

Title:

**Remote Covid Assessment in Primary Care (RECAP) risk prediction tool: derivation and real-world validation studies.**

Authors.

Ana Espinosa-Gonzalez\* PhD<sup>1</sup>, Denys Prociuk\*, MSc<sup>1</sup>, Francesca Fiorentino PhD<sup>1,7</sup>, Christian Ramtale MSc<sup>1</sup>, Ella Mi MBBChir<sup>1</sup>, Emma Mi MBBChir<sup>1</sup>, Ben Glampson<sup>2</sup>, Ana Luisa Neves PhD<sup>1</sup>, Cecilia Okusi, MRes<sup>3</sup>, Laiba Hussain<sup>3</sup>, Jack Macartney,<sup>3</sup> Martina Brown, MSc<sup>4</sup>, Ben Browne BM<sup>4</sup>, Caroline Warren<sup>4</sup>, Rachna Chowla MBA<sup>5</sup>, Jonty Heaversedge MBBS<sup>6</sup>, Trisha Greenhalgh PhD<sup>3</sup>, Simon de Lusignan MD<sup>3</sup>, Erik Mayer PhD<sup>1</sup>, Brendan Delaney MD<sup>1</sup>

Institutions

1. Institute of Global Health Innovation, Dept of Surgery and Cancer, Imperial College London, UK
2. Imperial College Healthcare NHS Trust, London, UK
3. Nuffield Dept of Primary Care, University of Oxford, UK
4. South Central Ambulance Service NHS Trust, UK
5. King's Health Partners, London, UK
6. South East London NHS Clinical Commissioning Group, London UK
7. Nightingale-Saunders Clinical Trials & Epidemiology Unit, King's Clinical Trials Unit, King's College London

\*Joint first authors

Corresponding author

Prof Brendan C Delaney  
Chair in Medical Informatics and Decision Making  
Imperial College  
Dept of Cancer and Surgery  
Room 506  
5th Floor Medical School  
St Mary's Hospital,  
Norfolk Place,  
London W2 1NY.  
[brendan.delaney@imperial.ac.uk](mailto:brendan.delaney@imperial.ac.uk)  
<http://www.imperial.ac.uk/people/brendan.delaney>

## Abstract

### Background

Accurate assessment of COVID-19 severity in the community is essential for best patient care and efficient use of services and requires a risk prediction score that is COVID-19 specific and adequately validated in a community setting. Following a qualitative phase to identify signs, symptoms and risk factors, we sought to develop and validate two COVID-19-specific risk prediction scores RECAP-GP (without peripheral oxygen saturation (SpO<sub>2</sub>)) and RECAP-O2 (with SpO<sub>2</sub>).

### Methods

Prospective cohort study using multivariable logistic regression for model development. Data on signs and symptoms (model predictors) were collected on community-based patients with suspected COVID-19 via primary care electronic health records systems and linked with secondary data on hospital admission (primary outcome) within 28 days of symptom onset. Data sources: RECAP-GP: Oxford-Royal College of General Practitioners Research and Surveillance Centre (RSC) primary care practices (development), Northwest London (NWL) primary care practices, NHS COVID-19 Clinical Assessment Service (CCAS) (validation). RECAP-O2: Doctaly Assist platform (development, and validation in subsequent sample). Estimated sample size was 2,880 per model.

### Findings

Data were available from 8,311 individuals. Observations, such as SpO<sub>2</sub>, were mostly missing in NWL, RSC, and CCAS data; however, SpO<sub>2</sub> was available for around 70% of Doctaly patients. In the final predictive models, RECAP-GP included sex, age, degree of breathlessness, temperature symptoms, and presence of hypertension (Area Under the Curve (AUC): 0.802, Validation Negative Predictive Value (NPV) of 'low risk' 98.8%. RECAP-O2 included age, degree of breathlessness, fatigue, and SpO<sub>2</sub> at rest (AUC: 0.843), Validation NPV of 'low risk' 99.4%.

### Interpretation

Both RECAP models are a valid tool in the assessment of COVID-19 patients in the community. RECAP-GP can be used initially, without need for observations, to identify patients who require monitoring. If the patient is monitored at home and SpO<sub>2</sub> is available, RECAP-O2 is useful to assess the need for further treatment escalation.

Keywords: risk prediction score; COVID-19; hospitalisation; general practice; primary care; remote consulting; out of hospital care

## **Funding**

The study was funded by the Community Jameel and the Imperial College President's Excellence Fund, the Economic and Social Research Council (ES/V010069/1), Health Data Research UK-Office of National Statistics Covid19 Awards. Infrastructure support from: the NIHR Imperial Biomedical Research Centre, the NIHR Oxford Biomedical Research Centre and the NIHR Imperial Patient Safety Translational Research Centre. This research was in part enabled by the Imperial Clinical Analytics Research and Evaluation (iCARE) environment and Whole System Integrated Care (WSIC) and used the iCARE and WSIC team and data resources.

## Research in context panel

### *Evidence before the study*

This study was conceived during the first COVID-19 wave in the UK (March - April 2020), as members of the research team contributed to the development of national clinical guidelines for COVID-19 management in the community and to the Oxford COVID-19 rapid review to track signs and symptoms of COVID-19 internationally. The review was carried out according to Cochrane Collaboration standards for rapid reviews and identified systematic reviews and large-scale observational studies describing the signs and symptoms of COVID-19. Evidence gathered showed worsening of COVID-19 symptoms around the 7th day of disease and challenges in identifying patients with higher likelihood of severity to increase their monitoring. To this end, tools such NEWS2 have been used in the UK to assess COVID-19 patients in primary care, but they do not capture the characteristics of COVID-19 infection and/or are not suitable for community remote assessment. Several COVID-19 risk scores have been developed. QCOVID provides a risk of mortality considering patients' existing risk factors but does not include acute signs and symptoms. ISARIC 4C Deterioration model has been specifically developed for hospital settings. In England, the NHS has implemented the Oximetry @home strategy to monitor patients with acute COVID-19 deemed at risk (older than 64 years old or with comorbidities) by providing pulse oximeters; however, the criteria for monitoring or for escalation of care have not been validated. There is, therefore, the need to develop a risk prediction score to establish COVID-19 patients' risk of deterioration to be used in the community for both face to face or remote consultation.

### *Added value of this study*

We developed and validated two COVID-19 specific risk prediction scores. One to be used in the initial remote assessment of patients with acute COVID-19 to assess need for monitoring (RECAP-GP). The second one to assess the need for further treatment escalation and includes peripheral saturation of oxygen among the model predictors (RECAP-O2). To our knowledge, this is the first COVID-19 specific risk prediction score to assess and monitor COVID-19 patients' risk of deterioration remotely. This will be a valuable resource to complement the use of oximetry in the community clinical decision-making when assessing a patient with acute COVID-19.

### *Implications of all available evidence*

To manage pandemic waves and their demand on healthcare, acute COVID-19 patients require close monitoring in the community and prompt escalation of their treatment. Guidance available so far relies on unvalidated tools and clinician judgement to assess deterioration. COVID-19 specific community-based risk prediction scores such as RECAP may contribute to reducing the uncertainty in the assessment and monitoring of COVID-19 patients, increase safety in clinical practice and improve outcomes by facilitating appropriate treatment escalation.

