BMJ Open

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059111
Article Type:	Original research
Date Submitted by the Author:	10-Nov-2021
Complete List of Authors:	Boer, Arjen-Kars; Catharina Hospital Deneer, Ruben; Catharina Hospital, ; Eindhoven University of Technology, Maas, Maaike; Catharina Ziekenhuis, emergency department Ammerlaan, Heidi; Catharina Hospital van Balkom, Roland; Catharina Hospital Thijssen, Wendy; Catharina ziekenhuis, Emergency medicine Bennenbroek, Sophie; Catharina Hospital Leers, Mathie; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Martens, Remy; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Buijs, Madelon; Atalmedial Kerremans, Jos; Alrijne Hospital Location Leiderdorp Messchaert, Muriël ; Gelre Hospitals van Suijlen, Jeroen; Gelre Hospitals van Riel, Natal; Eindhoven University of Technology, ; Amsterdam University Medical Centres Scharnhorst, Volkher; Catharina Hospital; Eindhoven University of Technology,
Keywords:	COVID-19, STATISTICS & RESEARCH METHODS, ACCIDENT & EMERGENCY MEDICINE, Clinical chemistry < PATHOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Arjen-Kars Boer[1*], Ruben Deneer[1,2*†], Maaike Maas[3], Heidi S. M. Ammerlaan[4], Roland H. H. van Balkom[5], Wendy A.M.H. Thijssen[3], Sophie Bennenbroek[3] , Mathie P. G. Leers[6], Remy J.H. Martens[6], Madelon M. Buijs[7], Jos J. Kerremans[8], Muriël Messchaert[9], Jeroen D.E. van Suijlen[9], Natal A.W. van Riel[2,10] and Volkher Scharnhorst[1,2]

* Both authors contributed equally

† Corresponding author

1 Department of Laboratory Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene Loper 3, 5612 AE, Eindhoven, the Netherlands

3 Department of Emergency Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

4 Department of Internal Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands

7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the Netherlands

8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital, Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands

9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334 DZ Apeldoorn, the Netherlands

10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.

Word count: 3645

Keywords

COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction model, routine laboratory tests

Corresponding author

Ruben Deneer

Michelangelolaan 2

5623 EJ Eindhoven, the Netherlands

ruben.deneer@catharinaziekenhuis.nl

Abstract

Objectives: Identifying patients with a possible SARS-CoV-2 infection in the emergency department (ED) is challenging. Symptoms differ, incidence rates vary and test capacity may be limited. As PCR testing all ED patients is neither feasible nor effective in most centers, a rapid, objective, low-cost early warning score to triage ED patients for a possible infection is developed.

Design: Case-control study.

Setting: Secondary and tertiary hospitals in the Netherlands.

Participants: Patients presenting at the ED with venous blood sampling from July 2019 to July 2020 (N = 10417, 279 SARS-CoV-2 positive). The temporal validation cohort covered the period from July 2020 to October 2021 (N = 14080, 1093 SARS-CoV-2 positive). The external validation cohort consisted of patients presenting at the ED of three hospitals in the Netherlands (N = 12061, 652 SARS-CoV-2 positive).

Primary outcome measures The primary outcome was one or more positive SARS-CoV-2 PCR-test results, within one day prior to, or one week after, ED presentation.

Results: The resulting "CoLab-score" consists of 10 routine laboratory measurements, and age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945). The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991, specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were confirmed in temporal and external validation.

Conclusions: The CoLab-score is based on routine laboratory measurements and is available within one hour after presentation. Depending on the prevalence, COVID-19 may be safely

ruled-out in over one third of ED presentations. Highly suspect cases can be identified regardless of presenting symptoms. The CoLab-score is a valuable tool to guide PCR testing, triage ED patients, and is available to any center with access to routine laboratory tests.

Article summary

Strengths and limitations of this study

- A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- Using regression analysis, a simple score was developed consisting of only 10 routine ED laboratory tests and age.
- The score was temporally and externally validation in 3 other centers, is available within 1 hour after presentation and can be used to triage patients with a possible SARS-CoV-2 infection in the ED.
- No evidence was found that the performance was affected by vaccinations and new SARS-CoV-2 variants.
- The score is not a replacement for PCR-testing, but can be used to guide PCR-testing.

Introduction

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has evolved into a global pandemic in 2020 [1]. For emergency department (ED) physicians, identifying presenting patients with a possible COVID-19 infection remains challenging since symptoms like fever, shortness of breath or coughing overlap with other illnesses [2,3]. It is crucial however, to identify a possible COVID-19 infection as early as possible. Early identification prevents further spreading and protects hospital staff by isolating a suspected patient, pending the results of a SARS-COV-2 RNA PCR test and/or chest CT. Conversely, when PCR testing or isolation treatment capacity is limited, ruling-out COVID-19 as soon as possible can save valuable resources.

In the era of electronic health records and clinical prediction models, developing an early warning score that can assist ED physicians in identifying patients presenting at the ED with COVID-19 is of great value. Moreover, if only routine ED test results are required as input, the score can be easily adopted by EDs worldwide, potentially reduce diagnostic costs and accelerate patient triage.

Many COVID-19 prediction models have already been developed, the living systematic review by Wynants et. al [4] provides an extensive overview and critical appraisal. Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early models were based on relatively small sample sizes, hampered by selection bias or were over-fitted by selecting too many features [4–6]. Aside from methodological shortcomings, most models are not developed as an early warning score for all ED patients. Firstly, they require features from tests that are not routinely performed or logged for all ED patients (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED patient population.

Secondly, the population on which models are commonly based, are PCR-tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by physicians.

In this study we report the development and validation of an early warning score that, based on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in a patient presenting at the ED. The score can assist ED physicians in triaging patients and prevent further transmission of COVID-19 by quickly identifying possibly infected patients or ruling out a possible infection when resources are scarce.

Methods

Study design

This is a retrospective case-control study where routine laboratory test results, combined with age and gender, from all patient presenting at the emergency department (ED) of the Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-2 PCR test results in a development dataset. A model that could predict the presence of a COVID-19 infection was fit to this dataset. Performance of the model was assessed by i) internal validation, ii) temporal validation and iii) external validation by using data from the ED of three other centers. The study was reviewed by the Medical research Ethics Committees United (MEC-U) under study number W20.071, which confirmed that the Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this study. The study was thereafter reviewed and approved by the internal hospital review board.

Patient and Public Involvement

Patients were not involved in the design, conduct or reporting of this study.

Development dataset

All ED presentations at the Catharina Hospital Eindhoven from July 2019 to July 2020 were included in the development dataset, provided that routine laboratory testing had been requested by the attending ED physician. The rationale for this inclusion period is to limit the effect of seasonal variation in the ED patient population by including the summer, fall and winter season of 2019 (control patients) and the winter, spring and summer season of 2020 (case and control patients). The routine laboratory panel at the ED consists of 28 laboratory tests. In some cases not all tests in the routine panel were requested or one or more

quantitative results were not available due to analytical interference (hemolysis, lipemia or icterus). Presentations with one or more missing values in any of the 28 laboratory test in the routine ED panel, were excluded. Presentations with one or more extreme lab results (> 10 times standard deviation from the median) were also excluded to minimize the effect on the estimation of regression coefficients. After the first case of COVID-19 in the Netherlands, all patients with symptoms of COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were approved by the Dutch national institute of public health (RIVM). If a patient had a positive PCR result in the past, subsequent presentations were excluded as re-presentations might be clinically different from de novo presentations.

The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were performed in the time window and the presentation occurred after the first case of COVID-19 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the first case were labelled as "*Pre-COVID-19*".

Laboratory tests

The routine laboratory panel consisted of hemocytometric and chemical analyses. The hemocytometric tests were performed on Sysmex XN-10 instruments (Sysmex Corp., Kobe, Japan) and consisted of hemoglobin, hematocrit, erythrocytes, mean corpuscular volume (MCV), mean cellular hemoglobin (MCH), mean cellular hemoglobin concentration (MCHC), thrombocytes, leukocytes, neutrophils, eosinophils, basophils, lymphocytes and

monocytes. The chemical analyses were performed on a Cobas 8000 Pro (Roche Dx, Basel, Switzerland) instrument and consisted of glucose, total bilirubin, aspartate aminotransferase (ASAT), alanine aminotransferase (ALAT), lactate dehydrogenase (LD), creatine kinase (CK), alkaline phosphatase (ALP), gamma-glutamyltransferase (gGT), blood urea nitrogen (BUN), creatinine, CKD-epi estimated glomerular filtration rate (eGFR), potassium, sodium, chloride, albumin (bromocresol green) and C-reactive protein (CRP). These results were combined with age and gender.

Modelling

All data were processed and analyzed in R version 4.1.1 [9]. Laboratory results, combined with age and gender were used as covariates in a regression model. Cases were defined as ED presentations labelled as "*PCR-positive*", controls were all other presentations (i.e. "*PCR-negative*", "*Untested*" or "*Pre-COVID-19*"). To achieve predictive accuracy, limit overfitting and perform feature selection, penalized logistic regression with an adaptive lasso penalty was chosen [10,11]. To minimize missing data, all non-numeric results at the extremes of the measuring range, were converted to numeric results by removing the "<" and ">" signs. For eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m2 and <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD, CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model predictions, these variables were transformed logarithmically. More details regarding model fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the glmnet-package [12].

CoLab-score

Since this is a retrospective case-control study, the sample prevalence may not reflect the true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term in the model should be adjusted according to the current prevalence (details can be found in the document, **Supplemental Material 1**) [13]. However, adjusting the intercept term is not straightforward to implement in clinical practice, therefore the linear predictor of the model was categorized into a score, this score is hereafter referred to as the "CoLab-score". The categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using the intercept adjustment formula by King [13]. The intervals obtained through these breaks correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and score 5 reflects high-risk. More details regarding the rationale of the CoLab-score categorization can be found in the document, **Supplemental Material 1**.

Internal validation

To assess model performance while taking overfitting into account, bootstrapping was performed. 1000 bootstrap samples were generated from the original data. On each bootstrap sample, the full model fitting procedure and CoLab-score conversion were performed. Optimism adjusted performance measures of the CoLab-score were obtained by applying the 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [14]. Performance measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate performance measures [15]. Although the full inclusion period from July 2019 to July 2020 was used for model fitting, the performance was evaluated on the period starting from the first

COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain performance measures that would reflect real world performance.

Temporal validation

For temporal validation, results from our center were prospectively analyzed from July 2020 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period there was also more widespread external PCR testing by municipal health services. The results of external conducted PCR tests were not available to our study. To overcome this limitation, the outcome in the temporal validation cohort was chosen as a composite of the hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test result. This period also covers both the emergence of new SARS-CoV-2 variants as well as vaccine rollout. However, neither vaccination status nor genomic sequencing was available to determine whether a patient was vaccinated or which variant caused the infection. Therefore, data from the Dutch national institute of public health (RIVM) was used, to divide the temporal validation period into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold for the entire period, as well as for each phase separately to determine a possible effect of vaccination and new variants on performance (results in the **Supplemental Material 2**). Model calibration was assessed graphically using the rms-package [16].

External validation

For the external validation, several centers in the Netherlands were approached and assessed if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available. Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and Zuyderland Medical Center (center 3). The hematological parameters were measured with Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external validation was similar to the temporal validation and consisted of assessing the AUC sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was assessed graphically analogous to the temporal validation dataset.

Results

Development dataset

12879 emergency department (ED) presentations of 10327 patients from July 2019 to July 2020 were included. After excluding cases with an incomplete lab panel, patient presentations that occurred after a positive PCR test in the past (re-presentations) and presentations with extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of 8610 patients remained (**Figure 1 A**).

	Pre-COVID N = 5890	**Asymptomatic N = 3303**	**PCR negative N = 945**	**PCR positive N = 279**
Age in years	61 (21)	60 (21)	66 (18)	69 (15)
Female gender	2909 (49.4 %)	1659 (50.2 %)	466 (49.3 %)	95 (34.1 %)
Specialism				
Internal medicine	1648 (28.0 %)	896 (27.1 %)	244 (25.8 %)	71 (25.4 %)
Surgery	1007 (17.1 %)	679 (20.6 %)	51 (5.4 %)	5 (1.8 %)
Neurology	775 (13.2 %)	468 (14.2 %)	64 (6.8 %)	5 (1.8 %)
Pulmonary medicine	714 (12.1 %)	220 (6.7 %)	326 (34.5 %)	167 (59.9 %)
Cardiology	560 (9.5 %)	322 (9.7 %)	145 (15.3 %)	6 (2.2 %)
Urology	309 (5.2 %)	148 (4.5 %)	15 (1.6 %)	7 (2.5 %)
Gastroenterology	306 (5.2 %)	224 (6.8 %)	27 (2.9 %)	1 (0.4 %)
Geriatrics	189 (3.2 %)	95 (2.9 %)	52 (5.5 %)	15 (5.4 %)
Orthopedics	147 (2.5 %)	109 (3.3 %)	11 (1.2 %)	0 (0.0 %)
Gynecology	118 (2.0 %)	82 (2.5 %)	2 (0.2 %)	0 (0.0 %)
Other	117 (2.0 %)	60 (1.8 %)	8 (0.8 %)	2 (0.7 %)
Hemoglobin in mmol/L	8.2 (1.3)	8.3 (1.3)	8.2 (1.4)	8.6 (1.1)
Hematocrit in L/L	0.403 (0.059)	0.405 (0.056)	0.405 (0.062)	0.417 (0.047)
Erythrocytes in /pL	4.41 (0.69)	4.43 (0.66)	4.41 (0.72)	4.61 (0.60)
MCV in fl	91.8 (6.4)	91.9 (6.1)	92.4 (6.7)	90.7 (5.5)
MCH in mmol	1.859 (0.157)	1.876 (0.150)	1.874 (0.172)	1.869 (0.141)
MCHC in mmol/L	20.2 (0.9)	20.4 (0.9)	20.3 (1.0)	20.6 (0.8)
Thrombocytes in /nL	263 (99)	266 (100)	269 (105)	217 (123)
Leukocytes in /nL	9.30 [7.06, 12.16]	8.92 [7.01, 11.89]	9.66 [7.17, 12.94]	6.33 [4.74, 8.48]
Neutrophils in /nL	6.62 [4.51, 9.53]	6.10 [4.42, 8.94]	7.01 [4.79, 10.02]	4.71 [3.30, 6.94]
Eosinophils in /nL	0.09 [0.03, 0.17]	0.09 [0.03, 0.18]	0.08 [0.02, 0.17]	0.00 [0.00, 0.02]
Basophils in /nL	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.01 [0.01, 0.02]
Lymphocytes in /nL	1.47 [0.93, 2.13]	1.56 [1.05, 2.18]	1.31 [0.80, 2.03]	0.86 [0.59, 1.21]
Monocytes in /nL	0.70 [0.52, 0.93]	0.69 [0.52, 0.91]	0.74 [0.54, 1.01]	0.45 [0.32, 0.64]
Glucose in mmol/L	6.76 [5.83, 8.39]	6.68 [5.76, 8.14]	6.98 [5.95, 8.85]	6.77 [5.98, 8.48]
Bilirubin in umol/L	7.5 [5.0, 11.6]	7.4 [5.1, 10.9]	8.3 [5.6, 12.4]	8.2 [6.3, 11.4]
ASAT in U/L	24.0 [19.1, 32.2]	26.5 [21.6, 35.1]	27.7 [21.7, 39.2]	40.7 [30.2, 57.2]
ALAT in U/L	24.3 [17.8, 35.3]	25.3 [18.4, 36.2]	25.7 [18.4, 40.0]	33.7 [23.3, 50.0]
LD in U/L	201 [173, 240]	198 [170, 236]	215 [178, 263]	300 [238, 403]
CK in U/L	82 [51, 134]	83 [52, 136]	76 [51, 125]	124 [62, 222]
ALP in IU/L	83.0 [68.0, 105.0]	81.0 [65.8, 102.5]	86.9 [67.9, 110.0]	71.0 [58.8, 85.0]
gGT in U/L	27.0 [17.0, 53.0]	28.4 [18.4, 50.5]	37.0 [22.4, 68.9]	42.0 [28.0, 83.5]
BUN in mmol/L	5.7 [4.3, 8.0]	5.8 [4.3, 7.8]	6.2 [4.6, 9.4]	6.1 [4.7, 8.9]

CKD-epi in ml/min/m2	80.9 [58.0, 99.1]	85.0 [63.5, 103.3]	79.1 [52.1, 96.6]	76.6 [54.9, 91.2]
Creatinine in umol/L	79 [64, 100]	74 [61, 94]	78 [62, 105]	82 [68, 104]
Potassium in mmol/L	4.06 (0.50)	4.03 (0.49)	4.07 (0.55)	3.91 (0.47)
Sodium in mmol/L	139.2 (4.0)	138.5 (3.9)	138.0 (4.3)	136.4 (4.1)
Chloride in mmol/L	104.4 (4.6)	103.8 (4.5)	102.9 (4.8)	101.6 (4.4)
Albumin in g/L	42.4 (4.9)	42.3 (4.5)	40.8 (4.8)	38.4 (3.8)
CRP in mg/L	8 [2, 41]	5 [1, 30]	18 [3, 69]	77 [37, 136]

Table 1: Descriptive statistics of development dataset and laboratory concentrations.

Shown are the laboratory tests routinely requested at ED presentation and their mean/median results (in the development dataset) for the presentations before the first COVID-19 patient in the Netherlands ("Pre-COVID-19"), presentations thereafter that were not tested for COVID-19 ("Untested"), tested negatively ("PCR negative") and tested positive ("PCR positive"). For results with normal distributions, the mean value and standard deviation (in round brackets) are shown. For results that have skewed or heavy tailed distributions, the median value and the interquartile range is shown [in squared brackets]. Dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error).

Descriptive statistics of ED presentations are shown in **Table 1,** dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error [17]). For the PCR positives (N = 279), 91% (95% CI: 88 to 94%) of the cases were tested positive in their first PCR. The remaining 24 patients were positive in their second (N = 18), third (N = 5) or fourth (N = 1) PCR.

CoLab-score

The model obtained through adaptive lasso regression contained eleven variables, which are depicted with their regression coefficients (weights) in **Table 2**.

Variable	β	Exclusion limit	Relative importance
Intercept	-6.885		-
Erythrocytes /pL	0.9379	Erythrocytes < 2.9 /pL	52 %
Leukocytes /nL	-0.1298		46 %
Eosinophils /nL	-6.834		86 %
Basophils /nL	-47.70	Basophils >0.33 /nL	100 %
$\log_{10}$ of Bilirubin in µmol/L	-1.142	Bilirubin >169 µmol/L	26 %
$\log_{10}$ of LD in U/L	5.369	LD >1564 U/L	58 %
$\log_{10}$ of ALP in IU/L	-3.114	AF >1000 IU/L	45 %
$\log_{10}$ of gGT in U/L	0.3605	gGT >1611 U/L	11 %
Albumin in g/L	-0.1156		45 %
CRP in mg/L	0.002560		15 %
Age in years	0.002275		4 %

Table 2: Calculation of the CoLab-linear predictor (LP).

The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of the 11 variables with their corresponding coefficients (β's). CoLab-LP = – 6.885 + [erythrocytes] × 0.9379 – [leukocytes] × 0.1298 – [eosinophils] × 6.834 – [basophils] × 47.7 – log10([bilirubin]) × 1.142 + log10([LD]) × 5.369 – log10([ALP]) × 3.114 + log10([gGT]) × 0.3605 – [albumin] × 0.1156 + [CRP] × 0.02560 + [age] × 0.002275. The LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid if any of the variables exceed the limits in the third column.

A larger β-coefficient does not imply that a variable is more important in predicting the odds of testing positive for SARS-CoV-2, since variables are on different scales. Therefore, the relative importance is calculated based on scaled coefficients. The absolute basophil count has the highest relative importance, followed by eosinophil count.

As shown in **Figure 2**, the linear predictor clearly discriminates between COVID-19 and non-COVID-19. The linear predictor is converted to CoLab-scores 0 – 5 with the cut-points depicted in **Figure 2**.

Internal validation

The model was validated in the period starting from the first COVID-19 infection to July 2020, in this period the mean prevalence was 7.2%. The AUC of the CoLab-score is 0.930 (95% CI: 0.909 to 0.945).

CoLab-score	Sensitivity	Specificity	PPV	NPV	% of population
0	0.984 (0.970 - 0.991)	0.411 (0.285 - 0.520)	0.115 (0.0932 - 0.141)	0.997 (0.994 - 0.999)	38.4 (26.4 - 48.4)
≤ 1	0.909 (0.886 - 0.943)	0.793 (0.744 - 0.826)	0.255 (0.207 - 0.299)	0.991 (0.989 - 0.995)	74.4 (69.4 - 77.4)
≤ 2	0.859 (0.811 - 0.889)	0.887 (0.866 - 0.901)	0.371 (0.317 - 0.414)	0.988 (0.983 - 0.991)	83.2 (82.2 - 85.2)
≤ 3	0.750 (0.700 - 0.810)	0.953 (0.944 - 0.959)	0.551 (0.494 - 0.601)	0.980 (0.975 - 0.985)	90.1 (89.1 - 91.1)
≤ 4	0.608 (0.522 - 0.685)	0.978 (0.973 - 0.983)	0.682 (0.622 - 0.740)	0.970 (0.962 - 0.977)	93.8 (92.8 - 93.8)

Table 3: Diagnostic performance CoLab-score in the development dataset.

The development dataset was internally validation for the period March 2020 – July 2020 (N = 4.527). Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV) and fraction of patients (%) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4). The numbers in round brackets represent the 95% bootstrapped confidence intervals. The first column defines the threshold above which CoLab-score a patient is considered positive. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

Diagnostic performance is shown in **Table 3.** A CoLab-score of 0 has a negative predictive value (NPV) of 0.997 (95% CI: 0.994 to 0.999) and positive predictive value (PPV) of 0.115 (0.0932 - 0.141), one third (38.4%, 95% CI: 26.4 to 48.4%) of all ED presentations were assigned this score and can therefore be safely excluded. Conversely, 6.2% (95% CI: 6.3 to 7.2%) of the ED patients had a CoLab-score = 5. Given the PPV of this score (0.682, 95% CI: 0.622 to 0.740, NPV: 0.970, 95% CI: 0.962 - 0.977), subsequent PCR testing is advised.

Temporal validation

As the CoLab-score was developed in our center after the first COVID-19-wave in the Netherlands, the performance was evaluated in our center from July 2020 until October 2021. Lab results from 17489 ED presentations were collected. After applying the inclusion flow as shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a COVID-19 infection.

The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the development cohort, although sensitivity is slightly lower and specificity slightly higher, 95% CIs overlap (cf. **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental Material 2 Figure 1)**. The discriminative ability is not affected by phases with different dominant variants and/or vaccination status. Diagnostic performance is also preserved in terms of sensitivity and specificity, PPV and NPV are difficult to compare due to different prevalence/pre-test probabilities in each phase (see **Supplemental Material 2 Table 1**).

In terms of the predicted probabilities, model calibration shows that overall predicted probabilities are too low (see **Supplemental Material 3** for the calibration plot), which is expected since the prevalence differs and the intercept has to be adjusted to the prevalence. In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that initially did not present with COVID-specific symptoms. Most patients had neurological or orthopedic presenting symptoms.

External validation

For external validation, data obtained from three other centers were used, center 1 (N = 1284, 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545, 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers is ≥ 0.96. Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for external centers are shown in **Supplemental Material 3**, the observed fraction of COVID-19 positives is slightly lower than expected in centers 1 and 2. For center 3, low probabilities appear slightly underestimated and high probabilities slightly overestimated.

CoLab -score	Validation set	Sensitivity	Specificity	PPV	NPV
0	Temporal	0.967 (0.957 - 0.977)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)
	Center 1	1.000 (1.000 - 1.000)	0.333 (0.308 - 0.360)	0.059 (0.057 - 0.062)	1.000 (1.000 - 1.000)
	Center 2	0.960 (0.919 - 0.990)	0.351 (0.334 - 0.369)	0.050 (0.047 - 0.052)	0.996 (0.992 - 0.999)
	Center 3	0.973 (0.955 - 0.988)	0.322 (0.307 - 0.338)	0.131 (0.127 - 0.134)	0.991 (0.986 - 0.996)
≤1	Temporal	0.888 (0.869 - 0.907)	0.790 (0.783 - 0.798)	0.252 (0.245 - 0.260)	0.989 (0.987 - 0.991)
	Center 1	0.923 (0.846 - 0.981)	0.695 (0.670 - 0.722)	0.113 (0.102 - 0.126)	0.995 (0.991 - 0.999)
	Center 2	0.929 (0.879 - 0.970)	0.680 (0.663 - 0.697)	0.093 (0.087 - 0.100)	0.996 (0.994 - 0.998)
	Center 3	0.917 (0.887 - 0.946)	0.675 (0.659 - 0.691)	0.228 (0.218 - 0.238)	0.987 (0.983 - 0.992)
≤2	Temporal	0.820 (0.796 - 0.842)	0.894 (0.889 - 0.899)	0.381 (0.368 - 0.395)	0.984 (0.982 - 0.986)
	Center 1	0.808 (0.692 - 0.904)	0.812 (0.791 - 0.834)	0.154 (0.131 - 0.179)	0.990 (0.984 - 0.995)
	Center 2	0.869 (0.798 - 0.929)	0.802 (0.787 - 0.816)	0.135 (0.122 - 0.147)	0.994 (0.991 - 0.997)
	Center 3	0.893 (0.860 - 0.926)	0.795 (0.781 - 0.809)	0.314 (0.297 - 0.330)	0.986 (0.982 - 0.990)
≤3	Temporal	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.595 (0.573 - 0.618)	0.977 (0.974 - 0.979)
	Center 1	0.750 (0.635 - 0.865)	0.910 (0.893 - 0.926)	0.260 (0.216 - 0.309)	0.989 (0.983 - 0.994)
	Center 2	0.687 (0.596 - 0.778)	0.899 (0.887 - 0.910)	0.194 (0.168 - 0.222)	0.988 (0.984 - 0.991)
	Center 3	0.768 (0.726 - 0.812)	0.887 (0.876 - 0.898)	0.417 (0.392 - 0.445)	0.973 (0.969 - 0.978)
≤4	Temporal	0.585 (0.555 - 0.616)	0.984 (0.982 - 0.987)	0.749 (0.724 - 0.777)	0.968 (0.965 - 0.970)
	Center 1	0.654 (0.519 - 0.769)	0.952 (0.939 - 0.964)	0.366 (0.296 - 0.447)	0.985 (0.979 - 0.990)
	Center 2	0.556 (0.455 - 0.647)	0.953 (0.945 - 0.961)	0.295 (0.246 - 0.349)	0.984 (0.980 - 0.987)
	Center 3	0.667 (0.619 - 0.720)	0.931 (0.922 - 0.940)	0.502 (0.467 - 0.541)	0.964 (0.959 - 0.969)

Table 4: Diagnostic performance of the CoLab-score in the validation dataset (temporal) and three external hospitals.

Sensitivities, specificities, positive predictive values (PPV) and negative predictive values (NPV) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

Discussion

Given the impact of COVID-19 on society and healthcare, there is a need for simple and fast detection of patients with a possible COVID-19 infection in the ED. The CoLab-score described in this study, is a fast and accurate risk score to triage patients presenting at the ED based on ten routine blood biomarkers and age.

The main strength of this study is that this score can be used as an early-warning or triaging tool for the entire ED population, regardless of presenting symptoms. This is in contrast to the vast majority of COVID-19 diagnostic models that have been developed on a pre-selected population of PCR-tested patients [18–25]. Moreover, the CoLab-score requires only routine blood tests instead of (features from) imaging such as CT-scans or laboratory tests that are not routinely collected in the ED , e.g. interleukin-6 or 3-hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific) [26,27].

Two other studies have been published which are similar to this study [20,28]. Interestingly, the study by Soltan et al., ranked basophils and eosinophils as the two most important features in predicting the outcome, similar to our results [28]. Eosinophils were also seen as one of the most important features by Plante et al. [20]. However, both studies focus on an artificial intelligence/machine learning approach. While their approach likely results in higher predictive performance due to the ability of machine learning models to capture non-linear and interaction effects, the goal of this study was to develop a simple, fast and robust model that can easily be implemented in current hospital IT systems.

Since this is a retrospective case-control study, there is some unavoidable missing data. In our cohort 17.6% of the ED presentations could not be used due to one or more missing laboratory results. This is lower or equal to similar studies; 22% [22], 17% [19] and 11% [25]. We do not expect that presentations with missing data have led to severe inclusion bias, important to note is that 7.7% of missingness is due to analytical errors which should not cause bias. For the remaining 9.9% of missingness, the full lab panel was most frequently missing for pediatric, obstetric and surgery patients which are rarely COVID-19 patients. This is also the case for external validation centers 1 and 2, in these centers only internal medicine ED presentations were tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine ED presentations, which is reflected by the few PCR-positive patients excluded.

The performance of the CoLab-score is affected by the time between the onset of symptoms and ED presentations. The score increases with the duration of symptoms and gradually decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some COVID-19 patients with early or late presentation after onset of symptoms can be missed. Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and <10 days prior to ED presentation.

It was chosen to exclude re-presentations. Since the median time between initial presentation and re-presentation was 12 days, these patients were most likely not re-infected patients, but patients who deteriorated after initial presentation/treatment. Given that the CoLab-score follows the host-immune response, the score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients would impact the performance of the CoLab-score as

patients in a later phase of the disease show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate after treatment for COVID-19.

Finally, the CoLab-score could lead to false positives by other viral infections. However, in an historic patient cohort, the CoLab-score had only limited discriminative ability in separating influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4 Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-response to the virus, it is expected that the CoLab-score is also sensitive to future SARS-CoV-2 variants. This is supported by the fact that the diagnostic performance is sustained in periods with different dominant variants. Moreover, there is no evidence that the diagnostic performance is affected by vaccinations. Although vaccination status is not registered for all presenting patients, there is no evidence that performance is reduced under increasing degrees of vaccination. In a small subgroup of 12 patients for whom vaccination status was registered, and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5) (see **Supplemental Material 2 Figure 2**),

To conclude, the CoLab-score developed and validated in this study, based on 10 routine laboratory results and age, is available within 1 hour for any patient presenting at the ED. The score can be used by clinicians to guide PCR testing or triage patients and helps to identify COVID-19 in asymptomatic patients. The lowest CoLab-score can be used to effectively rule-out a possible SARS-CoV-2 infection, the highest score to alert physicians to a possible infection. Thus, the CoLab-score is a valuable tool to rule out COVID-19, guide PCR testing and is available to any center with access to routine laboratory tests.

Funding statement

This was an investigator-initiated study and no funding was received for this study.

Competing interests

A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR reports no conflict of interest. VS reports no conflict of interest.

Data sharing statement

Datasets with source data for Table 1, Figure 2, Table 3 and Table 4, as well the R-code to fit the model is available from the Dryad repository, DOI:[WILL BE PROVIDED WHEN UNDER REVIEW]. Technical appendix can be found in **Supplemental Material 1**.

Author contributorship statement

Arjen-Kars Boer: Conceptualization (Lead), Data curation (Lead), Funding acquisition (Lead), Investigation (Equal), Methodology (Equal), Supervision (Equal), Writing-original draft (Equal), Writing-review & editing (Equal).

Ruben Deneer: Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Lead), Software (Lead), Visualization (Lead), Writing-original draft (Equal), Writing-review & editing (Equal).

Maaike Maas: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Equal), Writing-review & editing (Equal).

Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal), Writing-review & editing (Equal).

Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing (Equal).

References

1 Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 4 Feb 2021).

2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *https://doi.org/101056/NEJMoa2002032* 2020;**382**:1708–20. doi:10.1056/NEJMOA2002032

3 Vetter P, Vu DL, L'Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**. doi:10.1136/BMJ.M1470

4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18. doi:10.1136/bmj.m1328

5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review. J. Med. Syst. 2020;**44**:122. doi:10.1007/s10916-020-01582-x

6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417

7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19—Definition and Evaluation. *https://doi.org/101148/radiol2020201473* 2020;**296**:E97–104. doi:10.1148/RADIOL.2020201473

8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort

study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840

9 R Core Team. R: A Language and Environment for Statistical Computing. 2020.https://www.r-project.org/

10 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29. doi:10.1198/016214506000000735

11 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

12 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01

13 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63. doi:10.1093/oxfordjournals.pan.a004868

14 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973

15 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-2105-12-77

16 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.https://cran.r-project.org/package=rms

17 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons and progress. Scand. J. Clin. Lab. Invest. 1999;**59**:491–500. doi:10.1080/00365519950185229

18 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from

Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst* 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4

19 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502

20 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: A large, multicenter, real-world study. *J Med Internet Res* 2020;**22**:e24048. doi:10.2196/24048

21 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807. doi:10.1007/s00330-020-07022-1

22 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. *Clin Chem Lab Med* 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593

23 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305

24 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in the emergency department: derivation and validation of the LDH-CRP-Lymphocyte (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336

25 Tordjman M, Mekki A, Mali RD, *et al.* Pre-test probability for SARS-Cov-2-related infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342.

doi:10.1371/journal.pone.0243342

26 Peto T, Affron D, Afrough B, *et al.* COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine* 2021;**36**:100924. doi:10.1016/J.ECLINM.2021.100924

27 García-Fiñana M, Hughes DM, Cheyne CP, *et al.* Performance of the Innova SARS-CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot: population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637

28 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87. doi:10.1016/S2589-7500(20)30274-0

Figure legends

Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B) dataset.

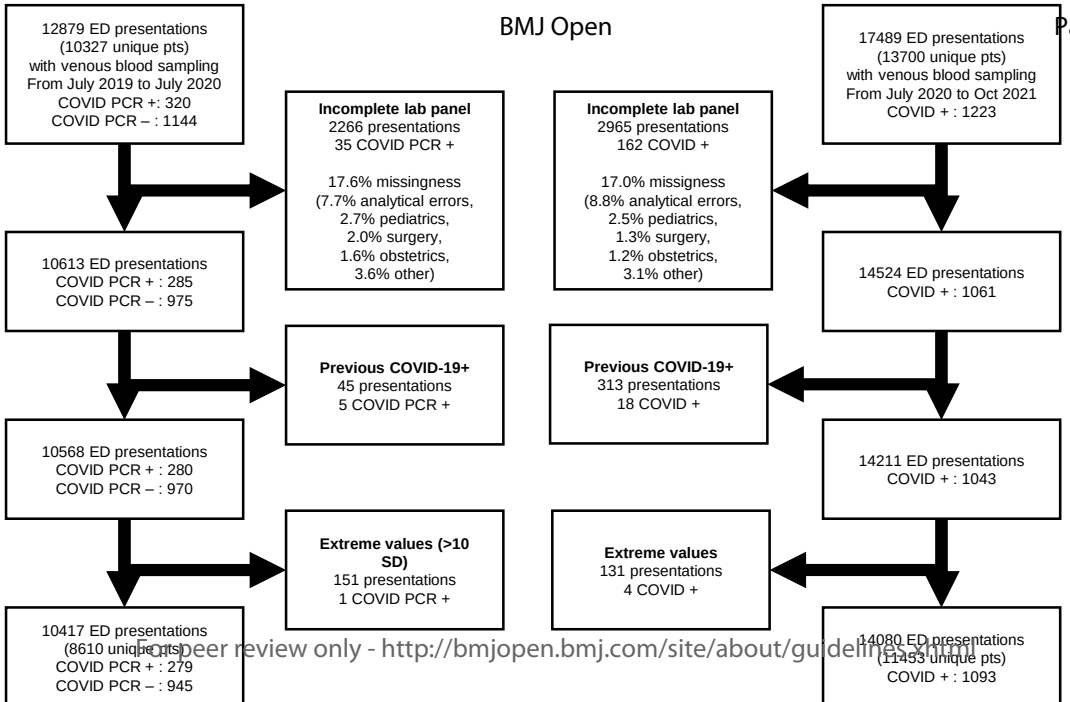
All patient admissions with routine venous blood sampling at the emergency department (ED) were included. For the development dataset, completeness of the lab panel was assessed for all the 29 laboratory tests (see Table 1), for the temporal validation dataset this was only necessary for 10 laboratory tests (see Table 2). The major causes of missingness are described in the text. In the development dataset, presentations with extreme values (>10 SD) were excluded. The same limits were applied to the temporal validation dataset (see Table 2 for limits).

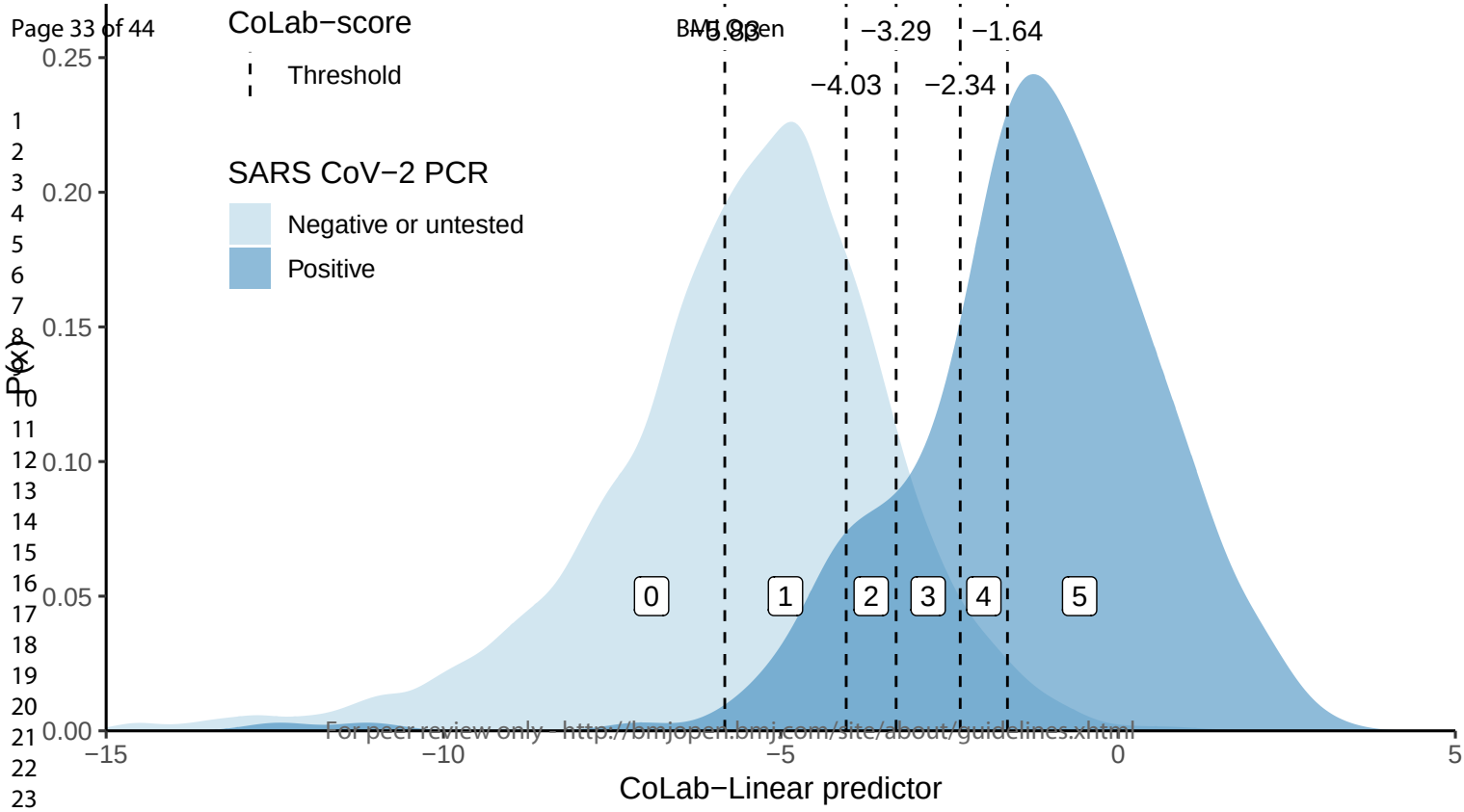
Figure 2: Probability density plot of the CoLab-linear predictor.

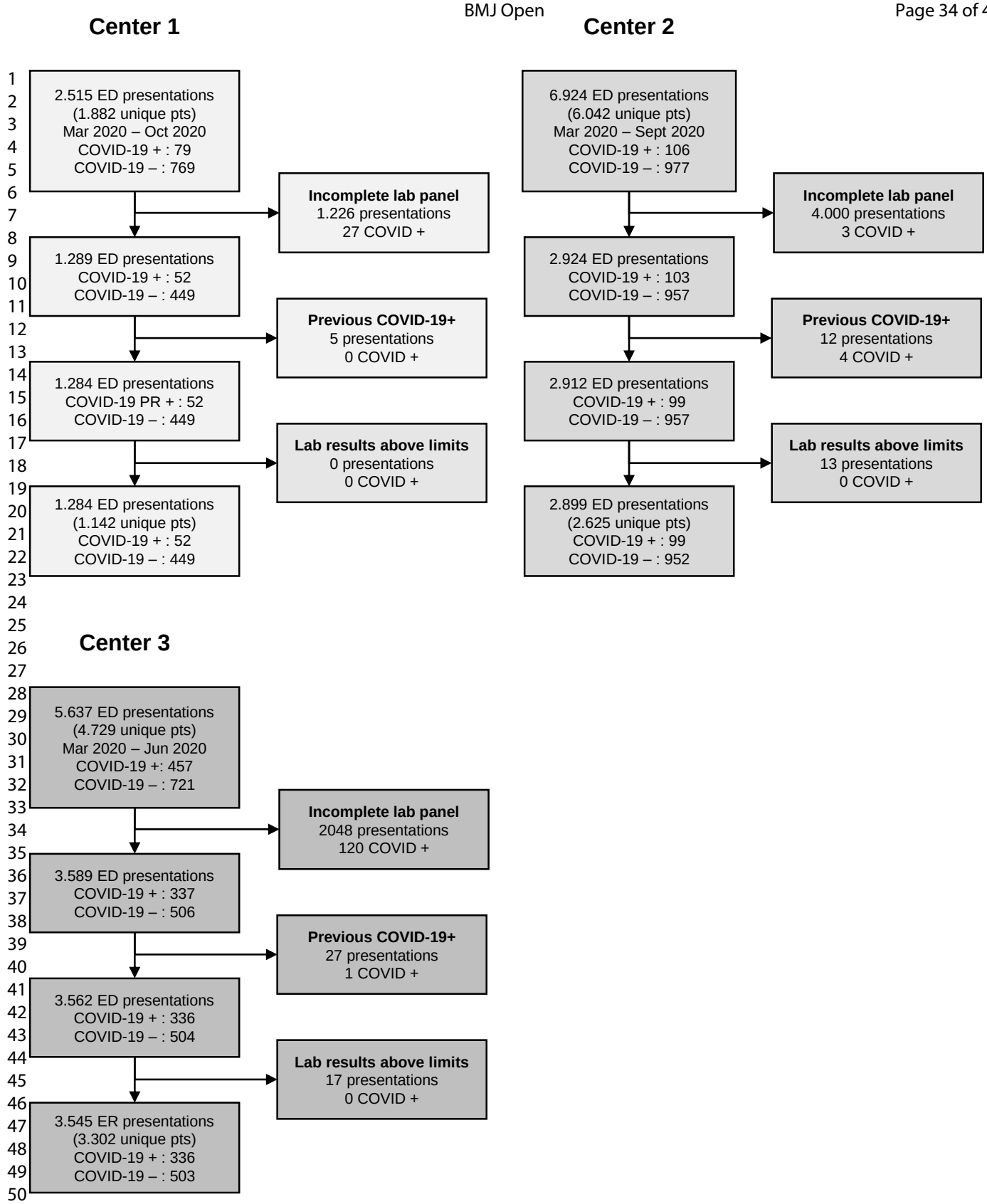
The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are plotted against the linear predictor (see table 2). The CoLab-score cut-offs (–5.83, –4.02, –3.29, –2.34 and –1.64) are depicted with vertical dashed lines. The white-boxed numbers (between the cut-offs) represent the corresponding CoLab-score. Note that while the area under both curves is identical (since these are probability density functions), in absolute numbers the "negative or untested"-group is about 36 times larger than the PCR positive group.

