

BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email info.bmjopen@bmj.com

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:	<i>BMJ Open</i>
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, Maaïke; 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

SCHOLARONE™
Manuscripts



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.

1
2
3
4 1 **Development and validation of an early warning score to identify**
5
6 2 **COVID-19 in the emergency department based on routine laboratory**
7
8
9 3 **tests: a multicenter case-control study**
10
11
12 4

15 5 Arjen-Kars Boer^{1*}, Ruben Deneer^{1,2**}, Maaïke Maas³, Heidi S. M. Ammerlaan⁴, Roland H. H.
16
17 6 van Balkom⁵, Wendy A.M.H. Thijssen³, Sophie Bennenbroek³, Mathie P. G. Leers⁶, Remy
18
19 7 J.H. Martens⁶, Madelon M. Buijs⁷, Jos J. Kerremans⁸, Muriël Messchaert⁹, Jeroen D.E. van
20
21
22 8 Suijlen⁹, Natal A.W. van Riel^{2,10} and Volkher Scharnhorst^{1,2}
23
24
25 9

26
27
28 10 * Both authors contributed equally
29

30
31 11 † Corresponding author
32
33
34 12

35
36
37 13 1 Department of Laboratory Medicine, Catharina Hospital Eindhoven,
38
39 14 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

40
41 15 2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene
42
43 16 Loper 3, 5612 AE, Eindhoven, the Netherlands

44
45
46 17 3 Department of Emergency Medicine, Catharina Hospital Eindhoven,
47
48 18 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

49
50 19 4 Department of Internal Medicine, Catharina Hospital Eindhoven,
51
52 20 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

53
54
55 21 5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven,
56
57 22 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands
58
59
60

- 1
2
3 23 6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der
4
5 24 Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands
6
7 25 7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the
8
9 Netherlands
10 26
11
12 27 8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital,
13
14 28 Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands
15
16
17 29 9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334
18
19 30 DZ Apeldoorn, the Netherlands
20
21 31 10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef
22
23 32 9, 1105 AZ Amsterdam, the Netherlands.
24
25
26
27 33
28
29

30 **Word count: 3645**
31
32
33
34

35 **Keywords**

36
37 COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction
38
39 model, routine laboratory tests
40
41
42
43
44
45

46 **Corresponding author**

47
48
49 Ruben Deneer

50
51
52 Michelangelolaan 2

53
54
55 5623 EJ Eindhoven, the Netherlands

56
57
58 ruben.deneer@catharinaziekenhuis.nl
59
60

45 **Abstract**

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

51 **Design:** Case-control study.

52 **Setting:** Secondary and tertiary hospitals in the Netherlands.

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

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

60 **Results:** The resulting “CoLab-score” consists of 10 routine laboratory measurements, and
61 age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945).
62 The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991,
63 specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity
64 (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were
65 confirmed in temporal and external validation.

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

1
2
3 68 ruled-out in over one third of ED presentations. Highly suspect cases can be identified
4
5 69 regardless of presenting symptoms. The CoLab-score is a valuable tool to guide PCR testing,
6
7 70 triage ED patients, and is available to any center with access to routine laboratory tests.
8
9
10
11 71

12 13 72 **Article summary**

14 15 16 73 Strengths and limitations of this study

- 17
18
19 74 • A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency
20
21 75 department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- 22
23 76 • Using regression analysis, a simple score was developed consisting of only 10 routine
24
25 77 ED laboratory tests and age.
- 26
27 78 • The score was temporally and externally validation in 3 other centers, is available
28
29 79 within 1 hour after presentation and can be used to triage patients with a possible SARS-
30
31 80 CoV-2 infection in the ED.
- 32
33 81 • No evidence was found that the performance was affected by vaccinations and new
34
35 82 SARS-CoV-2 variants.
- 36
37 83 • The score is not a replacement for PCR-testing, but can be used to guide PCR-testing.
38
39
40
41
42
43 84

85 Introduction

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

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

100 Many COVID-19 prediction models have already been developed, the living systematic
101 review by Wynants et. al [4] provides an extensive overview and critical appraisal.