## Introduction

Since the start of the COVID-19 pandemic in 2020, a priority in health systems globally has been rapidly to develop and validate data-driven algorithms to predict risk and guide care. [1-3] This includes a requirement to accurately distinguish between COVID-19 positive patients who can be safely managed in the community setting and those whose care should be escalated to hospital. In primary care, patients with acute COVID-19 have been typically categorised as 'reassure and safety net', 'monitor' or 'admit to hospital' (referred to as green, amber and red risk categories respectively),[4] based on clinical judgment and consensus on using a few clinical parameters such as heart rate, pulse rate, and peripheral oxygen saturation (SpO<sub>2</sub>).

Several risk prediction models have been used to support the management of acute COVID-19. However, these were derived and validated in hospital inpatient populations.[5] The QCOVID score, based on primary care data, predicts risk of death from COVID-19 based on demographic factors and pre-existing medical conditions, but it does not predict acute deterioration based on current clinical observations.[6] The NEWS2 score is widely used in UK emergency care as an early warning score for sepsis. However, the parameters in NEWS2 (such as tachycardia, fever and hypotension) are usually very late signs of clinical deterioration; have been found to perform poorly in the acute assessment of suspected COVID-19 in hospital inpatients,[7] and have not been evaluated outside hospital.[8] As initial patient contacts with health services are increasingly carried out remotely, a score was needed that could be administered over the telephone or other remote means.[9] Instead of vital signs, such a score could be based on clinical symptoms which the patient or a relative could assess (e.g., perceived breathlessness or confusion). A Delphi study (with qualitative and survey components) using 112 primary care clinicians and 50 patients derived a set of data items comprising symptoms and vital signs that might be included in a putative 'RECAP: Remote Covid Assessment in Primary care' prediction model. [10] Templates for collection of these RECAP data elements (known as RECAP-V0), using appropriate SNOMED clinical terms, were developed for Electronic Health Record (EHR) systems so that a model could be derived and then validated.

This study aimed to develop and validate two prediction models: firstly, a score incorporating observable vital signs (heart rate, temperature, respiratory rate, and oxygen saturation); secondly, a score for use when these parameters cannot be measured for lack of equipment or patient familiarity. Two cut-off values, for need of monitoring (green/amber) and consideration of hospital admission (amber/red), for acute COVID-19 were derived and validated.

## Methods

### Study design

RECAP, a prospective cohort and observational study, recruited patients presenting in primary care with symptoms of acute COVID-19. Patients were followed for 28 days from onset of symptoms to determine the occurrence of COVID-related hospital admissions. The previously published study protocol and statistical analysis plan are summarised below. [11, 12]

### Data sources and settings

To allow for parallel derivation and validation of the RECAP scores on different cohorts, we used four different UK primary care settings as shown in Figure 1 below.

*Figure 1 here*

### Data

We used primary data on patients' signs and symptoms collected in the community at the point of consultation linked to secondary data on hospital outcomes. For data collection the RECAP-V0 electronic template, with selected SNOMED codes [10] was used in EHR systems (EMIS, TPP SystmOne and Aداstra) and completed by clinicians when assessing patients with signs and symptoms of COVID-19. Additional data on comorbidities, ethnicity, age, and sex, that were included in the development of the risk prediction model were available from the EHR. The Doctaly-Assist platform collected RECAP-V0 data elements via patient self-report (see below). Since the clinical question that needs to be supported is 'does this patient require care escalation?', hospital admission i.e, escalation of care from the community, defined as a night's hospital stay, was the main outcome to be predicted by the model. To derive outcomes, all data were linked with Hospital Episode Statistics (HES) on admissions from NHS Digital in a Trusted Research Environment (TRE), Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID) at Oxford University, or in iCARE (Imperial Clinical Analytics, Research and Evaluation) environments at Imperial College NHS Trust, where data was linked to the NWL 'SitRep' data on COVID-19 admissions. A study-specific SNOMED code carrying the RECAP National Institute of Health Research Portfolio Number was used to identify the relevant records for the study in iCARE and ORCHID.

**Data collection in general practice (NWL and RSC):** NWL and Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) [13, 14] primary care practices completed the RECAP-V0 electronic template in EMIS or SystmOne and captured the verbal consent of patients upon completion of the template. Data from NWL practices are routinely included in the 'Whole Systems Integrated Care' record and data were accessed in the iCARE

system. Data for the RSC practices are routinely extracted and made available for analysis in ORCHID.

**Data collection by COVID-19 Clinical Assessment Service (CCAS):** The CCAS was commissioned from South Central Ambulance Service by NHS England and Improvement to conduct COVID-19 patient remote assessment. A version of the RECAP-V0 electronic template was deployed for use by CCAS clinicians in the Adastra system. Consent was managed in the same way as in primary care practices. Data was transferred to ORCHID for linkage and analysis.

**Data collection by patients using Doctaly Assist:** The Doctaly Assist platform is a service provided to patients in South East (SE) London to monitor patients with COVID-19 through a smartphone and with provision of pulse oximeter. [15] Patients are onboarded to a platform which interacts through WhatsApp (see workflow in supplementary material) to ask patients questions based on RECAP-V0 items. [15-17] Data collected were also transferred to ORCHID. Patient consent was not required in this case since data access was granted under Health Service (Control of Patient Information) Regulations 2002 (COPI) notice.

### **Sample size calculation**

We estimated a minimum sample size of 1,317 participants for model development and 1,400 for model external validation assuming 10% hospitalisation rate for COVID-19, a maximum of 24 predictor variables, a binary outcome (hospital admission), and a minimum 85% model specificity on validation. We aimed to recruit at least 2,880 participants per setting assuming 5-6% loss to follow-up. [12] Patients were recruited according to the following inclusion criteria: having symptoms of acute COVID-19 (within 14 days of onset of symptoms) based on clinical judgment, being 18 years of age or older, and able to provide informed consent (except for Doctaly Assist).

### **Statistical analysis**

All real-world studies are prone to data missingness, indeed the degree of missingness is a test of suitability of an item for use in clinical settings. The extent of missing data for each variable (outcome and predictors) was assessed on degree of missingness, patterns (at random or not at random), and possible reasons. If the degree of missingness was above 50%, the predictor variable was excluded from the model. If the degree of missingness was less than 50%, the data were imputed using multiple imputation chain equations (MICE). The outcome variable (i.e., hospital admission) was used for the imputation of the predictor variables. We used regression for continuous variables (normally distributed or transformed), and logit or ordinal logit for categorical variables. A total of five imputations were performed and aggregated based on Rubin's rules.[18] We compared observed and imputed values, especially for variables for



which the fraction of missing data was large. A summary of the full methods is provided in the Supplementary Materials and published statistical analysis plan [12]).