Figure 3: Inclusion flow of ED patients in three external centers.

All emergency department (ED) presentations with routine venous blood sampling were included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were excluded as "previous COVID-19+". Presentations with any laboratory result above the limits of the CoLab-score (see Table 2) were excluded.






Center 1
2.515 ED presentations
(1.882 unique pts)
Mar 2020 – Oct 2020
COVID-19 + : 79
COVID-19 – : 769
Incomplete lab panel
1.226 presentations
27 COVID +
1.289 ED presentations
COVID-19 + : 52
COVID-19 – : 449
Previous COVID-19+
5 presentations
0 COVID +
1.284 ED presentations
COVID-19 PR + : 52
COVID-19 – : 449
Lab results above limits
0 presentations
0 COVID +
1.284 ED presentations
(1.142 unique pts)
COVID-19 + : 52
COVID-19 – : 449
Center 2
6.924 ED presentations
(6.042 unique pts)
Mar 2020 – Sept 2020
COVID-19 + : 106
COVID-19 – : 977
Incomplete lab panel
4.000 presentations
3 COVID +
2.924 ED presentations
COVID-19 + : 103
COVID-19 – : 957
Previous COVID-19+
12 presentations
4 COVID +
2.912 ED presentations
COVID-19 + : 99
COVID-19 – : 957
Lab results above limits
13 presentations
0 COVID +
2.899 ED presentations
(2.625 unique pts)
COVID-19 + : 99
COVID-19 – : 952
Center 3
5.637 ED presentations
(4.729 unique pts)
Mar 2020 – Jun 2020
COVID-19 +: 457
COVID-19 – : 721
Incomplete lab panel
2048 presentations
120 COVID +
3.589 ED presentations
COVID-19 + : 337
COVID-19 – : 506
Previous COVID-19+
27 presentations
1 COVID +
3.562 ED presentations
COVID-19 + : 336
COVID-19 – : 504
Lab results above limits
17 presentations
0 COVID +
3.545 ER presentations
(3.302 unique pts)
COVID-19 + : 336
COVID-19 – : 503

Supplemental material 1

Model fitting

Prior to model fitting, covariates were scaled to zero mean and unit variance, after model fitting coefficients were unscaled to obtain regression coefficients on the original scale. In adaptive lasso, weights are applied to each of the covariates present in the lasso constraint, the weight vector has to be calculated before the adaptive lasso regression is performed. Due to multicollinearity between laboratory tests in the routine lab panel, weights in the adaptive lasso were based on ridge regression estimates ($\hat{\beta}_{ridge}$) as recommended by Zou. To obtain $\hat{\beta}_{ridge}$ the optimal penalty (λ) for the ridge regression was chosen using 10 fold cross-validation (CV) with area under the ROC curve (AUC) as the loss function. The λ corresponding to the maximum AUC was selected to obtain $\hat{\beta}_{ridge}$. The weight vector ($\widehat{w}$) was calculated by $\widehat{w} = 1/\left|\hat{\beta}_{ridge}\right|^2$. This weight vector was then used to fit an adaptive lasso regression where λ was chosen by the criterion ±1 SE of the maximum AUC.

Model intercept correction

The linear predictor for a patient *i* is calculated as follows: $lp_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$
Where *n* is the number of variables in the final model, x_{in} are the observed predictor variables for subject *i* and $\beta_n$the model coefficients. The linear predictor can then be converted to a probability for patient *i* (P_i) by the logistic function: $P_i = \frac{1}{1+e^{-lp_i}}$
The intercept term β_0 is sensitive to the fraction of cases versus controls in the dataset/population. Since the model is fitted to a case-control dataset where the number cases is fixed (all patients tested positive for COVID-19) and the number of controls is randomly chosen (a 6-month period pre-COVID), the intercept term β_0 is a result of this choice and will likely not be generalizable to the real-world setting. Prior correction is a method to correct the estimate of the intercept based on the true fraction of positives in the population, τ (prevalence of COVID-19 in the ED) and the fraction of cases in the development dataset, $\bar{y}$. The intercept term β_0 can then be corrected to obtain $\beta_{0corrected}$ using the following formula:

$$\beta_{0corrected} = \beta_0 + \beta_{adj}$$

$$\beta_{adj} = -ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right]$$

In our dataset $\bar{y} = 0.02675$ therefore:

$$\beta_{adj} = -ln\left(\frac{1-\tau}{\tau}\right) + 3.594$$

An estimate $\bar{\tau}$ can be used for the prevalence τ to obtain $\bar{\beta}_{adj}$ which can be plugged in the original linear predictor formula to obtain calibrated probabilities:

$$lp_i(\tau) = \beta_0 - ln\left(\frac{1-\tau}{\tau}\right) + 3.594 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$$

CoLab-score

An alternative, which is the basis of the CoLab-score, is to choose a fixed probability P_i above which one considers a patient eligible for further testing. The probability can be expressed as a number needed to test. If one is willing to test 10 patients to find one positive, all patients with $P_i \geq 0.1$ should be considered positive. In this study a number needed to test of 15 is used, therefore all patients with a $P_i \geq 0.067$ should be considered positive. On the linear predictor scale this translates to $\text{logit}(0.067) = -2.639$. To determine the cutoffs for difference prevalence thresholds one solves the following equation:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639$$
$$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$
$$lp_i(\tau) \geq ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

Choosing values for τ yields the cutoffs for the CoLab score:

$lp_i(\tau = 0.4) \geq -5.83$ (CoLab-score = 1)
$lp_i(\tau = 0.1) \geq -4.03$ (CoLab-score = 2)
$lp_i(\tau = 0.05) \geq -3.29$ (CoLab-score = 3)
$lp_i(\tau = 0.02) \geq -2.34$ (CoLab-score = 4)
$lp_i(\tau = 0.01) \geq -1.64$ (CoLab-score = 5)

These thresholds correspond to CoLab-scores 0 to 5. The interpretation of these scores is as follows; if the prevalence is <1%, only CoLab-score 5 should be classified as positive and CoLab-score 0 till 4 as negative. If the prevalence is 1% – 2%, CoLab-score 4 and 5 should be classified as positive and 1 – 3 negative. Similarly, with a prevalence of 2 – 5% the split is between CoLab-score 2 and 3 and with prevalence of 5 – 10% between CoLab-score 1 – 2. If the prevalence is higher than 10% only CoLab-score 0 is classified as negative. Using the CoLab-score in this fashion, aims to preserve a number need to test of 15.

Supplemental material 2

Vaccination status and COVID-19 ED prevalence plot

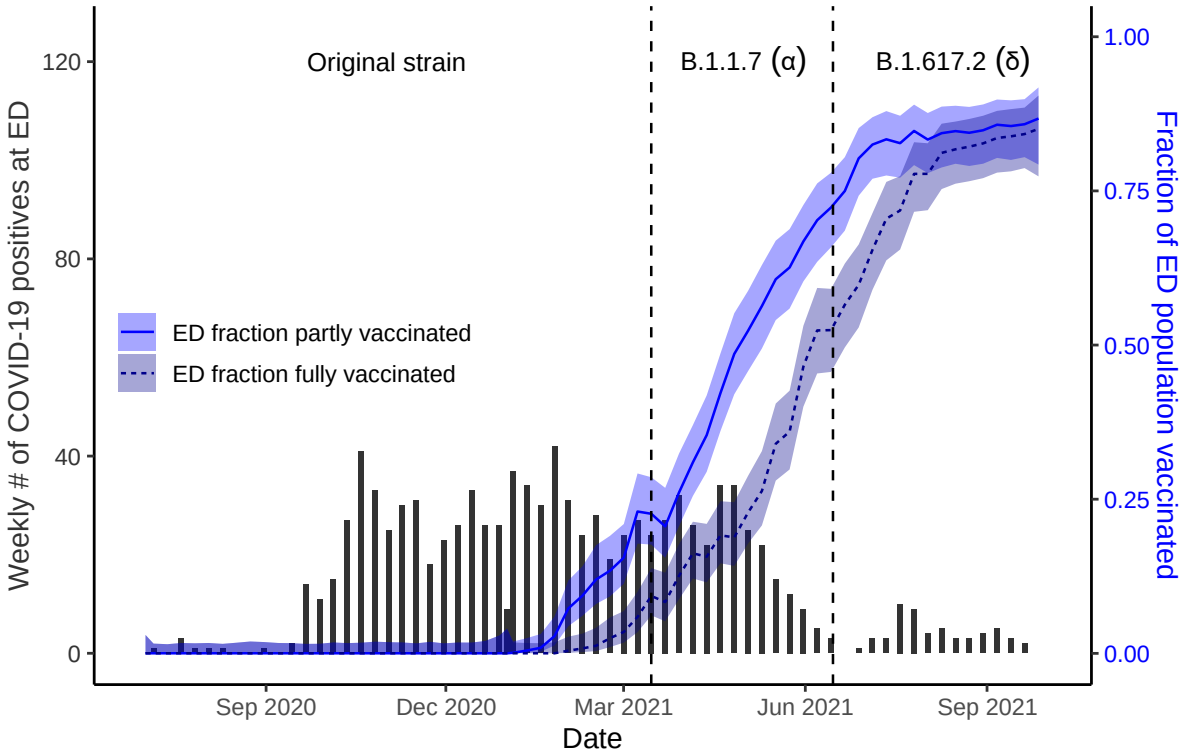


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

The temporal validation dataset consists of ED presentations from July 2020 until October 2021. As stated in the "Materials and Methods" section, this period was split into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. The ED fraction vaccinated is estimated by merging data from the Dutch national institute of public health by the date of the ED presentation and the year of birth of the patient. The gray bars depict weekly number of new COVID-19 cases at the ED, the blue lines the estimated fraction of ED patients fully or partially vaccinated.

CoLab-score performance

Phase	Cases/controls (prevalence)	AUC
Original strain & no vaccinations	694/7999 (8.6%)	0.909 (0.896 - 0.923)
B.1.1.7 strain & partial vaccination	287/2845 (10.1%)	0.937 (0.921 - 0.953)
B.1.617.2 strain & full vaccination	58/3236 (1.8%)	0.898 (0.857 - 0.939)

CoLab-score	Phase	Sensitivity	Specificity	PPV	NPV
0	Original strain & no vaccinations	0.960 (0.944 - 0.974)	0.418 (0.407 - 0.429)	0.135 (0.133 - 0.138)	0.991 (0.987 - 0.994)
	B.1.1.7 strain & partial vaccination	0.983 (0.969 - 0.997)	0.432 (0.413 - 0.450)	0.162 (0.158 - 0.168)	0.996 (0.992 - 0.999)
	B.1.617.2 strain & full vaccination	0.983 (0.948 - 1.000)	0.415 (0.396 - 0.432)	0.030 (0.028 - 0.031)	0.999 (0.998 - 1.000)
≤1	Original strain & no vaccinations	0.879 (0.854 - 0.902)	0.789 (0.779 - 0.798)	0.283 (0.273 - 0.294)	0.986 (0.983 - 0.988)
	B.1.1.7 strain & partial vaccination	0.916 (0.885 - 0.948)	0.809 (0.793 - 0.824)	0.350 (0.332 - 0.370)	0.989 (0.984 - 0.993)
	B.1.617.2 strain & full vaccination	0.862 (0.776 - 0.948)	0.780 (0.765 - 0.794)	0.067 (0.059 - 0.074)	0.997 (0.995 - 0.999)
≤2	Original strain & no vaccinations	0.813 (0.784 - 0.842)	0.894 (0.887 - 0.901)	0.421 (0.404 - 0.441)	0.980 (0.978 - 0.983)
	B.1.1.7 strain & partial vaccination	0.864 (0.826 - 0.902)	0.897 (0.885 - 0.908)	0.484 (0.455 - 0.516)	0.983 (0.979 - 0.988)
	B.1.617.2 strain & full vaccination	0.690 (0.569 - 0.810)	0.892 (0.881 - 0.902)	0.104 (0.086 - 0.123)	0.994 (0.991 - 0.996)
≤3	Original strain & no vaccinations	0.697 (0.661 - 0.731)	0.962 (0.957 - 0.966)	0.634 (0.605 - 0.662)	0.971 (0.968 - 0.974)
	B.1.1.7 strain & partial vaccination	0.760 (0.711 - 0.812)	0.963 (0.955 - 0.970)	0.696 (0.650 - 0.739)	0.973 (0.967 - 0.978)
	B.1.617.2 strain & full vaccination	0.621 (0.483 - 0.741)	0.960 (0.954 - 0.967)	0.222 (0.178 - 0.268)	0.993 (0.990 - 0.995)
≤4	Original strain & no vaccinations	0.566 (0.529 - 0.602)	0.984 (0.981 - 0.987)	0.775 (0.740 - 0.808)	0.960 (0.957 - 0.963)
	B.1.1.7 strain & partial vaccination	0.645 (0.589 - 0.704)	0.983 (0.978 - 0.988)	0.809 (0.762 - 0.856)	0.961 (0.955 - 0.967)
	B.1.617.2 strain & full vaccination	0.517 (0.397 - 0.638)	0.986 (0.982 - 0.990)	0.400 (0.319 - 0.500)	0.991 (0.989 - 0.993)

Table 2: Diagnostic performance of the CoLab-score in the temporal validation dataset, split by phase.

Sensitivities, specificities, positive predictive values (PPV) and negative predictive values (NPV) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is split into three phases according to dominant SARS-CoV-2 strains in the Netherlands and estimated fraction of ED patients vaccinated (see Figure above). Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

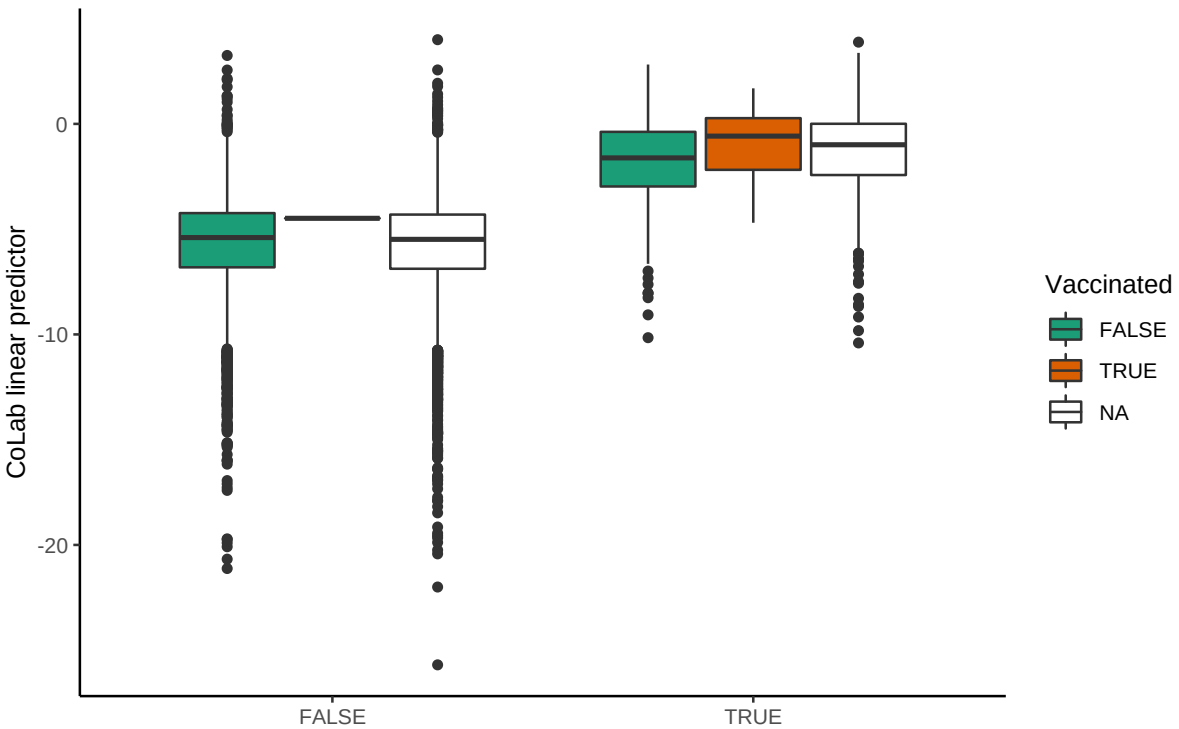


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

The CoLab linear predictor is calculated for all ED presentations in the temporal validation set. Presentations who are registered as vaccinated are labeled TRUE (N = 13). Presentations before vaccine roll-out are labeled FALSE (N = 5855). Presentations during vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13 presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative.

Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or considered positive until proven otherwise, therefore there is only one COVID-19 negative patient with a registered vaccination status.

Supplemental material 3

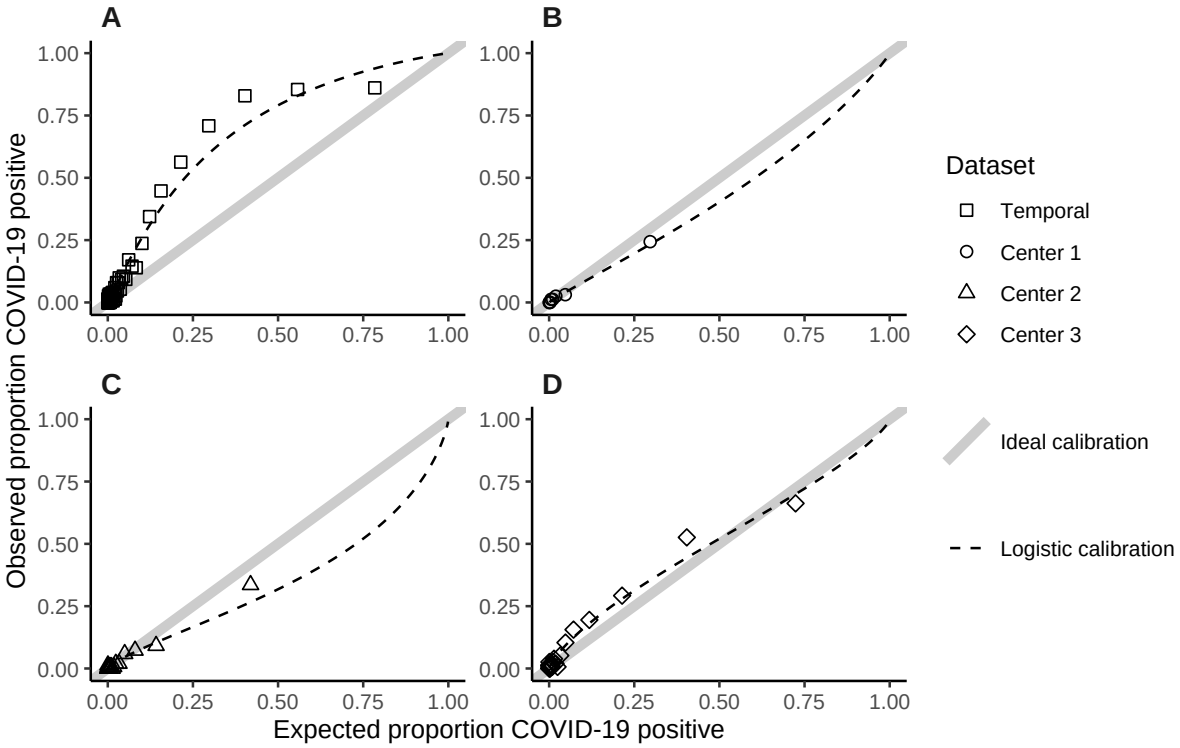


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

In the calibration plots, the proportion of observed COVID-19 positives versus expected probabilities are plotted. Observations are grouped with an average of 150 observations per group. The expected probabilities follow from applying the inverse logit function to the CoLab-linear predictor calculated from Table 2. If the observed proportion in an external dataset is lower than the expected proportion, this means risks are over-estimated, if the observed fraction is higher, risks are under-estimated. Ideally, observed proportions are equal to expected proportions, this ideal-calibration-line is shown as a straight line through the origin with a slope of 1. The logistic calibration line is a logistic regression fit of the predicted probabilities. [Intercept, slope] for plots A-D: A [1.34, 1.08], B [-0.39, 0.92], C [-0.76, 0.77], D [0.08, 0.79]. Although no validation datasets show perfect calibration, this is the result of differences in COVID-19 prevalence in the temporal validation dataset (7.4% versus 2.2%) and differences in calibration of laboratory equipment in the three external centers.

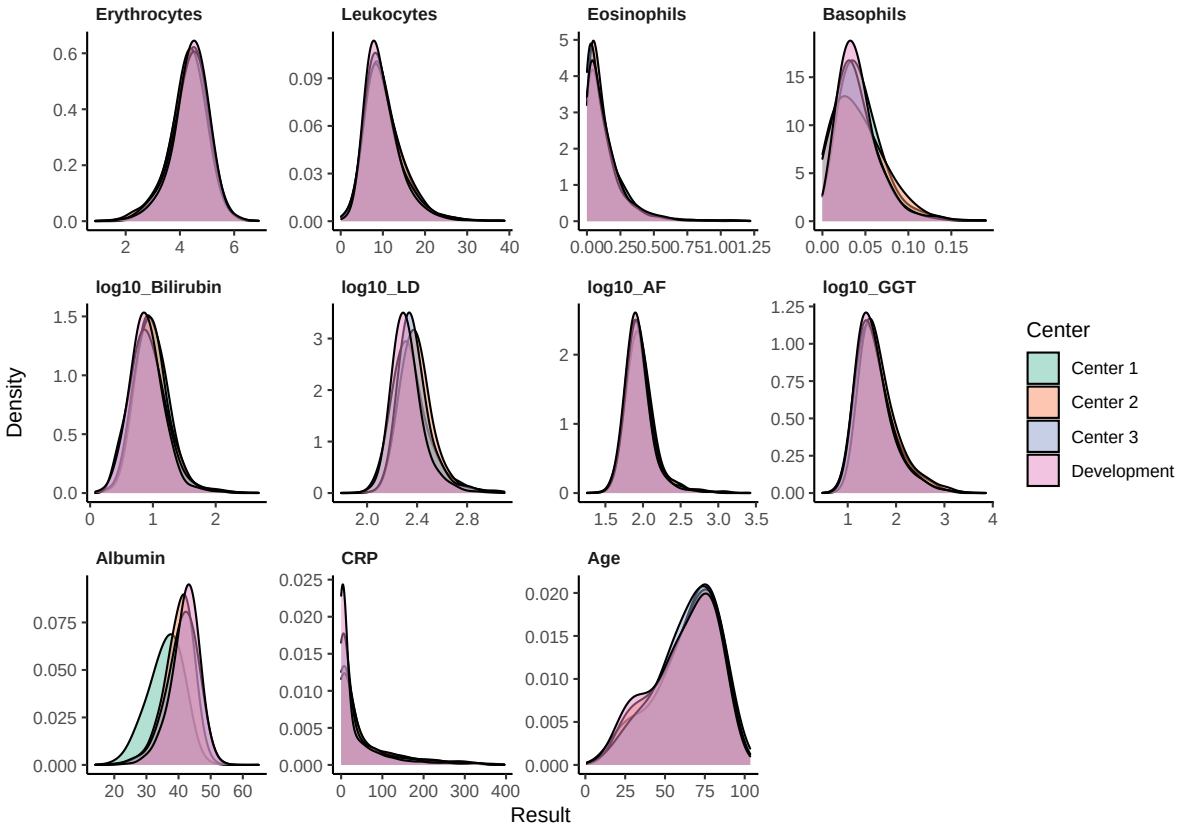


Figure 2: Probability density plots of laboratory parameters.

Probability density plots are shown for all control patients of the development dataset and the three external centers. Ideally all distributions should overlap since this implies that control patient populations are most likely similar in the development dataset to the external datasets. When comparing the distribution of the CoLab variables for all control-patients across different external validation datasets, albumin and LD show the largest deviations.

Supplemental material 4

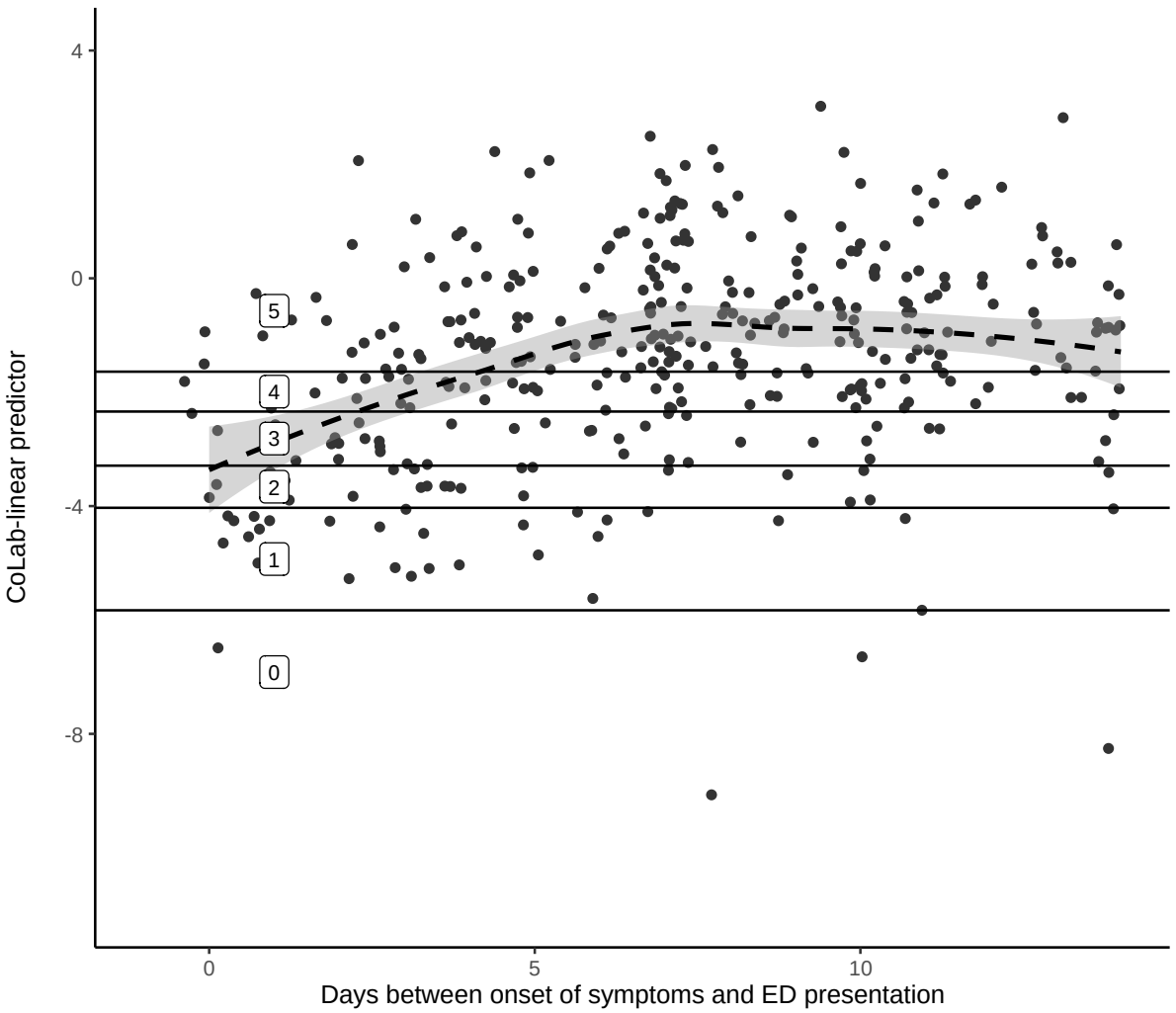


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predict is plotted against the duration of COVID-related symptoms as registered in the electronic patient records. Patients with unknown duration are not plotted. Patients without symptoms were plotted at 0 days. The solid horizontal lines represent the CoLab-score thresholds, the dashed line is a LOESS regression curve with 95% CI. As the duration of symptoms is an integer, some random jitter was added to the days, for visualization purposes. Note that only the first 14 days are shown in this graph.

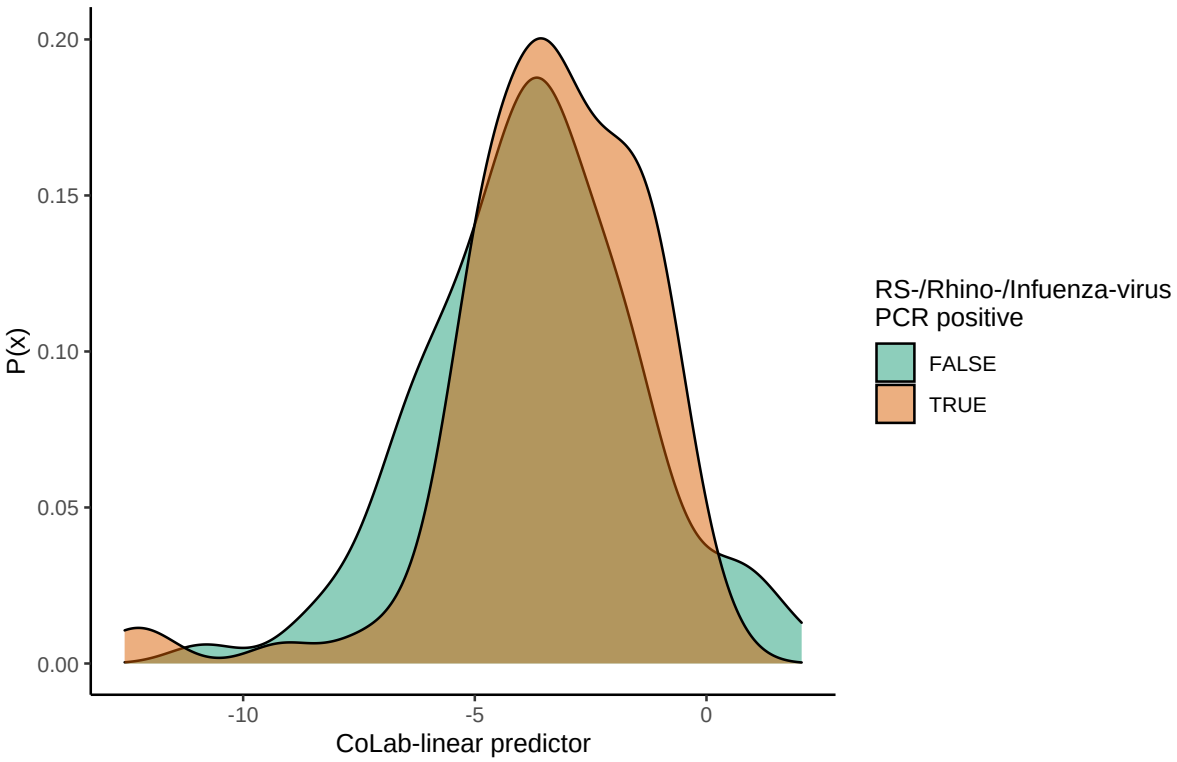


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

For 183 ED presentations that were PCR tested for either RS-, Rhino- and Influenza-virus the CoLab-score was calculated. 91 presentations were PCR positive, 92 were PCR negative. The CoLab-score is only marginally elevated for PCR positive patients, the area under the ROC-curve in separating both groups is 0.573 (95% CI: 4896-0.6563).

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3, 4
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6, 7
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	8, 11-12
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	8
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	8
	5b	D;V	Describe eligibility criteria for participants.	8, 9, S1
	5c	D;V	Give details of treatments received, if relevant.	N/A
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	9
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8, 9
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	N/A
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	10
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	10-12, S1
	10c	V	For validation, describe how the predictions were calculated.	16
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	11-13
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N/A
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	22
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	F1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	T1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	S3
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	F1, F3
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	T2
	15b	D	Explain how to the use the prediction model.	T2, S1
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	T3, T4
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21-23
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	19-20
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	19-20
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	20-21
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	N/A
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	N/A

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document. S = Supplemental material, F = Figure, T = Table.

BMJ Open

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059111.R1
Article Type:	Original research
Date Submitted by the Author:	01-Mar-2022
Complete List of Authors:	Boer, Arjen-Kars; Catharina Hospital Deneer, Ruben; Catharina Hospital, ; Eindhoven University of Technology, Maas, Maaike; Catharina Ziekenhuis, emergency department Ammerlaan, Heidi; Catharina Hospital van Balkom, Roland; Catharina Hospital Thijssen, Wendy; Catharina ziekenhuis, Emergency medicine Bennenbroek, Sophie; Catharina Hospital Leers, Mathie; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Martens, Remy; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Buijs, Madelon; Atalmedial Kerremans, Jos; Alrijne Hospital Location Leiderdorp Messchaert, Muriël ; Gelre Hospitals van Suijlen, Jeroen; Gelre Hospitals van Riel, Natal; Eindhoven University of Technology, ; Amsterdam University Medical Centres Scharnhorst, Volkher; Catharina Hospital; Eindhoven University of Technology,
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Health informatics, Infectious diseases
Keywords:	COVID-19, STATISTICS & RESEARCH METHODS, ACCIDENT & EMERGENCY MEDICINE, Clinical chemistry < PATHOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Arjen-Kars Boer[1*], Ruben Deneer[1,2*†], Maaike Maas[3], Heidi S. M. Ammerlaan[4], Roland H. H. van Balkom[5], Wendy A.M.H. Thijssen[3], Sophie Bennenbroek[3] , Mathie P. G. Leers[6], Remy J.H. Martens[6], Madelon M. Buijs[7], Jos J. Kerremans[8], Muriël Messchaert[9], Jeroen D.E. van Suijlen[9], Natal A.W. van Riel[2,10] and Volkher Scharnhorst[1,2]

* Both authors contributed equally

† Corresponding author

1 Department of Laboratory Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene Loper 3, 5612 AE, Eindhoven, the Netherlands

3 Department of Emergency Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

4 Department of Internal Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands

7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the Netherlands

8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital, Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands

9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334 DZ Apeldoorn, the Netherlands

10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.

Keywords

COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction model, routine laboratory tests

Corresponding author

Ruben Deneer

Michelangelolaan 2

5623 EJ Eindhoven, the Netherlands

ruben.deneer@catharinaziekenhuis.nl

Abstract

Objectives: Identifying patients with a possible SARS-CoV-2 infection in the emergency department (ED) is challenging. Symptoms differ, incidence rates vary and test capacity may be limited. As PCR testing all ED patients is neither feasible nor effective in most centers, a rapid, objective, low-cost early warning score to triage ED patients for a possible infection is developed.

Design: Case-control study.

Setting: Secondary and tertiary hospitals in the Netherlands.

Participants: Patients presenting at the ED with venous blood sampling from July 2019 to July 2020 (N = 10417, 279 SARS-CoV-2 positive). The temporal validation cohort covered the period from July 2020 to October 2021 (N = 14080, 1093 SARS-CoV-2 positive). The external validation cohort consisted of patients presenting at the ED of three hospitals in the Netherlands (N = 12061, 652 SARS-CoV-2 positive).

Primary outcome measures The primary outcome was one or more positive SARS-CoV-2 PCR-test results, within one day prior to, or one week after, ED presentation.

Results: The resulting "CoLab-score" consists of 10 routine laboratory measurements, and age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945). The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991, specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were confirmed in temporal and external validation.

Conclusions: The CoLab-score is based on routine laboratory measurements and is available within one hour after presentation. Depending on the prevalence, COVID-19 may be safely

ruled-out in over one third of ED presentations. Highly suspect cases can be identified regardless of presenting symptoms. The CoLab-score is continuous, in contrast to the binary outcome of lateral flow testing, and can guide PCR testing and triage ED patients.

Article summary

Strengths and limitations of this study

- A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- Using adaptive lasso regression analysis, the panel of 28 laboratory tests was reduced to a single score consisting of a subset of 10 routine ED laboratory tests and age.
- The score was temporally validated from July 2020 to October 2021, in the presence of vaccine roll-out and emergence of new SARS-CoV-2 variants.
- The score was externally validated in 3 other centers in the Netherlands.
- Missingness in the panel of laboratory tests varied between external centers, limiting generalizability of the score to the ED population for which the complete panel of laboratory tests was available.

Introduction

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has evolved into a global pandemic in 2020 [1]. For emergency department (ED) physicians, identifying presenting patients with a possible COVID-19 infection remains challenging since symptoms like fever, shortness of breath or coughing overlap with other illnesses [2,3]. It is crucial however, to identify a possible COVID-19 infection as early as possible. Early identification prevents further spreading and protects hospital staff by isolating a suspected patient, pending the results of a SARS-COV-2 RNA PCR test and/or chest CT. Conversely, when PCR testing or isolation treatment capacity is limited, ruling-out COVID-19 as soon as possible can save valuable resources.

In the era of electronic health records and clinical prediction models, developing an early warning score that can assist ED physicians in identifying patients presenting at the ED with COVID-19 is of great value. Moreover, if only routine ED test results are required as input, the score can be easily adopted by EDs worldwide, potentially reduce diagnostic costs and accelerate patient triage.

Many COVID-19 prediction models have already been developed, the living systematic review by Wynants et. al [4] provides an extensive overview and critical appraisal. Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early models were based on relatively small sample sizes, hampered by selection bias or were over-fitted by selecting too many features [4–6]. Aside from methodological shortcomings of early models, most models are not developed as an early warning score for all ED patients. Firstly, they require features from tests that are not routinely performed or logged for all ED patients (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED patient population. Secondly, the population on which models are commonly based, are PCR-

tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by physicians.

Only two studies were identified that focus on patients presenting at the ED, include unsuspected (and pre-pandemic) patients as controls, and rely solely on routine (laboratory) tests [9,10].

In this study we report the development and validation of an early warning score that, based on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in patients who undergo routine laboratory testing at presentation. The score can assist ED physicians in triaging patients and prevent further transmission of COVID-19 by quickly identifying possibly infected patients or ruling out a possible infection when resources are scarce.

Methods

Study design

This is a retrospective case-control study where routine laboratory test results, combined with age and gender, from all patient presenting at the emergency department (ED) of the Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-2 PCR test results in a development dataset. A model that could predict the presence of a COVID-19 infection was fit to this dataset. Performance of the model was assessed by i) internal validation, ii) temporal validation and iii) external validation by using data from the ED of three other centers. The study was reviewed by the Medical research Ethics Committees United (MEC-U) under study number W20.071, which confirmed that the Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this study. The study was thereafter reviewed and approved by the internal hospital review board.

Patient and Public Involvement

Patients were not involved in the design, conduct or reporting of this study.

Development dataset

All ED presentations at the Catharina Hospital Eindhoven from July 2019 to July 2020 were included in the development dataset, provided that routine laboratory testing had been requested by the attending ED physician. The rationale for this inclusion period is to limit the effect of seasonal variation in the ED patient population by including the summer, fall and winter season of 2019 (control patients) and the winter, spring and summer season of 2020 (case and control patients). The routine laboratory panel at the ED consists of 28 laboratory tests. In some cases not all tests in the routine panel were requested or one or more

quantitative results were not available due to analytical interference (hemolysis, lipemia or icterus). The routine ED laboratory panel is requested for (adult) patients presenting with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints, or for patients (including non-adult patients) presenting with specific complaints where a suspected diagnosis has to be ruled-in or ruled-out. Presentations with one or more missing values in any of the 28 laboratory test in the routine ED panel, were excluded. Presentations with one or more extreme lab results, > 10 times standard deviation from the median, were also excluded to minimize the effect on the estimation of regression coefficients. The median was chosen as a measure of central tendency due to its resistance for outliers. After the first case of COVID-19 in the Netherlands, all patients with symptoms of COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were approved by the Dutch national institute of public health (RIVM). If a patient had a positive PCR result in the past, subsequent presentations were excluded as re-presentations might be clinically different from de novo presentations.

The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were performed in the time window and the presentation occurred after the first case of COVID-19 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the first case were labelled as "*Pre-COVID-19*".

Laboratory tests

The routine laboratory panel consisted of hemocytometric and chemical analyses. The hemocytometric tests were performed on Sysmex XN-10 instruments (Sysmex Corp., Kobe, Japan) and consisted of hemoglobin, hematocrit, erythrocytes, mean corpuscular volume (MCV), mean cellular hemoglobin (MCH), mean cellular hemoglobin concentration (MCHC), thrombocytes, leukocytes, neutrophils, eosinophils, basophils, lymphocytes and monocytes. The chemical analyses were performed on a Cobas 8000 Pro (Roche Dx, Basel, Switzerland) instrument and consisted of glucose, total bilirubin, aspartate aminotransferase (ASAT), alanine aminotransferase (ALAT), lactate dehydrogenase (LD), creatine kinase (CK), alkaline phosphatase (ALP), gamma-glutamyltransferase (gGT), blood urea nitrogen (BUN), creatinine, CKD-epi estimated glomerular filtration rate (eGFR), potassium, sodium, chloride, albumin (bromocresol green) and C-reactive protein (CRP). These results were combined with age and gender.

Modelling

All data were processed and analyzed in R version 4.1.1 [11]. Laboratory results, combined with age and gender were used as covariates in a regression model. Cases were defined as ED presentations labelled as "*PCR-positive*", controls were all other presentations (i.e. "*PCR-negative*", "*Untested*" or "*Pre-COVID-19*"). To achieve predictive accuracy, limit overfitting and perform feature selection, penalized logistic regression with an adaptive lasso penalty was chosen [12,13]. To minimize missing data, all non-numeric results at the extremes of the measuring range, were converted to numeric results by removing the "<" and ">" signs. For eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m2 and <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD, CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model

predictions, these variables were transformed logarithmically. More details regarding model fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the glmnet-package [14].

CoLab-score

Since this is a retrospective case-control study, the sample prevalence may not reflect the true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term in the model should be adjusted according to the current prevalence (details can be found in the document, **Supplemental Material 1**) [15]. However, adjusting the intercept term is not straightforward to implement in clinical practice, therefore the linear predictor of the model was categorized into a score, this score is hereafter referred to as the "CoLab-score". The categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using the intercept adjustment formula by King [15]. The intervals obtained through these breaks correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and score 5 reflects high-risk. More details regarding the rationale of the CoLab-score categorization can be found in the document, **Supplemental Material 1**.

