imperial College London (IS-BRC-1215-20013), EU Platform for European Preparedness Against (Re-)emerging Epidemics (PREPARE; FP7 project 602525), and NIHR Clinical Research Network for providing infrastructure support for this research.

RKG is funded by National Institute for Health Research (DRF-2018-11-ST2-004 to RKG). MN is funded by the Wellcome Trust (207511/Z/17/Z). IA acknowledges NIHR Senior Investigator Award (NF-SI-0616-10037) funding. ML acknowledges NIHR (HTA 16/88/06) funding. PJMO is supported by a NIHR senior investigator award (201385). LT is supported by the Wellcome Trust (award 205228/Z/16/Z). MN is funded by a WT investigator award (207511/Z/17/Z) and the NIHR University College London Hospitals Biomedical Research Centre.

The views expressed are those of the authors and not necessarily those of the Department of Health and Social Care, DID, NIHR, MRC, Wellcome Trust, or PHE. The funders had no role in the study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the article for publication.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: support from the National Institute for Health Research (NIHR), the Medical Research Council (MRC), the NIHR Health Protection Research Unit (HPRU) in Emerging and Zoonotic Infections at University of Liverpool, Public Health England (PHE), Liverpool School of Tropical Medicine, University of Oxford, NIHR HPRU in Respiratory Infections at Imperial College London, Wellcome Trust, Department for International Development, the Bill and Melinda Gates Foundation,

Liverpool Experimental Cancer Medicine Centre, NIHR Biomedical Research Centre at Imperial College London, EU Platform for European Preparedness Against (Re-)emerging Epidemics (PREPARE), and NIHR Clinical Research Network for the submitted work; ABD reports grants from Department of Health and Social Care (DHSC), during the conduct of the study, grants from Wellcome Trust, outside the submitted work; CAG reports grants from DHSC National Institute of Health Research (NIHR) UK, during the conduct of the study; PWH reports grants from Wellcome Trust, Department for International Development, Bill and Melinda Gates Foundation, NIHR, during the conduct of the study; JSN-V-T reports grants from DHSC, England, during the conduct of the study, and is seconded to DHSC, England; MN is supported by a Wellcome Trust investigator award and the NIHR University College London Hospitals Biomedical Research Centre (BRC); PJMO reports personal fees from consultancies and from European Respiratory Society, grants from MRC, MRC Global Challenge Research Fund, EU, NIHR BRC, MRC/GSK, Wellcome Trust, NIHR (Health Protection Research Unit (HPRU) in Respiratory Infection), and is NIHR senior investigator outside the submitted work; his role as President of the British Society for Immunology was unpaid but travel and accommodation at some meetings was provided by the Society; JKB reports grants from MRC UK; MGS reports grants from DHSC NIHR UK, grants from MRC UK, grants from HPRU in Emerging and Zoonotic Infections, University of Liverpool, during the conduct of the study, other from Integrum Scientific LLC, Greensboro, NC, USA, outside the submitted work; LCWT reports grants from HPRU in Emerging and Zoonotic Infections, University of Liverpool, during the conduct of the study, and grants from Wellcome Trust outside the submitted work; EMH, HEH, JD, RG, RP, LN, KAH, GC, LM, SH, CJ, and CG, all declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

Data sharing statement

We welcome applications for data and material access via our Independent Data and Material Access Committee (https://isaric4c.net).

Code availability statement

The final prognostic model developed in this study will be made freely available at the point of peer-reviewed publication, to enable implementation in clinical practice and independent external validation in new datasets. The code underlying the prediction tool will also be made available.

References

1. World Health Organization. Coronavirus disease 2019 (COVID-19) Weekly Epidemiological Update. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports (2020).
2. Furlow, B. COVACTA trial raises questions about tocilizumab's benefit in COVID-19. *Lancet Rheumatol.* (2020) doi:10.1016/S2665-9913(20)30313-1.
3. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Preliminary Report. *N. Engl. J. Med.* NEJMoa2007764 (2020) doi:10.1056/NEJMoa2007764.
4. Sterne, J. A. C. *et al.* Association Between Administration of Systemic Corticosteroids and Mortality Among Critically Ill Patients With COVID-19. *JAMA* (2020) doi:10.1001/jama.2020.17023.
5. Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *N. Engl. J. Med.* NEJMoa2021436 (2020) doi:10.1056/NEJMoa2021436.
6. Wynants, L. *et al.* Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* **369**, (2020).
7. Wolff, R. F. *et al.* PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Ann. Intern. Med.* **170**, 51 (2019).
8. Gupta, R. K. *et al.* Systematic evaluation and external validation of 22 prognostic models among hospitalised adults with COVID-19: An observational cohort study. *Eur. Respir. J.* (2020) doi:10.1183/13993003.03498-2020.
9. Knight, S. 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* m3339 (2020) doi:10.1136/bmj.m3339.
10. Docherty, A. B. *et al.* Features of 20 133 UK patients in hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: prospective observational cohort study. *BMJ* **369**, m1985 (2020).
11. Collins, G. S., Reitsma, J. B., Altman, D. G. & Moons, K. G. M. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ* **350**, (2015).
12. COVID-19 Clinical Research Resources - ISARIC. https://isaric.tghn.org/covid-19-clinical-research-resources/.
13. NHS Regional Teams. https://www.england.nhs.uk/about/regional-area-teams/.
14. WHO Working Group on the Clinical Characterisation and Management of COVID-19 infection, J. C. *et al.* A minimal common outcome measure set for COVID-19 clinical research. *Lancet. Infect. Dis.* **0**, (2020).
15. European Centre for Disease Prevention and Control. Surveillance definitions for COVID-19. https://www.ecdc.europa.eu/en/covid-19/surveillance/surveillance-definitions.
16. Charlson, M. E., Pompei, P., Ales, K. L. & MacKenzie, C. R. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.* **40**, 373–83 (1987).
17. Frank E Harrell. Biostatistical Modeling. http://biostat.mc.vanderbilt.edu/wiki/pub/Main/BioMod/notes.pdf (2004).
18. Seymour, C. W. *et al.* Assessment of Clinical Criteria for Sepsis. *JAMA* **315**, 762 (2016).
19. Royal College of Physicians. National Early Warning Score (NEWS) 2 | RCP London. https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2.
20. Lim, W. S. *et al.* Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. *Thorax* **58**, 377–382 (2003).
21. Riley, R. D. *et al.* Minimum sample size for developing a multivariable prediction model: PART II - binary and time-to-event outcomes. *Stat. Med.* **38**, 1276–1296 (2019).
22. Riley, R. D. *et al.* Calculating the sample size required for developing a clinical prediction model. *BMJ* **368**, m441 (2020).
23. White, I. R., Royston, P. & Wood, A. M. Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.* **30**, 377–399 (2011).
24. Buuren, S. van & Groothuis-Oudshoorn, K. mice: Multivariate Imputation by Chained Equations in R. *J. Stat. Softw.* **45(3)**, 1–67 (2011).
25. Rubin, D. B. *Multiple imputation for nonresponse in surveys*. (Wiley-Interscience, 2004).
26. Debray, T. P. A. *et al.* Individual Participant Data (IPD) Meta-analyses of Diagnostic and Prognostic Modeling Studies: Guidance on Their Use. *PLOS Med.* **12**, e1001886 (2015).
27. Steyerberg, E. W. & Vergouwe, Y. Towards better clinical prediction models: seven steps for