102 Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early
103 models were based on relatively small sample sizes, hampered by selection bias or were over-
104 fitted by selecting too many features [4–6]. Aside from methodological shortcomings, most
105 models are not developed as an early warning score for all ED patients. Firstly, they require
106 features from tests that are not routinely performed or logged for all ED patients (e.g. the CO-
107 RADS score from a CT-scan [7] or non-lab based clinical variables in the PRIEST EWS [8])
108 and are therefore not straightforward to implement or scale to a large ED patient population.

1
2
3 109 Secondly, the population on which models are commonly based, are PCR-tested patients, i.e.
4
5 110 a pre-selection of a possible COVID-19 infection has already been done by physicians.
6
7
8 111 In this study we report the development and validation of an early warning score that, based
9
10 112 on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in a
11
12 113 patient presenting at the ED. The score can assist ED physicians in triaging patients and
13
14 114 prevent further transmission of COVID-19 by quickly identifying possibly infected patients or
15
16 115 ruling out a possible infection when resources are scarce.
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

116 **Methods**

117 *Study design*

118 This is a retrospective case-control study where routine laboratory test results, combined with
119 age and gender, from all patient presenting at the emergency department (ED) of the
120 Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-
121 2 PCR test results in a development dataset. A model that could predict the presence of a
122 COVID-19 infection was fit to this dataset. Performance of the model was assessed by i)
123 internal validation, ii) temporal validation and iii) external validation by using data from the
124 ED of three other centers. The study was reviewed by the Medical research Ethics
125 Committees United (MEC-U) under study number W20.071, which confirmed that the
126 Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this
127 study. The study was thereafter reviewed and approved by the internal hospital review board.

128

129 *Patient and Public Involvement*

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

131

132 *Development dataset*

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

1
2
3 140 quantitative results were not available due to analytical interference (hemolysis, lipemia or
4
5 141 icterus). Presentations with one or more missing values in any of the 28 laboratory test in the
6
7 142 routine ED panel, were excluded. Presentations with one or more extreme lab results (> 10
8
9 143 times standard deviation from the median) were also excluded to minimize the effect on the
10
11 144 estimation of regression coefficients. After the first case of COVID-19 in the Netherlands, all
12
13 145 patients with symptoms of COVID-19 (either fever and/or respiratory symptoms) were
14
15 146 subjected to nasopharyngeal PCR testing for SARS-CoV-2 RNA. PCR testing was performed
16
17 147 by commercial tests that were approved by the Dutch national institute of public health
18
19 148 (RIVM). If a patient had a positive PCR result in the past, subsequent presentations were
20
21 149 excluded as re-presentations might be clinically different from de novo presentations.
22
23
24
25
26 150 The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying
27
28 151 nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal
29
30 152 at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was
31
32 153 positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time
33
34 154 window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were
35
36 155 performed in the time window and the presentation occurred after the first case of COVID-19
37
38 156 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the
39
40 157 first case were labelled as "*Pre-COVID-19*".
41
42
43
44
45

158

159 ***Laboratory tests***

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

8

1
2
3 165 monocytes. The chemical analyses were performed on a Cobas 8000 Pro (Roche Dx, Basel,
4
5 166 Switzerland) instrument and consisted of glucose, total bilirubin, aspartate aminotransferase
6
7 167 (ASAT), alanine aminotransferase (ALAT), lactate dehydrogenase (LD), creatine kinase
8
9 168 (CK), alkaline phosphatase (ALP), gamma-glutamyltransferase (gGT), blood urea nitrogen
10
11 169 (BUN), creatinine, CKD-epi estimated glomerular filtration rate (eGFR), potassium, sodium,
12
13 170 chloride, albumin (bromocresol green) and C-reactive protein (CRP). These results were
14
15 171 combined with age and gender.
16
17
18
19
20 172