Internal validation

To assess model performance while taking overfitting into account, bootstrapping was performed. 1000 bootstrap samples were generated from the original data. On each bootstrap sample, the full model fitting procedure and CoLab-score conversion were performed. Optimism adjusted performance measures of the CoLab-score were obtained by applying the 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [16]. Performance

measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate performance measures [17]. Although the full inclusion period from July 2019 to July 2020 was used for model fitting, the performance was evaluated on the period starting from the first COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain performance measures that would reflect real world performance.

Temporal validation

For temporal validation, results from our center were prospectively analyzed from July 2020 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period there was also more widespread external PCR testing by municipal health services. The results of external conducted PCR tests were not available to our study. To overcome this limitation, the outcome in the temporal validation cohort was chosen as a composite of the hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test result. This period also covers both the emergence of new SARS-CoV-2 variants as well as vaccine rollout. However, neither vaccination status nor genomic sequencing was available to determine whether a patient was vaccinated or which variant caused the infection. Therefore, data from the Dutch national institute of public health (RIVM) was used, to divide the temporal validation period into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold

for the entire period, as well as for each phase separately to determine a possible effect of vaccination and new variants on performance (results in the **Supplemental Material 2**). Model calibration was assessed graphically using the rms-package [18].

External validation

For the external validation, several centers in the Netherlands were approached and assessed if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available. Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and Zuyderland Medical Center (center 3). The hematological parameters were measured with Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external validation was similar to the temporal validation and consisted of assessing the AUC sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was assessed graphically analogous to the temporal validation dataset.

Results

Development dataset

12879 emergency department (ED) presentations of 10327 patients from July 2019 to July 2020 were included. After excluding cases with an incomplete lab panel, patient presentations that occurred after a positive PCR test in the past (re-presentations) and presentations with extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of 8610 patients remained (**Figure 1 A**).

	Pre-COVID N = 5890	**Untested N = 3303**	**PCR negative N = 945**	**PCR positive N = 279**
Age in years	61 (21)	60 (21)	66 (18)	69 (15)
Female gender	2909 (49.4 %)	1659 (50.2 %)	466 (49.3 %)	95 (34.1 %)
Specialism				
Internal medicine	1648 (28.0 %)	896 (27.1 %)	244 (25.8 %)	71 (25.4 %)
Surgery	1007 (17.1 %)	679 (20.6 %)	51 (5.4 %)	5 (1.8 %)
Neurology	775 (13.2 %)	468 (14.2 %)	64 (6.8 %)	5 (1.8 %)
Pulmonary medicine	714 (12.1 %)	220 (6.7 %)	326 (34.5 %)	167 (59.9 %)
Cardiology	560 (9.5 %)	322 (9.7 %)	145 (15.3 %)	6 (2.2 %)
Urology	309 (5.2 %)	148 (4.5 %)	15 (1.6 %)	7 (2.5 %)
Gastroenterology	306 (5.2 %)	224 (6.8 %)	27 (2.9 %)	1 (0.4 %)
Geriatrics	189 (3.2 %)	95 (2.9 %)	52 (5.5 %)	15 (5.4 %)
Orthopedics	147 (2.5 %)	109 (3.3 %)	11 (1.2 %)	0 (0.0 %)
Gynecology	118 (2.0 %)	82 (2.5 %)	2 (0.2 %)	0 (0.0 %)
Other	117 (2.0 %)	60 (1.8 %)	8 (0.8 %)	2 (0.7 %)
Hemoglobin in mmol/L	8.2 (1.3)	8.3 (1.3)	8.2 (1.4)	8.6 (1.1)
Hematocrit in L/L	0.403 (0.059)	0.405 (0.056)	0.405 (0.062)	0.417 (0.047)
Erythrocytes in /pL	4.41 (0.69)	4.43 (0.66)	4.41 (0.72)	4.61 (0.60)
MCV in fl	91.8 (6.4)	91.9 (6.1)	92.4 (6.7)	90.7 (5.5)
MCH in mmol	1.859 (0.157)	1.876 (0.150)	1.874 (0.172)	1.869 (0.141)
MCHC in mmol/L	20.2 (0.9)	20.4 (0.9)	20.3 (1.0)	20.6 (0.8)
Thrombocytes in /nL	263 (99)	266 (100)	269 (105)	217 (123)
Leukocytes in /nL	9.30 [7.06, 12.16]	8.92 [7.01, 11.89]	9.66 [7.17, 12.94]	6.33 [4.74, 8.48]
Neutrophils in /nL	6.62 [4.51, 9.53]	6.10 [4.42, 8.94]	7.01 [4.79, 10.02]	4.71 [3.30, 6.94]
Eosinophils in /nL	0.09 [0.03, 0.17]	0.09 [0.03, 0.18]	0.08 [0.02, 0.17]	0.00 [0.00, 0.02]
Basophils in /nL	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.01 [0.01, 0.02]
Lymphocytes in /nL	1.47 [0.93, 2.13]	1.56 [1.05, 2.18]	1.31 [0.80, 2.03]	0.86 [0.59, 1.21]
Monocytes in /nL	0.70 [0.52, 0.93]	0.69 [0.52, 0.91]	0.74 [0.54, 1.01]	0.45 [0.32, 0.64]
Glucose in mmol/L	6.76 [5.83, 8.39]	6.68 [5.76, 8.14]	6.98 [5.95, 8.85]	6.77 [5.98, 8.48]
Bilirubin in umol/L	7.5 [5.0, 11.6]	7.4 [5.1, 10.9]	8.3 [5.6, 12.4]	8.2 [6.3, 11.4]
ASAT in U/L	24.0 [19.1, 32.2]	26.5 [21.6, 35.1]	27.7 [21.7, 39.2]	40.7 [30.2, 57.2]
ALAT in U/L	24.3 [17.8, 35.3]	25.3 [18.4, 36.2]	25.7 [18.4, 40.0]	33.7 [23.3, 50.0]
LD in U/L	201 [173, 240]	198 [170, 236]	215 [178, 263]	300 [238, 403]
CK in U/L	82 [51, 134]	83 [52, 136]	76 [51, 125]	124 [62, 222]
ALP in IU/L	83.0 [68.0, 105.0]	81.0 [65.8, 102.5]	86.9 [67.9, 110.0]	71.0 [58.8, 85.0]
gGT in U/L	27.0 [17.0, 53.0]	28.4 [18.4, 50.5]	37.0 [22.4, 68.9]	42.0 [28.0, 83.5]
BUN in mmol/L	5.7 [4.3, 8.0]	5.8 [4.3, 7.8]	6.2 [4.6, 9.4]	6.1 [4.7, 8.9]

CKD-epi in ml/min/m2	80.9 [58.0, 99.1]	85.0 [63.5, 103.3]	79.1 [52.1, 96.6]	76.6 [54.9, 91.2]
Potassium in mmol/L	4.06 (0.50)	4.03 (0.49)	4.07 (0.55)	3.91 (0.47)
Sodium in mmol/L	139.2 (4.0)	138.5 (3.9)	138.0 (4.3)	136.4 (4.1)
Chloride in mmol/L	104.4 (4.6)	103.8 (4.5)	102.9 (4.8)	101.6 (4.4)
Albumin in g/L	42.4 (4.9)	42.3 (4.5)	40.8 (4.8)	38.4 (3.8)
CRP in mg/L	8 [2, 41]	5 [1, 30]	18 [3, 69]	77 [37, 136]

Table 1: Descriptive statistics of development dataset and laboratory concentrations.

Shown are the laboratory tests routinely requested at ED presentation and their mean/median results (in the development dataset) for the presentations before the first COVID-19 patient in the Netherlands ("Pre-COVID-19"), presentations thereafter that were not tested for COVID-19 ("Untested"), tested negatively ("PCR negative") and tested positive ("PCR positive"). For results with normal distributions, the mean value and standard deviation (in round brackets) are shown. For results that have skewed or heavy tailed distributions, the median value and the interquartile range is shown [in squared brackets]. Dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error).

Descriptive statistics of ED presentations are shown in **Table 1,** dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error [19]). For the PCR positives (N = 279), 91% (95% CI: 88 to 94%) of the cases were tested positive in their first PCR. The remaining 24 patients were positive in their second (N = 18), third (N = 5) or fourth (N = 1) PCR.

CoLab-score

The model obtained through adaptive lasso regression contained eleven variables, which are depicted with their regression coefficients (weights) in **Table 2**.

Variable	β	Exclusion limit	Relative importance
Intercept	-6.885		-
Erythrocytes /pL	0.9379	Erythrocytes < 2.9 /pL	52 %
Leukocytes /nL	-0.1298		46 %
Eosinophils /nL	-6.834		86 %
Basophils /nL	-47.70	Basophils >0.33 /nL	100 %
$\log_{10}$ of Bilirubin in µmol/L	-1.142	Bilirubin >169 µmol/L	26 %
$\log_{10}$ of LD in U/L	5.369	LD >1564 U/L	58 %
$\log_{10}$ of ALP in IU/L	-3.114	AF >1000 IU/L	45 %
$\log_{10}$ of gGT in U/L	0.3605	gGT >1611 U/L	11 %
Albumin in g/L	-0.1156		45 %
CRP in mg/L	0.002560		15 %
Age in years	0.002275		4 %

Table 2: Calculation of the CoLab-linear predictor (LP).

The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of the 11 variables with their corresponding coefficients (β's). CoLab-LP = – 6.885 + [erythrocytes] × 0.9379 – [leukocytes] × 0.1298 – [eosinophils] × 6.834 – [basophils] × 47.7 – log10([bilirubin]) × 1.142 + log10([LD]) × 5.369 – log10([ALP]) × 3.114 + log10([gGT]) × 0.3605 – [albumin] × 0.1156 + [CRP] × 0.02560 + [age] × 0.002275. The LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid if any of the variables exceed the limits in the third column. The relative importance ranks the importance of variables in predicting the outcome, relative to the most important variable (in this case basophils).

A larger β-coefficient does not imply that a variable is more important in predicting the odds of testing positive for SARS-CoV-2, since variables are on different scales. The most important variables are basophiles, eosinophils and lactate dehydrogenase (LD).

As shown in **Figure 2**, the linear predictor clearly discriminates between COVID-19 and non-COVID-19. The linear predictor is converted to CoLab-scores 0 – 5 with the cut-points depicted in **Figure 2**.

Internal validation

The model was validated in the period starting from the first COVID-19 infection to July 2020, in this period the mean prevalence was 7.2%. The AUC of the CoLab-score is 0.930 (95% CI: 0.909 to 0.945).

CoLab-score	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN	% of population
0	0.984 (0.969 - 0.991)	0.410 (0.303 - 0.542)	0.115 (0.094 - 0.147)	0.997 (0.993 - 0.999)	133 (165 - 195)	485 (876 - 1360)	799 (1280 - 1660)	0 (2 - 6)	28 (38 - 51)
≤ 1	0.912 (0.892 - 0.952)	0.785 (0.741 - 0.827)	0.248 (0.208 - 0.300)	0.991 (0.989 - 0.995)	126 (152 - 185)	1520 (1690 - 1850)	314 (464 - 627)	4 (15 - 21)	69 (73 - 77)
≤ 2	0.856 (0.816 - 0.894)	0.880 (0.864 - 0.900)	0.357 (0.316 - 0.415)	0.988 (0.984 - 0.991)	114 (143 - 173)	1800 (1900 - 2010)	187 (259 - 317)	12 (24 - 36)	81 (83 - 84)
≤ 3	0.757 (0.706 - 0.809)	0.951 (0.945 - 0.959)	0.546 (0.496 - 0.603)	0.981 (0.976 - 0.985)	99 (127 - 157)	1960 (2050 - 2150)	77 (105 - 130)	24 (40 - 57)	89 (90 - 91)
≤ 4	0.612 (0.530 - 0.706)	0.978 (0.972 - 0.983)	0.683 (0.628 - 0.746)	0.970 (0.963 - 0.978)	74 (103 - 137)	2010 (2110 - 2210)	29 (48 - 69)	35 (64 - 90)	92 (94 - 94)

Table 3: Diagnostic performance CoLab-score in the development dataset.

The development dataset was internally validation for the period March 2020 – July 2020 (N = 4.527). Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) and fraction of presentations (%) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4). The numbers in round brackets represent the 95% optimism adjusted bootstrapped confidence intervals. The first column defines the threshold above which CoLab-score a patient is

considered positive. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

Diagnostic performance is shown in **Table 3.** A CoLab-score of 0 has a negative predictive value (NPV) of 0.997 (95% CI: 0.994 to 0.999) and positive predictive value (PPV) of 0.115 (0.0932 - 0.141), one third (38.4%, 95% CI: 26.4 to 48.4%) of all ED presentations were assigned this score and can therefore be safely excluded. Conversely, 6.2% (95% CI: 6.3 to 7.2%) of the ED patients had a CoLab-score = 5. Given the PPV of this score (0.682, 95% CI: 0.622 to 0.740, NPV: 0.970, 95% CI: 0.962 - 0.977), subsequent PCR testing is advised.

Temporal validation

As the CoLab-score was developed in our center after the first COVID-19-wave in the Netherlands, the performance was evaluated in our center from July 2020 until October 2021. Lab results from 17489 ED presentations were collected. After applying the inclusion flow as shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a COVID-19 infection.

The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the development cohort, although sensitivity is slightly lower and specificity slightly higher (cf. **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental Material 2 Figure 1)**. The discriminative ability was not lower in the second or third phase, compared to the first phase. Diagnostic performance is preserved in terms of sensitivity and specificity, except a moderately reduced sensitivity of scores ≥ 3 in the third phase as

compared to the first phase. PPV and NPV are incomparable due to different prevalence/pre-test probabilities in each phase (see **Supplemental Material 2 Table 1**).

In terms of the predicted probabilities, model calibration shows that overall predicted probabilities are too low (see **Supplemental Material 3** for the calibration plot), which is expected since the prevalence differs and the intercept has to be adjusted to the prevalence. In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that initially did not present with COVID-specific symptoms. Most patients had neurological or orthopedic presenting symptoms.

External validation

For external validation, data obtained from three other centers were used, center 1 (N = 1284, 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545, 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers is ≥ 0.96. Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for external centers are shown in **Supplemental Material 3**, the observed fraction of COVID-19 positives is slightly lower than expected in centers 1 and 2. For center 3, low probabilities appear slightly underestimated and high probabilities slightly overestimated.

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
0	Temporal	0.967 (0.956 - 0.978)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)	1005 (993 - 1016)	5476 (5366 - 5587)	7565 (7454 - 7675)	34 (23 - 46)
	Center 1	1.000 (1.000 - 1.000)	0.331 (0.307 - 0.358)	0.059 (0.057 - 0.061)	1.000 (1.000 - 1.000)	52 (52 - 52)	410 (380 - 443)	827 (794 - 857)	0 (0 - 0)
	Center 2	0.961 (0.922 - 0.990)	0.351 (0.333 - 0.369)	0.052 (0.049 - 0.054)	0.996 (0.992 - 0.999)	99 (95 - 102)	985 (935 - 1035)	1823 (1773 - 1873)	4 (1 - 8)
	Center 3	0.970 (0.950 - 0.988)	0.322 (0.306 - 0.338)	0.130 (0.126 - 0.133)	0.991 (0.984 - 0.996)	327 (320 - 333)	1042 (991 - 1092)	2193 (2143 - 2244)	10 (4 - 17)
≤ 1	Temporal	0.888 (0.870 - 0.908)	0.791 (0.783 - 0.798)	0.253 (0.245 - 0.261)	0.989 (0.987 - 0.991)	923 (904 - 943)	10311 (10215 - 10401)	2730 (2640 - 2826)	116 (96 - 135)
	Center 1	0.923 (0.846 - 0.981)	0.694 (0.669 - 0.720)	0.113 (0.101 - 0.124)	0.995 (0.991 - 0.999)	48 (44 - 51)	858 (828 - 891)	379 (346 - 409)	4 (1 - 8)
	Center 2	0.913 (0.854 - 0.961)	0.678 (0.661 - 0.696)	0.094 (0.087 - 0.101)	0.995 (0.992 - 0.998)	94 (88 - 99)	1905 (1857 - 1953)	903 (855 - 951)	9 (4 - 15)
	Center 3	0.914 (0.881 - 0.944)	0.674 (0.657 - 0.691)	0.226 (0.216 - 0.236)	0.987 (0.982 - 0.991)	308 (297 - 318)	2180 (2126 - 2234)	1055 (1001 - 1109)	29 (19 - 40)
≤ 2	Temporal	0.820 (0.796 - 0.843)	0.894 (0.889 - 0.899)	0.382 (0.367 - 0.396)	0.984 (0.982 - 0.986)	852 (827 - 876)	11661 (11591 - 11729)	1380 (1312 - 1450)	187 (163 - 212)
	Center 1	0.808 (0.692 - 0.904)	0.811 (0.788 - 0.832)	0.152 (0.129 - 0.176)	0.990 (0.984 - 0.995)	42 (36 - 47)	1003 (975 - 1029)	234 (208 - 262)	10 (5 - 16)
	Center 2	0.845 (0.777 - 0.913)	0.801 (0.785 - 0.815)	0.135 (0.122 - 0.147)	0.993 (0.990 - 0.996)	87 (80 - 94)	2248 (2205 - 2289)	560 (519 - 603)	16 (9 - 23)
	Center 3	0.890 (0.855 - 0.923)	0.794 (0.779 - 0.808)	0.311 (0.294 - 0.328)	0.986 (0.981 - 0.990)	300 (288 - 311)	2569 (2521 - 2615)	666 (620 - 714)	37 (26 - 49)
≤ 3	Temporal	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.596 (0.573 - 0.618)	0.977 (0.974 - 0.979)	738 (709 - 767)	12540 (12496 - 12582)	501 (459 - 545)	301 (272 - 330)
	Center 1	0.750 (0.635 - 0.865)	0.909 (0.892 - 0.925)	0.257 (0.213 - 0.306)	0.989 (0.983 - 0.994)	39 (33 - 45)	1124 (1104 - 1144)	113 (93 - 133)	13 (7 - 19)
	Center 2	0.660 (0.563 - 0.748)	0.897 (0.885 - 0.908)	0.190 (0.163 - 0.218)	0.986 (0.983 - 0.990)	68 (58 - 77)	2519 (2486 - 2549)	289 (259 - 322)	35 (26 - 45)
	Center 3	0.766 (0.718 - 0.810)	0.887 (0.876 - 0.898)	0.413 (0.386 - 0.442)	0.973 (0.968 - 0.978)	258 (242 - 273)	2869 (2835 - 2905)	366 (330 - 400)	79 (64 - 95)
≤ 4	Temporal	0.585 (0.556 - 0.615)	0.984 (0.982 - 0.987)	0.750 (0.724 - 0.778)	0.968 (0.965 - 0.970)	608 (578 - 639)	12838 (12811 - 12866)	203 (175 - 230)	431 (400 - 461)

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
	Center 1	0.654 (0.519 - 0.788)	0.951 (0.939 - 0.962)	0.359 (0.293 - 0.435)	0.985 (0.979 - 0.991)	34 (27 - 41)	1176 (1161 - 1190)	61 (47 - 76)	18 (11 - 25)
	Center 2	0.534 (0.437 - 0.621)	0.952 (0.943 - 0.959)	0.287 (0.239 - 0.339)	0.982 (0.979 - 0.986)	55 (45 - 64)	2672 (2649 - 2693)	136 (115 - 159)	48 (39 - 58)
	Center 3	0.665 (0.611 - 0.718)	0.930 (0.921 - 0.938)	0.497 (0.462 - 0.534)	0.964 (0.958 - 0.969)	224 (206 - 242)	3008 (2980 - 3036)	227 (199 - 255)	113 (95 - 131)

Table 4: Diagnostic performance of the CoLab-score in the validation dataset (temporal) and three external hospitals.

Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

Discussion

Given the impact of COVID-19 on society and healthcare, there is a need for simple and fast detection of patients with a possible COVID-19 infection in the ED. The CoLab-score described in this study, is a fast and accurate risk score to triage patients presenting at the ED based on ten routine blood biomarkers and age.

The main strength of this study is that this score can be used as an early-warning or triaging tool for the ED population presenting with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints where a routine blood panel is requested. This is in contrast to the vast majority of COVID-19 diagnostic models that have been developed on a pre-selected population of PCR-tested patients [9,20–26]. Moreover, the CoLab-score requires only routine blood tests, instead of (features from) imaging such as CT-scans or laboratory tests that are not routinely collected in the ED, e.g. interleukin-6 or 3-hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific) [27,28].

Two other studies have been published which are similar to this study [9,10]. Interestingly, the study by Soltan et al., ranked basophils and eosinophils as the two most important features in predicting the outcome, similar to our results [10]. Eosinophils were also seen as one of the most important features by Plante et al. [9]. However, both studies focus on an artificial intelligence/machine learning approach. While their approach likely results in higher predictive performance due to the ability of machine learning models to capture non-linear and interaction effects, the goal of this study was to develop a simple, fast and robust model that can easily be implemented in current hospital IT systems.

Since this is a retrospective case-control study, there is some unavoidable missing data. In our cohort 17.6% of the ED presentations could not be used due to one or more missing laboratory results. This is lower or equal to similar studies; 22% [23], 17% [21] and 11% [26]. Important to note is that 7.7% of missingness is due to analytical errors which can be assumed to be missing completely at random. For the remaining 9.9% of missingness, the full lab panel was most frequently missing for pediatric, obstetric and surgery patients. These patients are presenting with specific complaints for which specific laboratory tests are requested, and hence do not match the inclusion criteria for a routine blood panel. Overall the missingness was significantly lower in the PCR-tested group versus the untested group (χ2-test p-value <0.001).

In the external centers, there is a high level of missingness as a result of an incomplete laboratory panel. In the case of centers 1 and 2, only internal medicine ED presentations were tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine ED presentations, which is reflected by the few PCR-positive patients excluded. Due to these high levels of missingness, the results of the external centers cannot be used to show that the CoLab-score generalizes to the entire ED population. Rather, the results show that for the majority of COVID-19 positive patients presenting at the ED, a routine laboratory panel is available from which the CoLab-score can be calculated, and that the performance of the CoLab-score in this population is comparable to the development population.

The performance of the CoLab-score is affected by the time between the onset of symptoms and ED presentations. The score increases with the duration of symptoms and gradually decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some

COVID-19 patients with early or late presentation after onset of symptoms can be missed. Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and <10 days prior to ED presentation. Chemotherapy that causes myeloid suppression, will decrease neutrophilic, basophilic and eosinophilic counts and thereby "falsely" increasing the CoLab-score. Conversely, COVID-19 patients with severe anemia could have "falsely" lowered CoLab-scores. To minimize false negatives, we have therefore advised to report CoLab-scores only when the concentration of erythrocytes is ≥ 2.9 /pL.

It was chosen to exclude re-presentations after a previous presentation with COVID-19. Since the median time between initial presentation and re-presentation was 12 days, these patients were most likely not re-infected patients, but patients who deteriorated after initial presentation/treatment. Given that the CoLab-score follows the host-immune response, the score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients would impact the performance of the CoLab-score as patients in a later phase of the disease show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate after treatment for COVID-19. Other re-presentations were not excluded, which results in some patients appearing multiple times in a dataset. This was not corrected for in the regression model since the assumption was made that ED presentations are independent observations. The median time between re-presentations is 38 days, most likely resulting in variations in laboratory results between presentations, and hence, little to no correlation between presentations. A sensitivity analysis was performed whereby only the first presentation was included for each patient (**Supplemental Material 4 Table 1**), but no difference was found in performance in terms of sensitivity, specificity and AUC.

The CoLab-score does not serve as a replacement for PCR-testing or LFTs, and can be used to guide PCR-testing when routine blood tests are available. Note the performance of the

CoLab-score in a suspected/PCR-tested cohort is not equal to the (see **Supplemental Material 4 Table 1)**.

Finally, the CoLab-score could lead to false positives by other viral infections. However, in an historic patient cohort, the CoLab-score had only limited discriminative ability in separating influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4 Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-response to the virus, it is expected that the CoLab-score is also sensitive to future SARS-CoV-2 variants. This is supported by the fact that the diagnostic performance is sustained in periods with different dominant variants. Moreover, there is no evidence that the discriminative ability of the CoLab-score is lowered by a change in the ED patient population as a result of widespread vaccination. Although vaccination status is not registered for all presenting patients, in a small subgroup of 12 patients for whom vaccination status was registered, and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5) (see **Supplemental Material 2 Figure 2**),

To conclude, the CoLab-score developed and validated in this study, based on 10 routine laboratory results and age, is available within 1 hour for any patient presenting at the ED. The score can be used by clinicians to guide PCR testing or triage patients and helps to identify COVID-19 in patients presenting at the ED with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints where a routine blood panel is requested. The lowest CoLab-score can be used to effectively rule-out a possible SARS-CoV-2 infection, the highest score to alert physicians to a possible infection. The CoLab-score is therefore a valuable tool to rule out COVID-19, guide PCR testing and is available to any center with access to routine laboratory tests.

Funding statement

This was an investigator-initiated study and no funding was received for this study.

Competing interests

A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR reports no conflict of interest. VS reports no conflict of interest.

Data sharing statement

Datasets with source data for Table 1, Figure 2, Table 3 and Table 4, as well the R-code to fit the model is available from the Dryad repository, DOI:[WILL BE PROVIDED WHEN UNDER REVIEW]. Technical appendix can be found in **Supplemental Material 1**.

Author contributorship statement

Arjen-Kars Boer: Conceptualization (Lead), Data curation (Lead), Funding acquisition (Lead), Investigation (Equal), Methodology (Equal), Supervision (Equal), Writing-original draft (Equal), Writing-review & editing (Equal).

Ruben Deneer: Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Lead), Software (Lead), Visualization (Lead), Writing-original draft (Equal), Writing-review & editing (Equal).

Maaike Maas: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Equal), Writing-review & editing (Equal).

Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal), Writing-review & editing (Equal).

Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing (Equal).

References

1 Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 4 Feb 2021).

2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *https://doi.org/101056/NEJMoa2002032* 2020;**382**:1708–20. doi:10.1056/NEJMOA2002032

3 Vetter P, Vu DL, L'Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**. doi:10.1136/BMJ.M1470

4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18. doi:10.1136/bmj.m1328

5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review. J. Med. Syst. 2020;**44**:122. doi:10.1007/s10916-020-01582-x

6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417

7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19—Definition and Evaluation. *https://doi.org/101148/radiol2020201473* 2020;**296**:E97–104. doi:10.1148/RADIOL.2020201473

8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort

study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840

9 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: A large, multicenter, real-world study. *J Med Internet Res* 2020;**22**:e24048. doi:10.2196/24048

10 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87. doi:10.1016/S2589-7500(20)30274-0

11 R Core Team. R: A Language and Environment for Statistical Computing. 2020.https://www.r-project.org/

12 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29. doi:10.1198/016214506000000735

13 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

14 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01

15 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63. doi:10.1093/oxfordjournals.pan.a004868

16 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973

17 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-

2105-12-77

18 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.https://cran.r-project.org/package=rms

19 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons and progress. Scand. J. Clin. Lab. Invest. 1999;**59**:491–500. doi:10.1080/00365519950185229

20 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst* 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4

21 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502

22 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807. doi:10.1007/s00330-020-07022-1

23 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. *Clin Chem Lab Med* 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593

24 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305

25 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in the emergency department: derivation and validation of the LDH-CRP-Lymphocyte

(LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336

26 Tordjman M, Mekki A, Mali RD, *et al.* Pre-test probability for SARS-Cov-2-related infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342. doi:10.1371/journal.pone.0243342

27 Peto T, Affron D, Afrough B, *et al.* COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine* 2021;**36**:100924. doi:10.1016/J.ECLINM.2021.100924

28 García-Fiñana M, Hughes DM, Cheyne CP, *et al.* Performance of the Innova SARS-CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot: population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637

Figure legends

Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B) dataset.

All patient admissions with routine venous blood sampling at the emergency department (ED) were included. For the development dataset, completeness of the lab panel was assessed for all 28 laboratory tests , for the temporal validation dataset this was only necessary for 10 laboratory tests. The major causes of missingness are described in the text. In the development dataset, presentations with extreme values (>10 SD) were excluded. The same limits were applied to the temporal validation dataset (see Table 2 for limits).

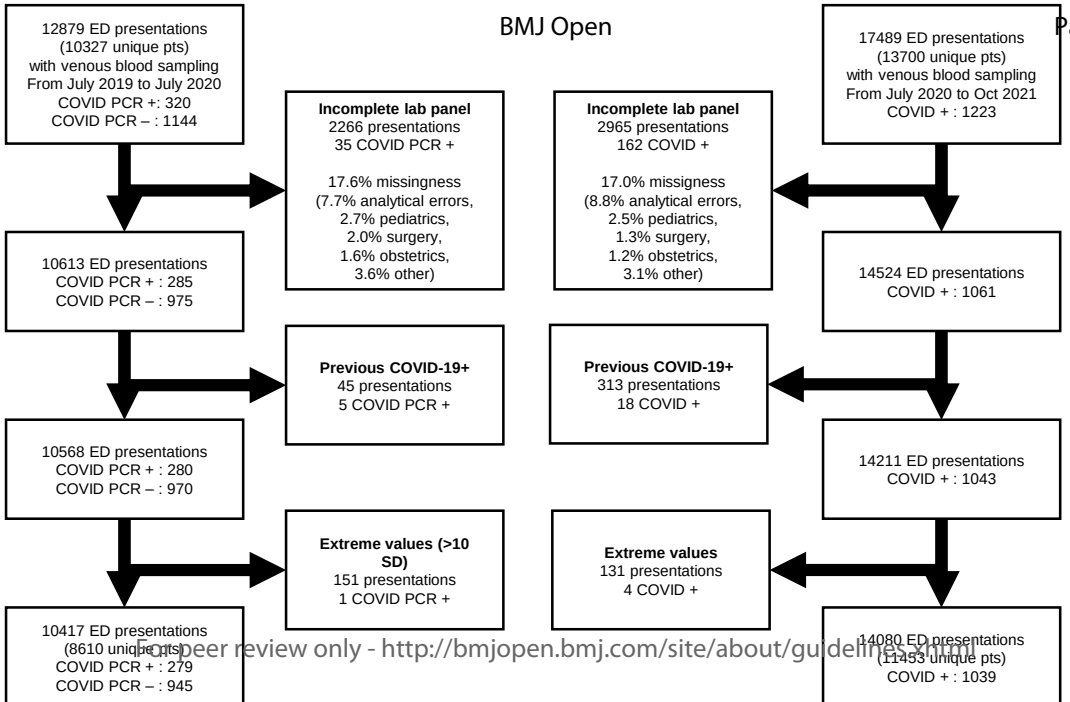
Figure 2: Probability density plot of the CoLab-linear predictor.

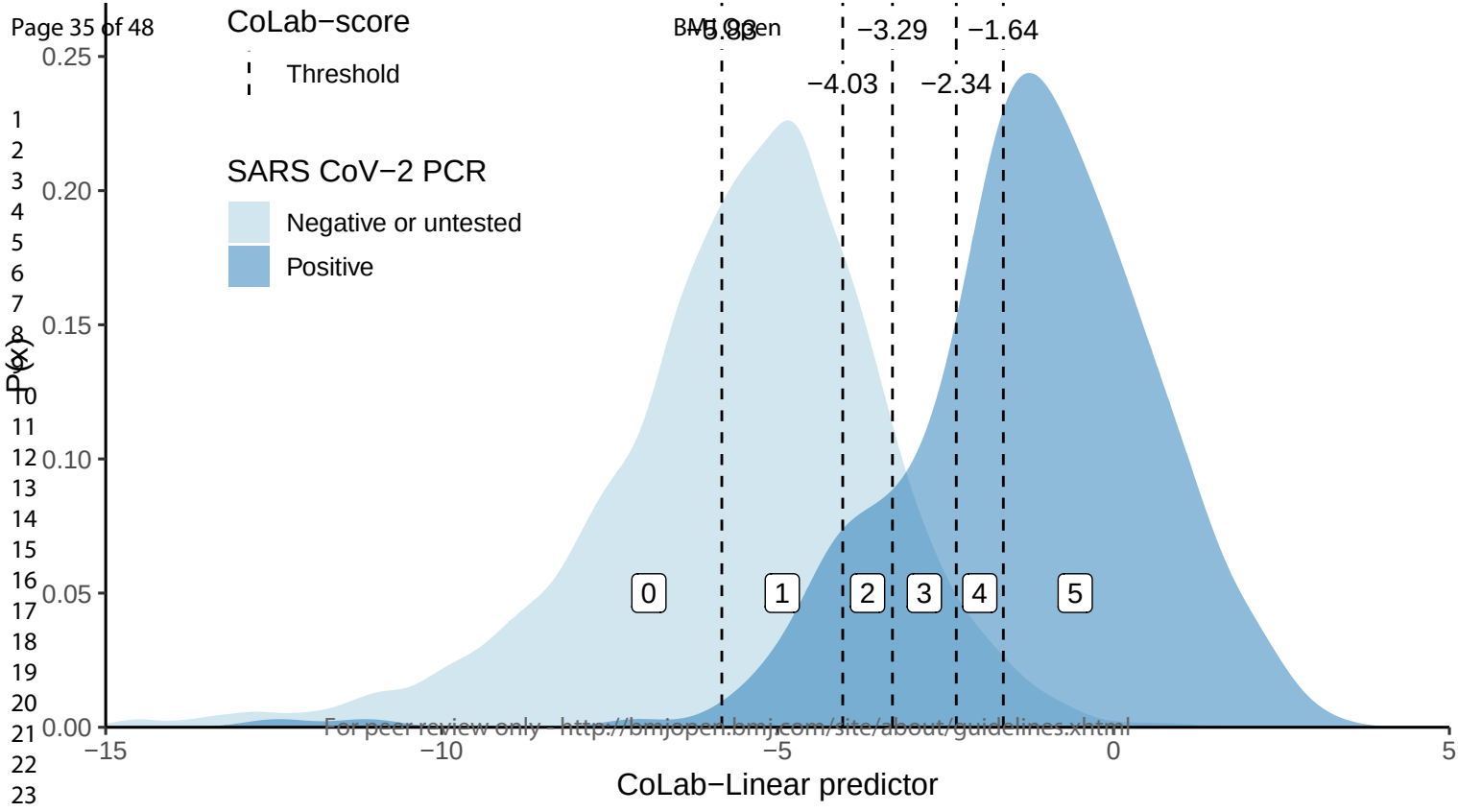
The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are plotted against the linear predictor (see table 2). The CoLab-score cut-offs (–5.83, –4.02, –3.29, –2.34 and –1.64) are depicted with vertical dashed lines. The white-boxed numbers (between the cut-offs) represent the corresponding CoLab-score. Note that while the area under both curves is identical (since these are probability density functions), in absolute numbers the "negative or untested"-group is about 36 times larger than the PCR positive group.

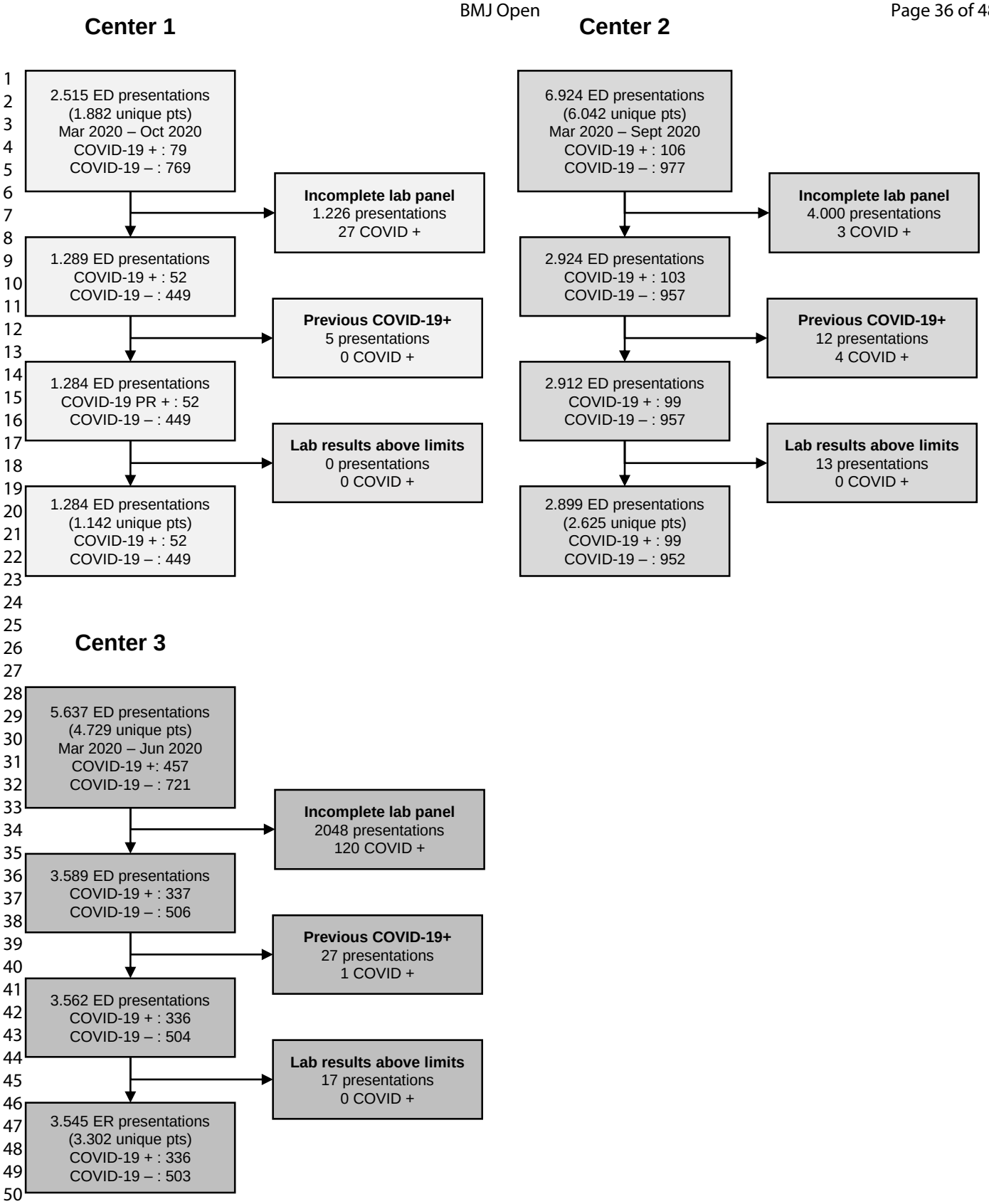
Figure 3: Inclusion flow of ED patients in three external centers.

All emergency department (ED) presentations with routine venous blood sampling were included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see

Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were excluded as "previous COVID-19+". Presentations with any laboratory result above the limits of the CoLab-score (see Table 2) were excluded.






Center 1
2.515 ED presentations
(1.882 unique pts)
Mar 2020 – Oct 2020
COVID-19 + : 79
COVID-19 – : 769
Incomplete lab panel
1.226 presentations
27 COVID +
1.289 ED presentations
COVID-19 + : 52
COVID-19 – : 449
Previous COVID-19+
5 presentations
0 COVID +
1.284 ED presentations
COVID-19 PR + : 52
COVID-19 – : 449
Lab results above limits
0 presentations
0 COVID +
1.284 ED presentations
(1.142 unique pts)
COVID-19 + : 52
COVID-19 – : 449
Center 2
6.924 ED presentations
(6.042 unique pts)
Mar 2020 – Sept 2020
COVID-19 + : 106
COVID-19 – : 977
Incomplete lab panel
4.000 presentations
3 COVID +
2.924 ED presentations
COVID-19 + : 103
COVID-19 – : 957
Previous COVID-19+
12 presentations
4 COVID +
2.912 ED presentations
COVID-19 + : 99
COVID-19 – : 957
Lab results above limits
13 presentations
0 COVID +
2.899 ED presentations
(2.625 unique pts)
COVID-19 + : 99
COVID-19 – : 952
Center 3
5.637 ED presentations
(4.729 unique pts)
Mar 2020 – Jun 2020
COVID-19 +: 457
COVID-19 – : 721
Incomplete lab panel
2048 presentations
120 COVID +
3.589 ED presentations
COVID-19 + : 337
COVID-19 – : 506
Previous COVID-19+
27 presentations
1 COVID +
3.562 ED presentations
COVID-19 + : 336
COVID-19 – : 504
Lab results above limits
17 presentations
0 COVID +
3.545 ER presentations
(3.302 unique pts)
COVID-19 + : 336
COVID-19 – : 503

Supplemental material 1

Model fitting

Prior to model fitting, covariates were scaled to zero mean and unit variance, after model fitting coefficients were unscaled to obtain regression coefficients on the original scale. In adaptive lasso, weights are applied to each of the covariates present in the lasso constraint, the weight vector has to be calculated before the adaptive lasso regression is performed. Due to multicollinearity between laboratory tests in the routine lab panel, weights in the adaptive lasso were based on ridge regression estimates ($\hat{\beta}_{ridge}$) as recommended by Zou. To obtain $\hat{\beta}_{ridge}$ the optimal penalty (λ) for the ridge regression was chosen using 10 fold cross-validation (CV) with area under the ROC curve (AUC) as the loss function. The λ corresponding to the maximum AUC was selected to obtain $\hat{\beta}_{ridge}$. The weight vector ($\widehat{w}$) was calculated by $\widehat{w} = 1/\left|\hat{\beta}_{ridge}\right|^2$. This weight vector was then used to fit an adaptive lasso regression where λ was chosen by the criterion ±1 SE of the maximum AUC.

Model intercept correction

The linear predictor for a patient *i* is calculated as follows: $lp_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$
Where *n* is the number of variables in the final model, x_{in} are the observed predictor variables for subject *i* and $\beta_n$the model coefficients. The linear predictor can then be converted to a probability for patient *i* (P_i) by the logistic function: $P_i = \frac{1}{1+e^{-lp_i}}$
The intercept term β_0 is sensitive to the fraction of cases versus controls in the dataset/population. Since the model is fitted to a case-control dataset where the number cases is fixed (all patients tested positive for COVID-19) and the number of controls is randomly chosen (a 6-month period pre-COVID), the intercept term β_0 is a result of this choice and will likely not be generalizable to the real-world setting. Prior correction is a method to correct the estimate of the intercept based on the true fraction of positives in the population, τ (prevalence of COVID-19 in the ED) and the fraction of cases in the development dataset, $\bar{y}$. The intercept term β_0 can then be corrected to obtain $\beta_{0corrected}$ using the following formula:

$$\beta_{0corrected} = \beta_0 + \beta_{adj}$$

$$\beta_{adj} = -ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right]$$

In our dataset $\bar{y} = 0.02675$ therefore:

$$\beta_{adj} = -ln\left(\frac{1-\tau}{\tau}\right) + 3.594$$

An estimate $\bar{\tau}$ can be used for the prevalence τ to obtain $\bar{\beta}_{adj}$ which can be plugged in the original linear predictor formula to obtain calibrated probabilities:

$$lp_i(\tau) = \beta_0 - ln\left(\frac{1-\tau}{\tau}\right) + 3.594 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$$

CoLab-score

An alternative, which is the basis of the CoLab-score, is to choose a fixed probability P_i above which one considers a patient eligible for further testing. The probability can be expressed as a number needed to test. If one is willing to test 10 patients to find one positive, all patients with $P_i \geq 0.1$ should be considered positive. In this study a number needed to test of 15 is used, therefore all patients with a $P_i \geq 0.067$ should be considered positive. On the linear predictor scale this translates to $\text{logit}(0.067) = -2.639$. To determine the cutoffs for difference prevalence thresholds one solves the following equation:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639$$
$$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$
$$lp_i(\tau) \geq ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

Choosing values for τ yields the cutoffs for the CoLab score:

$lp_i(\tau = 0.4) \geq -5.83$ (CoLab-score = 1)
$lp_i(\tau = 0.1) \geq -4.03$ (CoLab-score = 2)
$lp_i(\tau = 0.05) \geq -3.29$ (CoLab-score = 3)
$lp_i(\tau = 0.02) \geq -2.34$ (CoLab-score = 4)
$lp_i(\tau = 0.01) \geq -1.64$ (CoLab-score = 5)

These thresholds correspond to CoLab-scores 0 to 5. The interpretation of these scores is as follows; if the prevalence is <1%, only CoLab-score 5 should be classified as positive and CoLab-score 0 till 4 as negative. If the prevalence is 1% – 2%, CoLab-score 4 and 5 should be classified as positive and 1 – 3 negative. Similarly, with a prevalence of 2 – 5% the split is between CoLab-score 2 and 3 and with prevalence of 5 – 10% between CoLab-score 1 – 2. If the prevalence is higher than 10% only CoLab-score 0 is classified as negative. Using the CoLab-score in this fashion, aims to preserve a number need to test of 15.