development and an ABCD for validation. *Eur. Heart J.* **35**, 1925–31 (2014).
28. Vickers, A. J., van Calster, B. & Steyerberg, E. W. A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic Progn. Res.* **3**, 18 (2019).
29. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
30. Harrell Jr, F. E. rms: Regression Modeling Strategies. (2019).
31. Brown, M. rmda: Risk Model Decision Analysis. (2018).
32. Royston, P., Altman, D. G. & Sauerbrei, W. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat. Med.* **25**, 127–41 (2006).
33. Steyerberg, E. W. & Harrell, F. E. Prediction models need appropriate internal, internal–external, and external validation. *J. Clin. Epidemiol.* **69**, 245–247 (2016).
34. Riley, R. D., Windt, D. van der, Croft, P. & Moons, K. G. M. *Prognosis research in healthcare□: concepts, methods, and impact*.
35. Booth, S., Riley, R. D., Ensor, J., Lambert, P. C. & Rutherford, M. J. Temporal recalibration for improving prognostic model development and risk predictions in settings where survival is improving over time. *Int. J. Epidemiol.* (2020) doi:10.1093/ije/dyaa030.
36. Del Valle, D. M. *et al.* An inflammatory cytokine signature predicts COVID-19 severity and survival. *Nat. Med.* (2020) doi:10.1038/s41591-020-1051-9.
37. Smith, G. B., Prytherch, D. R., Meredith, P., Schmidt, P. E. & Featherstone, P. I. The ability of the National Early Warning Score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* **84**, 465–470 (2013).
38. Zhang, H. *et al.* Risk prediction for poor outcome and death in hospital in-patients with COVID-19: derivation in Wuhan, China and external validation in London, UK. *medRxiv* 2020.04.28.20082222 (2020) doi:10.1101/2020.04.28.20082222.
39. Olsson, T., Terent, A. & Lind, L. Rapid Emergency Medicine score: a new prognostic tool for in-hospital mortality in nonsurgical emergency department patients. *J. Intern. Med.* **255**, 579–587 (2004).
40. Dwyer, R., Hedlund, J., Henriques-Normark, B. & Kalin, M. Improvement of CRB-65 as a prognostic tool in adult patients with community-acquired pneumonia. *BMJ Open Respir. Res.* **1**, e000038 (2014).
41. Miyashita, N., Matsushima, T., Oka, M. & Japanese Respiratory Society. The JRS guidelines for the management of community-acquired pneumonia in adults: an update and new recommendations. *Intern. Med.* **45**, 419–28 (2006).
42. Subbe, C. P., Kruger, M., Rutherford, P. & Gemmel, L. Validation of a modified Early Warning Score in medical admissions. *QJM* **94**, 521–526 (2001).
43. Seymour, C. W. *et al.* Assessment of Clinical Criteria for Sepsis. *JAMA* **315**, 762 (2016).
44. Lu, J. *et al.* ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. *medRxiv* 2020.02.20.20025510 (2020) doi:10.1101/2020.02.20.20025510.

Tables

Table 1: Baseline characteristics of the study cohort, stratified by outcome

Participants are shown by the first chronological deterioration category through which they met the composite primary outcome (high dependency unit (HDU) admission, intensive care (ICU) admission or ventilatory support, or death).

Characteristic	Overall, N = 75,016	Ventilatory support or HDU/ICU, N = 15,089[1]	Died, N = 16,904[1]	No deterioration, N = 42,025[1]	(Missing), N = 998[1]
Age (years)	75 (60, 84)	65 (55, 75)	83 (77, 89)	73 (57, 83)	75 (59, 84)
Sex					
Female	32,846 (44%)	5,156 (34%)	7,114 (42%)	20,142 (48%)	434 (44%)
Male	42,026 (56%)	9,910 (66%)	9,753 (58%)	21,800 (52%)	563 (56%)
Missing	144	23	37	83	1
Ethnicity					
White	55,079 (83%)	9,985 (75%)	13,628 (90%)	30,855 (83%)	611 (79%)
South Asian	3,520 (5.3%)	1,010 (7.6%)	479 (3.1%)	1,992 (5.4%)	39 (5.1%)
Black	2,554 (3.8%)	744 (5.6%)	345 (2.3%)	1,435 (3.9%)	30 (3.9%)
East Asian	492 (0.7%)	162 (1.2%)	71 (0.5%)	255 (0.7%)	4 (0.5%)
Other	4,848 (7.3%)	1,406 (11%)	699 (4.6%)	2,656 (7.1%)	87 (11%)
Missing	8,523	1,782	1,682	4,832	227
Number of comorbidities	1 (1, 2)	1 (0, 2)	2 (1, 3)	1 (0, 2)	1 (0, 2)
Missing	840	68	203	406	163
Nosocomial infection	7,385 (9.9%)	592 (3.9%)	2,111 (13%)	4,538 (11%)	144 (17%)
Missing	636	62	101	345	128
Radiographic infiltrates	29,594 (62%)	8,427 (77%)	6,965 (63%)	14,015 (55%)	187 (54%)
Missing	27,245	4,132	5,929	16,534	650
Temperature (°C)	37.2 (36.5, 38.1)	37.5 (36.8, 38.4)	37.1 (36.4, 38.0)	37.1 (36.5, 38.0)	37.0 (36.5, 38.0)
Missing	3,110	463	650	1,783	214
Heart rate (per min)	90 (78, 104)	95 (82, 109)	90 (77, 105)	88 (76, 102)	89 (78, 102)
Missing	3,389	424	721	2,021	223
Respiratory rate (per min)	20 (18, 26)	24 (20, 30)	22 (18, 28)	20 (18, 24)	20 (18, 24)
Missing	3,537	483	689	2,113	252
Systolic blood pressure (mmHg)	130 (114, 147)	129 (115, 145)	128 (110, 147)	130 (115, 147)	130 (114, 148)
Missing	3,192	428	651	1,891	222
Diastolic blood pressure (mmHg)	74 (64, 84)	74 (65, 83)	71 (61, 82)	75 (65, 84)	73 (65, 83)
Missing	3,335	460	693	1,955	227
SpO2 (%)	95 (92, 97)	94 (89, 96)	95 (91, 97)	96 (94, 97)	96 (93, 98)
Missing	3,760	538	802	2,203	217
Room air or oxygen					
Room air	48,626 (69%)	7,249 (50%)	9,991 (63%)	30,810 (79%)	576 (76%)
Oxygen	21,469 (31%)	7,141 (50%)	5,827 (37%)	8,324 (21%)	177 (24%)
Missing	4,921	699	1,086	2,891	245
Glasgow coma scale	15 (15, 15)	15 (15, 15)	15 (15, 15)	15 (15, 15)	15 (15, 15)
Missing	7,849	1,762	1,697	4,021	369