A probabilistic risk prediction based on a multivariable logistic regression model, including the variables in RECAP-V0 as factors, was performed for the RSC dataset (model termed RECAP-GP), and separately for the Doctaly Assist dataset (model termed RECAP-O2). The models allow estimation of the likelihood of a particular patient with a COVID-19 diagnosis being admitted to hospital with COVID-19 within 28 days of symptom onset. Variables were checked for independence from each other by including in the model interaction terms between age and respiratory rate where available. When identifying COVID-19 as a cause of admission, we searched for COVID-19 ICD-10 codes, U071 or U072, as first, second or third cause of admission in HES data. COVID-19 as second or third cause were included if the first cause of admission was pneumonia, dyspnea, pulmonary embolism, or chest pain. [19] Multivariable regression models were constructed including all the parameters from the RECAP-V0 template, and their performance assessed to check for significance. Elements of the RECAP-V0 that were constructed using alternative codes with levels of severity (See Table 2) had that relationship maintained in the model. Using backward elimination, a model using only the predictor factors that have shown to be statistically significant with a P value <0.05 was obtained. Internal model validation was conducted using bootstrapping and receiver operating characteristic (ROC) curves plotted. The model performance in sub-populations by age (above and below 65) and by sex was investigated by comparing the diagnostic measures Akaike Information Criterion (AIC), C-Statistic, and visual inspection of the ROC curves. Selection of upper and lower cut-points for creating a Red-Amber-Green categorisation of patients was based on clinical considerations of risk and model performance and consensus of the research team. External validation of the RECAP-GP model was conducted using data separately from NWL primary care practices and CCAS to verify the specificity of the model predictions, as well as the sensitivity, negative predictive value (NPV) and positive predictive value (PPV). The RECAP-O2 model was validated using a subsequent cohort of subjects monitored by the platform. [12]

### **Role of funding source**

The funding sources had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

### **Results**

The recruitment of patients by practices in NWL and RCGP RSC ran from 1 October 2020 to 28 February 2021. A total of 4,278 patients (NWL=2,415; RCGP RSC=1,863) were recruited by 170 practices (NWL=103; RSC=103). CCAS was actively recruiting from 15 March 2021 to 23 May

2021 and enrolled 2,674 patients. SE London’s Doctaly Assist has provided records for a total of 4,045 patients (1,948 for model development collected between 26 November 2020 and 5 May 2021 (Doctaly-1 dataset); 2,085 for model validation between 8 May 2021 and 26 October 2021 (Doctaly-2 dataset)). The patient populations are described in Table 1. While mean age and sex were similar across the four cohorts, there was a higher proportion of non-white ethnicity in London cohorts, i.e, NWL and Doctaly samples.

The RCGP RSC data and the Doctaly-1 datasets were used for model development (Figure 1). The models were subsequently validated in the NWL GP and Doctaly-2 data respectively, as described in the statistical analysis plan (SAP). [12] Table 2 outlines all the model predictor variables considered, and whether they were included in the model following assessment of patterns of missing data.

	RCGP RSC (N=1863)		NWL (N=2415)		CCAS (N=2674)		Doctaly-1 (N=1948)		Doctaly-2 (N=2085)	
		Complete (%)		Complete (%)		Complete (%)		Complete (%)		Complete (%)
Mean age (Years) (S.D)	49 (17.7)	100%	46 (18.3)	100%	42 (15.9)	100%	44 (13.1)	100%	39 (11.9)	100%
Sex (% Female)	57%	100%	56%	100%	55%	100%	64%	100%	61%	100%
Ethnicity (% non- white)	33.5 %	93.5%	74%	91.4%	78.7%	85.6%	68.5%	92.5%	51.1 %	96.8%

Table 1. Summary characteristics of demographic data with degree of data completeness.

Parameter	RCGP RSC (RECAP-GP Model)		NWL (RECAP-GP external validation)		CCAS (RECAP-GP external validation)		Doctaly-1 (RECAP-O2 model)		Doctaly-2 (RECAP-O2 external validation)	
	% Completeness	Considered in model	% Completeness	Considered in model	% Completeness	Considered in model	% Completeness	Considered in model	% Completeness	Considered in model
Shortness of breath *	86%	Yes	87%	Yes	96%	Yes	91.7%	Yes	100%	Yes
Feeling feverish/shivers *	96%	Yes	78%	Yes	96%	Yes	91.5%	Yes	100%	Yes
Temperature (Observed)	24%	No	7%	No	27%	No	43%	No	45%	No
Fatigue *	92%	Yes	87%	Yes	91%	Yes	91.4%	Yes	100%	Yes
Acute cognitive decline *	82%	Yes	70%	Yes	0%	No	91.4%	Yes	100%	Yes
Time from first symptoms (days)	93%	No	32%	No	96%	No	99%	Yes	100%	Yes
Respiratory rate	15%	No	7%	No	22%	No	72%	Yes	89%	Yes
Heart rate	22%	No	21%	No	42%	No	70%	No	1%	No

Oxygen saturation at rest	15%	No	16%	No	13%	No	70%	Yes	62%	Yes
Oxygen saturation after 40 steps	6%	No	8%	No	3%	No	66%	Yes	57%	Yes
Muscle aches	48%	No	39%	No	52%	No	NA	No	NA	No
Trajectory of breathlessness *	45%	No	52.5%	No	79%	No	1%	yes	0%	Yes
Diabetes mellitus	100% **	Yes	100% **	Yes	100% ***	Yes	NA	No	NA	No
Hypertension	100% **	Yes	100% **	Yes	100% ***	Yes	NA	No	NA	No
Coronary health disease	100% **	Yes	100% **	Yes	100% ***	Yes	NA	No	NA	No
Chronic Kidney disease	100% **	Yes	100% **	Yes	100% ***	Yes	NA	No	NA	No
Age	100% **	Yes	100% **	Yes	100%		100%		100%	

Sex	100% **	Yes	100%**	Yes	100%		100%		100%	
Ethnicity	93.5% **	Yes	91.4% **	Yes	83.6%		92.5%		3.2%	

Table 2. Predictor variables considered for inclusion in the model and whether they were included following missingness assessment.

NA - not available \* These items contain a proposed severity scale in RECAP-v0 and were captured in the EHR templates as a drop-down list of alternatives with appropriate per item SNOMED coding. This ordering was carried through to the analysis. \*\* Extracted from GPs EHR systems. \*\*\* Collected from the caller as part of the NHS111 pathway.

COVID-related hospitalisation rates were similar in all the datasets except Doctaly-2, 4.4% in the RSC 3.8% in NWL 3.1 % in CCAS, 3.3% in Doctaly-1, but only 0.9% in Doctaly-2 cohort, probably probably because of higher vaccination rates in the Doctaly-2 cohort.

All continuous data were found to be sufficiently normally distributed by visual inspection and the pattern of missingness for each variable was random. MICE was performed on 15 iterations to allow for convergence of the imputation chains. The frequency distributions of the variables in the imputed dataset were compared to the original dataset variables visually and found to be similar and not requiring further transformation (full details in Supplementary Materials).

### Model 1 (RECAP-GP): RCGP and NWL primary care practice data

The RECAP-GP model was built using the RCGP RSC data. The predictor variables used in the final model were sex, age, history of hypertension, degree of breathlessness, and temperature symptoms. (Coefficients and p values shown in Table 3). Fatigue, confusion, ethnicity, body mass index (BMI), diabetes, coronary heart disease, and chronic kidney disease were excluded after backward elimination ( $p > 0.05$ ). The model showed good performance to distinguish between risk levels (AUC=0.802) (Figure 2). There was no significant difference in the performance of the model when stratified for sex and age, although the prevalence of patients in the >65 years group was lower (19%) than those under 65, it was in line with the English population (18.5%)

Parameter	Coefficients	P Value
Intercept	-6.32	$<2 \cdot 10^{-16}$
Sex (male =1)	0.56	0.018
Age (years)	0.04	$1.15 \cdot 10^{-6}$
Hypertension history	0.56	0.04
Breathlessness:		
Breathlessness (can't complete sentences at rest)	1.69	$6.7 \cdot 10^{-5}$

Parameter	Coefficients	P Value
Intercept	-6.32	$<2 \cdot 10^{-16}$
Sex (male =1)	0.56	0.018
Age (years)	0.04	$1.15 \cdot 10^{-6}$
Hypertension history	0.56	0.04
Breathlessness:		
Breathlessness (can't complete sentences at rest)	1.69	$6.7 \cdot 10^{-5}$
Breathlessness on mild exertion	0.61	0.025
Breathlessness on moderate exertion	0.22	0.57*
Fever:		
Temperature – rigors	0.10	0.98*
Temperature - feeling feverish	0.75	0.002

*Table 3: The RECAP-GP model. \*For fever and breathlessness severity all levels were included if one level was significant. Absence of hypertension, breathlessness and fever are the base coefficients in the logistic regression, set to zero and not shown.*

*Figure 2 here*

The cut-off points for the green, amber, and red risk groups were chosen by the research team clinicians (BD, SdeL, ErM, EmM, EIM ,TG, AE-G), before validation, using the specificities and sensitivities obtained from the ROC curve shown in Figure 2. The selected cut-off points and related specificity, sensitivity and interval likelihood ratios (ILR) are shown in Table 4. We opted for maximising model sensitivity (90%) for the low to moderate risk threshold to ensure all patients needing monitoring were in the amber group and maximising specificity (90%) in the moderate to high risk threshold to try to limit the number of unnecessary hospital admissions from the amber group.