21 22 23 173 *Modelling*

24
25 174 All data were processed and analyzed in R version 4.1.1 [9]. Laboratory results, combined
26
27 175 with age and gender were used as covariates in a regression model. Cases were defined as ED
28
29 176 presentations labelled as “PCR-positive”, controls were all other presentations (i.e. “PCR-
30
31 177 negative”, “Untested” or “Pre-COVID-19”). To achieve predictive accuracy, limit overfitting
32
33 178 and perform feature selection, penalized logistic regression with an adaptive lasso penalty was
34
35 179 chosen [10,11]. To minimize missing data, all non-numeric results at the extremes of the
36
37 180 measuring range, were converted to numeric results by removing the “<” and “>” signs. For
38
39 181 eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m² and
40
41 182 <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD,
42
43 183 CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model
44
45 184 predictions, these variables were transformed logarithmically. More details regarding model
46
47 185 fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the
48
49 186 glmnet-package [12].
50
51
52
53
54
55
56
57
58
59
60

188 *CoLab-score*

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

201

202 *Internal validation*

203 To assess model performance while taking overfitting into account, bootstrapping was
204 performed. 1000 bootstrap samples were generated from the original data. On each bootstrap
205 sample, the full model fitting procedure and CoLab-score conversion were performed.
206 Optimism adjusted performance measures of the CoLab-score were obtained by applying the
207 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [14]. Performance
208 measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative
209 predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate
210 performance measures [15]. Although the full inclusion period from July 2019 to July 2020
211 was used for model fitting, the performance was evaluated on the period starting from the first

1
2
3 212 COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain
4
5 213 performance measures that would reflect real world performance.
6
7

8 214

9
10
11 215 *Temporal validation*
12

13 216 For temporal validation, results from our center were prospectively analyzed from July 2020
14
15 217 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-
16
17 218 19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period
18
19 219 there was also more widespread external PCR testing by municipal health services. The
20
21 220 results of external conducted PCR tests were not available to our study. To overcome this
22
23 221 limitation, the outcome in the temporal validation cohort was chosen as a composite of the
24
25 222 hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test
26
27 223 result. This period also covers both the emergence of new SARS-CoV-2 variants as well as
28
29 224 vaccine rollout. However, neither vaccination status nor genomic sequencing was available to
30
31 225 determine whether a patient was vaccinated or which variant caused the infection. Therefore,
32
33 226 data from the Dutch national institute of public health (RIVM) was used, to divide the
34
35 227 temporal validation period into three phases: i) from July 2020 until March 2021, no
36
37 228 vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial
38
39 229 vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until
40
41 230 October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant.
42
43 231 See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted
44
45 232 of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold
46
47 233 for the entire period, as well as for each phase separately to determine a possible effect of
48
49 234 vaccination and new variants on performance (results in the **Supplemental Material 2**).
50
51 235 Model calibration was assessed graphically using the rms-package [16].
52
53
54
55
56
57
58
59
60 236

1
2
3 237 *External validation*
4

5 238 For the external validation, several centers in the Netherlands were approached and assessed
6
7 239 if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available.
8
9
10 240 Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals
11
12 241 (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and
13
14 242 Zuyderland Medical Center (center 3). The hematological parameters were measured with
15
16 243 Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and
17
18 244 Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with
19
20 245 Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott
21
22 246 Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external
23
24 247 validation was similar to the temporal validation and consisted of assessing the AUC
25
26 248 sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was
27
28 249 assessed graphically analogous to the temporal validation dataset.
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

250 Results

251 Development dataset

252 12879 emergency department (ED) presentations of 10327 patients from July 2019 to July
 253 2020 were included. After excluding cases with an incomplete lab panel, patient presentations
 254 that occurred after a positive PCR test in the past (re-presentations) and presentations with
 255 extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of
 256 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/m ²	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]

257

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

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

268

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

274

275 **CoLab-score**

276 The model obtained through adaptive lasso regression contained eleven variables, which are
 277 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 μ mol/L	-1.142	Bilirubin >169 μ mol/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 %

278

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

280 *The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of*
 281 *the 11 variables with their corresponding coefficients (β 's). CoLab-LP = - 6.885 +*
 282 *[erythrocytes] \times 0.9379 - [leukocytes] \times 0.1298 - [eosinophils] \times 6.834 - [basophils] \times*
 283 *47.7 - log₁₀([bilirubin]) \times 1.142 + log₁₀([LD]) \times 5.369 - log₁₀([ALP]) \times 3.114 +*
 284 *log₁₀([gGT]) \times 0.3605 - [albumin] \times 0.1156 + [CRP] \times 0.02560 + [age] \times 0.002275. The*
 285 *LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence*
 286 *is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid*
 287 *if any of the variables exceed the limits in the third column.*