Characteristic	Overall, N = 75,016	Ventilatory support or HDU/ICU, N = 15,089[1]	Died, N = 16,904[1]	No deterioration, N = 42,025[1]	(Missing), N = 998[1]
Haemoglobin (g/L)	128 (112, 142)	132 (116, 145)	122 (105, 138)	129 (113, 142)	128 (112, 141)
Missing	11,781	1,422	2,568	7,449	342
White cell count (x10^9/L)	7.5 (5.4, 10.7)	8.0 (5.7, 11.2)	8.3 (5.8, 11.9)	7.1 (5.2, 9.9)	7.7 (5.4, 10.7)
Missing	12,163	1,515	2,675	7,624	349
Lymphocytes (x10^9/L)	0.90 (0.60, 1.30)	0.80 (0.60, 1.20)	0.80 (0.50, 1.16)	0.96 (0.67, 1.40)	0.90 (0.60, 1.30)
Missing	12,377	1,558	2,706	7,764	349
Neutrophils (x10^9/L)	5.8 (3.9, 8.7)	6.4 (4.3, 9.3)	6.6 (4.3, 9.9)	5.3 (3.6, 7.9)	5.8 (3.7, 8.7)
Missing	12,341	1,557	2,708	7,727	349
Platelets (x10^9/L)	221 (167, 291)	219 (166, 287)	209 (154, 284)	226 (173, 294)	218 (161, 285)
Missing	12,496	1,562	2,736	7,848	350
Alanine aminotransferase (IU/L)	25 (16, 43)	33 (21, 54)	22 (15, 37)	24 (15, 40)	24 (15, 43)
Missing	26,783	3,895	6,300	16,086	502
Bilirubin (mg/dL)	10 (7, 14)	10 (7, 15)	10 (7, 15)	9 (6, 13)	10 (7, 14)
Missing	22,979	3,143	5,317	14,026	493
Urea (mmol/L)	7 (5, 11)	7 (5, 11)	10 (7, 16)	6 (4, 9)	7 (5, 12)
Missing	18,554	2,794	3,974	11,345	441
Creatinine (µmol/L)	86 (67, 121)	86 (68, 118)	106 (76, 158)	81 (64, 107)	85 (68, 122)
Missing	12,693	1,636	2,706	7,999	352
Sodium (mmol/L)	137 (134, 140)	136 (133, 139)	138 (135, 143)	137 (134, 140)	137 (134, 140)
Missing	12,312	1,515	2,631	7,811	355
C-reactive protein (mg/L)	80 (33, 154)	126 (64, 210)	98 (48, 174)	58 (22, 119)	76 (30, 151)
Missing	16,353	2,223	3,515	10,203	412
NHS region					
East of England	7,865 (10%)	1,653 (11%)	1,935 (11%)	4,223 (10%)	54 (5.4%)
London	8,252 (11%)	2,287 (15%)	1,510 (8.9%)	4,400 (10%)	55 (5.5%)
Midlands	15,599 (21%)	2,559 (17%)	3,703 (22%)	9,068 (22%)	269 (27%)
North East and Yorkshire	10,310 (14%)	2,236 (15%)	2,225 (13%)	5,773 (14%)	76 (7.6%)
North West	12,930 (17%)	2,178 (14%)	3,296 (19%)	7,311 (17%)	145 (15%)
Scotland	3,068 (4.1%)	606 (4.0%)	573 (3.4%)	1,846 (4.4%)	43 (4.3%)
South East	9,450 (13%)	2,131 (14%)	1,974 (12%)	5,052 (12%)	293 (29%)
South West	3,919 (5.2%)	726 (4.8%)	795 (4.7%)	2,361 (5.6%)	37 (3.7%)
Wales	3,623 (4.8%)	713 (4.7%)	893 (5.3%)	1,991 (4.7%)	26 (2.6%)

[1]Statistics presented: median (IQR); n (%)

Table 2: Validation performance in held-out London region.

Models are shown for prediction of in-hospital clinical deterioration and are sorted by C-statistic (total sample size = 8252 participants). CITL = calibration-in-the-large. 'Original outcome' column indicates original intended outcome for each candidate model during development. CITL and slopes are not shown for points score models since they are not on probability scale.