For external validation, the prediction model was run using both the NWL data and the CCAS data separately. We decided not to build a separate CCAS model as originally intended, as the data was similar to the GP practice data, and lower than initially expected (2020) admission rates would have limited power. The selected cut-off points were used to assign risk categories to patients as shown in Table 4, along with the observed model sensitivity, specificity, NPV and PPV. As true negative and true positive patients for the amber group cannot be defined, only the number of final hospitalisations is noted in the results. In the NWL GP data, the probability of being categorised as a low risk (green) patient and not needing admission (i.e., NPV) was high (98.7%), but the probability of being in the high risk (red) group and being admitted (i.e., PPV), was low (13.9%). In the CCAS data, the NPV was 98.3%, equivalent to the GP data, but the PPV was lower (5.9%), the confidence intervals just separated from the NWL validation.

<b>Cut-off point based on clinical team decision</b>		
	Green to Amber	Amber to Red
Sensitivity	90%	40%
Specificity	40%	90%
Threshold (Logit transformed)	0.027	0.098
Interval Likelihood Ratio	0.16	6



<b>Risk group assigned by model in NWL GP data</b>			
	Green	Amber	Red
N of patients	1453	797	158
N of actual hospitalisations (as % of N of patients)	18 (1.2%)	50 (6.3%)	22 (13.9%)
Sensitivity of the model on validation % (95% CI)	61.9% (59.9-63.9)	-	24.4% (16.0-34.6)
Specificity of the model on validation % (95% CI)	80% (70.2-87.7)	-	94.1% (93.1-95.2)
PPV of Red designation % (95% CI) % (95% CI)	13.9% (9.8-19.4)		
NPV of Green designation % (95% CI)	98.8% (98.1-99.2)		
<b>Risk group assigned by model in CCAS data</b>			
	Green	Amber	Red
N of patients	1512	958	204
N of actual hospitalisations (as % of N of patients)	25 (1.6%)	45 (4.7%)	12 (5.9%)

Sensitivity of the model on validation % (95% CI)	57.4% (55.4-59.3)	-	14.6% (7.8-24.2)
Specificity of the model on validation % (95% CI)	69.5% (54.4-79.2)	-	93.0% (92.0-94.0)
PPV of Red designation % (95% CI)	5.9% (3.5-9.7)		
NPV of Green designation % (95% CI)	98.3% (97.7-98.8)		

*Table 4. Cut-off points for the RECAP-GP model along with sensitivity, specificity, negative and positive predictive values following external validation in the NWL and CCAS data. Positive predictive value (PPV): N of hospitalisations in red group /N of patients in red group. Negative predictive value (NPP): Number of patients non-admitted in green group/N of patients in green group*

### **Model 2 (RECAP-O2): Doctaly-1 and Doctaly-2**

Given the predicted availability of larger numbers of Oxygen Saturation (SpO<sub>2</sub>) readings in the Doctaly Assist data, we planned a second model using the Doctaly-1 dataset. Predictor variables used in the final model were age, degree of breathlessness, fatigue, and SpO<sub>2</sub> at rest (coefficients and p-values shown in Table 5). Sex, ethnicity, temperature, acute cognitive decline, days since onset of symptoms, respiratory rate, and trajectory of breathlessness were excluded after backwards elimination (significance >0.05%). SpO<sub>2</sub> after activity was found to be co-linear with SpO<sub>2</sub> at rest in the model and was thus excluded from the final model.

The model was internally validated and calibrated through bootstrapping. Figure 3 shows the ROC curve along with the model diagnostic analysis. The AUC was 0.843 and AIC 354, which suggests good model performance. Notably the bottom left of the ROC curve has a steeper slope than the RECAP-GP model, indicating better discrimination in sicker patients. The model performance in sub-populations by age (>65) and sex was investigated and was not found to be significantly different.

Parameter	coefficient	P Value
-----------	-------------	---------

Intercept	25.0	$3.85 \cdot 10^{-7}$
Age (years)	0.04	0.0002
Breathlessness:		
Breathlessness – Feeling Uncomfortable to breathe	0.92	0.03
Breathlessness when walking* round the room	-0.43	0.26
Fatigue:		
Fatigue - Difficult to wake up	1.5	0.068
Fatigue - Too tired to do usual activities	1.23	0.0007
Oxygen saturation at rest % (0-100%)	-0.33	$9.8 \cdot 10^{-11}$

Table 5: The RECAP-O2 model. Note ‘scales’ of severity are included if one element is significant and only the most severe level is used in the model. \*For fatigue and breathlessness severity all levels were included if one level was significant. Absence of hypertension, breathlessness and fatigue are the base coefficients in the logistic regression, set to zero and not shown.

*Figure 3 here*

For external validation the RECAP-O2 model was run using the Doctaly-2 data. Likewise, as seen in Table 6, the cut-off points were chosen by the clinical team based on the ROC curve in Figure 3 and were used to assign risk categories to patients. Table 6 shows the selected cut-off points, related specificity and sensitivity, and the actual sensitivity and specificity, negative and positive predictive values of the RECAP-O2 model following external validation. Although NPV of green designation is slightly higher than RECAP-GP (99.4%), PPV of red designation is lower (8.8%), on account of the lower rates of admission in the Doctaly 2 data (the population now having the

majority vaccinated), which means lower probability of being admitted when categorised as a high risk patient.

<b>Cut-off point based on clinical team decision</b>			
	Green to Amber	Amber to Red	
Sensitivity	90%	48%	
Specificity	52%	95%	
Threshold (Logit transformed)	0.013	0.096	
Interval Likelihood Ratio	0.19	10.4	
<b>Risk group assigned by model in Doctaly-2 validation data</b>			
	Green	Amber	Red
N of patients	1183	457	34
N of actual hospitalisations (as % of N of patients)	7 (0.6%)	9 (1.9%)	3 (8.8%)
Sensitivity of model on validation % (95% CI)	71.1% (68.8-73.2)	-	15.8% (3.4-39.6)
Specificity of model on validation % (95% CI)	63.1% (38.4-83.7)	-	98.1% (97.3-98.7)

PPV of Red designation % (95% CI)	8.8% (3.1-22.4)
NPV of Green designation % (95% CI)	99.4% (98.9-99.7)

*Table 6: Cut-off points for the RECAP-O2 model (values chosen by the clinical team) along with sensitivity, specificity, negative and positive predictive values following external validation. Positive predictive value (PPV): N of hospitalisations in red group /N of patients in red group Negative predictive value (NPP): Number of patients non-admitted in green group/N of patients in green group*