Relative importance of variables

Since the variables included in the model are on different scales, the magnitude of the unscaled coefficients cannot be used to compare the importance of variables to each other. To give some indication of the importance of the variables in predicting the outcome, the unscaled coefficients obtained from the adaptive lasso regression were used to calculate the relative importance. The variable with the highest unscaled coefficient was used as maximum ($\beta_{unscaled,max}$), and all other scaled coefficients were divided by this maximum and multiplied by 100 to obtain the relative importance in %: $\frac{\beta_{unscaled}}{\beta_{unscaled,max}} \cdot 100$.

Supplemental material 2

Vaccination status and COVID-19 ED prevalence plot

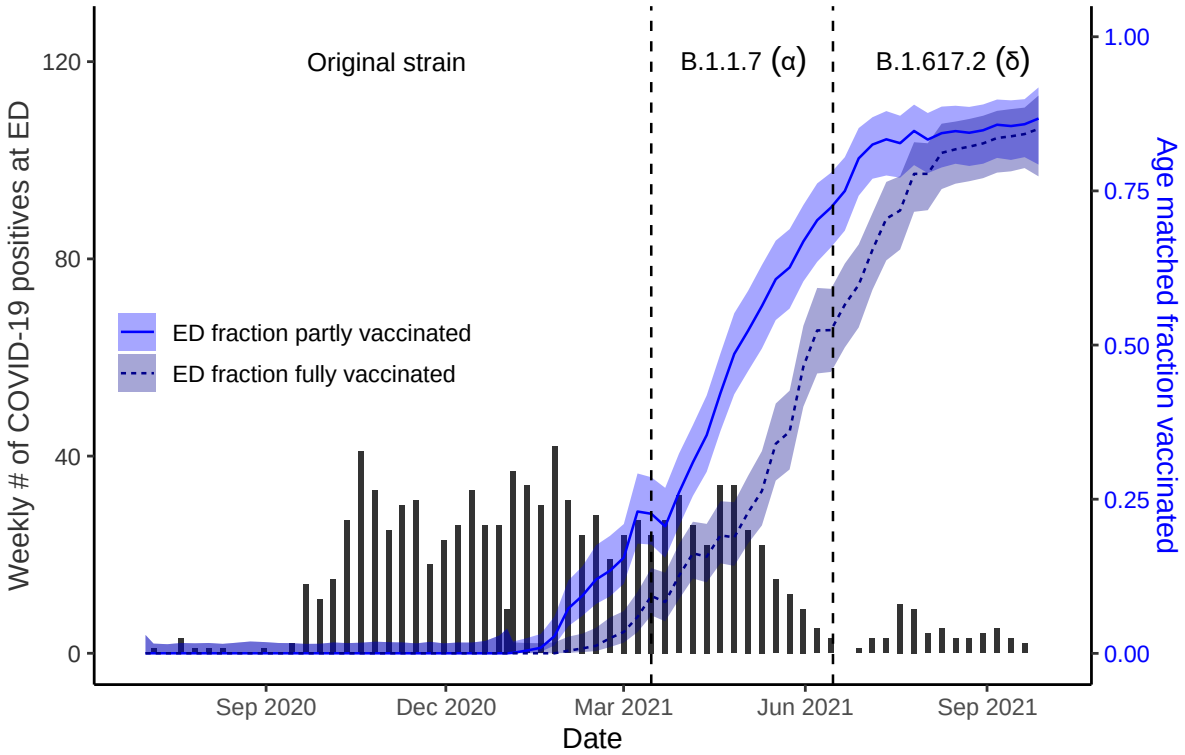


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

The temporal validation dataset consists of ED presentations from July 2020 until October 2021. As stated in the "Materials and Methods" section, this period was split into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. The ED fraction vaccinated is estimated by merging data from the Dutch national institute of public health by the date of the ED presentation and the year of birth of the patient. The gray bars depict weekly number of new COVID-19 cases at the ED, the blue lines the estimated fraction of ED patients fully or partially vaccinated.

CoLab-score performance

Phase	Cases/controls (prevalence)	AUC
Original strain & no vaccinations	694/7999 (8.6%)	0.909 (0.896 - 0.923)
B.1.1.7 strain & partial vaccination	287/2845 (10.1%)	0.937 (0.921 - 0.953)
B.1.617.2 strain & full vaccination	58/3236 (1.8%)	0.898 (0.857 - 0.939)

CoLab-score	Phase	Sensitivity	Specificity	PPV	NPV
0	Original strain & no vaccinations	0.960 (0.944 - 0.974)	0.418 (0.407 - 0.429)	0.135 (0.133 - 0.138)	0.991 (0.987 - 0.994)
	B.1.1.7 strain & partial vaccination	0.983 (0.969 - 0.997)	0.432 (0.413 - 0.450)	0.162 (0.158 - 0.168)	0.996 (0.992 - 0.999)
	B.1.617.2 strain & full vaccination	0.983 (0.948 - 1.000)	0.415 (0.396 - 0.432)	0.030 (0.028 - 0.031)	0.999 (0.998 - 1.000)
≤1	Original strain & no vaccinations	0.879 (0.854 - 0.902)	0.789 (0.779 - 0.798)	0.283 (0.273 - 0.294)	0.986 (0.983 - 0.988)
	B.1.1.7 strain & partial vaccination	0.916 (0.885 - 0.948)	0.809 (0.793 - 0.824)	0.350 (0.332 - 0.370)	0.989 (0.984 - 0.993)
	B.1.617.2 strain & full vaccination	0.862 (0.776 - 0.948)	0.780 (0.765 - 0.794)	0.067 (0.059 - 0.074)	0.997 (0.995 - 0.999)
≤2	Original strain & no vaccinations	0.813 (0.784 - 0.842)	0.894 (0.887 - 0.901)	0.421 (0.404 - 0.441)	0.980 (0.978 - 0.983)
	B.1.1.7 strain & partial vaccination	0.864 (0.826 - 0.902)	0.897 (0.885 - 0.908)	0.484 (0.455 - 0.516)	0.983 (0.979 - 0.988)
	B.1.617.2 strain & full vaccination	0.690 (0.569 - 0.810)	0.892 (0.881 - 0.902)	0.104 (0.086 - 0.123)	0.994 (0.991 - 0.996)
≤3	Original strain & no vaccinations	0.697 (0.661 - 0.731)	0.962 (0.957 - 0.966)	0.634 (0.605 - 0.662)	0.971 (0.968 - 0.974)
	B.1.1.7 strain & partial vaccination	0.760 (0.711 - 0.812)	0.963 (0.955 - 0.970)	0.696 (0.650 - 0.739)	0.973 (0.967 - 0.978)
	B.1.617.2 strain & full vaccination	0.621 (0.483 - 0.741)	0.960 (0.954 - 0.967)	0.222 (0.178 - 0.268)	0.993 (0.990 - 0.995)
≤4	Original strain & no vaccinations	0.566 (0.529 - 0.602)	0.984 (0.981 - 0.987)	0.775 (0.740 - 0.808)	0.960 (0.957 - 0.963)
	B.1.1.7 strain & partial vaccination	0.645 (0.589 - 0.704)	0.983 (0.978 - 0.988)	0.809 (0.762 - 0.856)	0.961 (0.955 - 0.967)
	B.1.617.2 strain & full vaccination	0.517 (0.397 - 0.638)	0.986 (0.982 - 0.990)	0.400 (0.319 - 0.500)	0.991 (0.989 - 0.993)

Table 2: Diagnostic performance of the CoLab-score in the temporal validation dataset, split by phase.

Sensitivities, specificities, positive predictive values (PPV) and negative predictive values (NPV) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is split into three phases according to dominant SARS-CoV-2 strains in the Netherlands and estimated fraction of ED patients vaccinated (see Figure above). Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5. The AUC was significantly higher in the second phase as compared to the first phase (DeLong test p-value: 0.0175), but did not differ significantly between the third and first phase (DeLong test p-value: 0.3903).

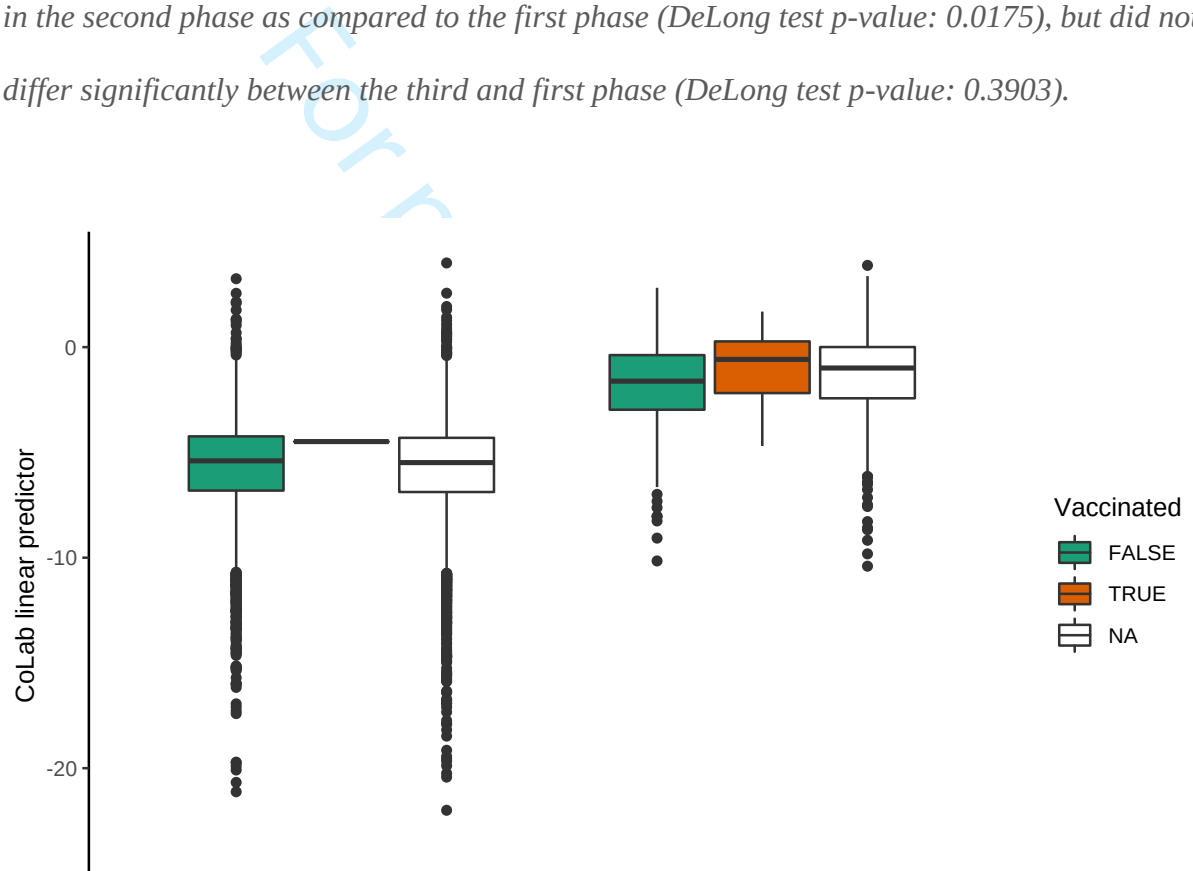


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

The CoLab linear predictor is calculated for all ED presentations in the temporal validation set. Presentations who are registered as vaccinated are labeled TRUE (N = 13). Presentations before vaccine roll-out are labeled FALSE (N = 5855). Presentations during

vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13 presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative. Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or considered positive until proven otherwise, therefore there is only one COVID-19 negative patient with a registered vaccination status.

Supplemental material 3

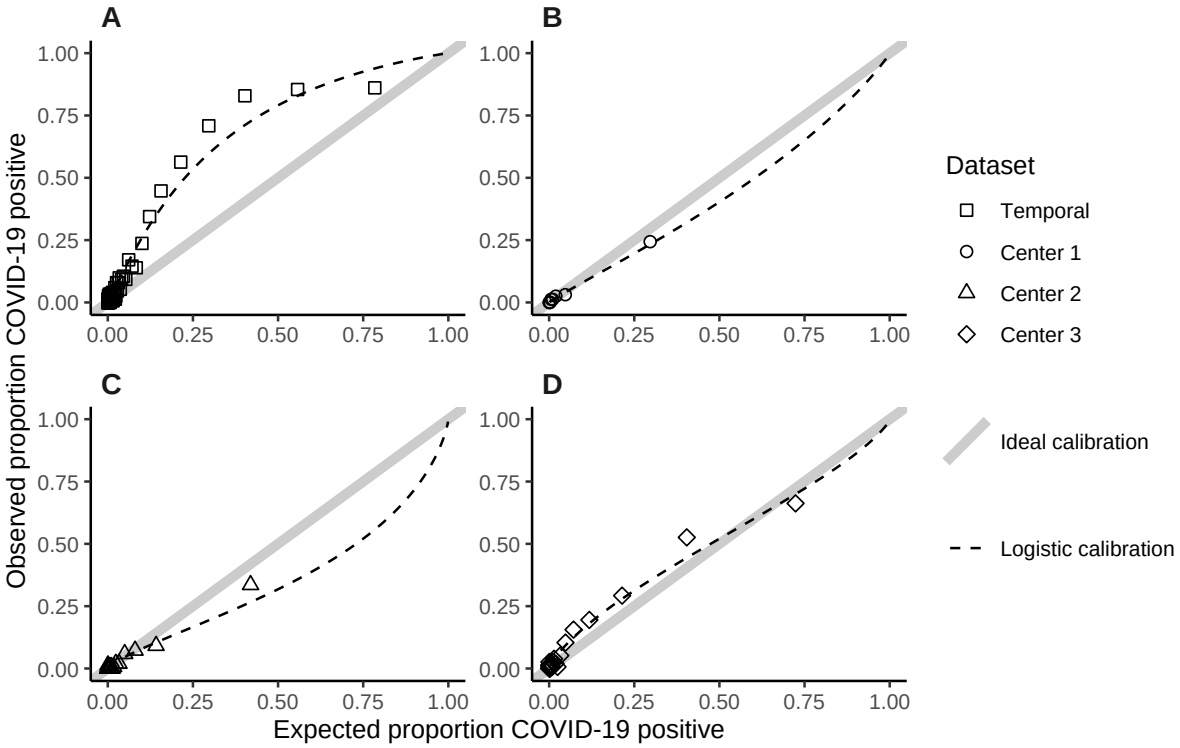


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

In the calibration plots, the proportion of observed COVID-19 positives versus expected probabilities are plotted. Observations are grouped with an average of 150 observations per group. The expected probabilities follow from applying the inverse logit function to the CoLab-linear predictor calculated from Table 2. If the observed proportion in an external dataset is lower than the expected proportion, this means risks are over-estimated, if the observed fraction is higher, risks are under-estimated. Ideally, observed proportions are equal to expected proportions, this ideal-calibration-line is shown as a straight line through the origin with a slope of 1. The logistic calibration line is a logistic regression fit of the predicted probabilities. [Intercept, slope] for plots A-D: A [1.34, 1.08], B [-0.39, 0.92], C [-0.76, 0.77], D [0.08, 0.79]. Although no validation datasets show perfect calibration, this is the result of differences in COVID-19 prevalence in the temporal validation dataset (7.4% versus 2.2%) and differences in calibration of laboratory equipment in the three external centers.

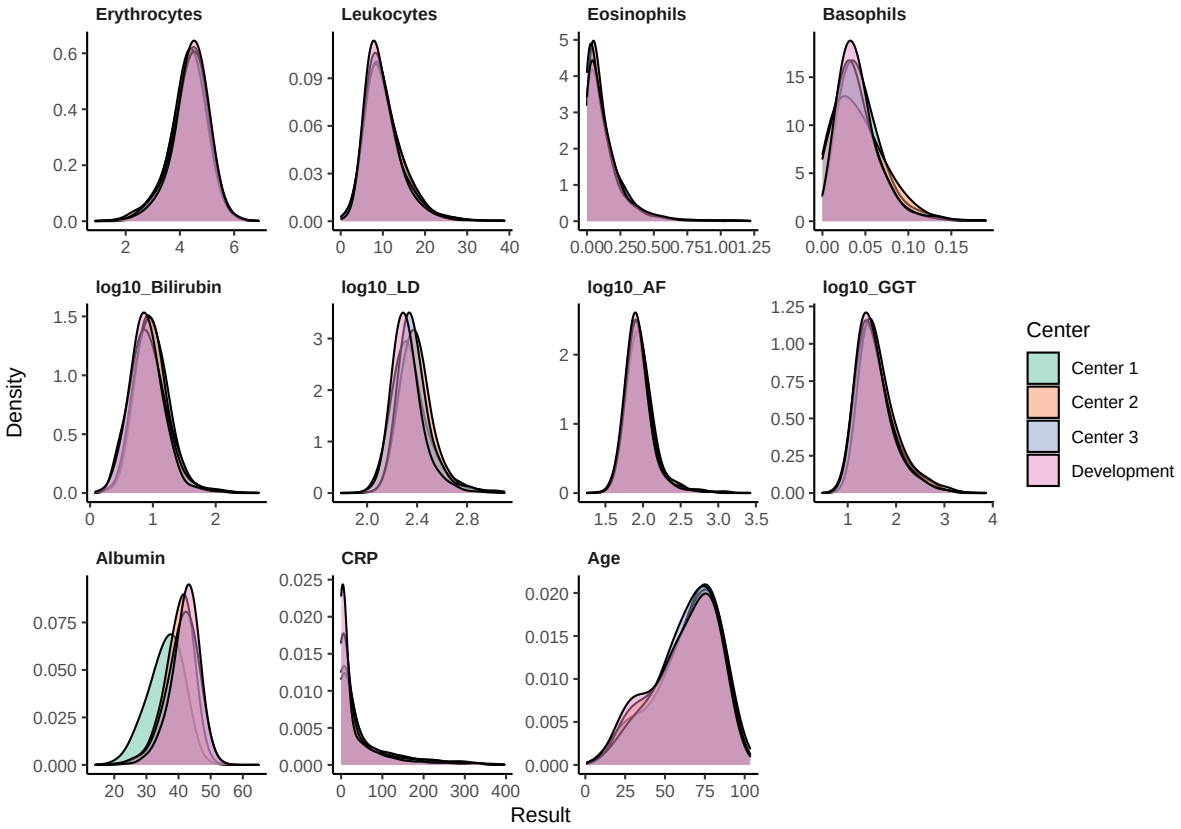


Figure 2: Probability density plots of laboratory parameters.

Probability density plots are shown for all control patients of the development dataset and the three external centers. Ideally all distributions should overlap since this implies that control patient populations are most likely similar in the development dataset to the external datasets. When comparing the distribution of the CoLab variables for all control-patients across different external validation datasets, albumin and LD show the largest deviations.

Supplemental material 4

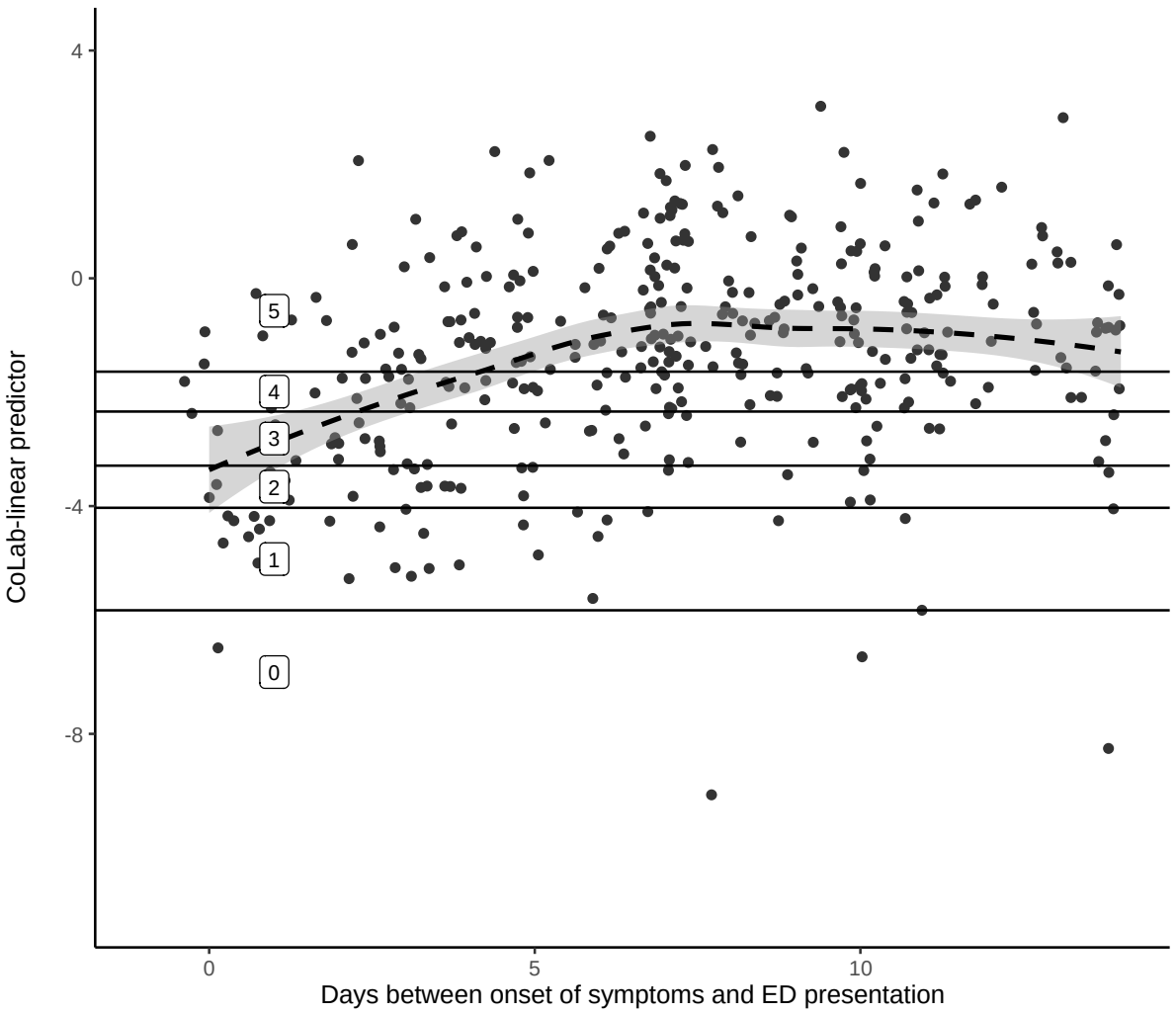


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predict is plotted against the duration of COVID-related symptoms as registered in the electronic patient records. Patients with unknown duration are not plotted. Patients without symptoms were plotted at 0 days. The solid horizontal lines represent the CoLab-score thresholds, the dashed line is a LOESS regression curve with 95% CI. As the duration of symptoms is an integer, some random jitter was added to the days, for visualization purposes. Note that only the first 14 days are shown in this graph.

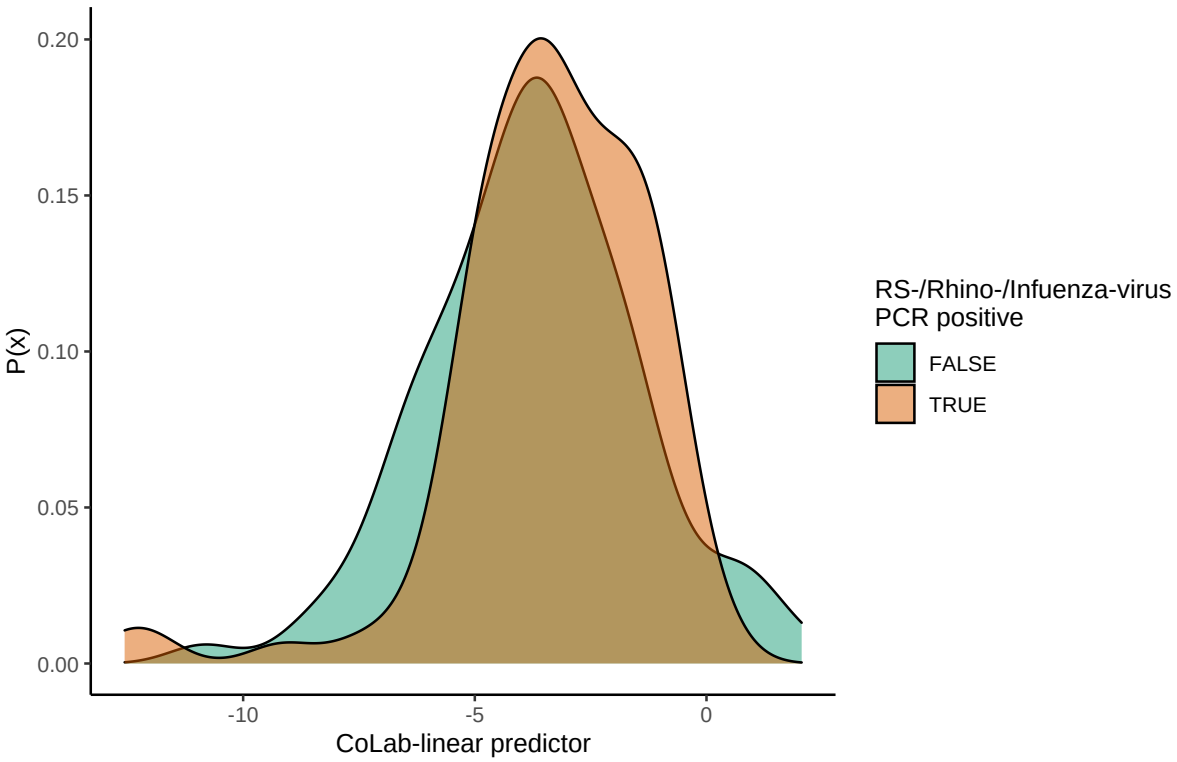


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

For 183 ED presentations that were PCR tested for either RS-, Rhino- and Influenza-virus the CoLab-score was calculated. 91 presentations were PCR positive, 92 were PCR negative. The CoLab-score is only marginally elevated for PCR positive patients, the area under the ROC-curve in separating both groups is 0.573 (95% CI: 4896-0.6563).

Inclusion criterion	Cases/controls (prevalence)	AUC
Temporal validation (reference)	1039/14080 (7.4%)	0.916 (0.906 - 0.927)
Only first presentations, re-presentations are excluded	937/11166 (8.4%)	0.919 (0.909 - 0.930)
Only PCR-tested presentations	372/4062 (9.2%)	0.840 (0.817 - 0.862)

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
0	Reference	0.967 (0.956 - 0.978)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)	1005 (993 - 1016)	5476 (5366 - 5587)	7565 (7454 - 7675)	34 (23 - 46)
	First presentations	0.968 (0.956 - 0.979)	0.416 (0.406 - 0.426)	0.132 (0.130 - 0.134)	0.993 (0.990 - 0.995)	907 (896 - 917)	4259 (4156 - 4353)	5970 (5876 - 6073)	30 (20 - 41)
	PCR-tested presentations	0.946 (0.922 - 0.968)	0.353 (0.338 - 0.368)	0.129 (0.125 - 0.132)	0.985 (0.979 - 0.991)	352 (343 - 360)	1303 (1246 - 1359)	2387 (2331 - 2444)	20 (12 - 29)
≤ 1	Reference	0.888 (0.870 - 0.908)	0.791 (0.783 - 0.798)	0.253 (0.245 - 0.261)	0.989 (0.987 - 0.991)	923 (904 - 943)	10311 (10215 - 10401)	2730 (2640 - 2826)	116 (96 - 135)
	First presentations	0.890 (0.870 - 0.908)	0.793 (0.785 - 0.801)	0.282 (0.273 - 0.292)	0.987 (0.985 - 0.990)	834 (815 - 851)	8112 (8030 - 8194)	2117 (2035 - 2199)	103 (86 - 122)
	PCR-tested presentations	0.852 (0.817 - 0.887)	0.671 (0.656 - 0.686)	0.207 (0.197 - 0.217)	0.978 (0.973 - 0.983)	317 (304 - 330)	2477 (2421 - 2533)	1213 (1157 - 1269)	55 (42 - 68)
≤2	Reference	0.820 (0.796 - 0.843)	0.894 (0.889 - 0.899)	0.382 (0.367 - 0.396)	0.984 (0.982 - 0.986)	852 (827 - 876)	11661 (11591 - 11729)	1380 (1312 - 1450)	187 (163 - 212)
	First presentations	0.824 (0.798 - 0.845)	0.898 (0.892 - 0.904)	0.426 (0.410 - 0.441)	0.982 (0.980 - 0.985)	772 (748 - 792)	9187 (9127 - 9249)	1042 (980 - 1102)	165 (145 - 189)
	PCR-tested presentations	0.734 (0.688 - 0.777)	0.800 (0.786 - 0.812)	0.270 (0.252 - 0.287)	0.968 (0.962 - 0.973)	273 (256 - 289)	2951 (2902 - 2997)	739 (693 - 788)	99 (83 - 116)
≤ 3	Reference	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.596 (0.573 - 0.618)	0.977 (0.974 - 0.979)	738 (709 - 767)	12540 (12496 - 12582)	501 (459 - 545)	301 (272 - 330)
	First presentations	0.716 (0.687 - 0.744)	0.966 (0.962 - 0.969)	0.658 (0.633 - 0.682)	0.974 (0.971 - 0.976)	671 (644 - 697)	9880 (9844 - 9915)	349 (314 - 385)	266 (240 - 293)
	PCR-tested presentations	0.591 (0.540 - 0.640)	0.911 (0.902 - 0.921)	0.403 (0.370 - 0.433)	0.957 (0.952 - 0.962)	220 (201 - 238)	3363 (3328 - 3397)	327 (293 - 362)	152 (134 - 171)
≤4	Reference	0.585 (0.556 - 0.615)	0.984 (0.982 - 0.987)	0.750 (0.724 - 0.778)	0.968 (0.965 - 0.970)	608 (578 - 639)	12838 (12811 - 12866)	203 (175 - 230)	431 (400 - 461)
	First presentations	0.590 (0.558 - 0.621)	0.987 (0.985 - 0.989)	0.805 (0.776 - 0.832)	0.963 (0.961 - 0.966)	553 (523 - 582)	10095 (10071 - 10117)	134 (112 - 158)	384 (355 - 414)
	PCR-tested presentations	0.452 (0.401 - 0.503)	0.959 (0.953 - 0.965)	0.526 (0.480 - 0.575)	0.945 (0.941 - 0.950)	168 (149 - 187)	3539 (3516 - 3562)	151 (128 - 174)	204 (185 - 223)

Table 1: Sensitivity analysis of the CoLab-score in the temporal validation dataset using different inclusion criteria.

Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is used to compare the performance of the CoLab-score with inclusion criteria that differ from the development dataset. The first line shows the performance of the temporal validation dataset with the original inclusion criteria as specified in Figure 1B. The second line shows the performance of the CoLab-score when all re-presentations are excluded (i.e. no repeated presentations). The third line shows the performance of the CoLab-score in the subgroup of patients that underwent PCR-testing.

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3, 4
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6, 7
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	8, 11-12
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	8
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	8
	5b	D;V	Describe eligibility criteria for participants.	8, 9, S1
	5c	D;V	Give details of treatments received, if relevant.	N/A
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	9
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8, 9
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	N/A
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	10
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	10-12, S1
	10c	V	For validation, describe how the predictions were calculated.	16
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	11-13
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N/A
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	22
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	F1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	T1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	S3
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	F1, F3
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	T2
	15b	D	Explain how to the use the prediction model.	T2, S1
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	T3, T4
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21-23
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	19-20
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	19-20
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	20-21
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	N/A
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	N/A

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document. S = Supplemental material, F = Figure, T = Table.

BMJ Open

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059111.R2
Article Type:	Original research
Date Submitted by the Author:	19-May-2022
Complete List of Authors:	Boer, Arjen-Kars; Catharina Hospital Deneer, Ruben; Catharina Hospital, ; Eindhoven University of Technology, Maas, Maaike; Catharina Ziekenhuis, emergency department Ammerlaan, Heidi; Catharina Hospital van Balkom, Roland; Catharina Hospital Thijssen, Wendy; Catharina ziekenhuis, Emergency medicine Bennenbroek, Sophie; Catharina Hospital Leers, Mathie; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Martens, Remy; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Buijs, Madelon; Atalmedial Kerremans, Jos; Alrijne Hospital Location Leiderdorp Messchaert, Muriël ; Gelre Hospitals van Suijlen, Jeroen; Gelre Hospitals van Riel, Natal; Eindhoven University of Technology, ; Amsterdam University Medical Centres Scharnhorst, Volkher; Catharina Hospital; Eindhoven University of Technology,
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Health informatics, Infectious diseases
Keywords:	COVID-19, STATISTICS & RESEARCH METHODS, ACCIDENT & EMERGENCY MEDICINE, Clinical chemistry < PATHOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Arjen-Kars Boer[1*], Ruben Deneer[1,2*†], Maaike Maas[3], Heidi S. M. Ammerlaan[4], Roland H. H. van Balkom[5], Wendy A.M.H. Thijssen[3], Sophie Bennenbroek[3] , Mathie P. G. Leers[6], Remy J.H. Martens[6], Madelon M. Buijs[7], Jos J. Kerremans[8], Muriël Messchaert[9], Jeroen D.E. van Suijlen[9], Natal A.W. van Riel[2,10] and Volkher Scharnhorst[1,2]

* Both authors contributed equally

† Corresponding author

1 Department of Laboratory Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene Loper 3, 5612 AE, Eindhoven, the Netherlands

3 Department of Emergency Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

4 Department of Internal Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands

7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the Netherlands

8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital, Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands

9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334 DZ Apeldoorn, the Netherlands

10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.

Keywords

COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction model, routine laboratory tests

Corresponding author

Ruben Deneer

Michelangelolaan 2

5623 EJ Eindhoven, the Netherlands

ruben.deneer@catharinaziekenhuis.nl

Abstract

Objectives: Identifying patients with a possible SARS-CoV-2 infection in the emergency department (ED) is challenging. Symptoms differ, incidence rates vary and test capacity may be limited. As PCR testing all ED patients is neither feasible nor effective in most centers, a rapid, objective, low-cost early warning score to triage ED patients for a possible infection is developed.

Design: Case-control study.

Setting: Secondary and tertiary hospitals in the Netherlands.

Participants: Patients presenting at the ED with venous blood sampling from July 2019 to July 2020 (N = 10417, 279 SARS-CoV-2 positive). The temporal validation cohort covered the period from July 2020 to October 2021 (N = 14080, 1093 SARS-CoV-2 positive). The external validation cohort consisted of patients presenting at the ED of three hospitals in the Netherlands (N = 12061, 652 SARS-CoV-2 positive).

Primary outcome measures The primary outcome was one or more positive SARS-CoV-2 PCR-test results, within one day prior to, or one week after, ED presentation.

Results: The resulting "CoLab-score" consists of 10 routine laboratory measurements, and age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945). The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991, specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were confirmed in temporal and external validation.

Conclusions: The CoLab-score is based on routine laboratory measurements and is available within one hour after presentation. Depending on the prevalence, COVID-19 may be safely

ruled-out in over one third of ED presentations. Highly suspect cases can be identified regardless of presenting symptoms. The CoLab-score is continuous, in contrast to the binary outcome of lateral flow testing, and can guide PCR testing and triage ED patients.

Article summary

Strengths and limitations of this study

- A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- Using adaptive lasso regression analysis, the panel of 28 laboratory tests was reduced to a single score consisting of a subset of 10 routine ED laboratory tests and age.
- The score was temporally validated from July 2020 to October 2021, in the presence of vaccine roll-out and emergence of new SARS-CoV-2 variants.
- The score was externally validated in 3 other centers in the Netherlands.
- Missingness in the panel of laboratory tests varied between external centers, limiting generalizability of the score to the ED population for which the complete panel of laboratory tests was available.
- The score was not directly compared to lateral flow testing.

Introduction

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has evolved into a global pandemic in 2020 [1]. For emergency department (ED) physicians, identifying presenting patients with a possible COVID-19 infection remains challenging since symptoms like fever, shortness of breath or coughing overlap with other illnesses [2,3]. It is crucial however, to identify a possible COVID-19 infection as early as possible. Early identification prevents further spreading and protects hospital staff by isolating a suspected patient, pending the results of a SARS-COV-2 RNA PCR test and/or chest CT. Conversely, when PCR testing or isolation treatment capacity is limited, ruling-out COVID-19 as soon as possible can save valuable resources.

In the era of electronic health records and clinical prediction models, developing an early warning score that can assist ED physicians in identifying patients presenting at the ED with COVID-19 is of great value. Moreover, if only routine ED test results are required as input, the score can be easily adopted by EDs worldwide, potentially reduce diagnostic costs and accelerate patient triage.

Many COVID-19 prediction models have already been developed, the living systematic review by Wynants et. al [4] provides an extensive overview and critical appraisal. Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early models were based on relatively small sample sizes, hampered by selection bias or were over-fitted by selecting too many features [4–6]. Aside from methodological shortcomings of early models, most models are not developed as an early warning score for all ED patients. Firstly, they require features from tests that are not routinely performed or logged for all ED patients (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED patient population. Secondly, the population on which models are commonly based, are PCR-

tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by physicians.

Only two studies were identified that focus on patients presenting at the ED, include unsuspected (and pre-pandemic) patients as controls, and rely solely on routine (laboratory) tests [9,10].

In this study we report the development and validation of an early warning score that, based on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in patients who undergo routine laboratory testing at presentation. The score can assist ED physicians in triaging patients and prevent further transmission of COVID-19 by quickly identifying possibly infected patients or ruling out a possible infection when resources are scarce.

Methods

Study design

This is a retrospective case-control study where routine laboratory test results, combined with age and gender, from all patient presenting at the emergency department (ED) of the Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-2 PCR test results in a development dataset. A model that could predict the presence of a COVID-19 infection was fit to this dataset. Performance of the model was assessed by i) internal validation, ii) temporal validation and iii) external validation by using data from the ED of three other centers. The study was reviewed by the Medical research Ethics Committees United (MEC-U) under study number W20.071, which confirmed that the Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this study. The study was thereafter reviewed and approved by the internal hospital review board.

Patient and Public Involvement

Patients were not involved in the design, conduct or reporting of this study.

Development dataset

All ED presentations at the Catharina Hospital Eindhoven from July 2019 to July 2020 were included in the development dataset, provided that routine laboratory testing had been requested by the attending ED physician. The rationale for this inclusion period is to limit the effect of seasonal variation in the ED patient population by including the summer, fall and winter season of 2019 (control patients) and the winter, spring and summer season of 2020 (case and control patients). The routine laboratory panel at the ED consists of 28 laboratory tests. In some cases not all tests in the routine panel were requested or one or more

quantitative results were not available due to analytical interference (hemolysis, lipemia or icterus). The routine ED laboratory panel is requested for (adult) patients presenting with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints, or for patients (including non-adult patients) presenting with specific complaints where a suspected diagnosis has to be ruled-in or ruled-out. Presentations with one or more missing values in any of the 28 laboratory test in the routine ED panel, were excluded. Presentations with one or more extreme lab results, > 10 times standard deviation from the median, were also excluded to minimize the effect on the estimation of regression coefficients. The median was chosen as a measure of central tendency due to its resistance for outliers. After the first case of COVID-19 in the Netherlands, all patients with symptoms of COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were approved by the Dutch national institute of public health (RIVM). If a patient had a positive PCR result in the past, subsequent presentations were excluded as re-presentations might be clinically different from de novo presentations.

The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were performed in the time window and the presentation occurred after the first case of COVID-19 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the first case were labelled as "*Pre-COVID-19*".

Laboratory tests

The routine laboratory panel consisted of hemocytometric and chemical analyses. The hemocytometric tests were performed on Sysmex XN-10 instruments (Sysmex Corp., Kobe, Japan) and consisted of hemoglobin, hematocrit, erythrocytes, mean corpuscular volume (MCV), mean cellular hemoglobin (MCH), mean cellular hemoglobin concentration (MCHC), thrombocytes, leukocytes, neutrophils, eosinophils, basophils, lymphocytes and monocytes. The chemical analyses were performed on a Cobas 8000 Pro (Roche Dx, Basel, Switzerland) instrument and consisted of glucose, total bilirubin, aspartate aminotransferase (ASAT), alanine aminotransferase (ALAT), lactate dehydrogenase (LD), creatine kinase (CK), alkaline phosphatase (ALP), gamma-glutamyltransferase (gGT), blood urea nitrogen (BUN), creatinine, CKD-epi estimated glomerular filtration rate (eGFR), potassium, sodium, chloride, albumin (bromocresol green) and C-reactive protein (CRP). These results were combined with age and gender.

Modelling

All data were processed and analyzed in R version 4.1.1 [11]. Laboratory results, combined with age and gender were used as covariates in a regression model. Cases were defined as ED presentations labelled as "*PCR-positive*", controls were all other presentations (i.e. "*PCR-negative*", "*Untested*" or "*Pre-COVID-19*"). To achieve predictive accuracy, limit overfitting and perform feature selection, penalized logistic regression with an adaptive lasso penalty was chosen [12,13]. To minimize missing data, all non-numeric results at the extremes of the measuring range, were converted to numeric results by removing the "<" and ">" signs. For eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m2 and <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD, CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model

predictions, these variables were transformed logarithmically. More details regarding model fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the glmnet-package [14].

CoLab-score

Since this is a retrospective case-control study, the sample prevalence may not reflect the true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term in the model should be adjusted according to the current prevalence (details can be found in the document, **Supplemental Material 1**) [15]. However, adjusting the intercept term is not straightforward to implement in clinical practice, therefore the linear predictor of the model was categorized into a score, this score is hereafter referred to as the "CoLab-score". The categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using the intercept adjustment formula by King [15]. The intervals obtained through these breaks correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and score 5 reflects high-risk. More details regarding the rationale of the CoLab-score categorization can be found in the document, **Supplemental Material 1**.