Score	Original outcome	C-statistic	CITL	Slope
4C Deterioration	Deterioration (in-hospital)	0.77 (0.75 - 0.78)	0.01 (-0.04 - 0.06)	0.96 (0.9 - 1.02)
NEWS2[37]	Deterioration (1 day)	0.69 (0.68 - 0.7)		
Zhang "death"[38]	Mortality (in-hospital)	0.67 (0.66 - 0.69)	2.26 (2.19 - 2.32)	0.19 (0.16 - 0.22)
4C Mortality[9]	Mortality (in-hospital)	0.67 (0.66 - 0.68)		
Zhang "poor"[38]	Deterioration (in-hospital)	0.67 (0.66 - 0.68)	0.54 (0.48 - 0.6)	0.15 (0.13 - 0.18)
REMS[39]	Mortality (in-hospital)	0.66 (0.65 - 0.68)		
DS-CURB65[40]	Mortality (30 days)	0.66 (0.65 - 0.67)		
A-DROP[41]	Mortality (30 days)	0.65 (0.64 - 0.66)		
CURB65[20]	Mortality (30 days)	0.65 (0.64 - 0.66)		
MEWS[42]	Deterioration (in-hospital)	0.63 (0.62 - 0.64)		
qSOFA[43]	Mortality (in-hospital)	0.63 (0.62 - 0.64)		
Lu[44]	Mortality (12 days)	0.62 (0.61 - 0.63)		

Figure 1: Multivariable associations between selected predictors and outcome in final model

Variable selection was done in each imputed dataset using backward elimination within each NHS region using AIC. Variables retained in >50% of multiply imputed datasets in >50% of NHS regions were selected. Continuous variables are modelled using restricted cubic splines. Final model parameters are pooled across multiply imputed datasets (total sample size for model development = 66764 participants). Black lines indicate point estimates; grey shaded regions indicate 95% confidence intervals.

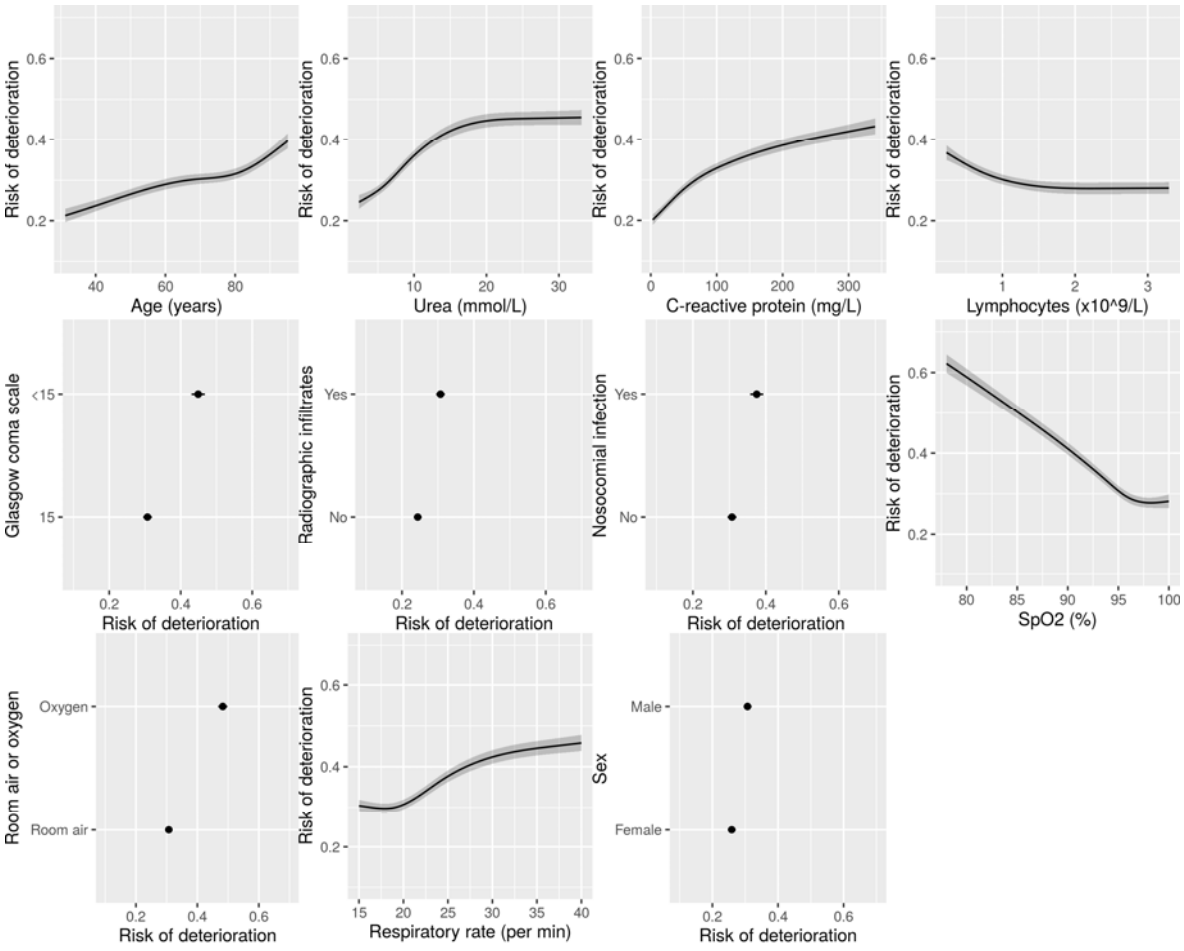


Figure 2: Internal-external cross validation of selected model by NHS region.

Pooled estimates are calculated through random-effects meta-analysis (total sample size = 66,764 participants). Dashed lines indicate lines of perfect calibration in the large (0) and slope (1), respectively. Black squares indicate point estimates; bars indicate 95% confidence intervals; diamonds indicate pooled random-effects meta-analysis estimates.