## Discussion

Assessment of severity of COVID-19 in the community is crucial to pandemic management worldwide. Our study provides a derivation and real-world validation of two risk scores specifically designed for COVID-19 patients in the community. The RECAP-GP model includes degree of breathlessness, temperature symptoms, presence of hypertension, sex, and age as hospital admission predictors, and can be used when a pulse oximeter is not available to the patient. This model provides a good prediction for risk of non-admission in the lowest risk group, e.g., from 2,415 patients in NW London, 60.1% were assigned to the green group with a NPV of 98.7%. The model performs less well at differentiating amber from red groups, with only 22 out of 158 red patients needing admission (PPV 13.9%). When validated in the CCAS data the RECAP-GP model performs slightly less well on PPV. However, performance on the green/amber cut is within the NWL confidence intervals with 57% patients being safely reassured. The RECAP-O2 model included degree of breathlessness, fatigue, SpO2 at rest, and age as predictors. This model can be used if pulse oximeters are available, such as in moderate risk patients under a monitoring service. Although the improvement of the slope on the ROC curves indicates an Interval LR increasing from 6 to 10 for the amber/red cut point, the model performs less well on validation with a PPV of 8.8%, due to the lower admission rate in the now majority-vaccinated validation set. However, its specificity is good (98.1 %), which supports its use to assess moderate risk patients' need for admission (only 9 of 457 amber patients required admission).

Much discussion has taken place around the potential role of SpO2 measurement as an early warning for disease severity. In the UK a national strategy for using home pulse oximetry, COVID Oximetry @home, was established during 2020 with the intention of identifying 'silent hypoxia' (without breathlessness). UK guidance recommends provision of pulse oximeters, arranged by local health authorities, to monitor patients with symptomatic COVID-19 older

than 65 years old or with risk factors. [16] The use of RECAP in the assessment of COVID-19 patients is a valuable addition to this strategy. Our results are aligned with the guidance since age and SpO<sub>2</sub> significantly predicts deterioration in our models, but RECAP also considers symptom severity so it can better support clinicians' judgment on who needs monitoring, particularly for younger patients without comorbidities. Moreover, it is better at identifying need for treatment escalation compared with SpO<sub>2</sub> alone, which is only one factor in the final model.

The RECAP models are founded on assessment of need for hospital admission in an observational dataset. In such a design it is impossible to eliminate 'incorporation' bias due to the admission decision being partly based on the parameters included in the model (e.g., SpO<sub>2</sub>). We mitigated this by counting admissions as at least one night in hospital, rather than only a review in the emergency department. Recruitment was conducted contemporaneously via GP practices, CCAS, and the Doctaly Assist service from fall of 2020 to spring of 2021. However, the Doctaly Assist validation dataset was collected during the summer of 2021 as the UK's COVID-19 vaccination programme covered up to 70% of the adult population. [20] This increasing vaccine coverage may account for the difference in validation of the RECAP-O2 model due to lower admissions being observed. Whilst it is unlikely that SARS-Cov-2 variants and vaccination status will affect the predictive signs of deterioration, these being driven by common pathophysiological processes, the overall lower event rates demand better performance in a model. However, all models should be subject to ongoing surveillance/calibration, especially with rapid changes in variants and vaccines.

In any study relying on data collection during routine practice there will be missing data. It is notable that remote consultations with GPs did not enable collection of vital signs from patients. Health monitoring devices, such as pulse oximeters and thermometers, are rarely available in the community unless patients are being provided with them. Therefore, these observations were largely missing in the GP and CCAS datasets and could not be included in the RECAP-GP model. In contrast, the Doctaly Assist datasets enabled us to assess the predictive value of SpO<sub>2</sub>, resting and on exercise, heart rate, respiratory rate, and temperature, determining, for the first time, the diagnostic value of SpO<sub>2</sub> monitoring in the community.

We took a careful approach to analysis with peer-review and prior publication of the protocol and Statistical Analysis Plan. The use of MICE and bootstrapping for the internal validation are standards in the development of diagnostic models. We also chose several relevant validation datasets. However, lack of availability of EHR data linkage for the Doctaly data limits the ability to include factors such as co-morbidity in the RECAP-O2 model. The age distributions, admission rates, ethnicity, and comorbidities are in line with UK population expectations,

except the Doctaly data has less over 75 yr olds than expected. SE London and NW London contain significant populations of black British and South Asian ethnicity, whilst the RSC network is more representative of the UK. [21] This supports the external generalisability of our findings. The thresholds used in the two RECAP scores can be adjusted to better suit local circumstances and, in the future, it may be necessary to consider re-testing a model with vaccination status as a potential predictor. We did not do this as data on vaccination status were not consistently available and policy was changing rapidly during the study.

The RECAP-VO template developed by the Delphi study contained 10 questions, the validated models contain four and five items only, significantly improving their fitness for use in the clinical setting.[22] The choice of logistic regression modelling means that most EHR systems using SNOMED codes will be able to recreate the RECAP electronic templates and integrate the score calculator into the system. The full models are provided as a downloadable code on Github.

The identification of cut-off points to differentiate between risk levels was based on clinical consensus. We intended to maximise the model sensitivity when deciding on the green-amber risk level (so as not to miss cases requiring monitoring), while intending to identify an amber-red cut-off point specificity that minimises unnecessary hospital admissions. The result is a tool that can be safely used for initial assessment and monitoring if appropriate monitoring mechanisms to identify deterioration are in place for the amber group. We suggest that the RECAP-GP score be used remotely in the initial assessment, without need for patient observations. Following this, if the patient is considered moderate to high risk, they could be provided with a home pulse oximeter for calculation of the RECAP-O2 score to detect deterioration. Given the low PPV of the Red categorisation on the amber/red cut point the RECAP-O2 model will tend to over alert. Care escalation should be considered with reference to national or local pathways, ability to monitor the patient in the community, hospital capacity and shared decision making with the patient.

Much has evolved since the first wave of the COVID-19 pandemic when we conceptualised this study. Mass vaccination has dramatically changed prognosis and oximetry is now much more widely used in the community than it was in early 2020, with 'at home' services available in many settings. Yet, new variants are triggering new pandemic waves across the globe, putting services under great strain. We believe that these two scores are likely to be a valuable resource to support clinical judgement, reduce uncertainty and improve safety in triage and monitoring of patients with suspected COVID-19 in health systems worldwide.





## **Contributors**

AEG coordinated recruitment across sites, helped with template development and elaboration of datasets in WSIC/iCARE and ORCHID and drafted the manuscript. DP and FF contributed to the design of the analysis, conducted the analysis and contributed to interpretation of the results. CR supported the ethics submission and amendments, communications with practices during the recruitment period, DPIA, COPI, and CAG applications. EIM and EmM contributed to code and template development, analysis and interpretation of results. BG contributed to the conceptualisation and design of the study, managed the data in iCare, contributed to the analysis and interpretation of the results. ALN contributed to the development of the protocol, helped obtain funding, and contributed to the interpretation of the results. CO managed the data in ORCHID and contributed to the analysis and interpretation of the results. LH contributed to obtaining funding, helped write the protocol and assisted in the management of data collection in practices. RC and JH assisted in the design of the SE London Covid management pathways, helped obtain funding, assisted data collection and interpretation of the analyses. MB, CW and BB coordinated the patient enrollment at CCAS. BD, ErM, S deL and TG developed the protocol and obtained funding. BD supervised the study and acts as corresponding author. All authors contributed to interpretation of results and drafting of the manuscript.

## **Declaration of interest**

The authors declare no relevant interests.