Internal validation

To assess model performance while taking overfitting into account, bootstrapping was performed. 1000 bootstrap samples were generated from the original data. On each bootstrap sample, the full model fitting procedure and CoLab-score conversion were performed. Optimism adjusted performance measures of the CoLab-score were obtained by applying the 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [16]. Performance

measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate performance measures [17]. Although the full inclusion period from July 2019 to July 2020 was used for model fitting, the performance was evaluated on the period starting from the first COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain performance measures that would reflect real world performance.

Temporal validation

For temporal validation, results from our center were prospectively analyzed from July 2020 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period there was also more widespread external PCR testing by municipal health services. The results of external conducted PCR tests were not available to our study. To overcome this limitation, the outcome in the temporal validation cohort was chosen as a composite of the hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test result. This period also covers both the emergence of new SARS-CoV-2 variants as well as vaccine rollout. However, neither vaccination status nor genomic sequencing was available to determine whether a patient was vaccinated or which variant caused the infection. Therefore, data from the Dutch national institute of public health (RIVM) was used, to divide the temporal validation period into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold

for the entire period, as well as for each phase separately to determine a possible effect of vaccination and new variants on performance (results in the **Supplemental Material 2**). Model calibration was assessed graphically using the rms-package [18].

External validation

For the external validation, several centers in the Netherlands were approached and assessed if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available. Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and Zuyderland Medical Center (center 3). The hematological parameters were measured with Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external validation was similar to the temporal validation and consisted of assessing the AUC sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was assessed graphically analogous to the temporal validation dataset.

Results

Development dataset

12879 emergency department (ED) presentations of 10327 patients from July 2019 to July 2020 were included. After excluding cases with an incomplete lab panel, patient presentations that occurred after a positive PCR test in the past (re-presentations) and presentations with extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of 8610 patients remained (**Figure 1 A**).

	Pre-COVID N = 5890	**Untested N = 3303**	**PCR negative N = 945**	**PCR positive N = 279**
Age in years	61 (21)	60 (21)	66 (18)	69 (15)
Female gender	2909 (49.4 %)	1659 (50.2 %)	466 (49.3 %)	95 (34.1 %)
Specialism				
Internal medicine	1648 (28.0 %)	896 (27.1 %)	244 (25.8 %)	71 (25.4 %)
Surgery	1007 (17.1 %)	679 (20.6 %)	51 (5.4 %)	5 (1.8 %)
Neurology	775 (13.2 %)	468 (14.2 %)	64 (6.8 %)	5 (1.8 %)
Pulmonary medicine	714 (12.1 %)	220 (6.7 %)	326 (34.5 %)	167 (59.9 %)
Cardiology	560 (9.5 %)	322 (9.7 %)	145 (15.3 %)	6 (2.2 %)
Urology	309 (5.2 %)	148 (4.5 %)	15 (1.6 %)	7 (2.5 %)
Gastroenterology	306 (5.2 %)	224 (6.8 %)	27 (2.9 %)	1 (0.4 %)
Geriatrics	189 (3.2 %)	95 (2.9 %)	52 (5.5 %)	15 (5.4 %)
Orthopedics	147 (2.5 %)	109 (3.3 %)	11 (1.2 %)	0 (0.0 %)
Gynecology	118 (2.0 %)	82 (2.5 %)	2 (0.2 %)	0 (0.0 %)
Other	117 (2.0 %)	60 (1.8 %)	8 (0.8 %)	2 (0.7 %)
Hemoglobin in mmol/L	8.2 (1.3)	8.3 (1.3)	8.2 (1.4)	8.6 (1.1)
Hematocrit in L/L	0.403 (0.059)	0.405 (0.056)	0.405 (0.062)	0.417 (0.047)
Erythrocytes in /pL	4.41 (0.69)	4.43 (0.66)	4.41 (0.72)	4.61 (0.60)
MCV in fl	91.8 (6.4)	91.9 (6.1)	92.4 (6.7)	90.7 (5.5)
MCH in mmol	1.859 (0.157)	1.876 (0.150)	1.874 (0.172)	1.869 (0.141)
MCHC in mmol/L	20.2 (0.9)	20.4 (0.9)	20.3 (1.0)	20.6 (0.8)
Thrombocytes in /nL	263 (99)	266 (100)	269 (105)	217 (123)
Leukocytes in /nL	9.30 [7.06, 12.16]	8.92 [7.01, 11.89]	9.66 [7.17, 12.94]	6.33 [4.74, 8.48]
Neutrophils in /nL	6.62 [4.51, 9.53]	6.10 [4.42, 8.94]	7.01 [4.79, 10.02]	4.71 [3.30, 6.94]
Eosinophils in /nL	0.09 [0.03, 0.17]	0.09 [0.03, 0.18]	0.08 [0.02, 0.17]	0.00 [0.00, 0.02]
Basophils in /nL	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.01 [0.01, 0.02]
Lymphocytes in /nL	1.47 [0.93, 2.13]	1.56 [1.05, 2.18]	1.31 [0.80, 2.03]	0.86 [0.59, 1.21]
Monocytes in /nL	0.70 [0.52, 0.93]	0.69 [0.52, 0.91]	0.74 [0.54, 1.01]	0.45 [0.32, 0.64]
Glucose in mmol/L	6.76 [5.83, 8.39]	6.68 [5.76, 8.14]	6.98 [5.95, 8.85]	6.77 [5.98, 8.48]
Bilirubin in umol/L	7.5 [5.0, 11.6]	7.4 [5.1, 10.9]	8.3 [5.6, 12.4]	8.2 [6.3, 11.4]
ASAT in U/L	24.0 [19.1, 32.2]	26.5 [21.6, 35.1]	27.7 [21.7, 39.2]	40.7 [30.2, 57.2]
ALAT in U/L	24.3 [17.8, 35.3]	25.3 [18.4, 36.2]	25.7 [18.4, 40.0]	33.7 [23.3, 50.0]
LD in U/L	201 [173, 240]	198 [170, 236]	215 [178, 263]	300 [238, 403]
CK in U/L	82 [51, 134]	83 [52, 136]	76 [51, 125]	124 [62, 222]
ALP in IU/L	83.0 [68.0, 105.0]	81.0 [65.8, 102.5]	86.9 [67.9, 110.0]	71.0 [58.8, 85.0]
gGT in U/L	27.0 [17.0, 53.0]	28.4 [18.4, 50.5]	37.0 [22.4, 68.9]	42.0 [28.0, 83.5]
BUN in mmol/L	5.7 [4.3, 8.0]	5.8 [4.3, 7.8]	6.2 [4.6, 9.4]	6.1 [4.7, 8.9]

CKD-epi in ml/min/m2	80.9 [58.0, 99.1]	85.0 [63.5, 103.3]	79.1 [52.1, 96.6]	76.6 [54.9, 91.2]
Potassium in mmol/L	4.06 (0.50)	4.03 (0.49)	4.07 (0.55)	3.91 (0.47)
Sodium in mmol/L	139.2 (4.0)	138.5 (3.9)	138.0 (4.3)	136.4 (4.1)
Chloride in mmol/L	104.4 (4.6)	103.8 (4.5)	102.9 (4.8)	101.6 (4.4)
Albumin in g/L	42.4 (4.9)	42.3 (4.5)	40.8 (4.8)	38.4 (3.8)
CRP in mg/L	8 [2, 41]	5 [1, 30]	18 [3, 69]	77 [37, 136]

Table 1: Descriptive statistics of development dataset and laboratory concentrations.

Shown are the laboratory tests routinely requested at ED presentation and their mean/median results (in the development dataset) for the presentations before the first COVID-19 patient in the Netherlands ("Pre-COVID-19"), presentations thereafter that were not tested for COVID-19 ("Untested"), tested negatively ("PCR negative") and tested positive ("PCR positive"). For results with normal distributions, the mean value and standard deviation (in round brackets) are shown. For results that have skewed or heavy tailed distributions, the median value and the interquartile range is shown [in squared brackets]. Dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error).

Descriptive statistics of ED presentations are shown in **Table 1,** dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error [19]). For the PCR positives (N = 279), 91% (95% CI: 88 to 94%) of the cases were tested positive in their first PCR. The remaining 24 patients were positive in their second (N = 18), third (N = 5) or fourth (N = 1) PCR.

CoLab-score

The model obtained through adaptive lasso regression contained eleven variables, which are depicted with their regression coefficients (weights) in **Table 2**.

Variable	β	Exclusion limit	Relative importance
Intercept	-6.885		-
Erythrocytes /pL	0.9379	Erythrocytes < 2.9 /pL	52 %
Leukocytes /nL	-0.1298		46 %
Eosinophils /nL	-6.834		86 %
Basophils /nL	-47.70	Basophils >0.33 /nL	100 %
$\log_{10}$ of Bilirubin in µmol/L	-1.142	Bilirubin >169 µmol/L	26 %
$\log_{10}$ of LD in U/L	5.369	LD >1564 U/L	58 %
$\log_{10}$ of ALP in IU/L	-3.114	AF >1000 IU/L	45 %
$\log_{10}$ of gGT in U/L	0.3605	gGT >1611 U/L	11 %
Albumin in g/L	-0.1156		45 %
CRP in mg/L	0.002560		15 %
Age in years	0.002275		4 %

Table 2: Calculation of the CoLab-linear predictor (LP).

The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of the 11 variables with their corresponding coefficients (β's). CoLab-LP = – 6.885 + [erythrocytes] × 0.9379 – [leukocytes] × 0.1298 – [eosinophils] × 6.834 – [basophils] × 47.7 – log10([bilirubin]) × 1.142 + log10([LD]) × 5.369 – log10([ALP]) × 3.114 + log10([gGT]) × 0.3605 – [albumin] × 0.1156 + [CRP] × 0.02560 + [age] × 0.002275. The LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid if any of the variables exceed the limits in the third column. The relative importance ranks the importance of variables in predicting the outcome, relative to the most important variable (in this case basophils).

A larger β-coefficient does not imply that a variable is more important in predicting the odds of testing positive for SARS-CoV-2, since variables are on different scales. The most important variables are basophiles, eosinophils and lactate dehydrogenase (LD).

As shown in **Figure 2**, the linear predictor clearly discriminates between COVID-19 and non-COVID-19. The linear predictor is converted to CoLab-scores 0 – 5 with the cut-points depicted in **Figure 2**.

Internal validation

The model was validated in the period starting from the first COVID-19 infection to July 2020, in this period the mean prevalence was 7.2%. The AUC of the CoLab-score is 0.930 (95% CI: 0.909 to 0.945).

CoLab-score	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN	% of population
0	0.984 (0.969 - 0.991)	0.410 (0.302 - 0.543)	0.115 (0.094 - 0.147)	0.997 (0.993 - 0.999)	273.4 (241.2 - 304.4)	1470.9 (1081.1 - 1950.9)	2119.1 (1633.5 - 2507.6)	4.6 (2.6 - 8.6)	38.0 (28.0 - 51.0)
≤ 1	0.912 (0.892 - 0.952)	0.785 (0.741 - 0.827)	0.248 (0.207 - 0.300)	0.991 (0.989 - 0.995)	253.5 (226.5 - 287.0)	2817.1 (2655.4 - 2961.2)	772.9 (623.2 - 934.5)	24.5 (13.4 - 30.2)	73.3 (69.3 - 77.3)
≤ 2	0.856 (0.816 - 0.895)	0.880 (0.864 - 0.900)	0.357 (0.315 - 0.415)	0.988 (0.984 - 0.991)	238.1 (209.6 - 267.9)	3160.8 (3100.7 - 3233.7)	429.1 (357.3 - 487.1)	39.9 (28.5 - 52.4)	82.9 (80.9 - 83.9)
≤ 3	0.757 (0.706 - 0.809)	0.951 (0.944 - 0.959)	0.546 (0.496 - 0.604)	0.981 (0.976 - 0.985)	210.4 (183.4 - 240.2)	3415.1 (3378.0 - 3456.4)	174.9 (147.0 - 199.3)	67.6 (51.9 - 84.9)	90.0 (89.0 - 91.0)
≤ 4	0.612 (0.530 - 0.706)	0.978 (0.972 - 0.983)	0.683 (0.628 - 0.746)	0.970 (0.963 - 0.978)	170.2 (141.6 - 204.9)	3510.6 (3476.8 - 3547.5)	79.4 (60.3 - 100.4)	107.9 (79.1 - 134.0)	93.7 (91.7 - 93.7)

Table 3: Bootstrapped diagnostic performance of the CoLab-score in the development dataset.

The development dataset was internally validated for the period March 2020 – July 2020 (N = 3868). The optimism-adjusted bootstrapped sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) and fraction of presentations (%) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4). The numbers in round brackets represent the 95% optimism-

adjusted bootstrapped confidence intervals. The first column defines the threshold above which CoLab-score a patient is considered positive. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5. Also note that TP, TN, FP and FN are not whole numbers, as these are obtained through bootstrapping and each bootstrap replicate contains a different number of controls and cases.

Diagnostic performance is shown in **Table 3.** A CoLab-score of 0 has a negative predictive value (NPV) of 0.997 (95% CI: 0.993 to 0.999) and positive predictive value (PPV) of 0.115 (0.0934 - 0.147), one third (38%, 95% CI: 28 to 514%) of all ED presentations were assigned this score and can therefore be safely excluded. Conversely, 6% (95% CI: 6 to 8%) of the ED patients had a CoLab-score = 5. Given the PPV of this score (0.683, 95% CI: 0.628 to 0.746, NPV: 0.970, 95% CI: 0.963 - 0.978), subsequent PCR testing is advised.

Temporal validation

As the CoLab-score was developed in our center after the first COVID-19-wave in the Netherlands, the performance was evaluated in our center from July 2020 until October 2021. Lab results from 17489 ED presentations were collected. After applying the inclusion flow as shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a COVID-19 infection.

The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the development cohort, although sensitivity is slightly lower and specificity slightly higher (cf. **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental**

Material 2 Figure 1). The discriminative ability was not lower in the second or third phase, compared to the first phase. Diagnostic performance is preserved in terms of sensitivity and specificity, except a moderately reduced sensitivity of scores ≥ 3 in the third phase as compared to the first phase. PPV and NPV are incomparable due to different prevalence/pre-test probabilities in each phase (see **Supplemental Material 2 Table 1**).

In terms of the predicted probabilities, model calibration shows that overall predicted probabilities are too low (see **Supplemental Material 3** for the calibration plot), which is expected since the prevalence differs and the intercept has to be adjusted to the prevalence. In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that initially did not present with COVID-specific symptoms. Most patients had neurological or orthopedic presenting symptoms.

External validation

For external validation, data obtained from three other centers were used, center 1 (N = 1284, 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545, 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers is ≥ 0.96. Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for external centers are shown in **Supplemental Material 3**, the observed fraction of COVID-19

positives is slightly lower than expected in centers 1 and 2. For center 3, low probabilities appear slightly underestimated and high probabilities slightly overestimated.

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
0	Temporal	0.967 (0.956 - 0.978)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)	1005 (993 - 1016)	5476 (5366 - 5587)	7565 (7454 - 7675)	34 (23 - 46)
	Center 1	1.000 (1.000 - 1.000)	0.331 (0.307 - 0.358)	0.059 (0.057 - 0.061)	1.000 (1.000 - 1.000)	52 (52 - 52)	410 (380 - 443)	827 (794 - 857)	0 (0 - 0)
	Center 2	0.961 (0.922 - 0.990)	0.351 (0.333 - 0.369)	0.052 (0.049 - 0.054)	0.996 (0.992 - 0.999)	99 (95 - 102)	985 (935 - 1035)	1823 (1773 - 1873)	4 (1 - 8)
	Center 3	0.970 (0.950 - 0.988)	0.322 (0.306 - 0.338)	0.130 (0.126 - 0.133)	0.991 (0.984 - 0.996)	327 (320 - 333)	1042 (991 - 1092)	2193 (2143 - 2244)	10 (4 - 17)
≤ 1	Temporal	0.888 (0.870 - 0.908)	0.791 (0.783 - 0.798)	0.253 (0.245 - 0.261)	0.989 (0.987 - 0.991)	923 (904 - 943)	10311 (10215 - 10401)	2730 (2640 - 2826)	116 (96 - 135)
	Center 1	0.923 (0.846 - 0.981)	0.694 (0.669 - 0.720)	0.113 (0.101 - 0.124)	0.995 (0.991 - 0.999)	48 (44 - 51)	858 (828 - 891)	379 (346 - 409)	4 (1 - 8)
	Center 2	0.913 (0.854 - 0.961)	0.678 (0.661 - 0.696)	0.094 (0.087 - 0.101)	0.995 (0.992 - 0.998)	94 (88 - 99)	1905 (1857 - 1953)	903 (855 - 951)	9 (4 - 15)
	Center 3	0.914 (0.881 - 0.944)	0.674 (0.657 - 0.691)	0.226 (0.216 - 0.236)	0.987 (0.982 - 0.991)	308 (297 - 318)	2180 (2126 - 2234)	1055 (1001 - 1109)	29 (19 - 40)
≤ 2	Temporal	0.820 (0.796 - 0.843)	0.894 (0.889 - 0.899)	0.382 (0.367 - 0.396)	0.984 (0.982 - 0.986)	852 (827 - 876)	11661 (11591 - 11729)	1380 (1312 - 1450)	187 (163 - 212)
	Center 1	0.808 (0.692 - 0.904)	0.811 (0.788 - 0.832)	0.152 (0.129 - 0.176)	0.990 (0.984 - 0.995)	42 (36 - 47)	1003 (975 - 1029)	234 (208 - 262)	10 (5 - 16)
	Center 2	0.845 (0.777 - 0.913)	0.801 (0.785 - 0.815)	0.135 (0.122 - 0.147)	0.993 (0.990 - 0.996)	87 (80 - 94)	2248 (2205 - 2289)	560 (519 - 603)	16 (9 - 23)
	Center 3	0.890 (0.855 - 0.923)	0.794 (0.779 - 0.808)	0.311 (0.294 - 0.328)	0.986 (0.981 - 0.990)	300 (288 - 311)	2569 (2521 - 2615)	666 (620 - 714)	37 (26 - 49)
≤ 3	Temporal	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.596 (0.573 - 0.618)	0.977 (0.974 - 0.979)	738 (709 - 767)	12540 (12496 - 12582)	501 (459 - 545)	301 (272 - 330)
	Center 1	0.750 (0.635 - 0.865)	0.909 (0.892 - 0.925)	0.257 (0.213 - 0.306)	0.989 (0.983 - 0.994)	39 (33 - 45)	1124 (1104 - 1144)	113 (93 - 133)	13 (7 - 19)
	Center 2	0.660 (0.563 - 0.748)	0.897 (0.885 - 0.908)	0.190 (0.163 - 0.218)	0.986 (0.983 - 0.990)	68 (58 - 77)	2519 (2486 - 2549)	289 (259 - 322)	35 (26 - 45)

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
	Center 3	0.766 (0.718 - 0.810)	0.887 (0.876 - 0.898)	0.413 (0.386 - 0.442)	0.973 (0.968 - 0.978)	258 (242 - 273)	2869 (2835 - 2905)	366 (330 - 400)	79 (64 - 95)
≤ 4	Temporal	0.585 (0.556 - 0.615)	0.984 (0.982 - 0.987)	0.750 (0.724 - 0.778)	0.968 (0.965 - 0.970)	608 (578 - 639)	12838 (12811 - 12866)	203 (175 - 230)	431 (400 - 461)
	Center 1	0.654 (0.519 - 0.788)	0.951 (0.939 - 0.962)	0.359 (0.293 - 0.435)	0.985 (0.979 - 0.991)	34 (27 - 41)	1176 (1161 - 1190)	61 (47 - 76)	18 (11 - 25)
	Center 2	0.534 (0.437 - 0.621)	0.952 (0.943 - 0.959)	0.287 (0.239 - 0.339)	0.982 (0.979 - 0.986)	55 (45 - 64)	2672 (2649 - 2693)	136 (115 - 159)	48 (39 - 58)
	Center 3	0.665 (0.611 - 0.718)	0.930 (0.921 - 0.938)	0.497 (0.462 - 0.534)	0.964 (0.958 - 0.969)	224 (206 - 242)	3008 (2980 - 3036)	227 (199 - 255)	113 (95 - 131)

Table 4: Diagnostic performance of the CoLab-score in the validation dataset (temporal) and three external hospitals.

Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

Discussion

Given the impact of COVID-19 on society and healthcare, there is a need for simple and fast detection of patients with a possible COVID-19 infection in the ED. The CoLab-score described in this study, is a fast and accurate risk score to triage patients presenting at the ED based on ten routine blood biomarkers and age.

The main strength of this study is that this score can be used as an early-warning or triaging tool for the ED population presenting with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints where a routine blood panel is requested. This is in contrast to the vast majority of COVID-19 diagnostic models that have been developed on a pre-selected population of PCR-tested patients [9,20–26]. Moreover, the CoLab-score requires only routine blood tests, instead of (features from) imaging such as CT-scans or laboratory tests that are not routinely collected in the ED, e.g. interleukin-6 or 3-hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific) [27,28].

Two other studies have been published which are similar to this study [9,10]. Interestingly, the study by Soltan et al., ranked basophils and eosinophils as the two most important features in predicting the outcome, similar to our results [10]. Eosinophils were also seen as one of the most important features by Plante et al. [9]. However, both studies focus on an artificial intelligence/machine learning approach. While their approach likely results in higher predictive performance, due to the ability of machine learning models to capture non-linear and interaction effects, the goal of this study was to develop a simple, fast and robust model that can easily be implemented in current hospital IT systems.

Since this is a retrospective case-control study, there is some unavoidable missing data. In our cohort 17.6% of the ED presentations could not be used due to one or more missing laboratory results. This is lower or equal to similar studies; 22% [23], 17% [21] and 11% [26]. Important to note is that 7.7% of missingness is due to analytical errors which can be assumed to be missing completely at random. For the remaining 9.9% of missingness, the full lab panel was most frequently missing for pediatric, obstetric and surgery patients. These patients are presenting with specific complaints for which specific laboratory tests are requested, and hence do not match the inclusion criteria for a routine blood panel. Overall the missingness was significantly lower in the PCR-tested group versus the untested group (χ^2-test p-value <0.001). It is assumed that all presentations in the untested group are COVID-19 negative. However, some presentations with asymptomatic COVID-19 could be present in the untested control group. The impact of these 'false controls' is most likely small as other studies indicate that there is a very low positivity rate among asymptomatic ED presentations (only a few in over a thousand tested asymptomatic cases) [29,30].

In the external centers, there is a high level of missingness as a result of an incomplete laboratory panel. In the case of centers 1 and 2, only internal medicine ED presentations were tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine ED presentations, which is reflected by the few PCR-positive patients excluded. Due to these high levels of missingness, the results of the external centers cannot be used to show that the CoLab-score generalizes to the entire ED population. Rather, the results show that for the majority of COVID-19 positive patients presenting at the ED, a routine laboratory panel is available from which the CoLab-score can be calculated, and that the performance of the CoLab-score in this population is comparable to the development population.

The performance of the CoLab-score is affected by the time between the onset of symptoms and ED presentations. The score increases with the duration of symptoms and gradually decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some COVID-19 patients with early or late presentation after onset of symptoms can be missed. Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and <10 days prior to ED presentation. Chemotherapy that causes myeloid suppression, will decrease neutrophilic, basophilic and eosinophilic counts and thereby "falsely" increasing the CoLab-score. Conversely, COVID-19 patients with severe anemia could have "falsely" lowered CoLab-scores. To minimize false negatives, we have therefore advised to report CoLab-scores only when the concentration of erythrocytes is ≥ 2.9 /pL.

It was chosen to exclude re-presentations after a previous presentation with COVID-19. Since the median time between initial presentation and re-presentation was 12 days, these patients were most likely not re-infected patients, but patients who deteriorated after initial presentation/treatment. Given that the CoLab-score follows the host-immune response, the score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients would impact the performance of the CoLab-score as patients in a later phase of the disease show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate after treatment for COVID-19. Other re-presentations were not excluded, which results in some patients appearing multiple times in a dataset. This was not corrected for in the regression model since the assumption was made that ED presentations are independent observations. The median time between re-presentations is 38 days, most likely resulting in variations in laboratory results between presentations, and hence, little to no correlation between presentations. A sensitivity analysis was performed whereby only the first

presentation was included for each patient (**Supplemental Material 4 Table 1**), but no difference was found in performance in terms of sensitivity, specificity and AUC.

The CoLab-score does not serve as a replacement for PCR-testing or LFTs, and can be used to guide PCR-testing when routine blood tests are available. Important to note is that the CoLab-score is only valid for ED presentations where routine blood testing is requested, and as a consequence does not generalize to the ED population who is otherwise well and does not undergo routine blood testing. Using the CoLab-score in a symptomatic/PCR-tested cohort also results in different diagnostic performance characteristics, as compared to using the score on the full ED cohort (see **Supplemental Material 4 Table 1)**.

Finally, the CoLab-score could lead to false positives by other viral infections. However, in an historic patient cohort, the CoLab-score had only limited discriminative ability in separating influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4 Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-response to the virus, it is expected that the CoLab-score is also sensitive to future SARS-CoV-2 variants. This is supported by the fact that the discriminative ability is sustained in periods with different dominant variants. Moreover, there is no evidence that the discriminative ability of the CoLab-score is lowered by a change in the ED patient population as a result of widespread vaccination. Although vaccination status is not registered for all presenting patients, in a small subgroup of 12 patients for whom vaccination status was registered, and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5) (see **Supplemental Material 2 Figure 2**),

To conclude, the CoLab-score developed and validated in this study, based on 10 routine laboratory results and age, is available within 1 hour for any patient presenting at the ED where routine blood testing is requested. The score can be used by clinicians to guide PCR testing or triage patients and helps to identify COVID-19 in patients presenting at the ED with

abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints where a routine blood panel is requested. The lowest CoLab-score can be used to effectively rule-out a possible SARS-CoV-2 infection, the highest score to alert physicians to a possible infection. The CoLab-score is therefore a valuable tool to rule out COVID-19, guide PCR testing and is available to any center with access to routine laboratory tests.

Funding statement

This was an investigator-initiated study and no funding was received for this study.

Competing interests

A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR reports no conflict of interest. VS reports no conflict of interest.

Data sharing statement

Datasets with source data for Table 1, Figure 2, Table 3 and Table 4, as well the R-code to fit the model is available from the Dryad repository, DOI:[WILL BE PROVIDED WHEN UNDER REVIEW]. Technical appendix can be found in **Supplemental Material 1**.

Author contributorship statement

Arjen-Kars Boer: Conceptualization (Lead), Data curation (Lead), Funding acquisition (Lead), Investigation (Equal), Methodology (Equal), Supervision (Equal), Writing-original draft (Equal), Writing-review & editing (Equal).

Ruben Deneer: Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Lead), Software (Lead), Visualization (Lead), Writing-original draft (Equal), Writing-review & editing (Equal).

Maaike Maas: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Equal), Writing-review & editing (Equal).

Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal), Writing-review & editing (Equal).

Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing (Equal).

References

1 Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 4 Feb 2021).

2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *https://doi.org/101056/NEJMoa2002032* 2020;**382**:1708–20. doi:10.1056/NEJMOA2002032

3 Vetter P, Vu DL, L'Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**. doi:10.1136/BMJ.M1470

4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18. doi:10.1136/bmj.m1328

5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review. J. Med. Syst. 2020;**44**:122. doi:10.1007/s10916-020-01582-x

6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417

7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19—Definition and Evaluation. *https://doi.org/101148/radiol2020201473* 2020;**296**:E97–104. doi:10.1148/RADIOL.2020201473

8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity

score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840

9 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: A large, multicenter, real-world study. *J Med Internet Res* 2020;**22**:e24048. doi:10.2196/24048

10 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87. doi:10.1016/S2589-7500(20)30274-0

11 R Core Team. R: A Language and Environment for Statistical Computing. 2020.https://www.r-project.org/

12 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29. doi:10.1198/016214506000000735

13 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

14 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01

15 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63. doi:10.1093/oxfordjournals.pan.a004868

16 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973

17 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to

analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-2105-12-77

18 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.https://cran.r-project.org/package=rms

19 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons and progress. Scand. J. Clin. Lab. Invest. 1999;**59**:491–500. doi:10.1080/00365519950185229

20 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst* 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4

21 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502

22 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807. doi:10.1007/s00330-020-07022-1

23 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. *Clin Chem Lab Med* 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593

24 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305

25 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in

the emergency department: derivation and validation of the LDH-CRP-Lymphocyte (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336

26 Tordjman M, Mekki A, Mali RD, *et al.* Pre-test probability for SARS-Cov-2-related infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342. doi:10.1371/journal.pone.0243342

27 Peto T, Affron D, Afrough B, *et al.* COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine* 2021;**36**:100924. doi:10.1016/J.ECLINM.2021.100924

28 García-Fiñana M, Hughes DM, Cheyne CP, *et al.* Performance of the Innova SARS-CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot: population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637

29 Ford JS, Parikh A, Sandhu R, *et al.* Testing Asymptomatic Emergency Department Patients for Coronavirus Disease 2019 (COVID-19) in a Low-prevalence Region. Acad. Emerg. Med. 2020;**27**:771–4. doi:10.1111/acem.14044

30 Ravani P, Saxinger L, Chandran U, *et al.* COVID-19 screening of asymptomatic patients admitted through emergency departments in Alberta: a prospective quality-improvement study. *C Open* 2020;**8**:E887–94. doi:10.9778/cmajo.20200191

Figure legends

Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B) dataset.

All patient admissions with routine venous blood sampling at the emergency department (ED) were included. For the development dataset, completeness of the lab panel was assessed for all 28 laboratory tests , for the temporal validation dataset this was only necessary for 10 laboratory tests. The major causes of missingness are described in the text. In the development dataset, presentations with extreme values (>10 SD) were excluded. The same limits were applied to the temporal validation dataset (see Table 2 for limits).

Figure 2: Probability density plot of the CoLab-linear predictor.

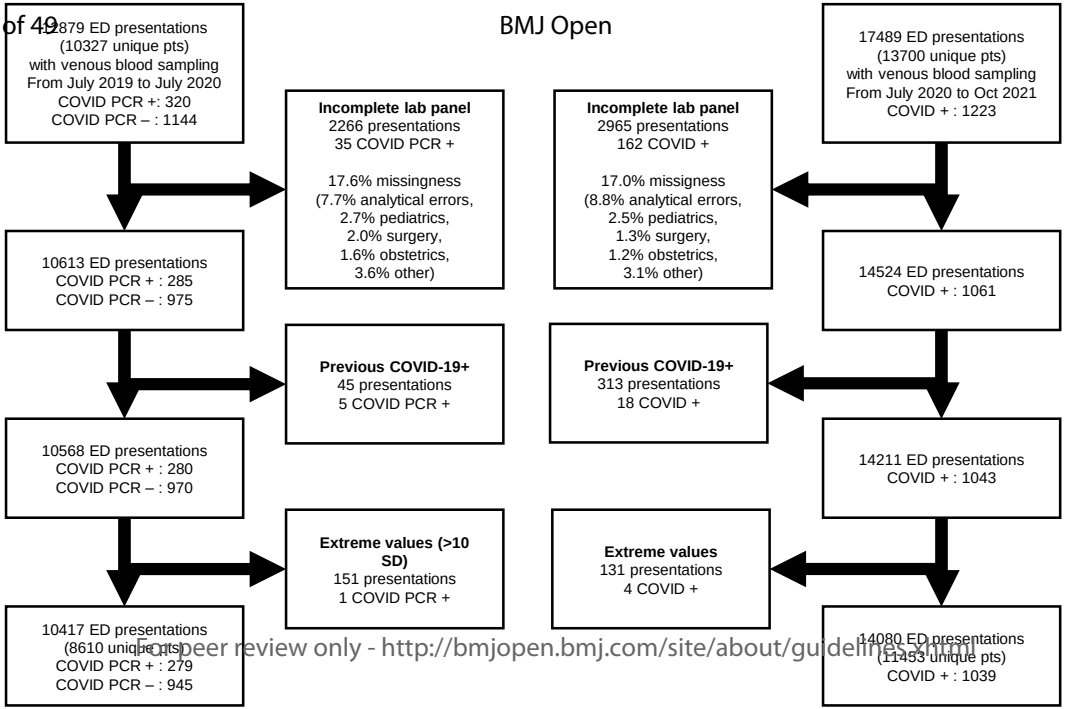
The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are plotted against the linear predictor (see table 2). The CoLab-score cut-offs (–5.83, –4.02, –3.29, –2.34 and –1.64) are depicted with vertical dashed lines. The white-boxed numbers (between the cut-offs) represent the corresponding CoLab-score. Note that while the area under both curves is identical (since these are probability density functions), in absolute numbers the "negative or untested"-group is about 36 times larger than the PCR positive group.

Figure 3: Inclusion flow of ED patients in three external centers.

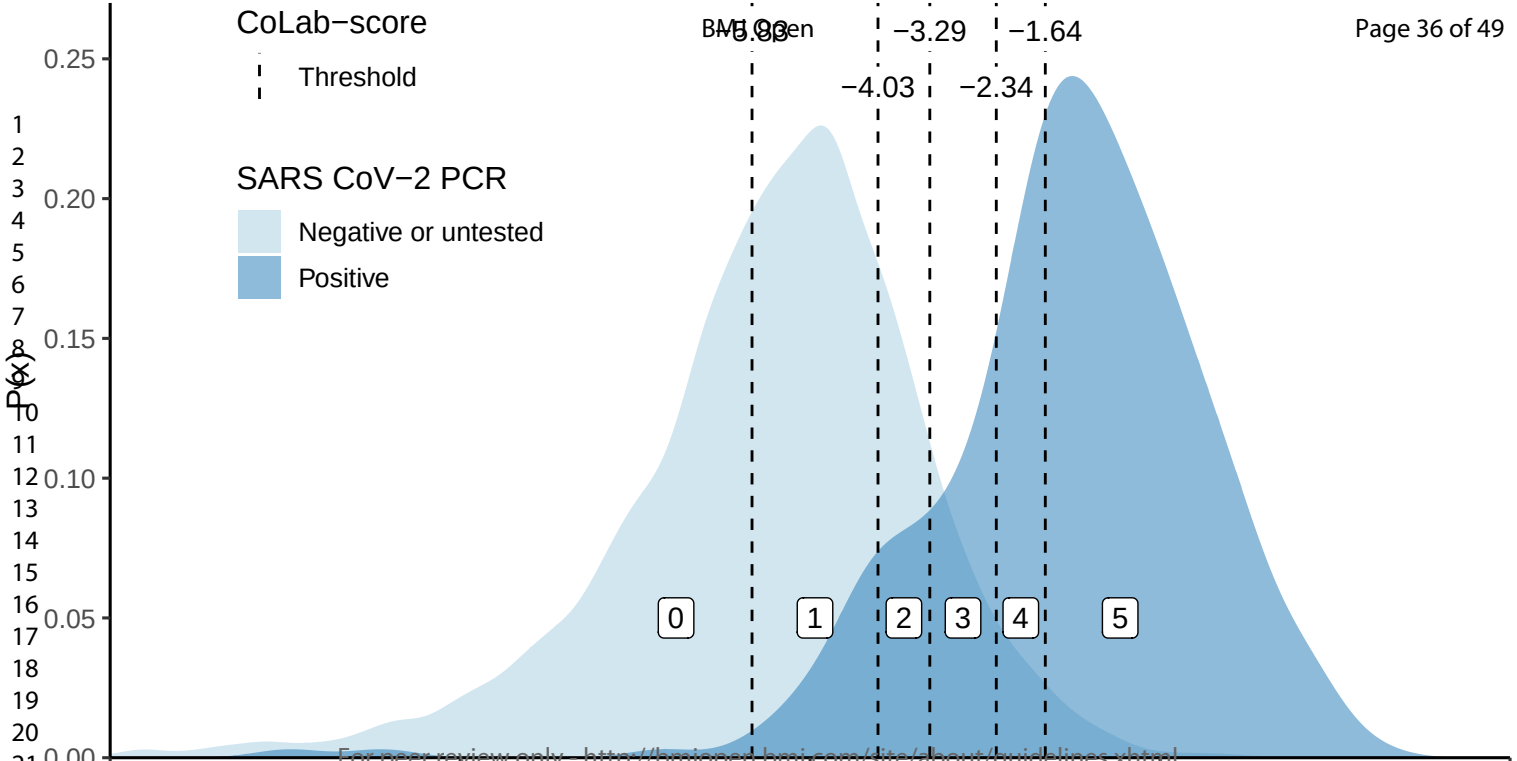
All emergency department (ED) presentations with routine venous blood sampling were included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see

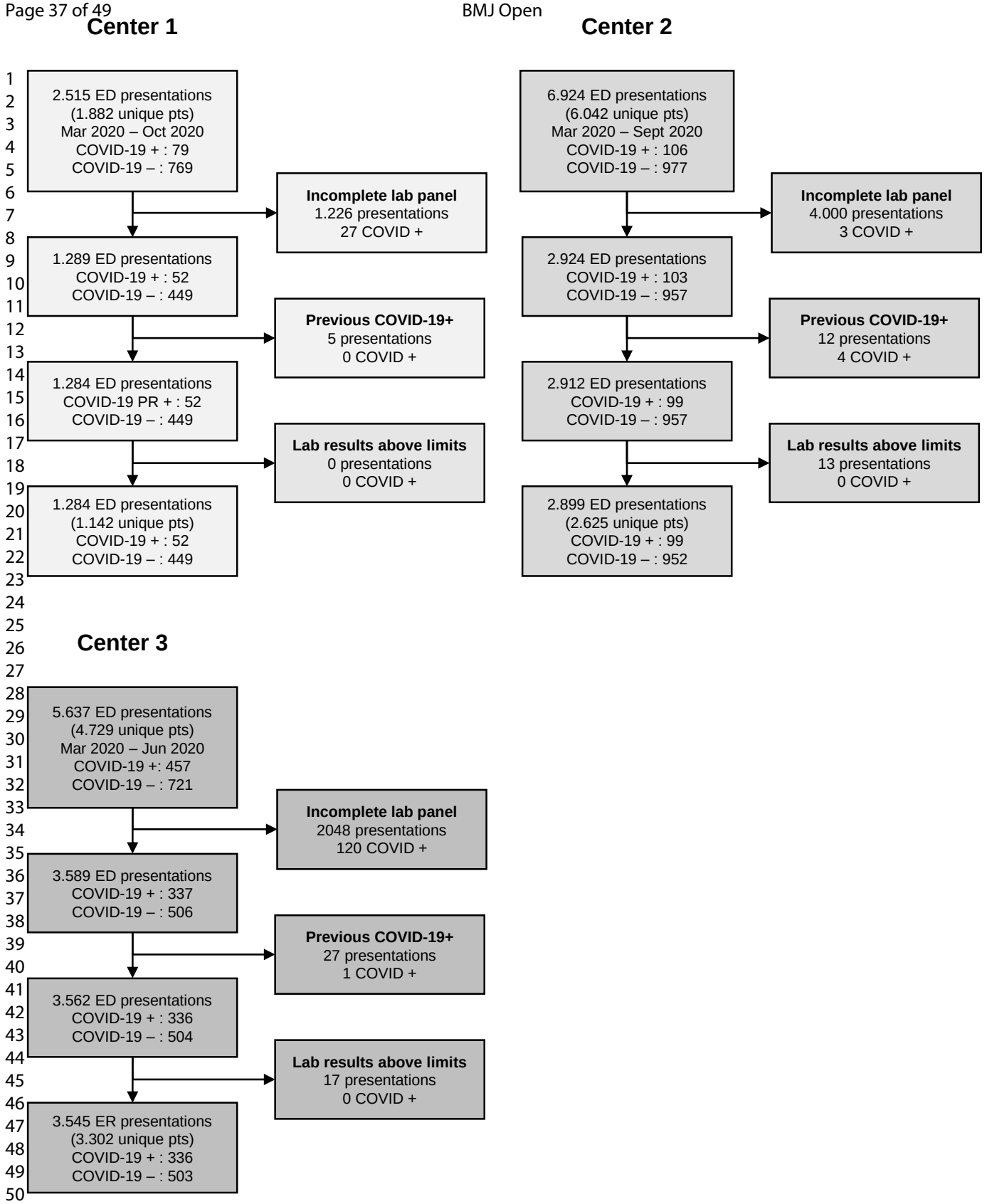
Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were excluded as "previous COVID-19+". Presentations with any laboratory result above the limits of the CoLab-score (see Table 2) were excluded.

A



B




Center 1
2.515 ED presentations
(1.882 unique pts)
Mar 2020 – Oct 2020
COVID-19 + : 79
COVID-19 – : 769
Incomplete lab panel
1.226 presentations
27 COVID +
1.289 ED presentations
COVID-19 + : 52
COVID-19 – : 449
Previous COVID-19+
5 presentations
0 COVID +
1.284 ED presentations
COVID-19 PR + : 52
COVID-19 – : 449
Lab results above limits
0 presentations
0 COVID +
1.284 ED presentations
(1.142 unique pts)
COVID-19 + : 52
COVID-19 – : 449
Center 2
6.924 ED presentations
(6.042 unique pts)
Mar 2020 – Sept 2020
COVID-19 + : 106
COVID-19 – : 977
Incomplete lab panel
4.000 presentations
3 COVID +
2.924 ED presentations
COVID-19 + : 103
COVID-19 – : 957
Previous COVID-19+
12 presentations
4 COVID +
2.912 ED presentations
COVID-19 + : 99
COVID-19 – : 957
Lab results above limits
13 presentations
0 COVID +
2.899 ED presentations
(2.625 unique pts)
COVID-19 + : 99
COVID-19 – : 952
Center 3
5.637 ED presentations
(4.729 unique pts)
Mar 2020 – Jun 2020
COVID-19 +: 457
COVID-19 – : 721
Incomplete lab panel
2048 presentations
120 COVID +
3.589 ED presentations
COVID-19 + : 337
COVID-19 – : 506
Previous COVID-19+
27 presentations
1 COVID +
3.562 ED presentations
COVID-19 + : 336
COVID-19 – : 504
Lab results above limits
17 presentations
0 COVID +
3.545 ER presentations
(3.302 unique pts)
COVID-19 + : 336
COVID-19 – : 503

Supplemental material 1

Model fitting

Prior to model fitting, covariates were scaled to zero mean and unit variance, after model fitting coefficients were unscaled to obtain regression coefficients on the original scale. In adaptive lasso, weights are applied to each of the covariates present in the lasso constraint, the weight vector has to be calculated before the adaptive lasso regression is performed. Due to multicollinearity between laboratory tests in the routine lab panel, weights in the adaptive lasso were based on ridge regression estimates ($\hat{\beta}_{ridge}$) as recommended by Zou. To obtain $\hat{\beta}_{ridge}$ the optimal penalty (λ) for the ridge regression was chosen using 10 fold cross-validation (CV) with area under the ROC curve (AUC) as the loss function. The λ corresponding to the maximum AUC was selected to obtain $\hat{\beta}_{ridge}$. The weight vector ($\widehat{w}$) was calculated by $\widehat{w} = 1/\left|\hat{\beta}_{ridge}\right|^2$. This weight vector was then used to fit an adaptive lasso regression where λ was chosen by the criterion ±1 SE of the maximum AUC.