288

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

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

296

297 *Internal validation*

298 The model was validated in the period starting from the first COVID-19 infection to July

299 2020, in this period the mean prevalence was 7.2%. The AUC of the CoLab-score is 0.930

300 (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)

301

302 **Table 3: Diagnostic performance CoLab-score in the development dataset.**303 *The development dataset was internally validation for the period March 2020 – July 2020 (N*304 *= 4.527). Sensitivities, specificities, positive predictive values (PPV), negative predictive*305 *values (NPV) and fraction of patients (%) are shown for fixed cut-offs (CoLab-score 0 till ≤*306 *4). The numbers in round brackets represent the 95% bootstrapped confidence intervals. The*307 *first column defines the threshold above which CoLab-score a patient is considered positive.*308 *Note that “0” lists the sensitivity and NPV of CoLab-score 0 and “≤ 4” lists the specificity*309 *and PPV of CoLab-score 5.*

310

311 Diagnostic performance is shown in **Table 3**. A CoLab-score of 0 has a negative predictive

312 value (NPV) of 0.997 (95% CI: 0.994 to 0.999) and positive predictive value (PPV) of 0.115

313 (0.0932 - 0.141), one third (38.4%, 95% CI: 26.4 to 48.4%) of all ED presentations were

314 assigned this score and can therefore be safely excluded. Conversely, 6.2% (95% CI: 6.3 to

315 7.2%) of the ED patients had a CoLab-score = 5. Given the PPV of this score (0.682, 95% CI:

316 0.622 to 0.740, NPV: 0.970, 95% CI: 0.962 - 0.977), subsequent PCR testing is advised.

1
2
3 317
4
56 318 *Temporal validation*
7

8 319 As the CoLab-score was developed in our center after the first COVID-19-wave in the
9
10 320 Netherlands, the performance was evaluated in our center from July 2020 until October 2021.
11
12 321 Lab results from 17489 ED presentations were collected. After applying the inclusion flow as
13
14 322 shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a
15
16 323 COVID-19 infection.
17

18
19
20 324 The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal
21
22 325 validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the
23
24 326 development cohort, although sensitivity is slightly lower and specificity slightly higher, 95%
25
26 327 CIs overlap (cf. **Table 3** and **Table 4**). The temporal validation dataset was also split into
27
28 328 three phases according to dominant SARS-CoV-2 variants and vaccine roll-out (see
29
30 329 **Supplemental Material 2 Figure 1**). The discriminative ability is not affected by phases with
31
32 330 different dominant variants and/or vaccination status. Diagnostic performance is also
33
34 331 preserved in terms of sensitivity and specificity, PPV and NPV are difficult to compare due to
35
36 332 different prevalence/pre-test probabilities in each phase (see **Supplemental Material 2 Table**
37
38 333 **1**).
39
40
41
42
43

44 334 In terms of the predicted probabilities, model calibration shows that overall predicted
45
46 335 probabilities are too low (see **Supplemental Material 3** for the calibration plot), which is
47
48 336 expected since the prevalence differs and the intercept has to be adjusted to the prevalence.
49
50 337 In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that
51
52 338 initially did not present with COVID-specific symptoms. Most patients had neurological or
53
54 339 orthopedic presenting symptoms.
55
56
57
58
59
60

340

341 *External validation*

342 For external validation, data obtained from three other centers were used, center 1 (N = 1284,
 343 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545,
 344 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19
 345 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was
 346 lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of
 347 the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891
 348 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

349 Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers
 350 is ≥ 0.96 . Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for
 351 external centers are shown in **Supplemental Material 3**, the observed fraction of COVID-19
 352 positives is slightly lower than expected in centers 1 and 2. For center 3, low probabilities
 353 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)

354

1
2
3 355 **Table 4: Diagnostic performance of the CoLab-score in the validation dataset (temporal)**
4
5 356 **and three external hospitals.**
6
7 357 *Sensitivities, specificities, positive predictive values (PPV) and negative predictive values*
8 358 *(NPV) are shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence*
9
10 359 *intervals in parentheses. Note that “0” lists the sensitivity and NPV of CoLab-score 0 and “ \leq*
11
12 360 *4” lists the specificity and PPV of CoLab-score 5.*
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