## **Data sharing**

The models are available for download from the following link: TBA

## **Acknowledgements**

Doctaly Assist (Phil Tyler and Dr Prad Velayuthan) provided the data extraction and SNOMED coding of their Chat-bot data. Dr Mark Ashworth and Dr Ibi Fakoya of King's College London evaluated the initial set-up of the Covid monitoring programme in SE London. Eamon O'Doherty of NW London Clinical Commissioning Groups extracted the relevant data tables from WSIC to enable the data analysis in iCare. Sneha Anand at Oxford University managed the governance, data transfer and linkage in ORCHID. Matt Widdows coordinated the development of the RECAP template with Adastra and extracted the relevant CCAS data for analysis. Dr Merlin Dunlop of Ardens assisted in the SNOMED coding of the RECAP-V0 templates and their distribution. Ashnee Dhondee of NWL CRN helped recruit practices to the study.

## Dissemination

The study was presented at the 1º Congresso Brasileiro de Evidências Clínicas na COVID-19 in May 2021 (reference) and at the SAPC ASM and FCI Scientific Conferences both in July 2021. The RECAP-VO model has informed the development of COVID-19 clinical assessment tools in other countries, such as the online assessment tool set up by the Fiocruz Oswaldo Cruz Foundation in Rio de Janeiro, Brazil (<https://redcap.ini.fiocruz.br/surveys/?s=HPHCHDEDHN>).

## Ethics

The study was sponsored by Imperial College London and approved by the North West-Greater Manchester East Research Ethics Committee and Health Research Authority in May 2020 (IRAS number: 283024, Research Ethics Committee reference number: 20/NW/0266). The study was badged as an Urgent Public Health Study by the National Institute of Health Research in October 2020.

The WSIC data analysis was undertaken within a research database that was given favorable ethics approval by the West Midlands Solihull Research Ethics Committee (reference 18/WM/0323; IRAS project ID 252449). All data used in this paper were fully anonymized before analysis. iCARE is a Trusted Research Environment and provides access to HRA REC approved anonymised data for research (REF 21/SW/0120, IRAS project ID: 282093)

At Oxford, analysis was undertaken within the secure data processing platform of the Oxford Royal College of General Practitioners Clinical Informatics Digital Hub (ORCHID) trusted research environment (TRE). (DIPA registration number: Z575783X, DARS number: DARS\_NIC\_431881\_N8BON, <https://orchid.phc.ox.ac.uk/index.php/orchid-privacy-notices/>).

The access to Doctaly Assist data was granted under COPI notice (and CAG Resolution 5 after expiration of COPI notice in March 2022) and, therefore, patient consent was not required. However, patients could opt-out using the National Opt-Out register.

## Funding

The study was funded by the Community Jameel and the Imperial College President's Excellence Fund, the Economic and Social Research Council, the UK Research and Innovation, Health Data Research UK. The authors gratefully acknowledge infrastructure support from the NIHR Imperial Patient Safety Translational Research Centre, the NIHR Imperial Biomedical Research Centre and the NIHR Oxford Biomedical Research Centre. This research was in part enabled by the Imperial Clinical Analytics Research and Evaluation (iCARE) environment and

Whole System Integrated Care (WSIC) and used the iCARE and WSIC team and data resources. The project was supported by the NIHR CRN Urgent Public Health Study.

## References

- [1] Greenhalgh T, Koh GCH, Car J. COVID-19: a remote assessment in primary care. *Bmj* 2020;368:m1182. doi: 10.1136/bmj.m1182 [published Online First: 2020/03/28]
- [2] Nunan D, Brassey J, Mahtani K, et al. COVID-19 Signs and Symptoms Tracker. Oxford: Centre for Evidence-Based Medicine 2020. Accessed 18th December 2021 at <https://www.cebm.net/COVID-19/COVID-19-signs-and-symptoms-tracker/>.
- [3] Wu Z, McGoogan JM. Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72 314 cases from the Chinese Center for Disease Control and Prevention. *JAMA* 2020;323:1239-42.
- [4] NHS England and NHS Improvement. Pulse oximetry to detect early deterioration of patients with COVID-19 in primary and community care settings. London: NHSE/I. Accessed 17th December 2021 at <https://www.england.nhs.uk/coronavirus/publication/pulse-oximetry-to-detect-early-deterioration-of-patients-with-covid-19-in-primary-and-community-care-settings/>; 2020 (updated 12th Jan 2021).
- [5] Knight, Stephen R, Antonia Ho, Riinu Pius, Iain Buchan, Gail Carson, Thomas M Drake, Jake Dunning, et al. 2020. "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". *The BMJ* 370: <http://dx.doi.org/10.1136/bmj.m3339>
- [6] Clift, Ash K, Carol A C Coupland, Ruth H Keogh, Karla Diaz-Ordaz, Elizabeth Williamson, Ewen M Harrison, Andrew Hayward, et al. 2020. "Living Risk Prediction Algorithm (QCOVID) for Risk of Hospital Admission and Mortality from Coronavirus 19 in Adults: National Derivation and Validation Cohort Study." *BMJ*. <https://doi.org/10.1136/bmj.m3731>.
- [7] Carr, Ewan, Rebecca Bendayan, Daniel Bean, Matt Stammers, Wenjuan Wang, Huayu Zhang, Thomas Searle, et al. 2021. "Evaluation and Improvement of the National Early Warning Score (NEWS2) for COVID-19: A Multi-Hospital Study." *BMC Medicine* 19 (1). <https://doi.org/10.1186/s12916-020-01893-3>.
- [8] Hodgson, Luke E, Jo Congleton, Richard Venn, C Lui, G Forni, and Paul J Roderick. 2018. "NEWS 2-Too Little Evidence to Implement?" *Clinical Medicine*. Vol. 18.

[9] Tankel JW, Ratcliffe D, Smith M, Mullarkey A, Pover J, Marsden Z, et al. Consequences of the emergency response to COVID-19: a whole health care system review in a single city in the United Kingdom. *BMC Emerg Med* [Internet]. 2021 [cited 2021 Dec 20];21(1). Available from: <https://doi.org/10.1186/s12873-021-00450-2>

[10] Greenhalgh, Trisha, Paul Thompson, Sietse Weiringa, Ana Luisa Neves, Laiba Husain, Merlin Dunlop, Alexander Rushforth, David Nunan, Simon De Lusignan, and Brendan Delaney. 2020. "What Items Should Be Included in an Early Warning Score for Remote Assessment of Suspected COVID-19? Qualitative and Delphi Study." *BMJ Open*. <https://doi.org/10.1136/bmjopen-2020-042626>.

[11] Espinosa-Gonzalez, Ana Belen, Ana Luisa Neves, Francesca Fiorentino, Denys Prociuk, Laiba Husain, Sonny Christian Ramtale, Emma Mi, et al. 2021. "Predicting Risk of Hospital Admission in Patients with Suspected COVID-19 in a Community Setting: Protocol for Development and Validation of a Multivariate Risk Prediction Tool." *JMIR Research Protocols* 10 (5). <https://doi.org/10.2196/29072>.

[12] Fiorentino, Francesca, Denys Prociuk, Ana Belen Espinosa Gonzalez, Ana Luisa Neves, Laiba Husain, Sonny Christian Ramtale, Emma Mi, et al. n.d. "An Early Warning Risk Prediction Tool (RECAP-V1) for Patients Diagnosed With COVID-19: Protocol for a Statistical Analysis Plan." Accessed October 31, 2021. <https://doi.org/10.2196/30083>.