Model intercept correction

The linear predictor for a patient *i* is calculated as follows: $lp_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$
Where *n* is the number of variables in the final model, x_{in} are the observed predictor variables for subject *i* and $\beta_n$the model coefficients. The linear predictor can then be converted to a probability for patient *i* (P_i) by the logistic function: $P_i = \frac{1}{1+e^{-lp_i}}$
The intercept term β_0 is sensitive to the fraction of cases versus controls in the dataset/population. Since the model is fitted to a case-control dataset where the number cases is fixed (all patients tested positive for COVID-19) and the number of controls is randomly chosen (a 6-month period pre-COVID), the intercept term β_0 is a result of this choice and will likely not be generalizable to the real-world setting. Prior correction is a method to correct the estimate of the intercept based on the true fraction of positives in the population, τ (prevalence of COVID-19 in the ED) and the fraction of cases in the development dataset, $\bar{y}$. The intercept term β_0 can then be corrected to obtain $\beta_{0corrected}$ using the following formula:

$$\beta_{0corrected} = \beta_0 + \beta_{adj}$$

$$\beta_{adj} = -ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right]$$

In our dataset $\bar{y} = 0.02675$ therefore:

$$\beta_{adj} = -ln\left(\frac{1-\tau}{\tau}\right) + 3.594$$

An estimate $\bar{\tau}$ can be used for the prevalence τ to obtain $\bar{\beta}_{adj}$ which can be plugged in the original linear predictor formula to obtain calibrated probabilities:

$$lp_i(\tau) = \beta_0 - ln\left(\frac{1-\tau}{\tau}\right) + 3.594 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$$

CoLab-score

An alternative, which is the basis of the CoLab-score, is to choose a fixed probability P_i above which one considers a patient eligible for further testing. The probability can be expressed as a number needed to test. If one is willing to test 10 patients to find one positive, all patients with $P_i \geq 0.1$ should be considered positive. In this study a number needed to test of 15 is used, therefore all patients with a $P_i \geq 0.067$ should be considered positive. On the linear predictor scale this translates to $\text{logit}(0.067) = -2.639$. To determine the cutoffs for difference prevalence thresholds one solves the following equation:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639$$
$$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$
$$lp_i(\tau) \geq ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

Choosing values for τ yields the cutoffs for the CoLab score:

$lp_i(\tau = 0.4) \geq -5.83$ (CoLab-score = 1)
$lp_i(\tau = 0.1) \geq -4.03$ (CoLab-score = 2)
$lp_i(\tau = 0.05) \geq -3.29$ (CoLab-score = 3)
$lp_i(\tau = 0.02) \geq -2.34$ (CoLab-score = 4)
$lp_i(\tau = 0.01) \geq -1.64$ (CoLab-score = 5)

These thresholds correspond to CoLab-scores 0 to 5. The interpretation of these scores is as follows; if the prevalence is <1%, only CoLab-score 5 should be classified as positive and CoLab-score 0 till 4 as negative. If the prevalence is 1% – 2%, CoLab-score 4 and 5 should be classified as positive and 1 – 3 negative. Similarly, with a prevalence of 2 – 5% the split is between CoLab-score 2 and 3 and with prevalence of 5 – 10% between CoLab-score 1 – 2. If the prevalence is higher than 10% only CoLab-score 0 is classified as negative. Using the CoLab-score in this fashion, aims to preserve a number need to test of 15.

Relative importance of variables

Since the variables included in the model are on different scales, the magnitude of the unscaled coefficients cannot be used to compare the importance of variables to each other. To give some indication of the importance of the variables in predicting the outcome, the unscaled coefficients obtained from the adaptive lasso regression were used to calculate the relative importance. The variable with the highest unscaled coefficient was used as maximum ($\beta_{unscaled,max}$), and all other scaled coefficients were divided by this maximum and multiplied by 100 to obtain the relative importance in %: $\frac{\beta_{unscaled}}{\beta_{unscaled,max}} \cdot 100$.

Supplemental material 2

Vaccination status and COVID-19 ED prevalence plot

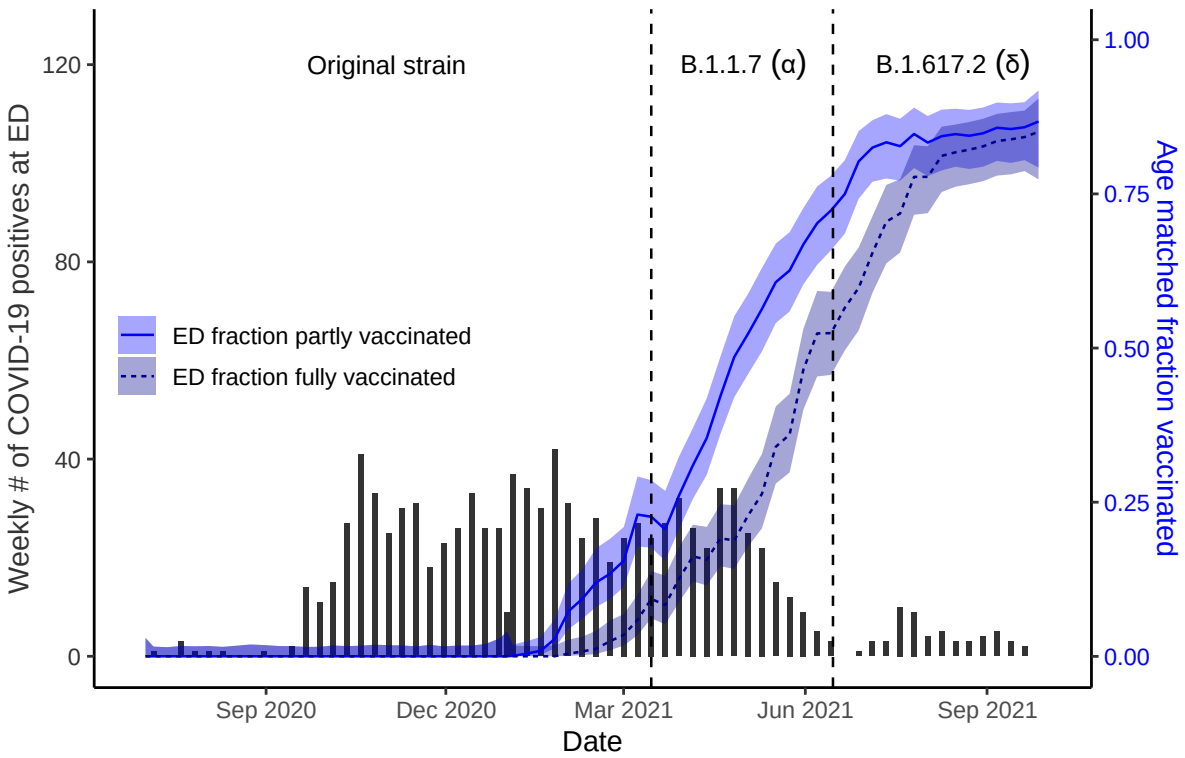


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

The temporal validation dataset consists of ED presentations from July 2020 until October 2021. As stated in the "Materials and Methods" section, this period was split into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. The ED fraction vaccinated is estimated by merging data from the Dutch national institute of public health by the date of the ED presentation and the year of birth of the patient. The gray bars depict weekly number of new COVID-19 cases at the ED, the blue lines the estimated fraction of ED patients fully or partially vaccinated.

CoLab-score performance

Phase	Cases/controls (prevalence)	AUC
Original strain & no vaccinations	694/7999 (8.6%)	0.909 (0.896 - 0.923)
B.1.1.7 strain & partial vaccination	287/2845 (10.1%)	0.937 (0.921 - 0.953)
B.1.617.2 strain & full vaccination	58/3236 (1.8%)	0.898 (0.857 - 0.939)

CoLab-score	Phase	Sensitivity	Specificity	PPV	NPV
0	Original strain & no vaccinations	0.960 (0.944 - 0.974)	0.418 (0.407 - 0.429)	0.135 (0.133 - 0.138)	0.991 (0.987 - 0.994)
	B.1.1.7 strain & partial vaccination	0.983 (0.969 - 0.997)	0.432 (0.413 - 0.450)	0.162 (0.158 - 0.168)	0.996 (0.992 - 0.999)
	B.1.617.2 strain & full vaccination	0.983 (0.948 - 1.000)	0.415 (0.396 - 0.432)	0.030 (0.028 - 0.031)	0.999 (0.998 - 1.000)
≤1	Original strain & no vaccinations	0.879 (0.854 - 0.902)	0.789 (0.779 - 0.798)	0.283 (0.273 - 0.294)	0.986 (0.983 - 0.988)
	B.1.1.7 strain & partial vaccination	0.916 (0.885 - 0.948)	0.809 (0.793 - 0.824)	0.350 (0.332 - 0.370)	0.989 (0.984 - 0.993)
	B.1.617.2 strain & full vaccination	0.862 (0.776 - 0.948)	0.780 (0.765 - 0.794)	0.067 (0.059 - 0.074)	0.997 (0.995 - 0.999)
≤2	Original strain & no vaccinations	0.813 (0.784 - 0.842)	0.894 (0.887 - 0.901)	0.421 (0.404 - 0.441)	0.980 (0.978 - 0.983)
	B.1.1.7 strain & partial vaccination	0.864 (0.826 - 0.902)	0.897 (0.885 - 0.908)	0.484 (0.455 - 0.516)	0.983 (0.979 - 0.988)
	B.1.617.2 strain & full vaccination	0.690 (0.569 - 0.810)	0.892 (0.881 - 0.902)	0.104 (0.086 - 0.123)	0.994 (0.991 - 0.996)
≤3	Original strain & no vaccinations	0.697 (0.661 - 0.731)	0.962 (0.957 - 0.966)	0.634 (0.605 - 0.662)	0.971 (0.968 - 0.974)
	B.1.1.7 strain & partial vaccination	0.760 (0.711 - 0.812)	0.963 (0.955 - 0.970)	0.696 (0.650 - 0.739)	0.973 (0.967 - 0.978)
	B.1.617.2 strain & full vaccination	0.621 (0.483 - 0.741)	0.960 (0.954 - 0.967)	0.222 (0.178 - 0.268)	0.993 (0.990 - 0.995)
≤4	Original strain & no vaccinations	0.566 (0.529 - 0.602)	0.984 (0.981 - 0.987)	0.775 (0.740 - 0.808)	0.960 (0.957 - 0.963)
	B.1.1.7 strain & partial vaccination	0.645 (0.589 - 0.704)	0.983 (0.978 - 0.988)	0.809 (0.762 - 0.856)	0.961 (0.955 - 0.967)
	B.1.617.2 strain & full vaccination	0.517 (0.397 - 0.638)	0.986 (0.982 - 0.990)	0.400 (0.319 - 0.500)	0.991 (0.989 - 0.993)

Table 2: Diagnostic performance of the CoLab-score in the temporal validation dataset, split by phase.

Sensitivities, specificities, positive predictive values (PPV) and negative predictive values (NPV) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is split into three phases according to dominant SARS-CoV-2 strains in the Netherlands and estimated fraction of ED patients vaccinated (see Figure above). Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5. The AUC was significantly higher in the second phase as compared to the first phase (DeLong test p-value: 0.0175), but did not differ significantly between the third and first phase (DeLong test p-value: 0.3903).

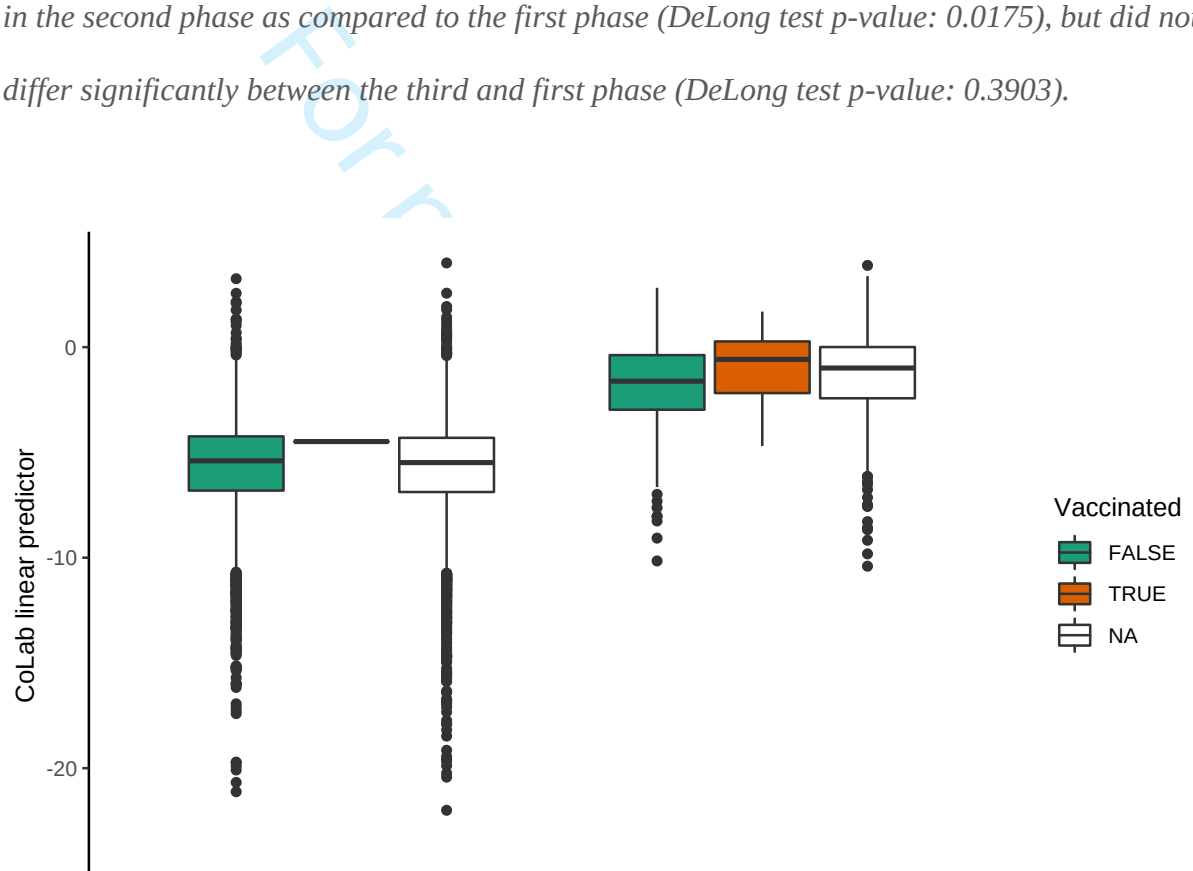


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

The CoLab linear predictor is calculated for all ED presentations in the temporal validation set. Presentations who are registered as vaccinated are labeled TRUE (N = 13). Presentations before vaccine roll-out are labeled FALSE (N = 5855). Presentations during

vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13 presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative. Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or considered positive until proven otherwise, therefore there is only one COVID-19 negative patient with a registered vaccination status.

Supplemental material 3

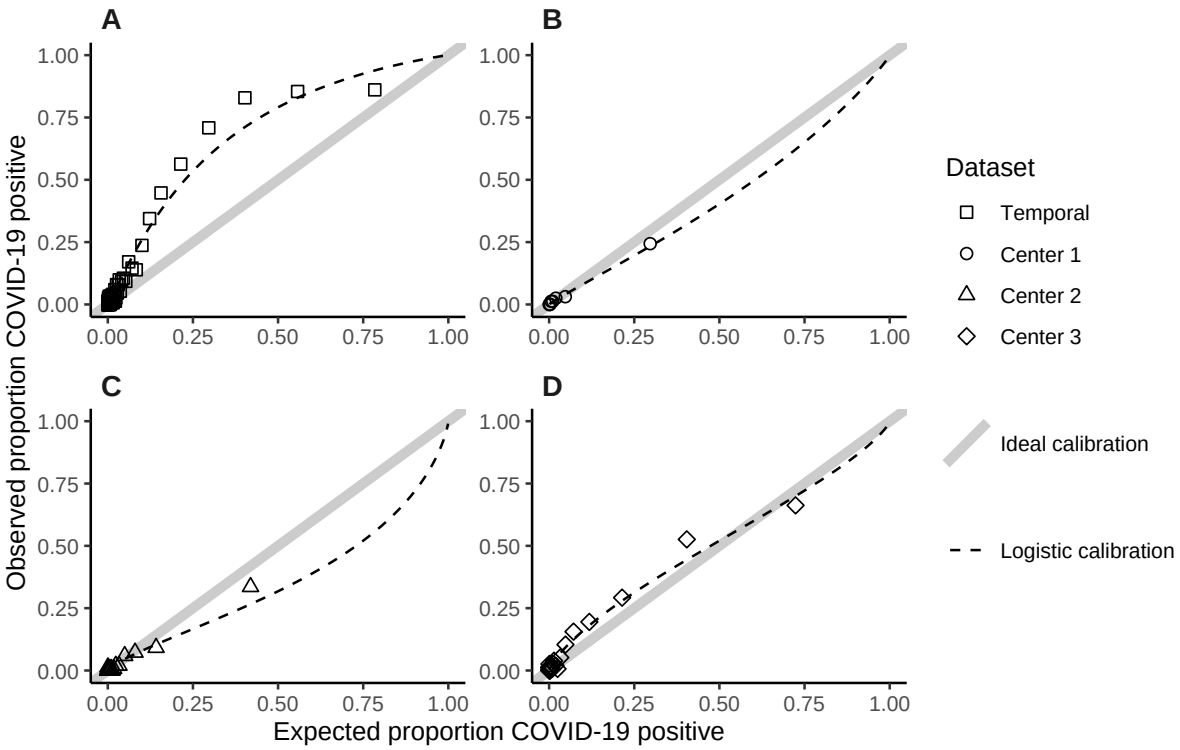


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

In the calibration plots, the proportion of observed COVID-19 positives versus expected probabilities are plotted. Observations are grouped with an average of 150 observations per group. The expected probabilities follow from applying the inverse logit function to the CoLab-linear predictor calculated from Table 2. If the observed proportion in an external dataset is lower than the expected proportion, this means risks are over-estimated, if the observed fraction is higher, risks are under-estimated. Ideally, observed proportions are equal to expected proportions, this ideal-calibration-line is shown as a straight line through the origin with a slope of 1. The logistic calibration line is a logistic regression fit of the predicted probabilities. [Intercept, slope] for plots A-D: A [1.34, 1.08], B [-0.39, 0.92], C [-0.76, 0.77], D [0.08, 0.79]. Although no validation datasets show perfect calibration, this is the result of differences in COVID-19 prevalence in the temporal validation dataset (7.4% versus 2.2%) and differences in calibration of laboratory equipment in the three external centers.

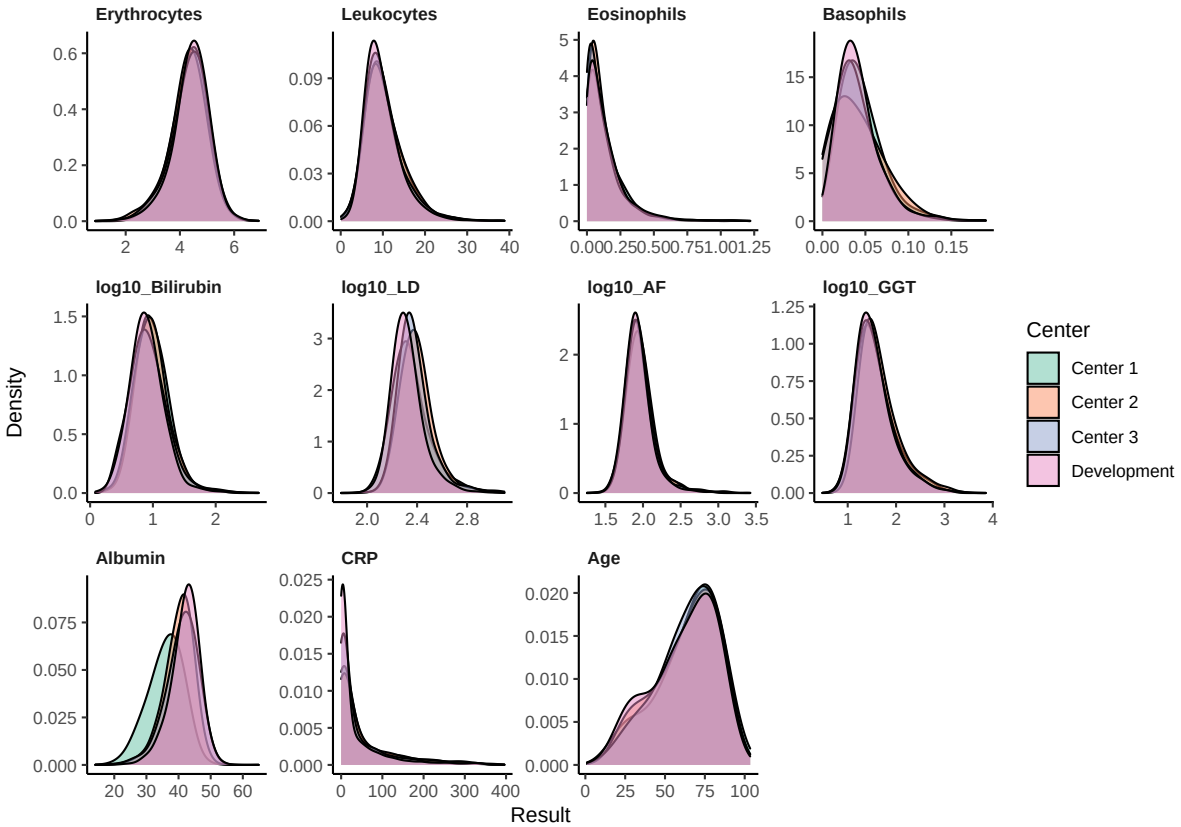


Figure 2: Probability density plots of laboratory parameters.

Probability density plots are shown for all control patients of the development dataset and the three external centers. Ideally all distributions should overlap since this implies that control patient populations are most likely similar in the development dataset to the external datasets. When comparing the distribution of the CoLab variables for all control-patients across different external validation datasets, albumin and LD show the largest deviations.

Supplemental material 4

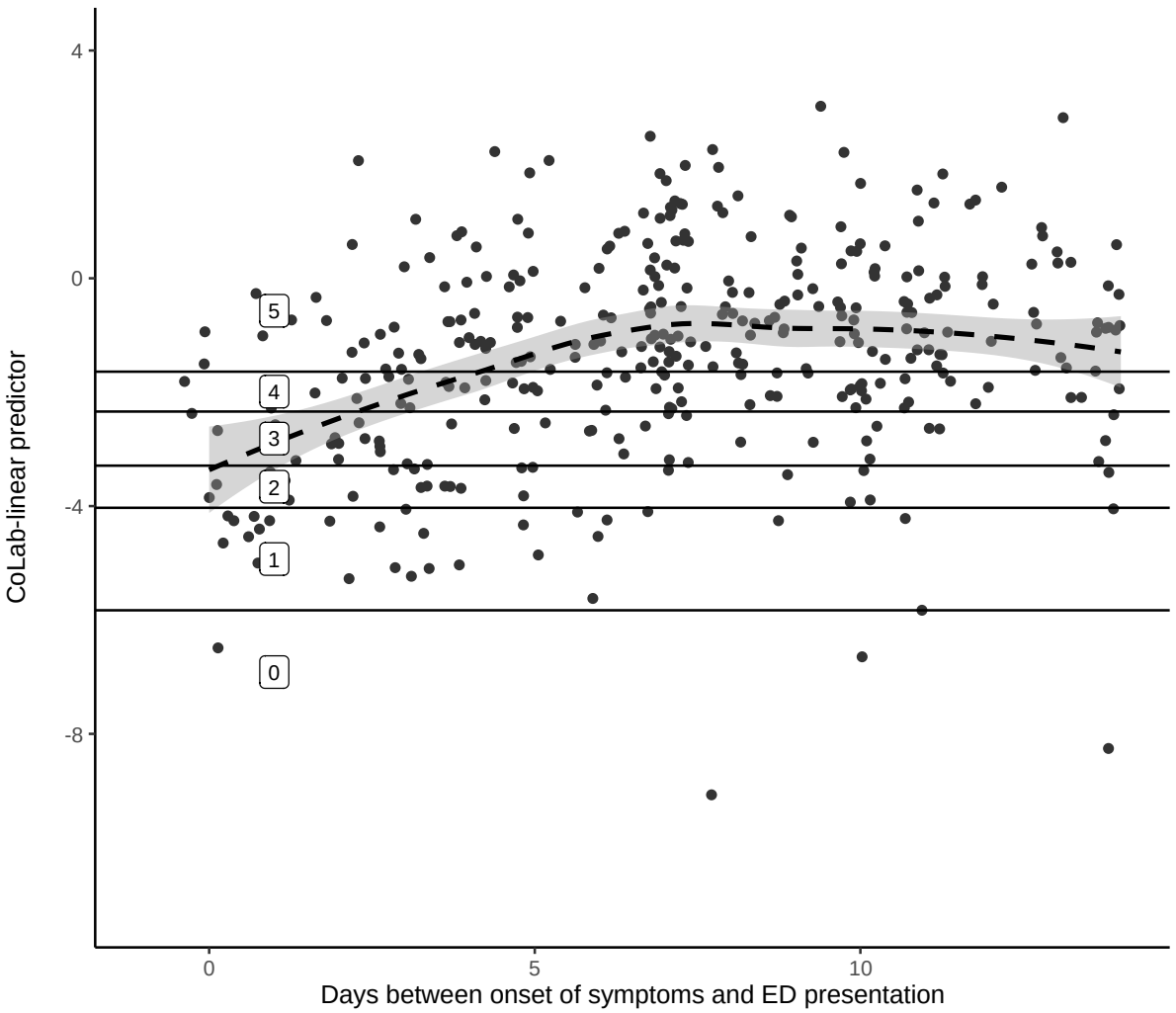


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predict is plotted against the duration of COVID-related symptoms as registered in the electronic patient records. Patients with unknown duration are not plotted. Patients without symptoms were plotted at 0 days. The solid horizontal lines represent the CoLab-score thresholds, the dashed line is a LOESS regression curve with 95% CI. As the duration of symptoms is an integer, some random jitter was added to the days, for visualization purposes. Note that only the first 14 days are shown in this graph.

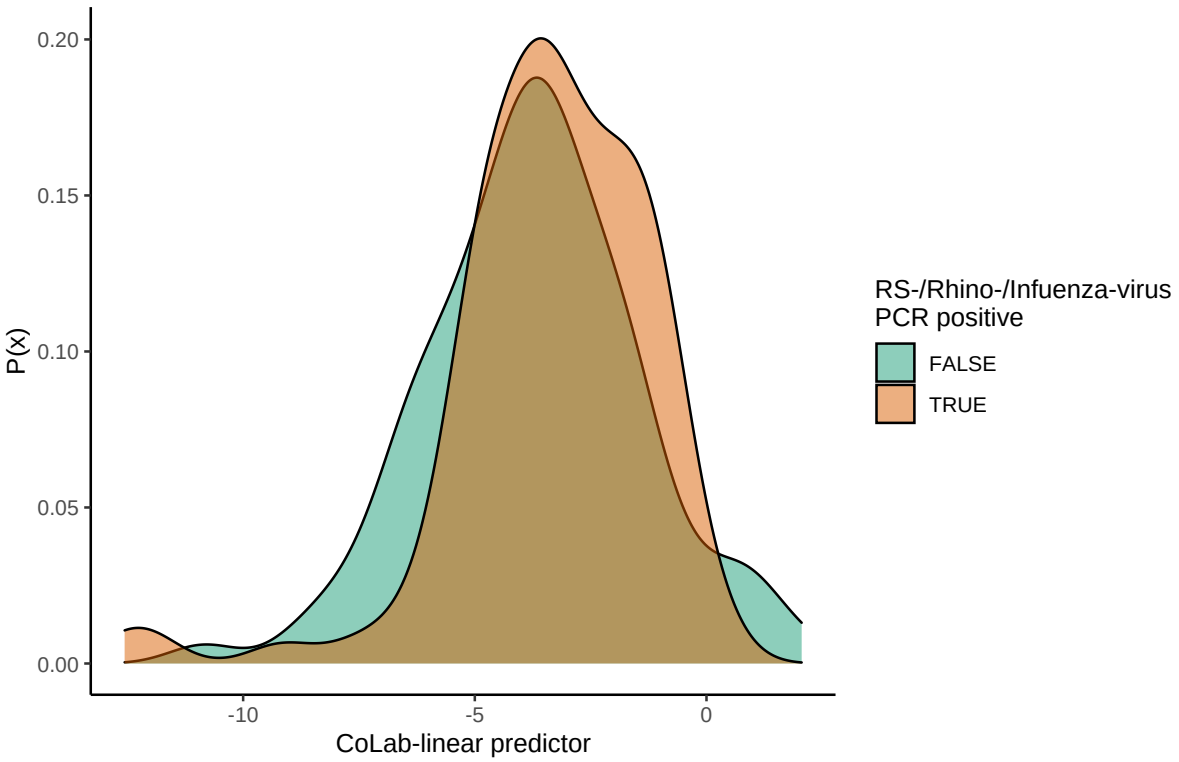


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

For 183 ED presentations that were PCR tested for either RS-, Rhino- and Influenza-virus the CoLab-score was calculated. 91 presentations were PCR positive, 92 were PCR negative. The CoLab-score is only marginally elevated for PCR positive patients, the area under the ROC-curve in separating both groups is 0.573 (95% CI: 4896-0.6563).

Inclusion criterion	Cases/controls (prevalence)	AUC
Temporal validation (reference)	1039/14080 (7.4%)	0.916 (0.906 - 0.927)
Only first presentations, re-presentations are excluded	937/11166 (8.4%)	0.919 (0.909 - 0.930)
Only PCR-tested presentations	372/4062 (9.2%)	0.840 (0.817 - 0.862)

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
0	Reference	0.967 (0.956 - 0.978)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)	1005 (993 - 1016)	5476 (5366 - 5587)	7565 (7454 - 7675)	34 (23 - 46)
	First presentations	0.968 (0.956 - 0.979)	0.416 (0.406 - 0.426)	0.132 (0.130 - 0.134)	0.993 (0.990 - 0.995)	907 (896 - 917)	4259 (4156 - 4353)	5970 (5876 - 6073)	30 (20 - 41)
	PCR-tested presentations	0.946 (0.922 - 0.968)	0.353 (0.338 - 0.368)	0.129 (0.125 - 0.132)	0.985 (0.979 - 0.991)	352 (343 - 360)	1303 (1246 - 1359)	2387 (2331 - 2444)	20 (12 - 29)
≤ 1	Reference	0.888 (0.870 - 0.908)	0.791 (0.783 - 0.798)	0.253 (0.245 - 0.261)	0.989 (0.987 - 0.991)	923 (904 - 943)	10311 (10215 - 10401)	2730 (2640 - 2826)	116 (96 - 135)
	First presentations	0.890 (0.870 - 0.908)	0.793 (0.785 - 0.801)	0.282 (0.273 - 0.292)	0.987 (0.985 - 0.990)	834 (815 - 851)	8112 (8030 - 8194)	2117 (2035 - 2199)	103 (86 - 122)
	PCR-tested presentations	0.852 (0.817 - 0.887)	0.671 (0.656 - 0.686)	0.207 (0.197 - 0.217)	0.978 (0.973 - 0.983)	317 (304 - 330)	2477 (2421 - 2533)	1213 (1157 - 1269)	55 (42 - 68)
≤2	Reference	0.820 (0.796 - 0.843)	0.894 (0.889 - 0.899)	0.382 (0.367 - 0.396)	0.984 (0.982 - 0.986)	852 (827 - 876)	11661 (11591 - 11729)	1380 (1312 - 1450)	187 (163 - 212)
	First presentations	0.824 (0.798 - 0.845)	0.898 (0.892 - 0.904)	0.426 (0.410 - 0.441)	0.982 (0.980 - 0.985)	772 (748 - 792)	9187 (9127 - 9249)	1042 (980 - 1102)	165 (145 - 189)
	PCR-tested presentations	0.734 (0.688 - 0.777)	0.800 (0.786 - 0.812)	0.270 (0.252 - 0.287)	0.968 (0.962 - 0.973)	273 (256 - 289)	2951 (2902 - 2997)	739 (693 - 788)	99 (83 - 116)
≤ 3	Reference	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.596 (0.573 - 0.618)	0.977 (0.974 - 0.979)	738 (709 - 767)	12540 (12496 - 12582)	501 (459 - 545)	301 (272 - 330)
	First presentations	0.716 (0.687 - 0.744)	0.966 (0.962 - 0.969)	0.658 (0.633 - 0.682)	0.974 (0.971 - 0.976)	671 (644 - 697)	9880 (9844 - 9915)	349 (314 - 385)	266 (240 - 293)
	PCR-tested presentations	0.591 (0.540 - 0.640)	0.911 (0.902 - 0.921)	0.403 (0.370 - 0.433)	0.957 (0.952 - 0.962)	220 (201 - 238)	3363 (3328 - 3397)	327 (293 - 362)	152 (134 - 171)
≤4	Reference	0.585 (0.556 - 0.615)	0.984 (0.982 - 0.987)	0.750 (0.724 - 0.778)	0.968 (0.965 - 0.970)	608 (578 - 639)	12838 (12811 - 12866)	203 (175 - 230)	431 (400 - 461)
	First presentations	0.590 (0.558 - 0.621)	0.987 (0.985 - 0.989)	0.805 (0.776 - 0.832)	0.963 (0.961 - 0.966)	553 (523 - 582)	10095 (10071 - 10117)	134 (112 - 158)	384 (355 - 414)
	PCR-tested presentations	0.452 (0.401 - 0.503)	0.959 (0.953 - 0.965)	0.526 (0.480 - 0.575)	0.945 (0.941 - 0.950)	168 (149 - 187)	3539 (3516 - 3562)	151 (128 - 174)	204 (185 - 223)

Table 1: Sensitivity analysis of the CoLab-score in the temporal validation dataset using different inclusion criteria.

Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is used to compare the performance of the CoLab-score with inclusion criteria that differ from the development dataset. The first line shows the performance of the temporal validation dataset with the original inclusion criteria as specified in Figure 1B. The second line shows the performance of the CoLab-score when all re-presentations are excluded (i.e. no repeated presentations). The third line shows the performance of the CoLab-score in the subgroup of patients that underwent PCR-testing.

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3, 4
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6, 7
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	8, 11-12
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	8
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	8
	5b	D;V	Describe eligibility criteria for participants.	8, 9, S1
	5c	D;V	Give details of treatments received, if relevant.	N/A
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	9
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8, 9
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	N/A
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	10
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	10-12, S1
	10c	V	For validation, describe how the predictions were calculated.	16
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	11-13
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N/A
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	22
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	F1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	T1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	S3
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	F1, F3
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	T2
	15b	D	Explain how to the use the prediction model.	T2, S1
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	T3, T4
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21-23
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	19-20
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	19-20
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	20-21
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	N/A
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	N/A

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document. S = Supplemental material, F = Figure, T = Table.

BMJ Open

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2021-059111.R3
Article Type:	Original research
Date Submitted by the Author:	09-Jun-2022
Complete List of Authors:	Boer, Arjen-Kars; Catharina Hospital Deneer, Ruben; Catharina Hospital, ; Eindhoven University of Technology, Maas, Maaike; Catharina Ziekenhuis, emergency department Ammerlaan, Heidi; Catharina Hospital van Balkom, Roland; Catharina Hospital Thijssen, Wendy; Catharina ziekenhuis, Emergency medicine Bennenbroek, Sophie; Catharina Hospital Leers, Mathie; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Martens, Remy; Zuyderland Medical Centre Heerlen, Clinical Chemistry & Hematology Buijs, Madelon; Atalmedial Kerremans, Jos; Alrijne Hospital Location Leiderdorp Messchaert, Muriël ; Gelre Hospitals van Suijlen, Jeroen; Gelre Hospitals van Riel, Natal; Eindhoven University of Technology, ; Amsterdam University Medical Centres Scharnhorst, Volkher; Catharina Hospital; Eindhoven University of Technology,
Primary Subject Heading:	Emergency medicine
Secondary Subject Heading:	Health informatics, Infectious diseases
Keywords:	COVID-19, STATISTICS & RESEARCH METHODS, ACCIDENT & EMERGENCY MEDICINE, Clinical chemistry < PATHOLOGY, Health informatics < BIOTECHNOLOGY & BIOINFORMATICS

I, the Submitting Author has the right to grant and does grant on behalf of all authors of the Work (as defined in the below author licence), an exclusive licence and/or a non-exclusive licence for contributions from authors who are: i) UK Crown employees; ii) where BMJ has agreed a CC-BY licence shall apply, and/or iii) in accordance with the terms applicable for US Federal Government officers or employees acting as part of their official duties; on a worldwide, perpetual, irrevocable, royalty-free basis to BMJ Publishing Group Ltd ("BMJ") its licensees and where the relevant Journal is co-owned by BMJ to the co-owners of the Journal, to publish the Work in this journal and any other BMJ products and to exploit all rights, as set out in our licence.

The Submitting Author accepts and understands that any supply made under these terms is made by BMJ to the Submitting Author unless you are acting as an employee on behalf of your employer or a postgraduate student of an affiliated institution which is paying any applicable article publishing charge ("APC") for Open Access articles. Where the Submitting Author wishes to make the Work available on an Open Access basis (and intends to pay the relevant APC), the terms of reuse of such Open Access shall be governed by a Creative Commons licence – details of these licences and which Creative Commons licence will apply to this Work are set out in our licence referred to above.

Other than as permitted in any relevant BMJ Author's Self Archiving Policies, I confirm this Work has not been accepted for publication elsewhere, is not being considered for publication elsewhere and does not duplicate material already published. I confirm all authors consent to publication of this Work and authorise the granting of this licence.

Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study

Arjen-Kars Boer[1*], Ruben Deneer[1,2*†], Maaike Maas[3], Heidi S. M. Ammerlaan[4], Roland H. H. van Balkom[5], Wendy A.M.H. Thijssen[3], Sophie Bennenbroek[3] , Mathie P. G. Leers[6], Remy J.H. Martens[6], Madelon M. Buijs[7], Jos J. Kerremans[8], Muriël Messchaert[9], Jeroen D.E. van Suijlen[9], Natal A.W. van Riel[2,10] and Volkher Scharnhorst[1,2]

* Both authors contributed equally

† Corresponding author

1 Department of Laboratory Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene Loper 3, 5612 AE, Eindhoven, the Netherlands

3 Department of Emergency Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

4 Department of Internal Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven, Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands

7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the Netherlands

8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital, Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands

9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334 DZ Apeldoorn, the Netherlands

10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9, 1105 AZ Amsterdam, the Netherlands.

Keywords

COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction model, routine laboratory tests

Corresponding author

Ruben Deneer

Michelangelolaan 2

5623 EJ Eindhoven, the Netherlands

ruben.deneer@catharinaziekenhuis.nl

Abstract

Objectives: Identifying patients with a possible SARS-CoV-2 infection in the emergency department (ED) is challenging. Symptoms differ, incidence rates vary and test capacity may be limited. As PCR testing all ED patients is neither feasible nor effective in most centers, a rapid, objective, low-cost early warning score to triage ED patients for a possible infection is developed.

Design: Case-control study.

Setting: Secondary and tertiary hospitals in the Netherlands.

Participants: Patients presenting at the ED with venous blood sampling from July 2019 to July 2020 (N = 10417, 279 SARS-CoV-2 positive). The temporal validation cohort covered the period from July 2020 to October 2021 (N = 14080, 1093 SARS-CoV-2 positive). The external validation cohort consisted of patients presenting at the ED of three hospitals in the Netherlands (N = 12061, 652 SARS-CoV-2 positive).

Primary outcome measures The primary outcome was one or more positive SARS-CoV-2 PCR-test results, within one day prior to, or one week after, ED presentation.

Results: The resulting "CoLab-score" consists of 10 routine laboratory measurements, and age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945). The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991, specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were confirmed in temporal and external validation.

Conclusions: The CoLab-score is based on routine laboratory measurements and is available within one hour after presentation. Depending on the prevalence, COVID-19 may be safely

ruled-out in over one third of ED presentations. Highly suspect cases can be identified regardless of presenting symptoms. The CoLab-score is continuous, in contrast to the binary outcome of lateral flow testing, and can guide PCR testing and triage ED patients.

Article summary

Strengths and limitations of this study

- A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- Using adaptive lasso regression analysis, the panel of 28 laboratory tests was reduced to a single score consisting of a subset of 10 routine ED laboratory tests and age.
- The score was temporally validated from July 2020 to October 2021, in the presence of vaccine roll-out and emergence of new SARS-CoV-2 variants.
- The score was externally validated in 3 other centers in the Netherlands.
- Missingness in the panel of laboratory tests varied between external centers, limiting generalizability of the score to the ED population for which the complete panel of laboratory tests was available.
- The score was not directly compared to lateral flow testing.

Introduction

COVID-19, caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has evolved into a global pandemic in 2020 [1]. For emergency department (ED) physicians, identifying presenting patients with a possible COVID-19 infection remains challenging since symptoms like fever, shortness of breath or coughing overlap with other illnesses [2,3]. It is crucial however, to identify a possible COVID-19 infection as early as possible. Early identification prevents further spreading and protects hospital staff by isolating a suspected patient, pending the results of a SARS-COV-2 RNA PCR test and/or chest CT. Conversely, when PCR testing or isolation treatment capacity is limited, ruling-out COVID-19 as soon as possible can save valuable resources.

In the era of electronic health records and clinical prediction models, developing an early warning score that can assist ED physicians in identifying patients presenting at the ED with COVID-19 is of great value. Moreover, if only routine ED test results are required as input, the score can be easily adopted by EDs worldwide, potentially reduce diagnostic costs and accelerate patient triage.

Many COVID-19 prediction models have already been developed, the living systematic review by Wynants et. al [4] provides an extensive overview and critical appraisal. Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early models were based on relatively small sample sizes, hampered by selection bias or were over-fitted by selecting too many features [4–6]. Aside from methodological shortcomings of early models, most models are not developed as an early warning score for all ED patients. Firstly, they require features from tests that are not routinely performed or logged for all ED patients (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED patient population. Secondly, the population on which models are commonly based, are PCR-

tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by physicians.

Only two studies were identified that focus on patients presenting at the ED, include unsuspected (and pre-pandemic) patients as controls, and rely solely on routine (laboratory) tests [9,10].

In this study we report the development and validation of an early warning score that, based on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in patients who undergo routine laboratory testing at presentation. The score can assist ED physicians in triaging patients and prevent further transmission of COVID-19 by quickly identifying possibly infected patients or ruling out a possible infection when resources are scarce.

Methods

Study design

This is a retrospective case-control study where routine laboratory test results, combined with age and gender, from all patient presenting at the emergency department (ED) of the Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-2 PCR test results in a development dataset. A model that could predict the presence of a COVID-19 infection was fit to this dataset. Performance of the model was assessed by i) internal validation, ii) temporal validation and iii) external validation by using data from the ED of three other centers. The study was reviewed by the Medical research Ethics Committees United (MEC-U) under study number W20.071, which confirmed that the Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this study. The study was thereafter reviewed and approved by the internal hospital review board.