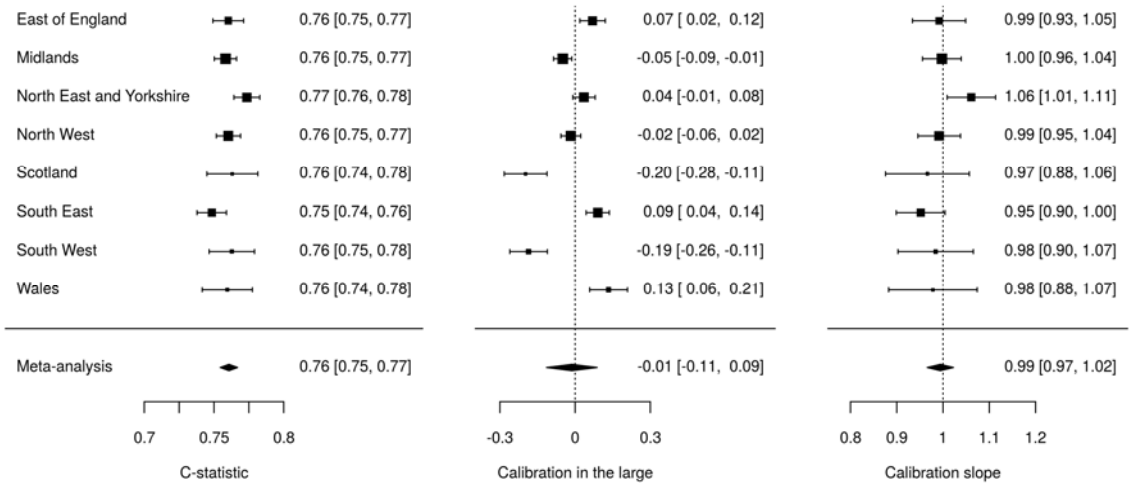


Figure 3: Calibration and decision curve analysis in held-out London region

Calibration (a) is shown using a loess-smoother across multiply imputed datasets. The rug plot indicates the distribution of predicted risk. Net benefit (b) is shown with loess smoothing for each candidate model compared to the 'treat all' and 'treat none' approaches. Points score models are recalibrated to the validation data, resulting in optimistic estimates of net benefit for these models. Total sample size = 8,252 participants).