[13] de Lusignan S, Dorward J, Correa A, Jones N, Akinyemi O, Amirthalingam G, Andrews N, Byford R, Dabrera G, Elliot A, Ellis J, Ferreira F, Lopez Bernal J, Okusi C, Ramsay M, Sherlock J, Smith G, Williams J, Hobbs R. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of General Practitioners Research and Surveillance Centre primary care network: a cross-sectional study. *The Lancet Infectious Diseases*. 2020 Sep;20(9):1034-1042. doi: 10.1016/S1473-3099(20)30371-6.

[14] de Lusignan S, Lopez Bernal J, Zambon M, Akinyemi O, Amirthalingam G, Andrews N, Borrow R, Byford R, Charlett A, Dabrera G, Ellis J, Elliot A, Feher M, Ferreira F, Krajenbrink E, Leach J, Linley E, Liyanage H, Okusi C, Ramsay M, Smith G, Sherlock J, Thomas N, Tripathy M, Williams J, Howsam G, Joy M, Hobbs R. Emergence of a Novel Coronavirus (COVID-19): Protocol for Extending Surveillance Used by the Royal College of General Practitioners Research and Surveillance Centre and Public Health England. *JMIR Public Health Surveillance*. 2020 Apr;6(2): e18606. doi: 10.2196/18606.

[15] Doctaly. 2021. "Doctaly Assist" <https://doctalyassist.com/lewisham/covid> (accessed on 11 December 2021)

[16] NHS. 2021. "COVID Oximetry @home" <https://www.england.nhs.uk/nhs-at-home/covid-oximetry-at-home/> (accessed on 11 December 2021)

[17] Clarke, Jonathan, Kelsey Flott, Roberto Fernandez Crespo, Hutan Ashrafian, Gianluca Fontana, Jonathan Benger, Ara Darzi, Sarah Elkin, and Fernandez R Crespo. 2021. "Assessing the Safety of Home Oximetry for COVID-19: A Multisite Retrospective Observational Study." *BMJ Open* 11: 49235. <https://doi.org/10.1136/bmjopen-2021-049235>.

[18] Rubin D. Multiple imputation for nonresponse in surveys. New York: Wiley, 1987

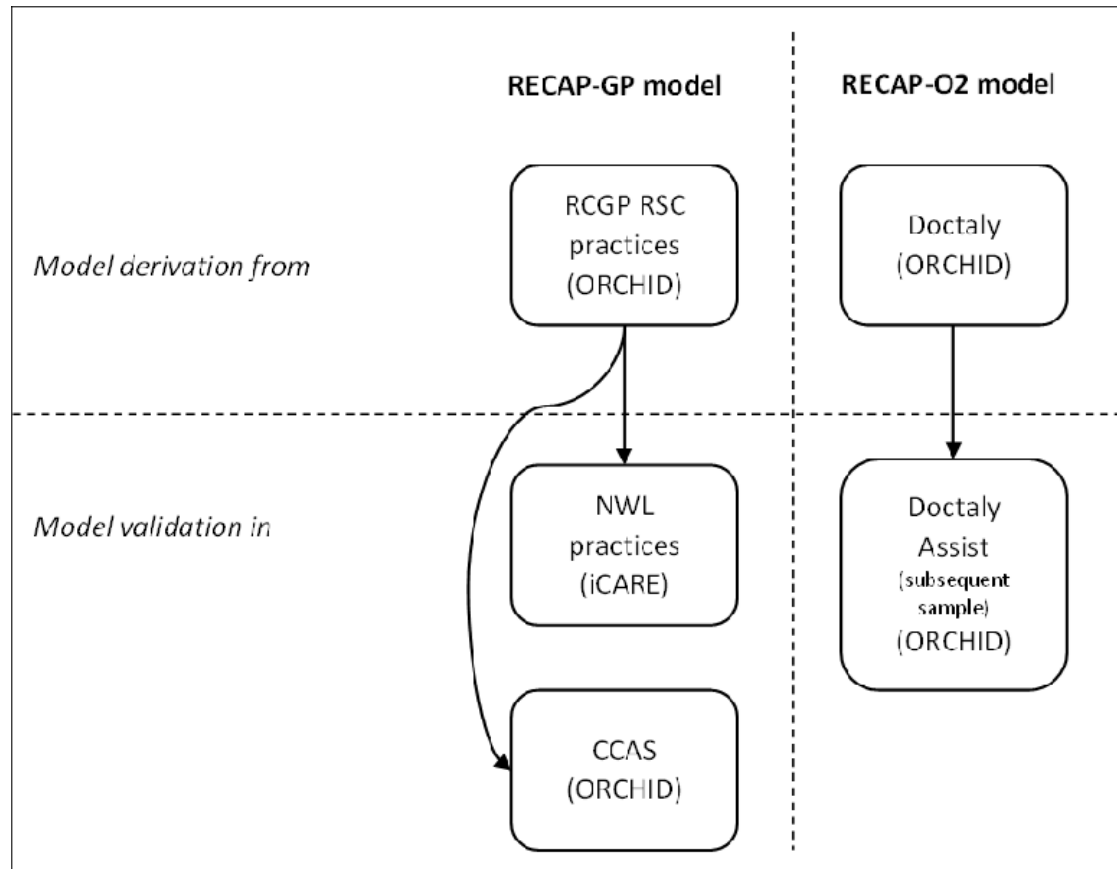
[19] Ramos-Casals, Manuel, Pilar Brito-Zerón, and Xavier Mariette. 2021. "Systemic and Organ-Specific Immune-Related Manifestations of COVID-19." *Nature Reviews Rheumatology*. <https://doi.org/10.1038/s41584-021-00608-z>.

[20] Gov.UK. 2021. "Coronavirus (COVID-19) in the UK" <https://coronavirus.data.gov.uk/details/vaccinations> (accessed on 12 December 2021)

[21] Correa, Ana, William Hinton, Andrew Mc Govern, Jeremy Van Vlymen, Ivelina Yonova, Simon Jones, and Simon De Lusignan. 2016. "Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) Sentinel Network: A Cohort Profile." *BMJ Open* 6 (4). <https://doi.org/10.1136/BMJOPEN-2016-011092>.