Patient and Public Involvement

Patients were not involved in the design, conduct or reporting of this study.

Development dataset

All ED presentations at the Catharina Hospital Eindhoven from July 2019 to July 2020 were included in the development dataset, provided that routine laboratory testing had been requested by the attending ED physician. The rationale for this inclusion period is to limit the effect of seasonal variation in the ED patient population by including the summer, fall and winter season of 2019 (control patients) and the winter, spring and summer season of 2020 (case and control patients). The routine laboratory panel at the ED consists of 28 laboratory tests. In some cases not all tests in the routine panel were requested or one or more

quantitative results were not available due to analytical interference (hemolysis, lipemia or icterus). The routine ED laboratory panel is requested for (adult) patients presenting with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints, or for patients (including non-adult patients) presenting with specific complaints where a suspected diagnosis has to be ruled-in or ruled-out. Presentations with one or more missing values in any of the 28 laboratory test in the routine ED panel, were excluded. Presentations with one or more extreme lab results, > 10 times standard deviation from the median, were also excluded to minimize the effect on the estimation of regression coefficients. The median was chosen as a measure of central tendency due to its resistance for outliers. After the first case of COVID-19 in the Netherlands, all patients with symptoms of COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were approved by the Dutch national institute of public health (RIVM). If a patient had a positive PCR result in the past, subsequent presentations were excluded as re-presentations might be clinically different from de novo presentations.

The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were performed in the time window and the presentation occurred after the first case of COVID-19 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the first case were labelled as "*Pre-COVID-19*".

Laboratory tests

The routine laboratory panel consisted of hemocytometric and chemical analyses. The hemocytometric tests were performed on Sysmex XN-10 instruments (Sysmex Corp., Kobe, Japan) and consisted of hemoglobin, hematocrit, erythrocytes, mean corpuscular volume (MCV), mean cellular hemoglobin (MCH), mean cellular hemoglobin concentration (MCHC), thrombocytes, leukocytes, neutrophils, eosinophils, basophils, lymphocytes and monocytes. The chemical analyses were performed on a Cobas 8000 Pro (Roche Dx, Basel, Switzerland) instrument and consisted of glucose, total bilirubin, aspartate aminotransferase (ASAT), alanine aminotransferase (ALAT), lactate dehydrogenase (LD), creatine kinase (CK), alkaline phosphatase (ALP), gamma-glutamyltransferase (gGT), blood urea nitrogen (BUN), creatinine, CKD-epi estimated glomerular filtration rate (eGFR), potassium, sodium, chloride, albumin (bromocresol green) and C-reactive protein (CRP). These results were combined with age and gender.

Modelling

All data were processed and analyzed in R version 4.1.1 [11]. Laboratory results, combined with age and gender were used as covariates in a regression model. Cases were defined as ED presentations labelled as "*PCR-positive*", controls were all other presentations (i.e. "*PCR-negative*", "*Untested*" or "*Pre-COVID-19*"). To achieve predictive accuracy, limit overfitting and perform feature selection, penalized logistic regression with an adaptive lasso penalty was chosen [12,13]. To minimize missing data, all non-numeric results at the extremes of the measuring range, were converted to numeric results by removing the "<" and ">" signs. For eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m2 and <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD, CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model

predictions, these variables were transformed logarithmically. More details regarding model fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the glmnet-package [14].

CoLab-score

Since this is a retrospective case-control study, the sample prevalence may not reflect the true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term in the model should be adjusted according to the current prevalence (details can be found in the document, **Supplemental Material 1**) [15]. However, adjusting the intercept term is not straightforward to implement in clinical practice, therefore the linear predictor of the model was categorized into a score, this score is hereafter referred to as the "CoLab-score". The categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using the intercept adjustment formula by King [15]. The intervals obtained through these breaks correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and score 5 reflects high-risk. More details regarding the rationale of the CoLab-score categorization can be found in the document, **Supplemental Material 1**.

Internal validation

To assess model performance while taking overfitting into account, bootstrapping was performed. 1000 bootstrap samples were generated from the original data. On each bootstrap sample, the full model fitting procedure and CoLab-score conversion were performed. Optimism adjusted performance measures of the CoLab-score were obtained by applying the 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [16]. Performance

measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate performance measures [17]. Although the full inclusion period from July 2019 to July 2020 was used for model fitting, the performance was evaluated on the period starting from the first COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain performance measures that would reflect real world performance.

Temporal validation

For temporal validation, results from our center were prospectively analyzed from July 2020 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period there was also more widespread external PCR testing by municipal health services. The results of external conducted PCR tests were not available to our study. To overcome this limitation, the outcome in the temporal validation cohort was chosen as a composite of the hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test result. This period also covers both the emergence of new SARS-CoV-2 variants as well as vaccine rollout. However, neither vaccination status nor genomic sequencing was available to determine whether a patient was vaccinated or which variant caused the infection. Therefore, data from the Dutch national institute of public health (RIVM) was used, to divide the temporal validation period into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold

for the entire period, as well as for each phase separately to determine a possible effect of vaccination and new variants on performance (results in the **Supplemental Material 2**). Model calibration was assessed graphically using the rms-package [18].

External validation

For the external validation, several centers in the Netherlands were approached and assessed if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available. Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and Zuyderland Medical Center (center 3). The hematological parameters were measured with Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external validation was similar to the temporal validation and consisted of assessing the AUC sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was assessed graphically analogous to the temporal validation dataset.

Results

Development dataset

12879 emergency department (ED) presentations of 10327 patients from July 2019 to July 2020 were included. After excluding cases with an incomplete lab panel, patient presentations that occurred after a positive PCR test in the past (re-presentations) and presentations with extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of 8610 patients remained (**Figure 1 A**).

	Pre-COVID N = 5890	**Untested N = 3303**	**PCR negative N = 945**	**PCR positive N = 279**
Age in years	61 (21)	60 (21)	66 (18)	69 (15)
Female gender	2909 (49.4 %)	1659 (50.2 %)	466 (49.3 %)	95 (34.1 %)
Specialism				
Internal medicine	1648 (28.0 %)	896 (27.1 %)	244 (25.8 %)	71 (25.4 %)
Surgery	1007 (17.1 %)	679 (20.6 %)	51 (5.4 %)	5 (1.8 %)
Neurology	775 (13.2 %)	468 (14.2 %)	64 (6.8 %)	5 (1.8 %)
Pulmonary medicine	714 (12.1 %)	220 (6.7 %)	326 (34.5 %)	167 (59.9 %)
Cardiology	560 (9.5 %)	322 (9.7 %)	145 (15.3 %)	6 (2.2 %)
Urology	309 (5.2 %)	148 (4.5 %)	15 (1.6 %)	7 (2.5 %)
Gastroenterology	306 (5.2 %)	224 (6.8 %)	27 (2.9 %)	1 (0.4 %)
Geriatrics	189 (3.2 %)	95 (2.9 %)	52 (5.5 %)	15 (5.4 %)
Orthopedics	147 (2.5 %)	109 (3.3 %)	11 (1.2 %)	0 (0.0 %)
Gynecology	118 (2.0 %)	82 (2.5 %)	2 (0.2 %)	0 (0.0 %)
Other	117 (2.0 %)	60 (1.8 %)	8 (0.8 %)	2 (0.7 %)
Hemoglobin in mmol/L	8.2 (1.3)	8.3 (1.3)	8.2 (1.4)	8.6 (1.1)
Hematocrit in L/L	0.403 (0.059)	0.405 (0.056)	0.405 (0.062)	0.417 (0.047)
Erythrocytes in /pL	4.41 (0.69)	4.43 (0.66)	4.41 (0.72)	4.61 (0.60)
MCV in fl	91.8 (6.4)	91.9 (6.1)	92.4 (6.7)	90.7 (5.5)
MCH in mmol	1.859 (0.157)	1.876 (0.150)	1.874 (0.172)	1.869 (0.141)
MCHC in mmol/L	20.2 (0.9)	20.4 (0.9)	20.3 (1.0)	20.6 (0.8)
Thrombocytes in /nL	263 (99)	266 (100)	269 (105)	217 (123)
Leukocytes in /nL	9.30 [7.06, 12.16]	8.92 [7.01, 11.89]	9.66 [7.17, 12.94]	6.33 [4.74, 8.48]
Neutrophils in /nL	6.62 [4.51, 9.53]	6.10 [4.42, 8.94]	7.01 [4.79, 10.02]	4.71 [3.30, 6.94]
Eosinophils in /nL	0.09 [0.03, 0.17]	0.09 [0.03, 0.18]	0.08 [0.02, 0.17]	0.00 [0.00, 0.02]
Basophils in /nL	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.04 [0.02, 0.05]	0.01 [0.01, 0.02]
Lymphocytes in /nL	1.47 [0.93, 2.13]	1.56 [1.05, 2.18]	1.31 [0.80, 2.03]	0.86 [0.59, 1.21]
Monocytes in /nL	0.70 [0.52, 0.93]	0.69 [0.52, 0.91]	0.74 [0.54, 1.01]	0.45 [0.32, 0.64]
Glucose in mmol/L	6.76 [5.83, 8.39]	6.68 [5.76, 8.14]	6.98 [5.95, 8.85]	6.77 [5.98, 8.48]
Bilirubin in umol/L	7.5 [5.0, 11.6]	7.4 [5.1, 10.9]	8.3 [5.6, 12.4]	8.2 [6.3, 11.4]
ASAT in U/L	24.0 [19.1, 32.2]	26.5 [21.6, 35.1]	27.7 [21.7, 39.2]	40.7 [30.2, 57.2]
ALAT in U/L	24.3 [17.8, 35.3]	25.3 [18.4, 36.2]	25.7 [18.4, 40.0]	33.7 [23.3, 50.0]
LD in U/L	201 [173, 240]	198 [170, 236]	215 [178, 263]	300 [238, 403]
CK in U/L	82 [51, 134]	83 [52, 136]	76 [51, 125]	124 [62, 222]
ALP in IU/L	83.0 [68.0, 105.0]	81.0 [65.8, 102.5]	86.9 [67.9, 110.0]	71.0 [58.8, 85.0]
gGT in U/L	27.0 [17.0, 53.0]	28.4 [18.4, 50.5]	37.0 [22.4, 68.9]	42.0 [28.0, 83.5]
BUN in mmol/L	5.7 [4.3, 8.0]	5.8 [4.3, 7.8]	6.2 [4.6, 9.4]	6.1 [4.7, 8.9]

CKD-epi in ml/min/m2	80.9 [58.0, 99.1]	85.0 [63.5, 103.3]	79.1 [52.1, 96.6]	76.6 [54.9, 91.2]
Potassium in mmol/L	4.06 (0.50)	4.03 (0.49)	4.07 (0.55)	3.91 (0.47)
Sodium in mmol/L	139.2 (4.0)	138.5 (3.9)	138.0 (4.3)	136.4 (4.1)
Chloride in mmol/L	104.4 (4.6)	103.8 (4.5)	102.9 (4.8)	101.6 (4.4)
Albumin in g/L	42.4 (4.9)	42.3 (4.5)	40.8 (4.8)	38.4 (3.8)
CRP in mg/L	8 [2, 41]	5 [1, 30]	18 [3, 69]	77 [37, 136]

Table 1: Descriptive statistics of development dataset and laboratory concentrations.

Shown are the laboratory tests routinely requested at ED presentation and their mean/median results (in the development dataset) for the presentations before the first COVID-19 patient in the Netherlands ("Pre-COVID-19"), presentations thereafter that were not tested for COVID-19 ("Untested"), tested negatively ("PCR negative") and tested positive ("PCR positive"). For results with normal distributions, the mean value and standard deviation (in round brackets) are shown. For results that have skewed or heavy tailed distributions, the median value and the interquartile range is shown [in squared brackets]. Dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error).

Descriptive statistics of ED presentations are shown in **Table 1,** dark grey marked figures indicate a clinically relevant difference from the Pre-COVID-19 category (based on the total allowable error [19]). For the PCR positives (N = 279), 91% (95% CI: 88 to 94%) of the cases were tested positive in their first PCR. The remaining 24 patients were positive in their second (N = 18), third (N = 5) or fourth (N = 1) PCR.

CoLab-score

The model obtained through adaptive lasso regression contained eleven variables, which are depicted with their regression coefficients (weights) in **Table 2**.

Variable	β	Exclusion limit	Relative importance
Intercept	-6.885		-
Erythrocytes /pL	0.9379	Erythrocytes < 2.9 /pL	52 %
Leukocytes /nL	-0.1298		46 %
Eosinophils /nL	-6.834		86 %
Basophils /nL	-47.70	Basophils >0.33 /nL	100 %
$\log_{10}$ of Bilirubin in µmol/L	-1.142	Bilirubin >169 µmol/L	26 %
$\log_{10}$ of LD in U/L	5.369	LD >1564 U/L	58 %
$\log_{10}$ of ALP in IU/L	-3.114	AF >1000 IU/L	45 %
$\log_{10}$ of gGT in U/L	0.3605	gGT >1611 U/L	11 %
Albumin in g/L	-0.1156		45 %
CRP in mg/L	0.002560		15 %
Age in years	0.002275		4 %

Table 2: Calculation of the CoLab-linear predictor (LP).

The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of the 11 variables with their corresponding coefficients (β's). CoLab-LP = – 6.885 + [erythrocytes] × 0.9379 – [leukocytes] × 0.1298 – [eosinophils] × 6.834 – [basophils] × 47.7 – log10([bilirubin]) × 1.142 + log10([LD]) × 5.369 – log10([ALP]) × 3.114 + log10([gGT]) × 0.3605 – [albumin] × 0.1156 + [CRP] × 0.02560 + [age] × 0.002275. The LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid if any of the variables exceed the limits in the third column. The relative importance ranks the importance of variables in predicting the outcome, relative to the most important variable (in this case basophils).

A larger β-coefficient does not imply that a variable is more important in predicting the odds of testing positive for SARS-CoV-2, since variables are on different scales. The most important variables are basophiles, eosinophils and lactate dehydrogenase (LD).

As shown in **Figure 2**, the linear predictor clearly discriminates between COVID-19 and non-COVID-19. The linear predictor is converted to CoLab-scores 0 – 5 with the cut-points depicted in **Figure 2**.

Internal validation

The model was validated in the period starting from the first COVID-19 infection to July 2020, in this period the mean prevalence was 7.2%. The AUC of the CoLab-score is 0.930 (95% CI: 0.909 to 0.945).

CoLab-score	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN	% of population
0	0.984 (0.969 - 0.991)	0.410 (0.302 - 0.543)	0.115 (0.094 - 0.147)	0.997 (0.993 - 0.999)	273.4 (241.2 - 304.4)	1470.9 (1081.1 - 1950.9)	2119.1 (1633.5 - 2507.6)	4.6 (2.6 - 8.6)	38.0 (28.0 - 51.0)
≤ 1	0.912 (0.892 - 0.952)	0.785 (0.741 - 0.827)	0.248 (0.207 - 0.300)	0.991 (0.989 - 0.995)	253.5 (226.5 - 287.0)	2817.1 (2655.4 - 2961.2)	772.9 (623.2 - 934.5)	24.5 (13.4 - 30.2)	73.3 (69.3 - 77.3)
≤ 2	0.856 (0.816 - 0.895)	0.880 (0.864 - 0.900)	0.357 (0.315 - 0.415)	0.988 (0.984 - 0.991)	238.1 (209.6 - 267.9)	3160.8 (3100.7 - 3233.7)	429.1 (357.3 - 487.1)	39.9 (28.5 - 52.4)	82.9 (80.9 - 83.9)
≤ 3	0.757 (0.706 - 0.809)	0.951 (0.944 - 0.959)	0.546 (0.496 - 0.604)	0.981 (0.976 - 0.985)	210.4 (183.4 - 240.2)	3415.1 (3378.0 - 3456.4)	174.9 (147.0 - 199.3)	67.6 (51.9 - 84.9)	90.0 (89.0 - 91.0)
≤ 4	0.612 (0.530 - 0.706)	0.978 (0.972 - 0.983)	0.683 (0.628 - 0.746)	0.970 (0.963 - 0.978)	170.2 (141.6 - 204.9)	3510.6 (3476.8 - 3547.5)	79.4 (60.3 - 100.4)	107.9 (79.1 - 134.0)	93.7 (91.7 - 93.7)

Table 3: Bootstrapped diagnostic performance of the CoLab-score in the development dataset.

The development dataset was internally validated for the period March 2020 – July 2020 (N = 3868). The optimism-adjusted bootstrapped sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) and fraction of presentations (%) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4). The numbers in round brackets represent the 95% optimism-

adjusted bootstrapped confidence intervals. The first column defines the threshold above which CoLab-score a patient is considered positive. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5. Also note that TP, TN, FP and FN are not whole numbers, as these are obtained through bootstrapping and each bootstrap replicate contains a different number of controls and cases.

Diagnostic performance is shown in **Table 3.** A CoLab-score of 0 has a negative predictive value (NPV) of 0.997 (95% CI: 0.993 to 0.999) and positive predictive value (PPV) of 0.115 (0.0934 - 0.147), one third (38%, 95% CI: 28 to 514%) of all ED presentations were assigned this score and can therefore be safely excluded. Conversely, 6% (95% CI: 6 to 8%) of the ED patients had a CoLab-score = 5. Given the PPV of this score (0.683, 95% CI: 0.628 to 0.746, NPV: 0.970, 95% CI: 0.963 - 0.978), subsequent PCR testing is advised.

Temporal validation

As the CoLab-score was developed in our center after the first COVID-19-wave in the Netherlands, the performance was evaluated in our center from July 2020 until October 2021. Lab results from 17489 ED presentations were collected. After applying the inclusion flow as shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a COVID-19 infection.

The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the development cohort, although sensitivity is slightly lower and specificity slightly higher (cf. **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental**

Material 2 Figure 1). The discriminative ability was not lower in the second or third phase, compared to the first phase. Diagnostic performance is preserved in terms of sensitivity and specificity, except a moderately reduced sensitivity of scores ≥ 3 in the third phase as compared to the first phase. PPV and NPV are incomparable due to different prevalence/pre-test probabilities in each phase (see **Supplemental Material 2 Table 1**).

In terms of the predicted probabilities, model calibration shows that overall predicted probabilities are too low (see **Supplemental Material 3 Figure 1** for the calibration plot), which is expected since the prevalence differs and the intercept has to be adjusted to the prevalence.

In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that initially did not present with COVID-specific symptoms. Most patients had neurological or orthopedic presenting symptoms.

External validation

For external validation, data obtained from three other centers were used, center 1 (N = 1284, 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545, 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers is ≥ 0.96. Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for external centers are shown in **Supplemental Material 3 Figure 1**, the observed fraction of

369 COVID-19 positives is slightly lower than expected in centers 1 and 2. For center 3, low

370 probabilities appear slightly underestimated and high probabilities slightly overestimated.

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
0	Temporal	0.967 (0.956 - 0.978)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)	1005 (993 - 1016)	5476 (5366 - 5587)	7565 (7454 - 7675)	34 (23 - 46)
	Center 1	1.000 (1.000 - 1.000)	0.331 (0.307 - 0.358)	0.059 (0.057 - 0.061)	1.000 (1.000 - 1.000)	52 (52 - 52)	410 (380 - 443)	827 (794 - 857)	0 (0 - 0)
	Center 2	0.961 (0.922 - 0.990)	0.351 (0.333 - 0.369)	0.052 (0.049 - 0.054)	0.996 (0.992 - 0.999)	99 (95 - 102)	985 (935 - 1035)	1823 (1773 - 1873)	4 (1 - 8)
	Center 3	0.970 (0.950 - 0.988)	0.322 (0.306 - 0.338)	0.130 (0.126 - 0.133)	0.991 (0.984 - 0.996)	327 (320 - 333)	1042 (991 - 1092)	2193 (2143 - 2244)	10 (4 - 17)
≤ 1	Temporal	0.888 (0.870 - 0.908)	0.791 (0.783 - 0.798)	0.253 (0.245 - 0.261)	0.989 (0.987 - 0.991)	923 (904 - 943)	10311 (10215 - 10401)	2730 (2640 - 2826)	116 (96 - 135)
	Center 1	0.923 (0.846 - 0.981)	0.694 (0.669 - 0.720)	0.113 (0.101 - 0.124)	0.995 (0.991 - 0.999)	48 (44 - 51)	858 (828 - 891)	379 (346 - 409)	4 (1 - 8)
	Center 2	0.913 (0.854 - 0.961)	0.678 (0.661 - 0.696)	0.094 (0.087 - 0.101)	0.995 (0.992 - 0.998)	94 (88 - 99)	1905 (1857 - 1953)	903 (855 - 951)	9 (4 - 15)
	Center 3	0.914 (0.881 - 0.944)	0.674 (0.657 - 0.691)	0.226 (0.216 - 0.236)	0.987 (0.982 - 0.991)	308 (297 - 318)	2180 (2126 - 2234)	1055 (1001 - 1109)	29 (19 - 40)
≤ 2	Temporal	0.820 (0.796 - 0.843)	0.894 (0.889 - 0.899)	0.382 (0.367 - 0.396)	0.984 (0.982 - 0.986)	852 (827 - 876)	11661 (11591 - 11729)	1380 (1312 - 1450)	187 (163 - 212)
	Center 1	0.808 (0.692 - 0.904)	0.811 (0.788 - 0.832)	0.152 (0.129 - 0.176)	0.990 (0.984 - 0.995)	42 (36 - 47)	1003 (975 - 1029)	234 (208 - 262)	10 (5 - 16)
	Center 2	0.845 (0.777 - 0.913)	0.801 (0.785 - 0.815)	0.135 (0.122 - 0.147)	0.993 (0.990 - 0.996)	87 (80 - 94)	2248 (2205 - 2289)	560 (519 - 603)	16 (9 - 23)
	Center 3	0.890 (0.855 - 0.923)	0.794 (0.779 - 0.808)	0.311 (0.294 - 0.328)	0.986 (0.981 - 0.990)	300 (288 - 311)	2569 (2521 - 2615)	666 (620 - 714)	37 (26 - 49)
≤ 3	Temporal	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.596 (0.573 - 0.618)	0.977 (0.974 - 0.979)	738 (709 - 767)	12540 (12496 - 12582)	501 (459 - 545)	301 (272 - 330)
	Center 1	0.750 (0.635 - 0.865)	0.909 (0.892 - 0.925)	0.257 (0.213 - 0.306)	0.989 (0.983 - 0.994)	39 (33 - 45)	1124 (1104 - 1144)	113 (93 - 133)	13 (7 - 19)
	Center 2	0.660 (0.563 - 0.748)	0.897 (0.885 - 0.908)	0.190 (0.163 - 0.218)	0.986 (0.983 - 0.990)	68 (58 - 77)	2519 (2486 - 2549)	289 (259 - 322)	35 (26 - 45)

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
≤ 4	Center 3	0.766 (0.718 - 0.810)	0.887 (0.876 - 0.898)	0.413 (0.386 - 0.442)	0.973 (0.968 - 0.978)	258 (242 - 273)	2869 (2835 - 2905)	366 (330 - 400)	79 (64 - 95)
	Temporal	0.585 (0.556 - 0.615)	0.984 (0.982 - 0.987)	0.750 (0.724 - 0.778)	0.968 (0.965 - 0.970)	608 (578 - 639)	12838 (12811 - 12866)	203 (175 - 230)	431 (400 - 461)
	Center 1	0.654 (0.519 - 0.788)	0.951 (0.939 - 0.962)	0.359 (0.293 - 0.435)	0.985 (0.979 - 0.991)	34 (27 - 41)	1176 (1161 - 1190)	61 (47 - 76)	18 (11 - 25)
	Center 2	0.534 (0.437 - 0.621)	0.952 (0.943 - 0.959)	0.287 (0.239 - 0.339)	0.982 (0.979 - 0.986)	55 (45 - 64)	2672 (2649 - 2693)	136 (115 - 159)	48 (39 - 58)
	Center 3	0.665 (0.611 - 0.718)	0.930 (0.921 - 0.938)	0.497 (0.462 - 0.534)	0.964 (0.958 - 0.969)	224 (206 - 242)	3008 (2980 - 3036)	227 (199 - 255)	113 (95 - 131)

Table 4: Diagnostic performance of the CoLab-score in the validation dataset (temporal) and three external hospitals.

Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5.

Discussion

Given the impact of COVID-19 on society and healthcare, there is a need for simple and fast detection of patients with a possible COVID-19 infection in the ED. The CoLab-score described in this study, is a fast and accurate risk score to triage patients presenting at the ED based on ten routine blood biomarkers and age.

The main strength of this study is that this score can be used as an early-warning or triaging tool for the ED population presenting with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints where a routine blood panel is requested. This is in contrast to the vast majority of COVID-19 diagnostic models that have been developed on a pre-selected population of PCR-tested patients [9,20–26]. Moreover, the CoLab-score requires only routine blood tests, instead of (features from) imaging such as CT-scans or laboratory tests that are not routinely collected in the ED, e.g. interleukin-6 or 3-hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific) [27,28].

Two other studies have been published which are similar to this study [9,10]. Interestingly, the study by Soltan et al., ranked basophils and eosinophils as the two most important features in predicting the outcome, similar to our results [10]. Eosinophils were also seen as one of the most important features by Plante et al. [9]. However, both studies focus on an artificial intelligence/machine learning approach. While their approach likely results in higher predictive performance, due to the ability of machine learning models to capture non-linear and interaction effects, the goal of this study was to develop a simple, fast and robust model that can easily be implemented in current hospital IT systems.

Since this is a retrospective case-control study, there is some unavoidable missing data. In our cohort 17.6% of the ED presentations could not be used due to one or more missing laboratory results. This is lower or equal to similar studies; 22% [23], 17% [21] and 11% [26]. Important to note is that 7.7% of missingness is due to analytical errors which can be assumed to be missing completely at random. For the remaining 9.9% of missingness, the full lab panel was most frequently missing for pediatric, obstetric and surgery patients. These patients are presenting with specific complaints for which specific laboratory tests are requested, and hence do not match the inclusion criteria for a routine blood panel. Overall the missingness was significantly lower in the PCR-tested group versus the untested group (χ^2-test p-value <0.001). It is assumed that all presentations in the untested group are COVID-19 negative. However, some presentations with asymptomatic COVID-19 could be present in the untested control group. The impact of these 'false controls' is most likely small as other studies indicate that there is a very low positivity rate among asymptomatic ED presentations (only a few in over a thousand tested asymptomatic cases) [29,30]. The vast majority of controls were not tested for COVID-19, because they were either pre-pandemic or untested patients (89% in the development dataset). Clinical data always contains some unavoidable 'noise' in the form of misregistrations, misdiagnoses or patients who were missed. We have tried to mitigate this by including a large pre-pandemic control group and including all PCR tests within 1 week after discharge.

In the external centers, there is a high level of missingness as a result of an incomplete laboratory panel. In the case of centers 1 and 2, only internal medicine ED presentations were tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine ED presentations, which is reflected by the few PCR-positive patients excluded. Due to these

high levels of missingness, the results of the external centers cannot be used to show that the CoLab-score generalizes to the entire ED population. Rather, the results show that for the majority of COVID-19 positive patients presenting at the ED, a routine laboratory panel is available from which the CoLab-score can be calculated, and that the performance of the CoLab-score in this population is comparable to the development population. Differences in the distribution of CoLab variables between centers are shown in **Supplemental Material 3 Figure 2**.

The performance of the CoLab-score is affected by the time between the onset of symptoms and ED presentations. The score increases with the duration of symptoms and gradually decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some COVID-19 patients with early or late presentation after onset of symptoms can be missed. Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and <10 days prior to ED presentation. Chemotherapy that causes myeloid suppression, will decrease neutrophilic, basophilic and eosinophilic counts and thereby "falsely" increasing the CoLab-score. Conversely, COVID-19 patients with severe anemia could have "falsely" lowered CoLab-scores. To minimize false negatives, we have therefore advised to report CoLab-scores only when the concentration of erythrocytes is ≥ 2.9 /pL.

It was chosen to exclude re-presentations after a previous presentation with COVID-19. Since the median time between initial presentation and re-presentation was 12 days, these patients were most likely not re-infected patients, but patients who deteriorated after initial presentation/treatment. Given that the CoLab-score follows the host-immune response, the score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients would impact the performance of the CoLab-score as patients in a later phase of the disease show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to

patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate after treatment for COVID-19. Other re-presentations were not excluded, which results in some patients appearing multiple times in a dataset. This was not adjusted for in the regression model since the assumption was made that ED presentations are independent observations. The median time between re-presentations is 38 days, most likely resulting in variations in laboratory results between presentations, and hence, little to no correlation between presentations. A sensitivity analysis was performed whereby only the first presentation was included for each patient (**Supplemental Material 4 Table 1**), but no difference was found in performance in terms of sensitivity, specificity and AUC.

The CoLab-score does not serve as a replacement for PCR-testing or LFTs, and can be used to guide PCR-testing when routine blood tests are available. Important to note is that the CoLab-score is only valid for ED presentations where routine blood testing is requested, and as a consequence does not generalize to the ED population who is otherwise well and does not undergo routine blood testing. Using the CoLab-score in a symptomatic/PCR-tested cohort also results in different diagnostic performance characteristics, as compared to using the score on the full ED cohort (see **Supplemental Material 4 Table 1)**.

Finally, the CoLab-score could lead to false positives by other viral infections. However, in an historic patient cohort, the CoLab-score had only limited discriminative ability in separating influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4 Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-response to the virus, it is hypothesized that the CoLab-score could also be sensitive to future SARS-CoV-2 variants. This is supported by the fact that the discriminative ability is sustained in periods with different dominant variants, although the sensitivity of scores ≥ 3 is somewhat lower in the third phase (see **Supplemental Material 2 Table 1**). Although vaccination status is not registered for all presenting patients, in a small subgroup of 12 patients for whom

vaccination status was registered, and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5) (see **Supplemental Material 2 Figure 2**). Continuous assessment of the performance of the CoLab-score is required due to the emergence of new variants and changes in the host's immune response.

To conclude, the CoLab-score developed and validated in this study, based on 10 routine laboratory results and age, is available within 1 hour for any patient presenting at the ED where routine blood testing is requested. The score can be used by clinicians to guide PCR testing or triage patients and helps to identify COVID-19 in patients presenting at the ED with abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific complaints where a routine blood panel is requested. The lowest CoLab-score can be used to effectively rule-out a possible SARS-CoV-2 infection, the highest score to alert physicians to a possible infection. The CoLab-score is therefore a valuable tool to rule out COVID-19, guide PCR testing and is available to any center with access to routine laboratory tests.

Data Availability Statement

Datasets with source data for Table 1, Figure 2 and Table 4, as well the R-code to fit the model is available from a Dryad repository [31].

Funding statement

This was an investigator-initiated study and no funding was received for this study.

Competing interests

A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR reports no conflict of interest. VS reports no conflict of interest.

Author contributorship statement

Arjen-Kars Boer: Conceptualization (Lead), Data curation (Lead), Funding acquisition (Lead), Investigation (Equal), Methodology (Equal), Supervision (Equal), Writing-original draft (Equal), Writing-review & editing (Equal).

Ruben Deneer: Data curation (Equal), Formal analysis (Equal), Investigation (Equal), Methodology (Lead), Software (Lead), Visualization (Lead), Writing-original draft (Equal), Writing-review & editing (Equal).

Maaike Maas: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Equal), Writing-review & editing (Equal).

Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).

Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing (Equal).

Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal), Writing-review & editing (Equal).

Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing (Equal).

References

1 Coronavirus Disease (COVID-19) Situation Reports. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/ (accessed 4 Feb 2021).

2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in China. *https://doi.org/101056/NEJMoa2002032* 2020;**382**:1708–20. doi:10.1056/NEJMOA2002032

3 Vetter P, Vu DL, L'Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**. doi:10.1136/BMJ.M1470

4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18. doi:10.1136/bmj.m1328

5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19): A Systematic Review. J. Med. Syst. 2020;**44**:122. doi:10.1007/s10916-020-01582-x

6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417

7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT Assessment Scheme for Patients Suspected of Having COVID-19—Definition and Evaluation. *https://doi.org/101148/radiol2020201473* 2020;**296**:E97–104. doi:10.1148/RADIOL.2020201473

8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840

9 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine learning tool to rule out COVID-19 among adults in the emergency department using routine blood tests: A large, multicenter, real-world study. *J Med Internet Res* 2020;**22**:e24048. doi:10.2196/24048

10 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical data for patients attending hospital: development and prospective validation of an artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87. doi:10.1016/S2589-7500(20)30274-0

11 R Core Team. R: A Language and Environment for Statistical Computing. 2020.https://www.r-project.org/

12 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29. doi:10.1198/016214506000000735

13 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x

14 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01

15 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63. doi:10.1093/oxfordjournals.pan.a004868

16 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation. *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973

17 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-2105-12-77

18 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.https://cran.r-project.org/package=rms

19 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons and progress. Scand. J. Clin. Lab. Invest. 1999;**59**:491–500. doi:10.1080/00365519950185229

20 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst* 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4

21 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502

22 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807. doi:10.1007/s00330-020-07022-1

23 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-infected patients at the emergency department using routine testing. *Clin Chem Lab Med* 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593

24 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305

25 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in the emergency department: derivation and validation of the LDH-CRP-Lymphocyte (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336

26 Tordjman M, Mekki A, Mali RD, *et al.* Pre-test probability for SARS-Cov-2-related infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342. doi:10.1371/journal.pone.0243342

27 Peto T, Affron D, Afrough B, *et al.* COVID-19: Rapid antigen detection for SARS-CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for mass-testing. *EClinicalMedicine* 2021;**36**:100924. doi:10.1016/J.ECLINM.2021.100924

28 García-Fiñana M, Hughes DM, Cheyne CP, *et al.* Performance of the Innova SARS-CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot: population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637

29 Ford JS, Parikh A, Sandhu R, *et al.* Testing Asymptomatic Emergency Department Patients for Coronavirus Disease 2019 (COVID-19) in a Low-prevalence Region. Acad. Emerg. Med. 2020;**27**:771–4. doi:10.1111/acem.14044

30 Ravani P, Saxinger L, Chandran U, *et al.* COVID-19 screening of asymptomatic patients admitted through emergency departments in Alberta: a prospective quality-improvement study. *C Open* 2020;**8**:E887–94. doi:10.9778/cmajo.20200191

[dataset] 31 Boer A-K, Deneer R. Source data for: Development and validation of an early warning score to identify COVID-19 in the emergency department based on routine laboratory tests: a multicenter case-control study. *Dryad Digit Repos* Published Online First: 2021. doi:https://doi.org/10.5061/dryad.5hqbzkh6x

Figure legends

Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B) dataset.

All patient admissions with routine venous blood sampling at the emergency department (ED) were included. For the development dataset, completeness of the lab panel was assessed for all 28 laboratory tests , for the temporal validation dataset this was only necessary for 10 laboratory tests. The major causes of missingness are described in the text. In the development dataset, presentations with extreme values (>10 SD) were excluded. The same limits were applied to the temporal validation dataset (see Table 2 for limits).

Figure 2: Probability density plot of the CoLab-linear predictor.

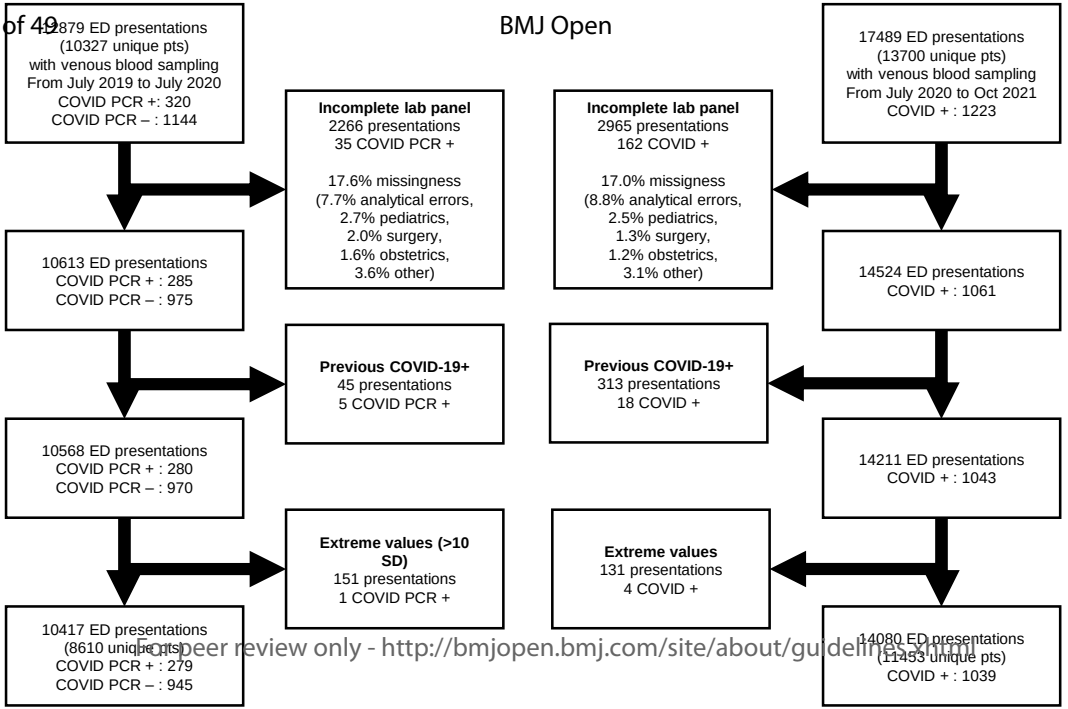
The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are plotted against the linear predictor (see table 2). The CoLab-score cut-offs (–5.83, –4.02, –3.29, –2.34 and –1.64) are depicted with vertical dashed lines. The white-boxed numbers (between the cut-offs) represent the corresponding CoLab-score. Note that while the area under both curves is identical (since these are probability density functions), in absolute numbers the "negative or untested"-group is about 36 times larger than the PCR positive group.

Figure 3: Inclusion flow of ED patients in three external centers.

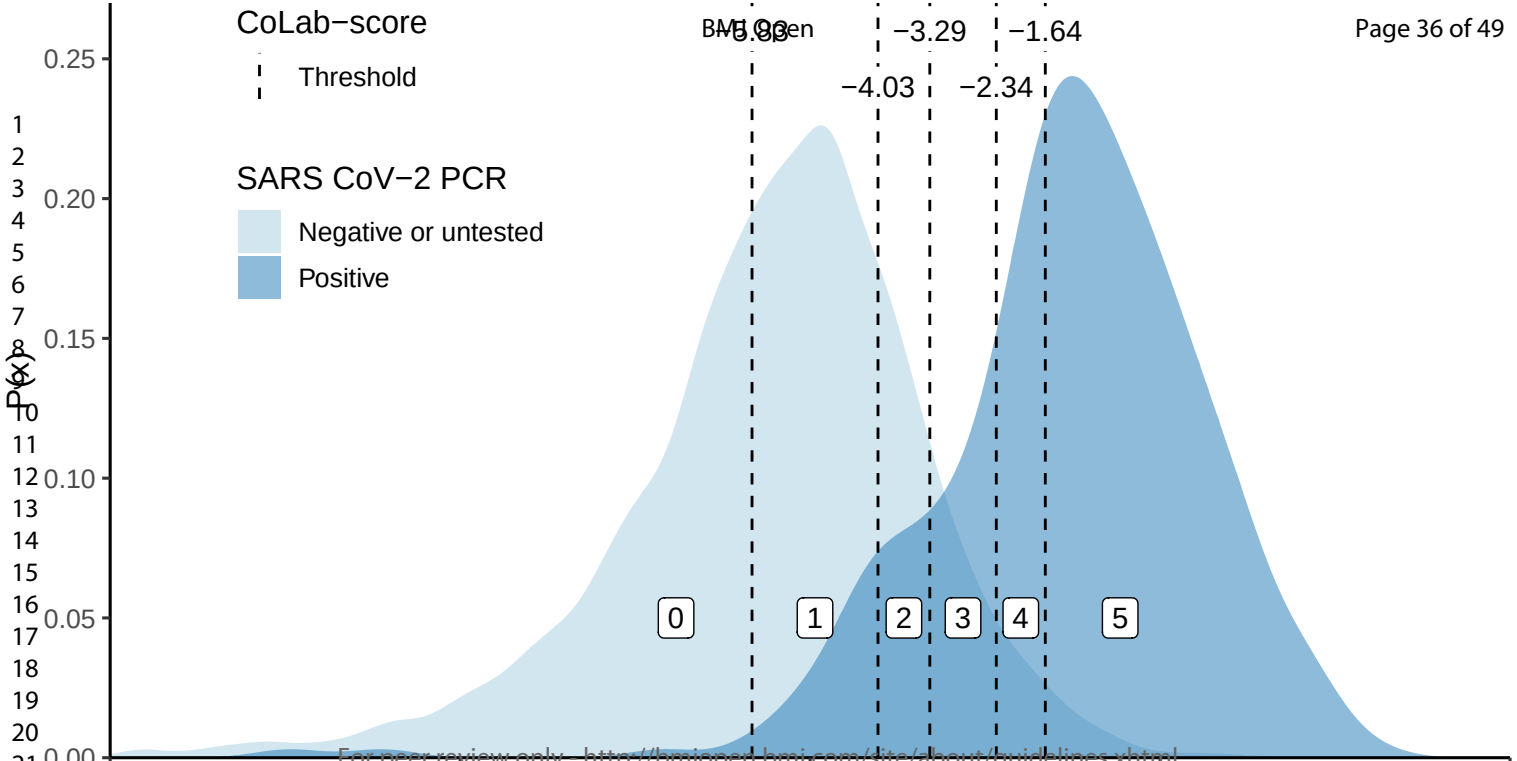
All emergency department (ED) presentations with routine venous blood sampling were included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see

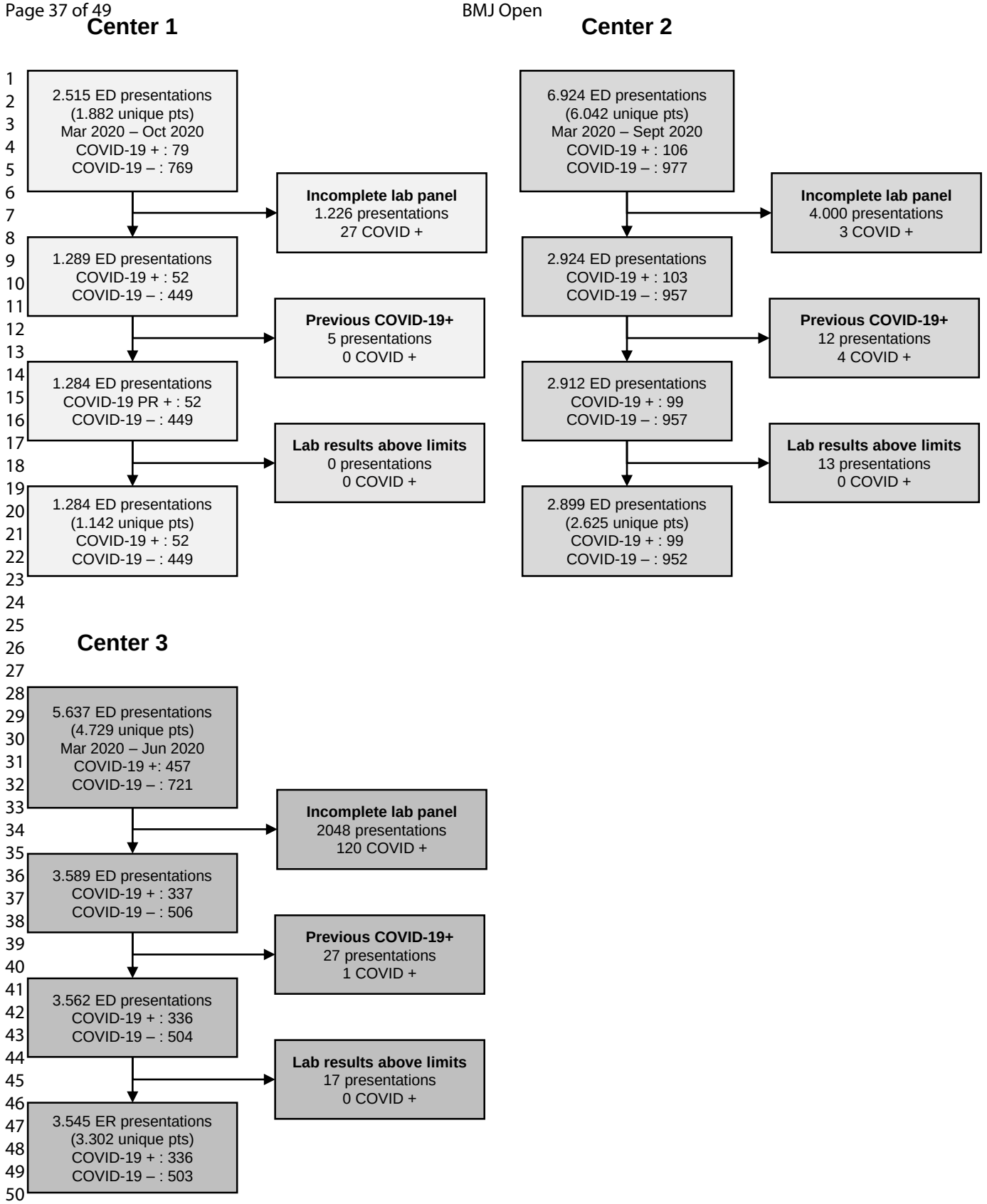
Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were excluded as "previous COVID-19+". Presentations with any laboratory result above the limits of the CoLab-score (see Table 2) were excluded.