361 Discussion

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

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

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

1
2
3 384 Since this is a retrospective case-control study, there is some unavoidable missing data. In our
4
5 385 cohort 17.6% of the ED presentations could not be used due to one or more missing
6
7 386 laboratory results. This is lower or equal to similar studies; 22% [22], 17% [19] and 11% [25].
8
9
10 387 We do not expect that presentations with missing data have led to severe inclusion bias,
11
12 388 important to note is that 7.7% of missingness is due to analytical errors which should not
13
14 389 cause bias. For the remaining 9.9% of missingness, the full lab panel was most frequently
15
16 390 missing for pediatric, obstetric and surgery patients which are rarely COVID-19 patients. This
17
18 391 is also the case for external validation centers 1 and 2, in these centers only internal medicine
19
20 392 ED presentations were tested with a laboratory panel containing the 10 tests required for the
21
22 393 CoLab-score. The ED lab panel of other disciplines (e.g. urology, surgery or pediatrics)
23
24 394 differed and did not contain the required tests. Nevertheless, the majority of COVID-19
25
26 395 patients were internal medicine ED presentations, which is reflected by the few PCR-positive
27
28 396 patients excluded.

29
30
31
32
33 397 The performance of the CoLab-score is affected by the time between the onset of symptoms
34
35 398 and ED presentations. The score increases with the duration of symptoms and gradually
36
37 399 decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of
38
39 400 COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some
40
41 401 COVID-19 patients with early or late presentation after onset of symptoms can be missed.
42
43 402 Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and
44
45 403 <10 days prior to ED presentation.

46
47
48
49
50 404 It was chosen to exclude re-presentations. Since the median time between initial presentation
51
52 405 and re-presentation was 12 days, these patients were most likely not re-infected patients, but
53
54 406 patients who deteriorated after initial presentation/treatment. Given that the CoLab-score
55
56 407 follows the host-immune response, the score is time sensitive (see **Supplemental Material 4**
57
58 408 **Figure 1**). Including these patients would impact the performance of the CoLab-score as

1
2
3 409 patients in a later phase of the disease show different biomarker profiles. The CoLab-score is
4
5 410 aimed towards alerting clinicians to patients presenting with a novel SARS-CoV-2 infection,
6
7 411 rather than patients who deteriorate after treatment for COVID-19.
8
9

10 412 Finally, the CoLab-score could lead to false positives by other viral infections. However, in an
11
12 413 historic patient cohort, the CoLab-score had only limited discriminative ability in separating
13
14 414 influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4**
15
16 415 **Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-
17
18 416 response to the virus, it is expected that the CoLab-score is also sensitive to future SARS-
19
20 417 CoV-2 variants. This is supported by the fact that the diagnostic performance is sustained in
21
22 418 periods with different dominant variants. Moreover, there is no evidence that the diagnostic
23
24 419 performance is affected by vaccinations. Although vaccination status is not registered for all
25
26 420 presenting patients, there is no evidence that performance is reduced under increasing degrees
27
28 421 of vaccination. In a small subgroup of 12 patients for whom vaccination status was registered,
29
30 422 and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5) (see
31
32 423 **Supplemental Material 2 Figure 2**),
33
34
35
36
37
38

39 424 To conclude, the CoLab-score developed and validated in this study, based on 10 routine
40
41 425 laboratory results and age, is available within 1 hour for any patient presenting at the ED. The
42
43 426 score can be used by clinicians to guide PCR testing or triage patients and helps to identify
44
45 427 COVID-19 in asymptomatic patients. The lowest CoLab-score can be used to effectively rule-
46
47 428 out a possible SARS-CoV-2 infection, the highest score to alert physicians to a possible
48
49 429 infection. Thus, the CoLab-score is a valuable tool to rule out COVID-19, guide PCR testing
50
51 430 and is available to any center with access to routine laboratory tests.
52
53
54
55

56 431
57
58
59
60

432 **Funding statement**

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

434

435 **Competing interests**

436 A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no
437 conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT
438 reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of
439 interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no
440 conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR
441 reports no conflict of interest. VS reports no conflict of interest.