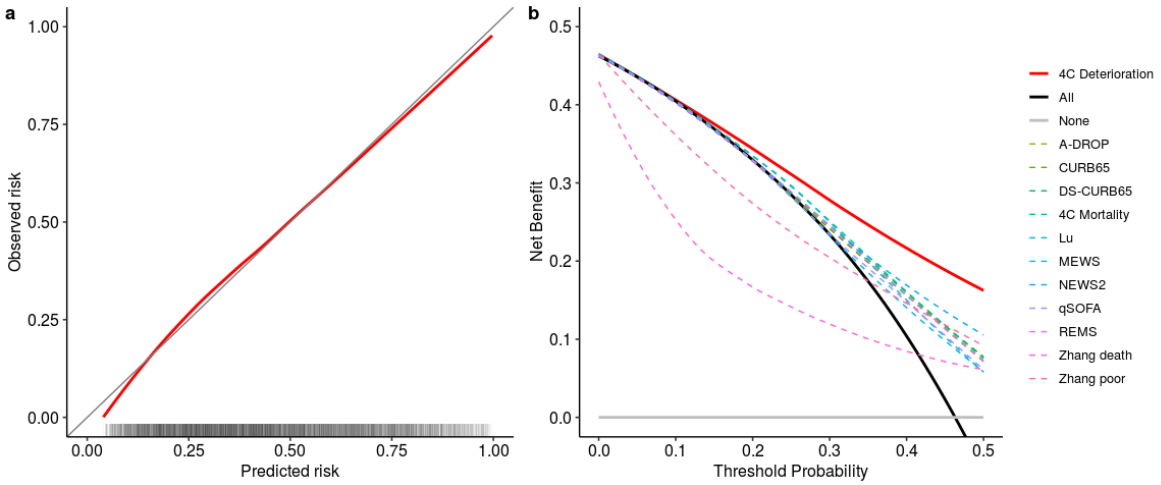
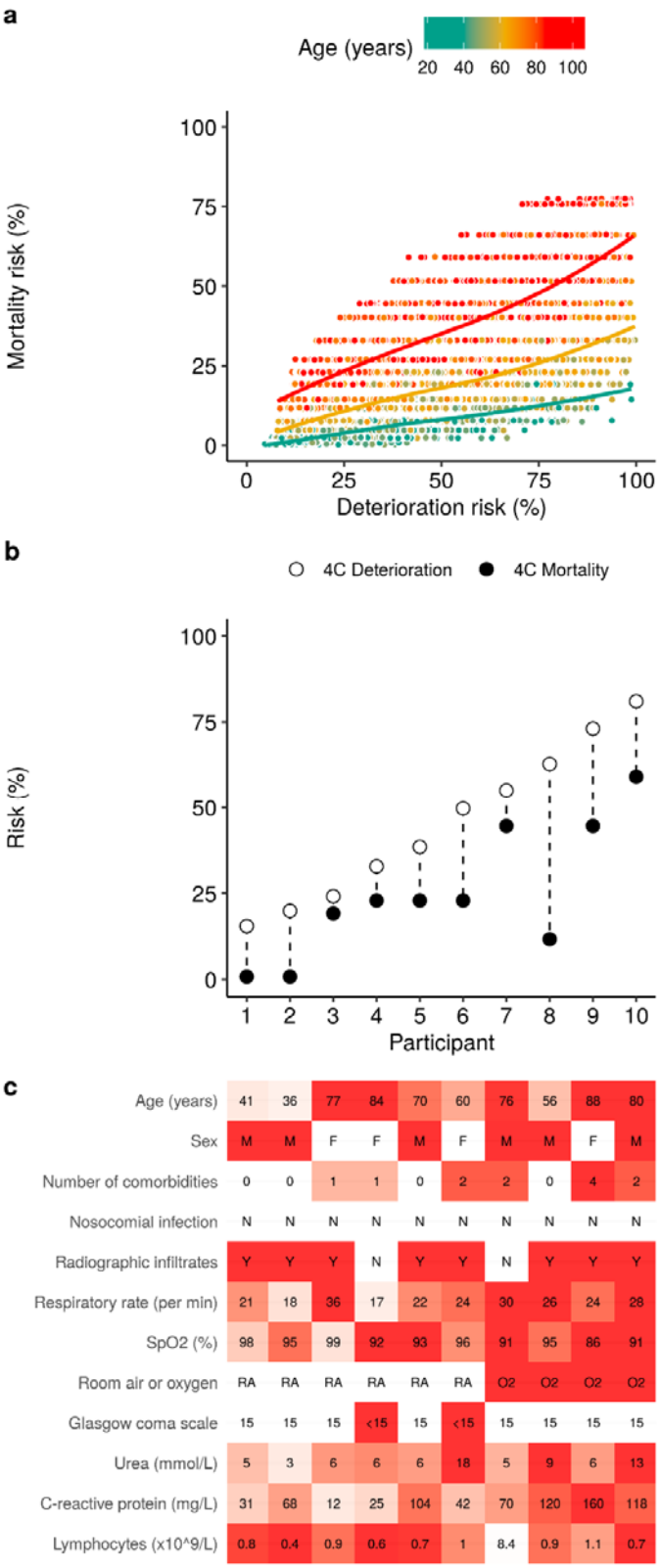
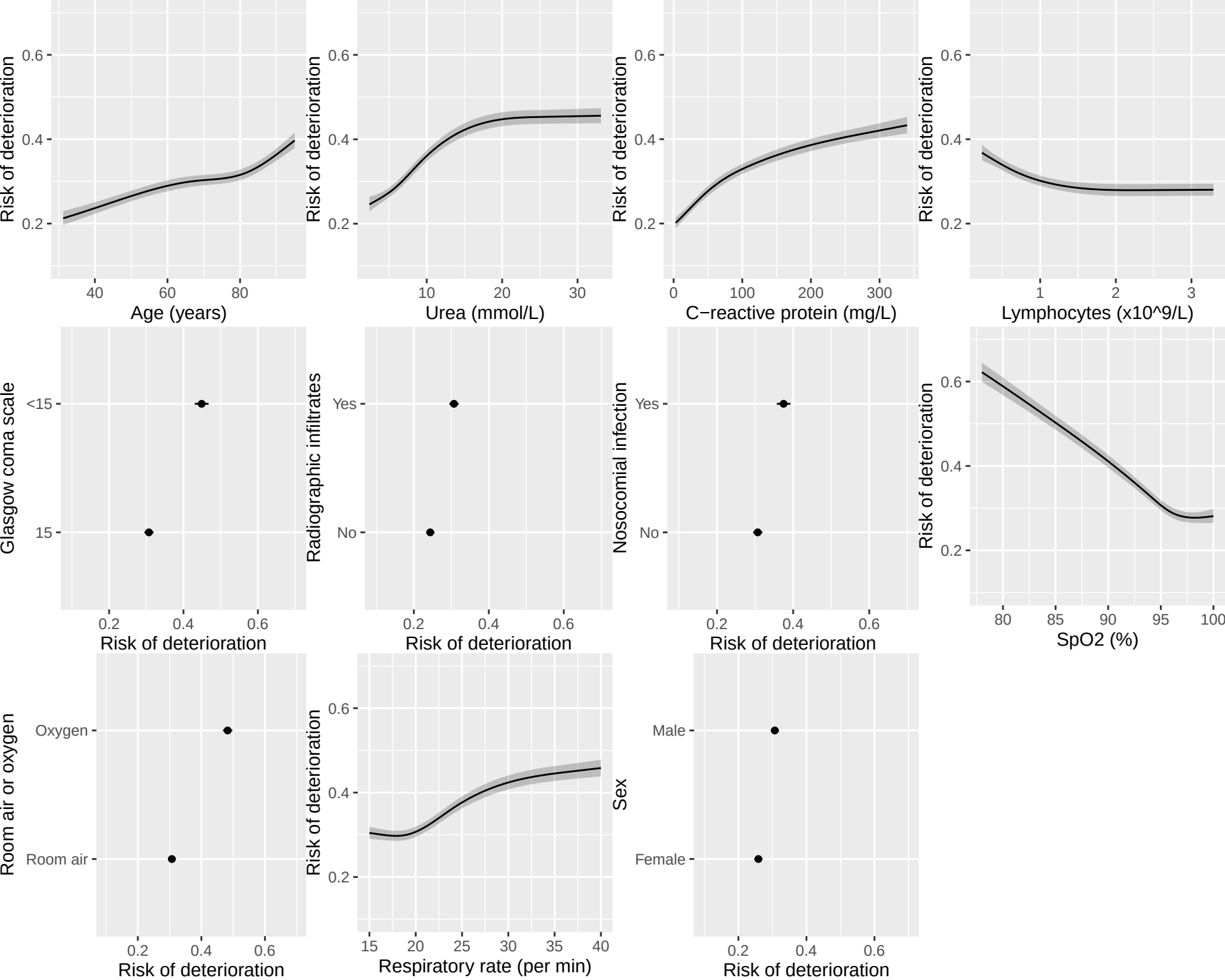
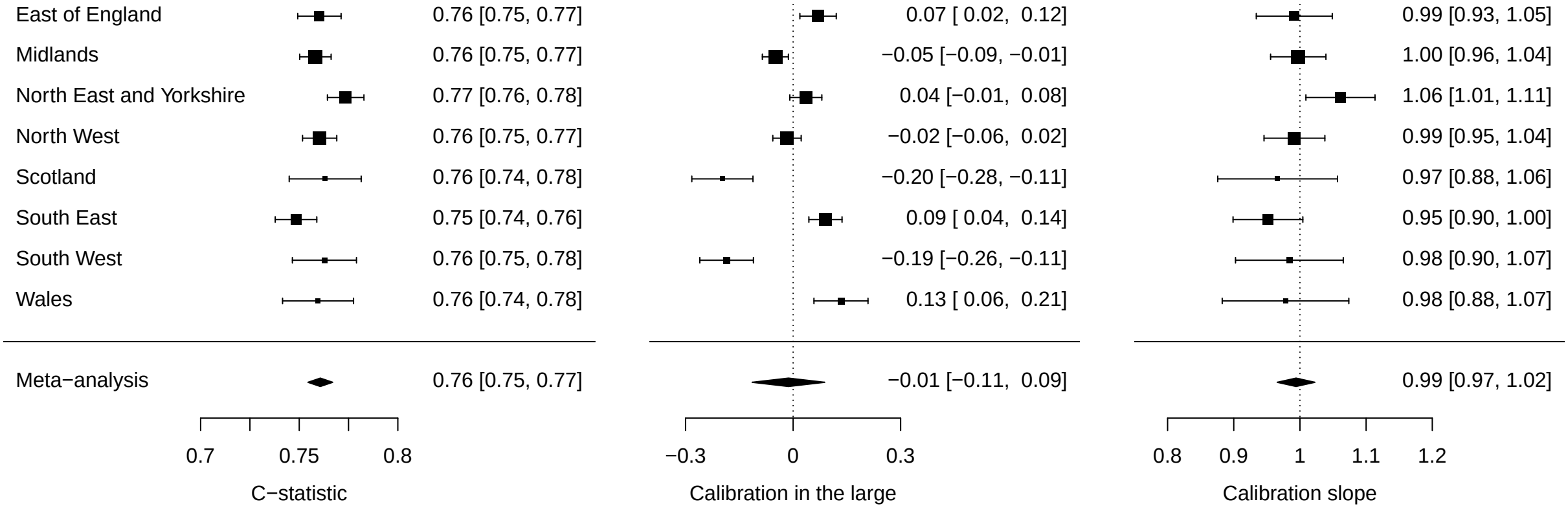