A



B




Center 1
2.515 ED presentations
(1.882 unique pts)
Mar 2020 – Oct 2020
COVID-19 + : 79
COVID-19 – : 769
Incomplete lab panel
1.226 presentations
27 COVID +
1.289 ED presentations
COVID-19 + : 52
COVID-19 – : 449
Previous COVID-19+
5 presentations
0 COVID +
1.284 ED presentations
COVID-19 PR + : 52
COVID-19 – : 449
Lab results above limits
0 presentations
0 COVID +
1.284 ED presentations
(1.142 unique pts)
COVID-19 + : 52
COVID-19 – : 449
Center 2
6.924 ED presentations
(6.042 unique pts)
Mar 2020 – Sept 2020
COVID-19 + : 106
COVID-19 – : 977
Incomplete lab panel
4.000 presentations
3 COVID +
2.924 ED presentations
COVID-19 + : 103
COVID-19 – : 957
Previous COVID-19+
12 presentations
4 COVID +
2.912 ED presentations
COVID-19 + : 99
COVID-19 – : 957
Lab results above limits
13 presentations
0 COVID +
2.899 ED presentations
(2.625 unique pts)
COVID-19 + : 99
COVID-19 – : 952
Center 3
5.637 ED presentations
(4.729 unique pts)
Mar 2020 – Jun 2020
COVID-19 +: 457
COVID-19 – : 721
Incomplete lab panel
2048 presentations
120 COVID +
3.589 ED presentations
COVID-19 + : 337
COVID-19 – : 506
Previous COVID-19+
27 presentations
1 COVID +
3.562 ED presentations
COVID-19 + : 336
COVID-19 – : 504
Lab results above limits
17 presentations
0 COVID +
3.545 ER presentations
(3.302 unique pts)
COVID-19 + : 336
COVID-19 – : 503

Supplemental material 1

Model fitting

Prior to model fitting, covariates were scaled to zero mean and unit variance, after model fitting coefficients were unscaled to obtain regression coefficients on the original scale. In adaptive lasso, weights are applied to each of the covariates present in the lasso constraint, the weight vector has to be calculated before the adaptive lasso regression is performed. Due to multicollinearity between laboratory tests in the routine lab panel, weights in the adaptive lasso were based on ridge regression estimates ($\hat{\beta}_{ridge}$) as recommended by Zou. To obtain $\hat{\beta}_{ridge}$ the optimal penalty (λ) for the ridge regression was chosen using 10 fold cross-validation (CV) with area under the ROC curve (AUC) as the loss function. The λ corresponding to the maximum AUC was selected to obtain $\hat{\beta}_{ridge}$. The weight vector ($\widehat{w}$) was calculated by $\widehat{w} = 1/\left|\hat{\beta}_{ridge}\right|^2$. This weight vector was then used to fit an adaptive lasso regression where λ was chosen by the criterion ±1 SE of the maximum AUC.

Model intercept correction

The linear predictor for a patient *i* is calculated as follows: $lp_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$
Where *n* is the number of variables in the final model, x_{in} are the observed predictor variables for subject *i* and $\beta_n$the model coefficients. The linear predictor can then be converted to a probability for patient *i* (P_i) by the logistic function: $P_i = \frac{1}{1+e^{-lp_i}}$
The intercept term β_0 is sensitive to the fraction of cases versus controls in the dataset/population. Since the model is fitted to a case-control dataset where the number cases is fixed (all patients tested positive for COVID-19) and the number of controls is randomly chosen (a 6-month period pre-COVID), the intercept term β_0 is a result of this choice and will likely not be generalizable to the real-world setting. Prior correction is a method to correct the estimate of the intercept based on the true fraction of positives in the population, τ (prevalence of COVID-19 in the ED) and the fraction of cases in the development dataset, $\bar{y}$. The intercept term β_0 can then be corrected to obtain $\beta_{0corrected}$ using the following formula:

$$\beta_{0corrected} = \beta_0 + \beta_{adj}$$

$$\beta_{adj} = -ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{y}}{1-\bar{y}}\right)\right]$$

In our dataset $\bar{y} = 0.02675$ therefore:

$$\beta_{adj} = -ln\left(\frac{1-\tau}{\tau}\right) + 3.594$$

An estimate $\bar{\tau}$ can be used for the prevalence τ to obtain $\bar{\beta}_{adj}$ which can be plugged in the original linear predictor formula to obtain calibrated probabilities:

$$lp_i(\tau) = \beta_0 - ln\left(\frac{1-\tau}{\tau}\right) + 3.594 + \beta_1 x_{i1} + \cdots + \beta_n x_{in}$$

CoLab-score

An alternative, which is the basis of the CoLab-score, is to choose a fixed probability P_i above which one considers a patient eligible for further testing. The probability can be expressed as a number needed to test. If one is willing to test 10 patients to find one positive, all patients with $P_i \geq 0.1$ should be considered positive. In this study a number needed to test of 15 is used, therefore all patients with a $P_i \geq 0.067$ should be considered positive. On the linear predictor scale this translates to $\text{logit}(0.067) = -2.639$. To determine the cutoffs for difference prevalence thresholds one solves the following equation:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639$$
$$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$
$$lp_i(\tau) \geq ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

Choosing values for τ yields the cutoffs for the CoLab score:

$lp_i(\tau = 0.4) \geq -5.83$ (CoLab-score = 1)
$lp_i(\tau = 0.1) \geq -4.03$ (CoLab-score = 2)
$lp_i(\tau = 0.05) \geq -3.29$ (CoLab-score = 3)
$lp_i(\tau = 0.02) \geq -2.34$ (CoLab-score = 4)
$lp_i(\tau = 0.01) \geq -1.64$ (CoLab-score = 5)

These thresholds correspond to CoLab-scores 0 to 5. The interpretation of these scores is as follows; if the prevalence is <1%, only CoLab-score 5 should be classified as positive and CoLab-score 0 till 4 as negative. If the prevalence is 1% – 2%, CoLab-score 4 and 5 should be classified as positive and 1 – 3 negative. Similarly, with a prevalence of 2 – 5% the split is between CoLab-score 2 and 3 and with prevalence of 5 – 10% between CoLab-score 1 – 2. If the prevalence is higher than 10% only CoLab-score 0 is classified as negative. Using the CoLab-score in this fashion, aims to preserve a number need to test of 15.

Relative importance of variables

Since the variables included in the model are on different scales, the magnitude of the unscaled coefficients cannot be used to compare the importance of variables to each other. To give some indication of the importance of the variables in predicting the outcome, the unscaled coefficients obtained from the adaptive lasso regression were used to calculate the relative importance. The variable with the highest unscaled coefficient was used as maximum ($\beta_{unscaled,max}$), and all other scaled coefficients were divided by this maximum and multiplied by 100 to obtain the relative importance in %: $\frac{\beta_{unscaled}}{\beta_{unscaled,max}} \cdot 100$.

Supplemental material 2

Vaccination status and COVID-19 ED prevalence plot

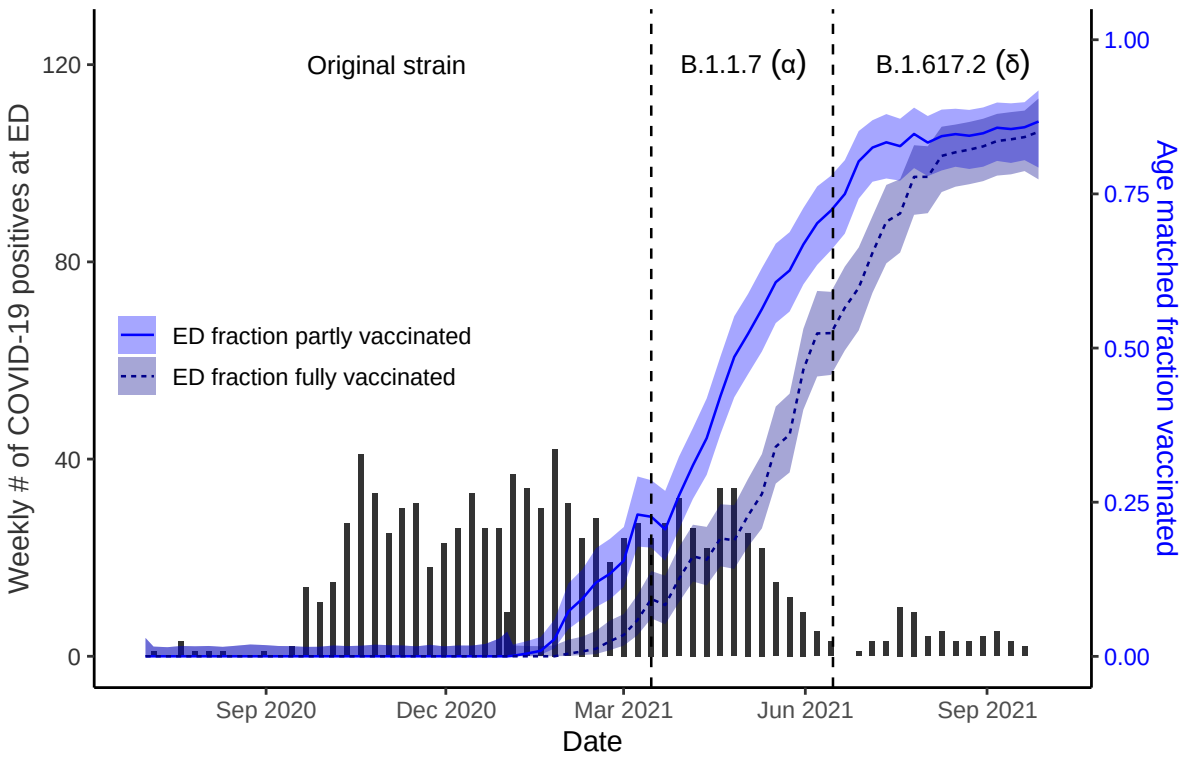


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

The temporal validation dataset consists of ED presentations from July 2020 until October 2021. As stated in the "Materials and Methods" section, this period was split into three phases: i) from July 2020 until March 2021, no vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant. The ED fraction vaccinated is estimated by merging data from the Dutch national institute of public health by the date of the ED presentation and the year of birth of the patient. The gray bars depict weekly number of new COVID-19 cases at the ED, the blue lines the estimated fraction of ED patients fully or partially vaccinated.

CoLab-score performance

Phase	Cases/controls (prevalence)	AUC
Original strain & no vaccinations	694/7999 (8.6%)	0.909 (0.896 - 0.923)
B.1.1.7 strain & partial vaccination	287/2845 (10.1%)	0.937 (0.921 - 0.953)
B.1.617.2 strain & full vaccination	58/3236 (1.8%)	0.898 (0.857 - 0.939)

CoLab-score	Phase	Sensitivity	Specificity	PPV	NPV
0	Original strain & no vaccinations	0.960 (0.944 - 0.974)	0.418 (0.407 - 0.429)	0.135 (0.133 - 0.138)	0.991 (0.987 - 0.994)
	B.1.1.7 strain & partial vaccination	0.983 (0.969 - 0.997)	0.432 (0.413 - 0.450)	0.162 (0.158 - 0.168)	0.996 (0.992 - 0.999)
	B.1.617.2 strain & full vaccination	0.983 (0.948 - 1.000)	0.415 (0.396 - 0.432)	0.030 (0.028 - 0.031)	0.999 (0.998 - 1.000)
≤1	Original strain & no vaccinations	0.879 (0.854 - 0.902)	0.789 (0.779 - 0.798)	0.283 (0.273 - 0.294)	0.986 (0.983 - 0.988)
	B.1.1.7 strain & partial vaccination	0.916 (0.885 - 0.948)	0.809 (0.793 - 0.824)	0.350 (0.332 - 0.370)	0.989 (0.984 - 0.993)
	B.1.617.2 strain & full vaccination	0.862 (0.776 - 0.948)	0.780 (0.765 - 0.794)	0.067 (0.059 - 0.074)	0.997 (0.995 - 0.999)
≤2	Original strain & no vaccinations	0.813 (0.784 - 0.842)	0.894 (0.887 - 0.901)	0.421 (0.404 - 0.441)	0.980 (0.978 - 0.983)
	B.1.1.7 strain & partial vaccination	0.864 (0.826 - 0.902)	0.897 (0.885 - 0.908)	0.484 (0.455 - 0.516)	0.983 (0.979 - 0.988)
	B.1.617.2 strain & full vaccination	0.690 (0.569 - 0.810)	0.892 (0.881 - 0.902)	0.104 (0.086 - 0.123)	0.994 (0.991 - 0.996)
≤3	Original strain & no vaccinations	0.697 (0.661 - 0.731)	0.962 (0.957 - 0.966)	0.634 (0.605 - 0.662)	0.971 (0.968 - 0.974)
	B.1.1.7 strain & partial vaccination	0.760 (0.711 - 0.812)	0.963 (0.955 - 0.970)	0.696 (0.650 - 0.739)	0.973 (0.967 - 0.978)
	B.1.617.2 strain & full vaccination	0.621 (0.483 - 0.741)	0.960 (0.954 - 0.967)	0.222 (0.178 - 0.268)	0.993 (0.990 - 0.995)
≤4	Original strain & no vaccinations	0.566 (0.529 - 0.602)	0.984 (0.981 - 0.987)	0.775 (0.740 - 0.808)	0.960 (0.957 - 0.963)
	B.1.1.7 strain & partial vaccination	0.645 (0.589 - 0.704)	0.983 (0.978 - 0.988)	0.809 (0.762 - 0.856)	0.961 (0.955 - 0.967)
	B.1.617.2 strain & full vaccination	0.517 (0.397 - 0.638)	0.986 (0.982 - 0.990)	0.400 (0.319 - 0.500)	0.991 (0.989 - 0.993)

Table 1: Diagnostic performance of the CoLab-score in the temporal validation dataset, split by phase.

Sensitivities, specificities, positive predictive values (PPV) and negative predictive values (NPV) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is split into three phases according to dominant SARS-CoV-2 strains in the Netherlands and estimated fraction of ED patients vaccinated (see Figure above). Note that "0" lists the sensitivity and NPV of CoLab-score 0 and "≤ 4" lists the specificity and PPV of CoLab-score 5. The AUC was significantly higher in the second phase as compared to the first phase (DeLong test p-value: 0.0175), but did not differ significantly between the third and first phase (DeLong test p-value: 0.3903).

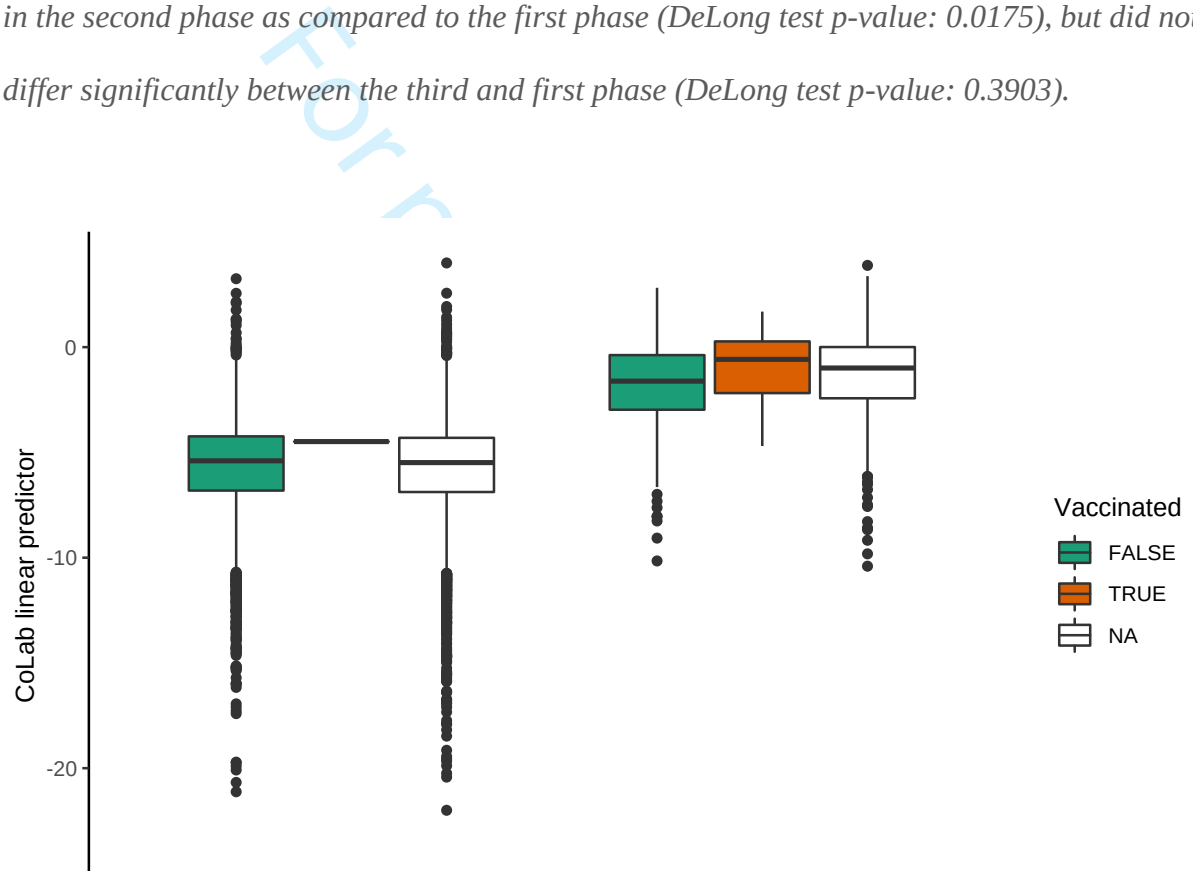


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

The CoLab linear predictor is calculated for all ED presentations in the temporal validation set. Presentations who are registered as vaccinated are labeled TRUE (N = 13). Presentations before vaccine roll-out are labeled FALSE (N = 5855). Presentations during

vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13 presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative. Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or considered positive until proven otherwise, therefore there is only one COVID-19 negative patient with a registered vaccination status.

Supplemental material 3

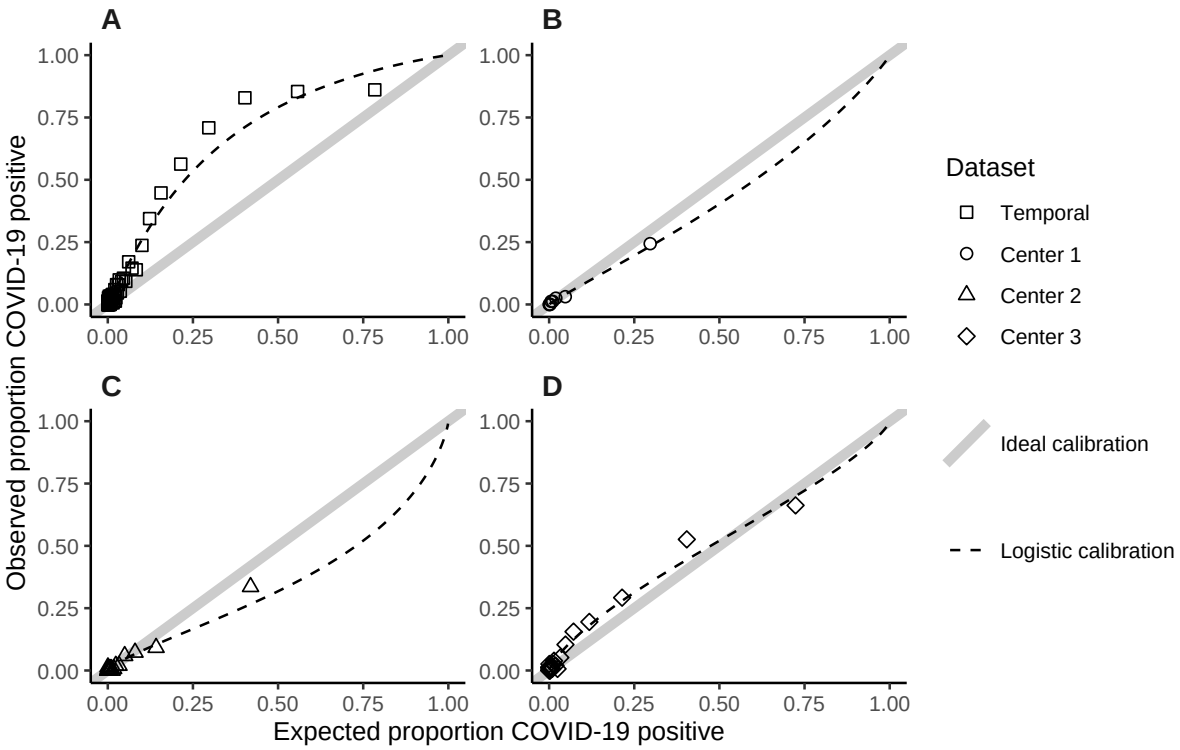


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

In the calibration plots, the proportion of observed COVID-19 positives versus expected probabilities are plotted. Observations are grouped with an average of 150 observations per group. The expected probabilities follow from applying the inverse logit function to the CoLab-linear predictor calculated from Table 2. If the observed proportion in an external dataset is lower than the expected proportion, this means risks are over-estimated, if the observed fraction is higher, risks are under-estimated. Ideally, observed proportions are equal to expected proportions, this ideal-calibration-line is shown as a straight line through the origin with a slope of 1. The logistic calibration line is a logistic regression fit of the predicted probabilities. [Intercept, slope] for plots A-D: A [1.34, 1.08], B [-0.39, 0.92], C [-0.76, 0.77], D [0.08, 0.79]. Although no validation datasets show perfect calibration, this is the result of differences in COVID-19 prevalence in the temporal validation dataset (7.4% versus 2.2%) and differences in calibration of laboratory equipment in the three external centers.

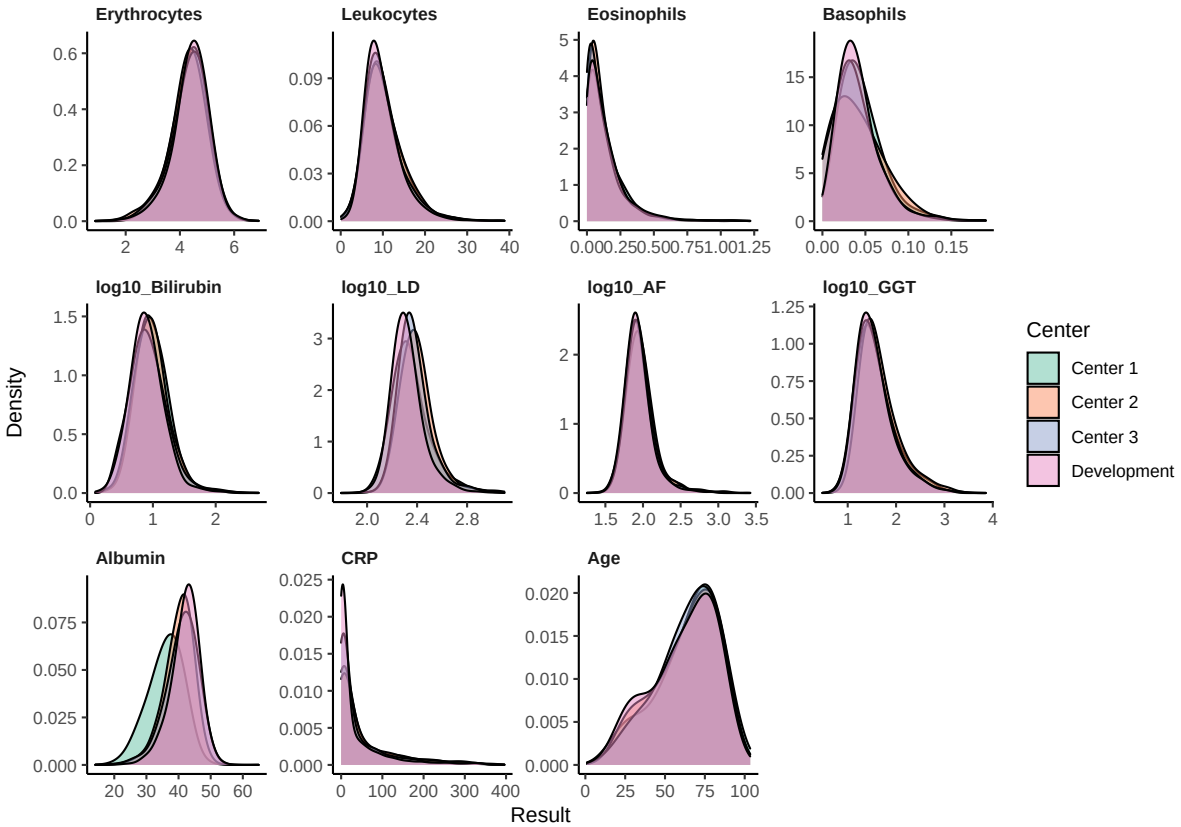


Figure 2: Probability density plots of laboratory parameters.

Probability density plots are shown for all control patients of the development dataset and the three external centers. Ideally all distributions should overlap since this implies that control patient populations are most likely similar in the development dataset to the external datasets. When comparing the distribution of the CoLab variables for all control-patients across different external validation datasets, albumin and LD show the largest deviations.

Supplemental material 4

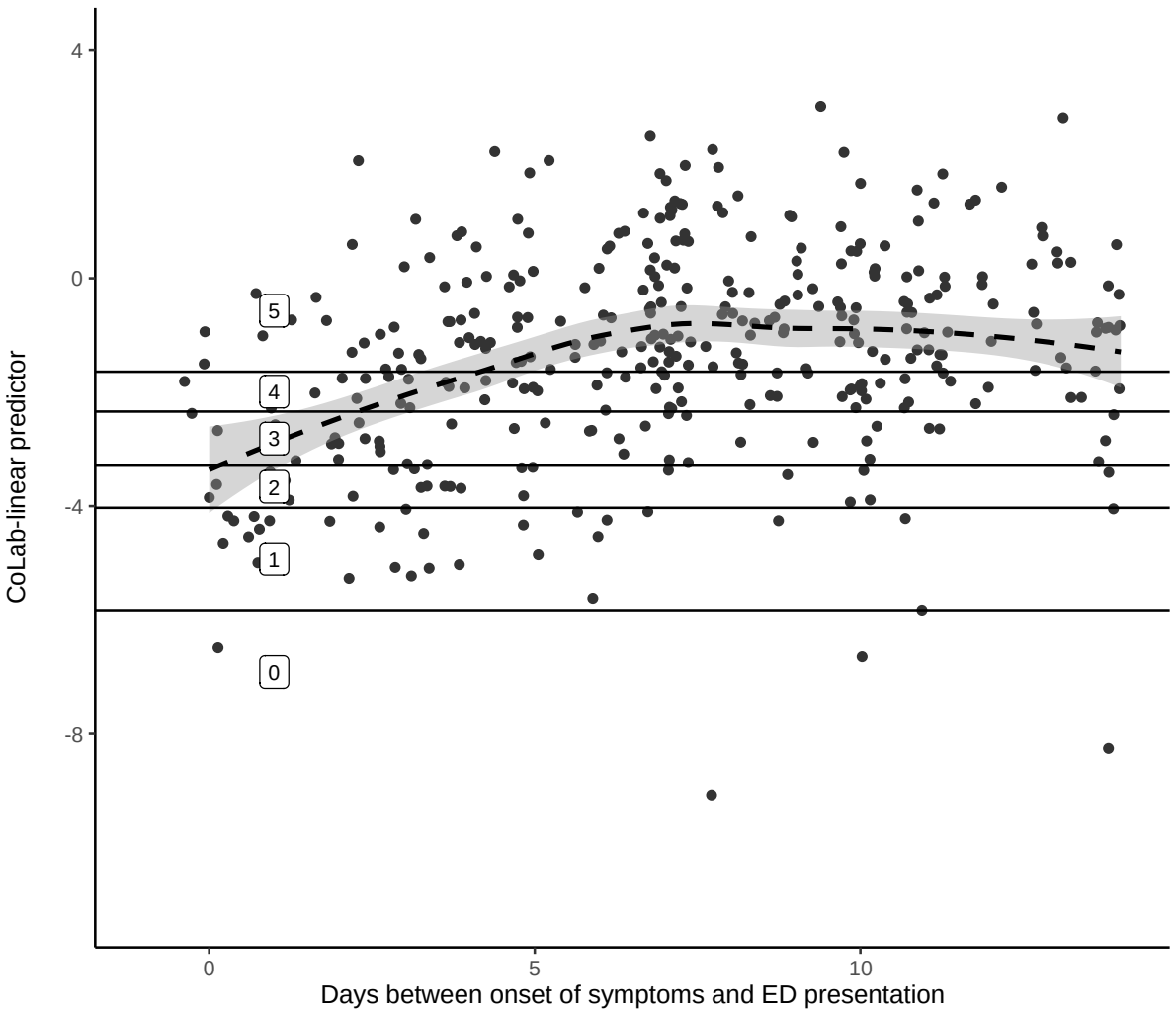


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predict is plotted against the duration of COVID-related symptoms as registered in the electronic patient records. Patients with unknown duration are not plotted. Patients without symptoms were plotted at 0 days. The solid horizontal lines represent the CoLab-score thresholds, the dashed line is a LOESS regression curve with 95% CI. As the duration of symptoms is an integer, some random jitter was added to the days, for visualization purposes. Note that only the first 14 days are shown in this graph.

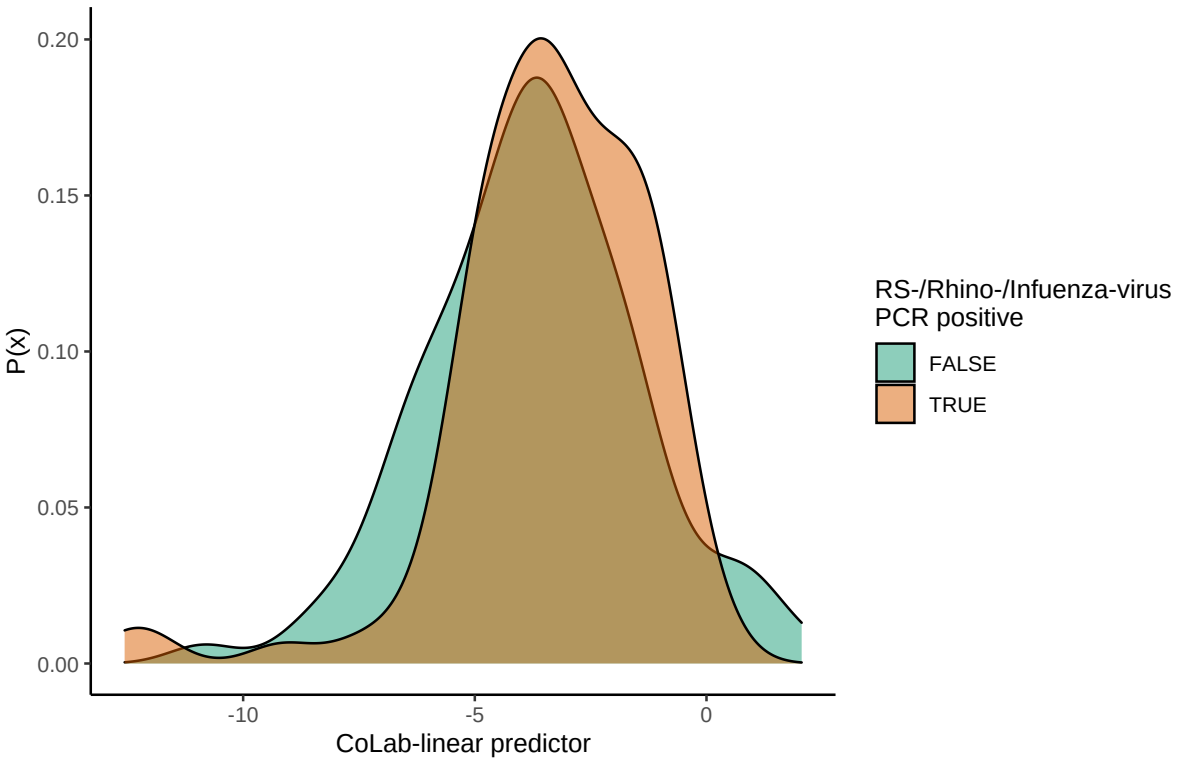


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

For 183 ED presentations that were PCR tested for either RS-, Rhino- and Influenza-virus the CoLab-score was calculated. 91 presentations were PCR positive, 92 were PCR negative. The CoLab-score is only marginally elevated for PCR positive patients, the area under the ROC-curve in separating both groups is 0.573 (95% CI: 4896-0.6563).

Inclusion criterion	Cases/controls (prevalence)	AUC
Temporal validation (reference)	1039/14080 (7.4%)	0.916 (0.906 - 0.927)
Only first presentations, re-presentations are excluded	937/11166 (8.4%)	0.919 (0.909 - 0.930)
Only PCR-tested presentations	372/4062 (9.2%)	0.840 (0.817 - 0.862)

CoLab-score	Validation set	Sensitivity	Specificity	PPV	NPV	TP	TN	FP	FN
0	Reference	0.967 (0.956 - 0.978)	0.420 (0.411 - 0.428)	0.117 (0.115 - 0.119)	0.994 (0.992 - 0.996)	1005 (993 - 1016)	5476 (5366 - 5587)	7565 (7454 - 7675)	34 (23 - 46)
	First presentations	0.968 (0.956 - 0.979)	0.416 (0.406 - 0.426)	0.132 (0.130 - 0.134)	0.993 (0.990 - 0.995)	907 (896 - 917)	4259 (4156 - 4353)	5970 (5876 - 6073)	30 (20 - 41)
	PCR-tested presentations	0.946 (0.922 - 0.968)	0.353 (0.338 - 0.368)	0.129 (0.125 - 0.132)	0.985 (0.979 - 0.991)	352 (343 - 360)	1303 (1246 - 1359)	2387 (2331 - 2444)	20 (12 - 29)
≤ 1	Reference	0.888 (0.870 - 0.908)	0.791 (0.783 - 0.798)	0.253 (0.245 - 0.261)	0.989 (0.987 - 0.991)	923 (904 - 943)	10311 (10215 - 10401)	2730 (2640 - 2826)	116 (96 - 135)
	First presentations	0.890 (0.870 - 0.908)	0.793 (0.785 - 0.801)	0.282 (0.273 - 0.292)	0.987 (0.985 - 0.990)	834 (815 - 851)	8112 (8030 - 8194)	2117 (2035 - 2199)	103 (86 - 122)
	PCR-tested presentations	0.852 (0.817 - 0.887)	0.671 (0.656 - 0.686)	0.207 (0.197 - 0.217)	0.978 (0.973 - 0.983)	317 (304 - 330)	2477 (2421 - 2533)	1213 (1157 - 1269)	55 (42 - 68)
≤2	Reference	0.820 (0.796 - 0.843)	0.894 (0.889 - 0.899)	0.382 (0.367 - 0.396)	0.984 (0.982 - 0.986)	852 (827 - 876)	11661 (11591 - 11729)	1380 (1312 - 1450)	187 (163 - 212)
	First presentations	0.824 (0.798 - 0.845)	0.898 (0.892 - 0.904)	0.426 (0.410 - 0.441)	0.982 (0.980 - 0.985)	772 (748 - 792)	9187 (9127 - 9249)	1042 (980 - 1102)	165 (145 - 189)
	PCR-tested presentations	0.734 (0.688 - 0.777)	0.800 (0.786 - 0.812)	0.270 (0.252 - 0.287)	0.968 (0.962 - 0.973)	273 (256 - 289)	2951 (2902 - 2997)	739 (693 - 788)	99 (83 - 116)
≤ 3	Reference	0.710 (0.682 - 0.738)	0.962 (0.958 - 0.965)	0.596 (0.573 - 0.618)	0.977 (0.974 - 0.979)	738 (709 - 767)	12540 (12496 - 12582)	501 (459 - 545)	301 (272 - 330)
	First presentations	0.716 (0.687 - 0.744)	0.966 (0.962 - 0.969)	0.658 (0.633 - 0.682)	0.974 (0.971 - 0.976)	671 (644 - 697)	9880 (9844 - 9915)	349 (314 - 385)	266 (240 - 293)
	PCR-tested presentations	0.591 (0.540 - 0.640)	0.911 (0.902 - 0.921)	0.403 (0.370 - 0.433)	0.957 (0.952 - 0.962)	220 (201 - 238)	3363 (3328 - 3397)	327 (293 - 362)	152 (134 - 171)
≤4	Reference	0.585 (0.556 - 0.615)	0.984 (0.982 - 0.987)	0.750 (0.724 - 0.778)	0.968 (0.965 - 0.970)	608 (578 - 639)	12838 (12811 - 12866)	203 (175 - 230)	431 (400 - 461)
	First presentations	0.590 (0.558 - 0.621)	0.987 (0.985 - 0.989)	0.805 (0.776 - 0.832)	0.963 (0.961 - 0.966)	553 (523 - 582)	10095 (10071 - 10117)	134 (112 - 158)	384 (355 - 414)
	PCR-tested presentations	0.452 (0.401 - 0.503)	0.959 (0.953 - 0.965)	0.526 (0.480 - 0.575)	0.945 (0.941 - 0.950)	168 (149 - 187)	3539 (3516 - 3562)	151 (128 - 174)	204 (185 - 223)

Table 1: Sensitivity analysis of the CoLab-score in the temporal validation dataset using different inclusion criteria.

Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals in parentheses. The temporal validation dataset is used to compare the performance of the CoLab-score with inclusion criteria that differ from the development dataset. The first line shows the performance of the temporal validation dataset with the original inclusion criteria as specified in Figure 1B. The second line shows the performance of the CoLab-score when all re-presentations are excluded (i.e. no repeated presentations). The third line shows the performance of the CoLab-score in the subgroup of patients that underwent PCR-testing.

TRIPOD Checklist: Prediction Model Development and Validation

Section/Topic	Item		Checklist Item	Page
Title and abstract				
Title	1	D;V	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	D;V	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	3, 4
Introduction				
Background and objectives	3a	D;V	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	6, 7
	3b	D;V	Specify the objectives, including whether the study describes the development or validation of the model or both.	7
Methods				
Source of data	4a	D;V	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	8, 11-12
	4b	D;V	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	8
Participants	5a	D;V	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	8
	5b	D;V	Describe eligibility criteria for participants.	8, 9, S1
	5c	D;V	Give details of treatments received, if relevant.	N/A
Outcome	6a	D;V	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	9
	6b	D;V	Report any actions to blind assessment of the outcome to be predicted.	N/A
Predictors	7a	D;V	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	8, 9
	7b	D;V	Report any actions to blind assessment of predictors for the outcome and other predictors.	N/A
Sample size	8	D;V	Explain how the study size was arrived at.	N/A
Missing data	9	D;V	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	9
Statistical analysis methods	10a	D	Describe how predictors were handled in the analyses.	10
	10b	D	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	10-12, S1
	10c	V	For validation, describe how the predictions were calculated.	16
	10d	D;V	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	11-13
	10e	V	Describe any model updating (e.g., recalibration) arising from the validation, if done.	N/A
Risk groups	11	D;V	Provide details on how risk groups were created, if done.	N/A
Development vs. validation	12	V	For validation, identify any differences from the development data in setting, eligibility criteria, outcome, and predictors.	22
Results				
Participants	13a	D;V	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	F1
	13b	D;V	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	T1
	13c	V	For validation, show a comparison with the development data of the distribution of important variables (demographics, predictors and outcome).	S3
Model development	14a	D	Specify the number of participants and outcome events in each analysis.	F1, F3
	14b	D	If done, report the unadjusted association between each candidate predictor and outcome.	N/A
Model specification	15a	D	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	T2
	15b	D	Explain how to the use the prediction model.	T2, S1
Model performance	16	D;V	Report performance measures (with CIs) for the prediction model.	T3, T4
Model-updating	17	V	If done, report the results from any model updating (i.e., model specification, model performance).	N/A
Discussion				
Limitations	18	D;V	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	21-23
Interpretation	19a	V	For validation, discuss the results with reference to performance in the development data, and any other validation data.	19-20
	19b	D;V	Give an overall interpretation of the results, considering objectives, limitations, results from similar studies, and other relevant evidence.	19-20
Implications	20	D;V	Discuss the potential clinical use of the model and implications for future research.	20-21
Other information				
Supplementary information	21	D;V	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	N/A
Funding	22	D;V	Give the source of funding and the role of the funders for the present study.	N/A

*Items relevant only to the development of a prediction model are denoted by D, items relating solely to a validation of a prediction model are denoted by V, and items relating to both are denoted D;V. We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document. S = Supplemental material, F = Figure, T = Table.