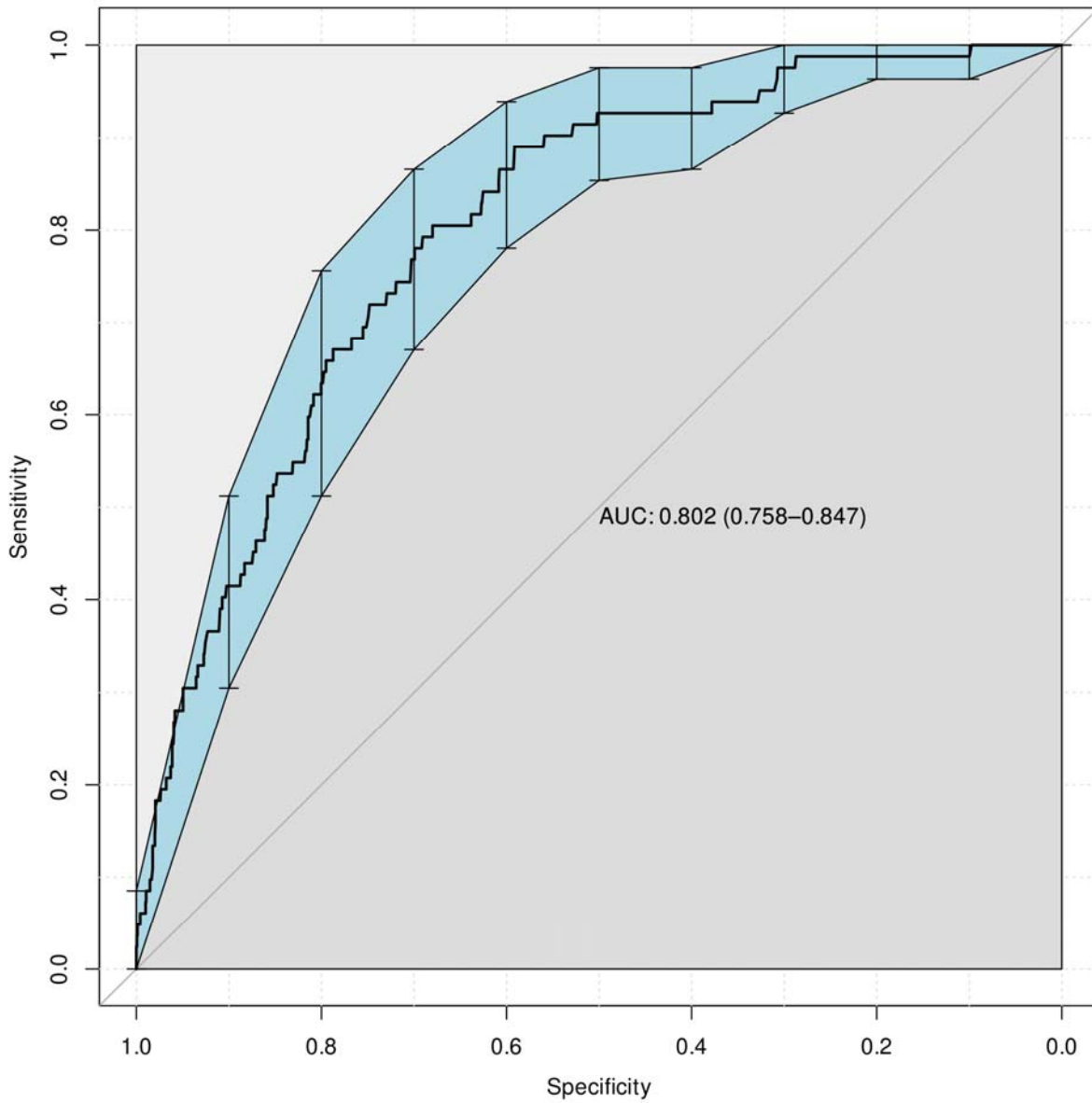
[22] Robert A. Jenders, Chapter 15 - Decision Rules and Expressions, Editor(s): Robert A. Greenes, *Clinical Decision Support (Second Edition)*, Academic Press, 2014, Pages 417-434, ISBN 9780123984760, <https://doi.org/10.1016/B978-0-12-398476-0.00015-4>.

## Figures

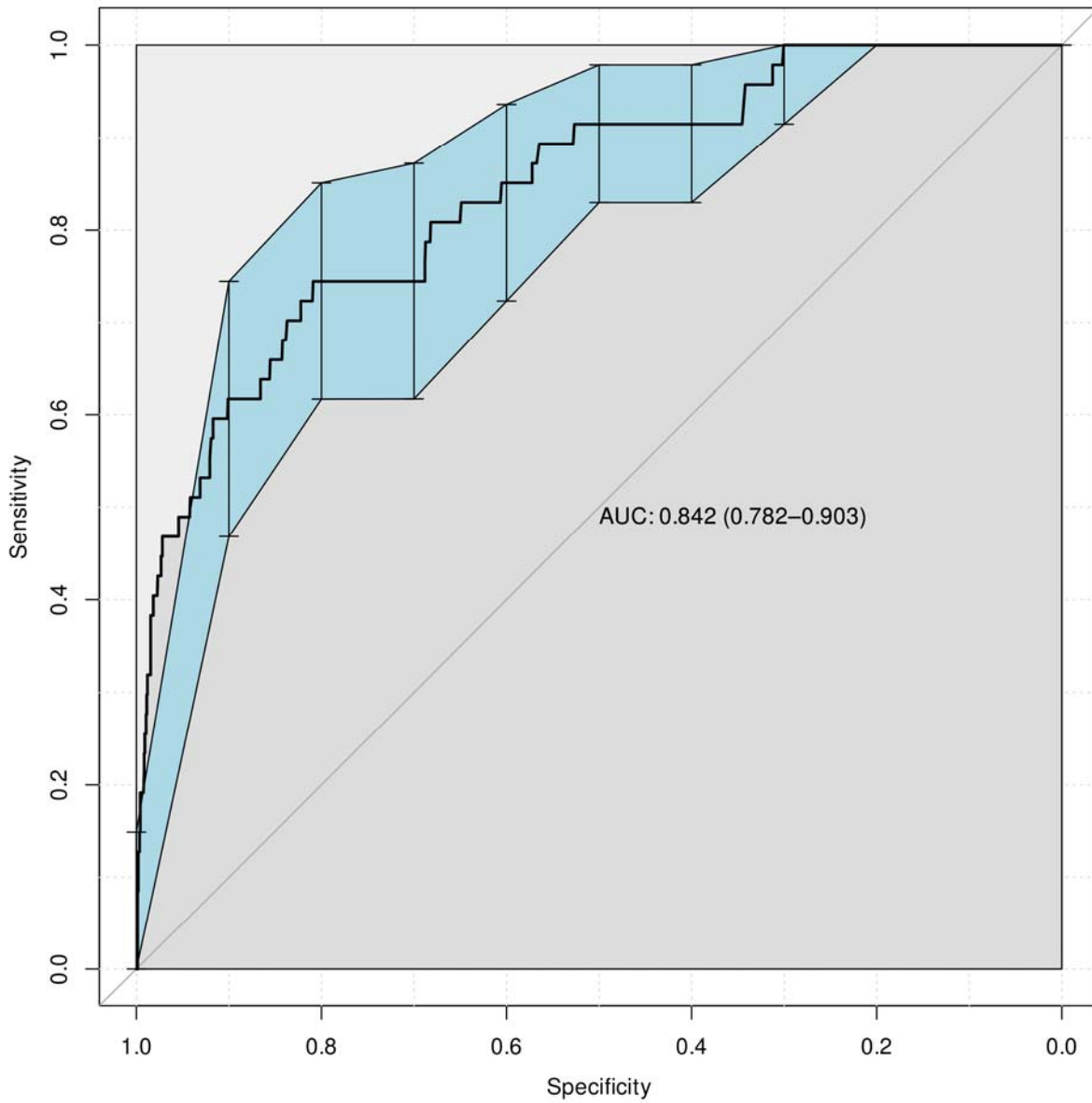


*Figure 1 Settings used for derivation and validation of the RECAP scores.*

*RCGP RSC: Royal College of General Practitioners Research and Surveillance Centre. NWL: North West London. ORCHID: Oxford-Royal College of General Practitioners Clinical Informatics Digital Hub environment. iCARE: Imperial Clinical Analytics, Research and Evaluation environment, CCAS: NHS111 Covid Clinical Assessment Service.*



*Figure 2. ROC curve of the RECAP-GP model performance following bootstrapping for internal validation along with model diagnostic measures obtained as part of model calibration and performance assessment. (AUC: area under the curve, AIC: Akaike Information Criterion)*



*Figure 3. ROC curve of the RECAP-O2 model performance following bootstrapping for internal validation along with model diagnostic measures obtained as part of model calibration and performance assessment.*