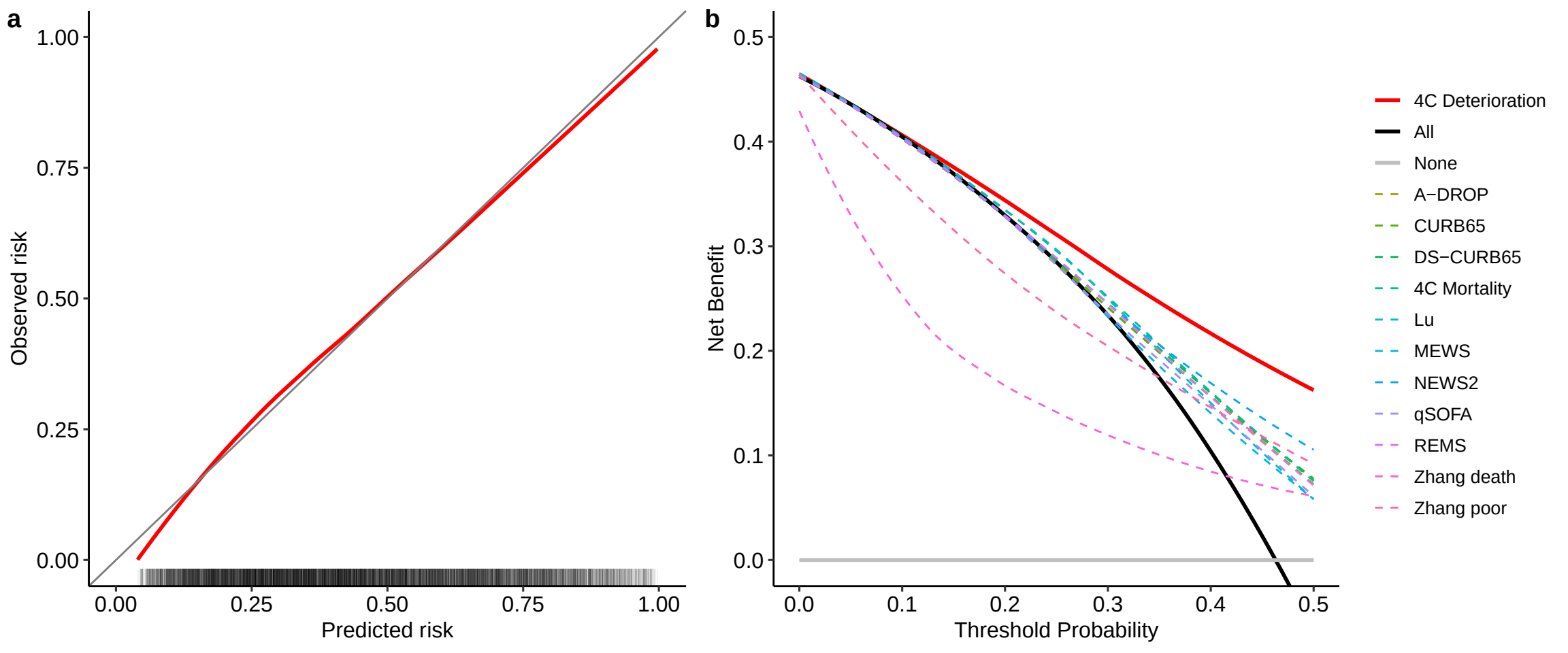
Figure 4: 4C Deterioration vs Mortality predictions for (a) London validation cohort (n = 8252) and (b-c) randomly sampled example patients.

4C Mortality probabilities are calculated from points scores, based on observed mortality risk for each score in the original validation data. In panel (a) smoothed plot reflects loess fit, stratified by age (green = under 50 years; yellow = 50 to 69; red = 70 or older). In panel (b), example patients are randomly sampled from the validation cohort, stratified by deciles of 4C Deterioration model predictions. Characteristics of each example participant are shown in (c), with red indicating characteristics associated with higher risk predictions. In (c), M = male; F = female; O2 = receiving oxygen therapy; RA = room air; Y = yes; N = no.

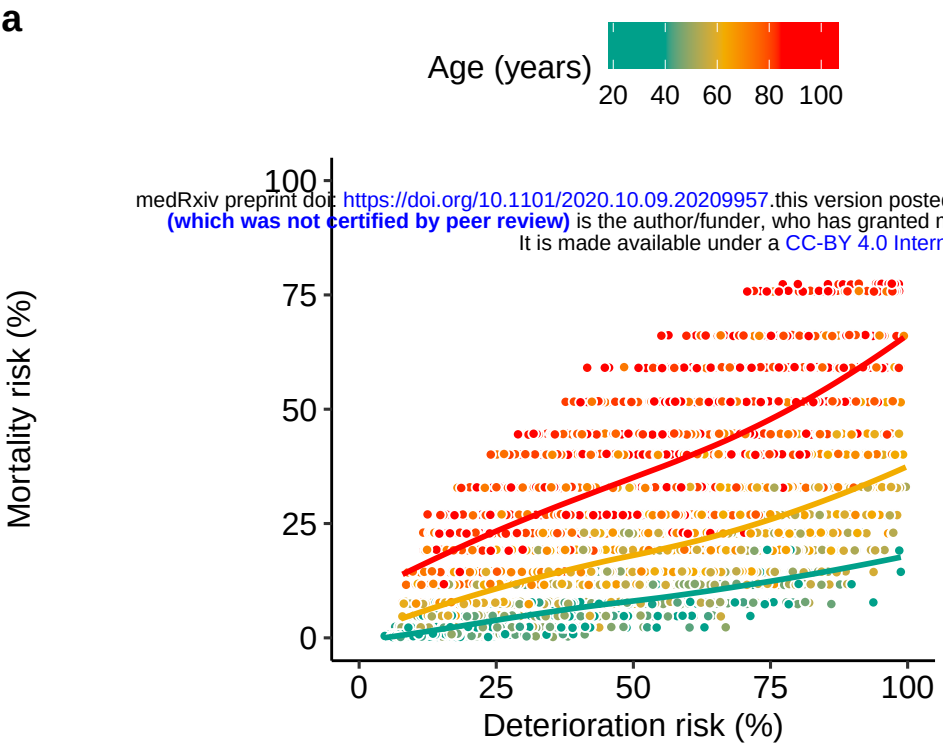



Risk of deterioration
0.6
0.4
0.2
40
60
80
Age (years)
Risk of deterioration
0.6
0.4
0.2
10
20
30
Urea (mmol/L)
Risk of deterioration
0.6
0.4
0.2
0
100
200
300
C−reactive protein (mg/L)
Risk of deterioration
0.6
0.4
0.2
1
2
3
Lymphocytes (x10^9/L)
Glasgow coma scale
<15
15
0.2
0.4
0.6
Risk of deterioration
Radiographic infiltrates
Yes
No
0.2
0.4
0.6
Risk of deterioration
Nosocomial infection
Yes
No
0.2
0.4
0.6
Risk of deterioration
Risk of deterioration
0.6
0.4
0.2
80
85
90
95
100
SpO2 (%)
Room air or oxygen
Oxygen
Room air
0.2
0.4
0.6
Risk of deterioration
Risk of deterioration
0.6
0.4
0.2
15
20
25
30
35
40
Respiratory rate (per min)
Sex
Male
Female
0.2
0.4
0.6
Risk of deterioration

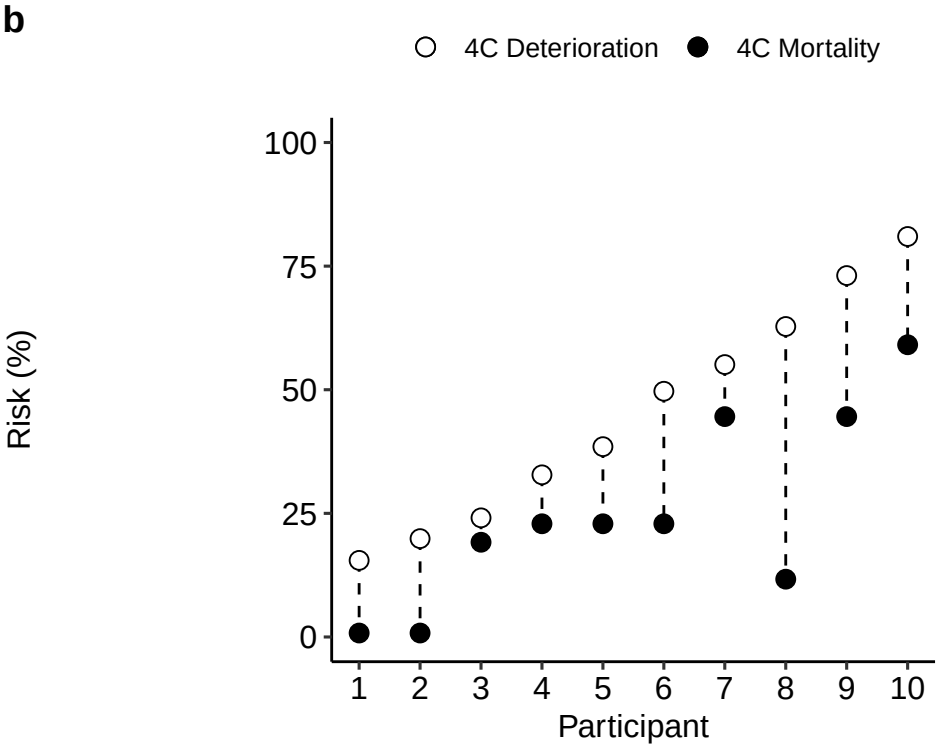

East of England
Midlands
North East and Yorkshire
North West
Scotland
South East
South West
Wales
Meta−analysis
0.76 [0.75, 0.77]
0.76 [0.75, 0.77]
0.77 [0.76, 0.78]
0.76 [0.75, 0.77]
0.76 [0.74, 0.78]
0.75 [0.74, 0.76]
0.76 [0.75, 0.78]
0.76 [0.74, 0.78]
0.76 [0.75, 0.77]
0.7
0.75
0.8
C−statistic
0.07 [0.02, 0.12]
−0.05 [−0.09, −0.01]
0.04 [−0.01, 0.08]
−0.02 [−0.06, 0.02]
−0.20 [−0.28, −0.11]
0.09 [0.04, 0.14]
−0.19 [−0.26, −0.11]
0.13 [0.06, 0.21]
−0.01 [−0.11, 0.09]
−0.3
0
0.3
Calibration in the large
0.99 [0.93, 1.05]
1.00 [0.96, 1.04]
1.06 [1.01, 1.11]
0.99 [0.95, 1.04]
0.97 [0.88, 1.06]
0.95 [0.90, 1.00]
0.98 [0.90, 1.07]
0.98 [0.88, 1.07]
0.99 [0.97, 1.02]
0.8
0.9
1
1.1
1.2
Calibration slope


a
1.00
0.75
0.50
0.25
0.00
Observed risk
0.00
0.25
0.50
0.75
1.00
Predicted risk
b
0.5
0.4
0.3
0.2
0.1
0.0
Net Benefit
0.0
0.1
0.2
0.3
0.4
0.5
Threshold Probability
4C Deterioration
All
None
A−DROP
CURB65
DS−CURB65
4C Mortality
Lu
MEWS
NEWS2
qSOFA
REMS
Zhang death
Zhang poor

a



b



c

Age (years)	41	36	77	84	70	60	76	56	88	80
Sex	M	M	F	F	M	F	M	M	F	M
Number of comorbidities	0	0	1	1	0	2	2	0	4	2
Nosocomial infection	N	N	N	N	N	N	N	N	N	N
Radiographic infiltrates	Y	Y	Y	N	Y	Y	N	Y	Y	Y
Respiratory rate (per min)	21	18	36	17	22	24	30	26	24	28
SpO2 (%)	98	95	99	92	93	96	91	95	86	91
Room air or oxygen	RA	RA	RA	RA	RA	RA	O2	O2	O2	O2
Glasgow coma scale	15	15	15	<15	15	<15	15	15	15	15
Urea (mmol/L)	5	3	6	6	6	18	5	9	6	13
C−reactive protein (mg/L)	31	68	12	25	104	42	70	120	160	118
Lymphocytes (x10^9/L)	0.8	0.4	0.9	0.6	0.7	1	8.4	0.9	1.1	0.7