442

443 **Data sharing statement**

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

447

448 **Author contributorship statement**

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

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

- 1
2
3 455 Maaïke Maas: Conceptualization (Supporting), Resources (Supporting), Supervision
4
5 456 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
6
7
8 457 Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision
9
10 458 (Supporting), Validation (Equal), Writing-review & editing (Equal).
11
12
13 459 Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision
14
15 460 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
16
17
18 461 Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision
19
20 462 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
21
22
23 463 Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision
24
25 464 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
26
27
28 465 Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
29
30
31 466 Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
32
33
34 467 Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
35
36
37 468 Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
38
39
40 469 Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
41
42
43 470 Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing
44
45 471 (Equal).
46
47
48 472 Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal),
49
50 473 Writing-review & editing (Equal).
51
52
53 474 Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project
54
55 475 administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing
56
57 476 (Equal).
58
59
60

477 **References**

- 478 1 Coronavirus Disease (COVID-19) Situation Reports.
479 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
480 (accessed 4 Feb 2021).
- 481 2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in
482 China. <https://doi.org/10.1056/NEJMoa2002032> 2020;**382**:1708–20.
483 doi:10.1056/NEJMOA2002032
- 484 3 Vetter P, Vu DL, L’Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**.
485 doi:10.1136/BMJ.M1470
- 486 4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and
487 prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18.
488 doi:10.1136/bmj.m1328
- 489 5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine
490 Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19):
491 A Systematic Review. *J. Med. Syst.* 2020;**44**:122. doi:10.1007/s10916-020-01582-x
- 492 6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical
493 Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417
- 494 7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT
495 Assessment Scheme for Patients Suspected of Having COVID-19—Definition
496 and Evaluation. <https://doi.org/10.1148/radiol2020201473> 2020;**296**:E97–104.
497 doi:10.1148/RADIOL.2020201473
- 498 8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity
499 score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort

- 1
2
3 500 study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840
4
5
6 501 9 R Core Team. R: A Language and Environment for Statistical Computing.
7
8 502 2020.<https://www.r-project.org/>
9
10
11 503 10 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29.
12
13 504 doi:10.1198/016214506000000735
14
15
16 505 11 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B*
17
18 506 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x
19
20
21
22 507 12 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models
23
24 508 via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01
25
26
27 509 13 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63.
28
29 510 doi:10.1093/oxfordjournals.pan.a004868
30
31
32 511 14 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation.
33
34 512 *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973
35
36
37 513 15 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to
38
39 514 analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-
40
41 515 2105-12-77
42
43
44
45 516 16 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.[https://cran.r-](https://cran.r-project.org/package=rms)
46
47 517 [project.org/package=rms](https://cran.r-project.org/package=rms)
48
49
50 518 17 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons
51
52 519 and progress. *Scand. J. Clin. Lab. Invest.* 1999;**59**:491–500.
53
54 520 doi:10.1080/00365519950185229
55
56
57
58 521 18 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from
59
60

- 1
2
3 522 Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst*
4
5 523 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4
6
7
8 524 19 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of
9
10 525 SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J*
11
12 526 *Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502
13
14
15 527 20 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine
16
17 528 learning tool to rule out COVID-19 among adults in the emergency department using
18
19 529 routine blood tests: A large, multicenter, real-world study. *J Med Internet Res*
20
21 530 2020;**22**:e24048. doi:10.2196/24048
22
23
24
25 531 21 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical
26
27 532 and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807.
28
29 533 doi:10.1007/s00330-020-07022-1
30
31
32
33 534 22 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-
34
35 535 infected patients at the emergency department using routine testing. *Clin Chem Lab Med*
36
37 536 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593
38
39
40
41 537 23 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a
42
43 538 diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int*
44
45 539 *J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305
46
47
48
49 540 24 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in
50
51 541 the emergency department: derivation and validation of the LDH-CRP-Lymphocyte
52
53 542 (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336
54
55
56 543 25 Tordjman M, Mekki A, Mali RD, *et al.* Pre-test probability for SARS-Cov-2-related
57
58 544 infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342.
59
60

1
2
3 545 doi:10.1371/journal.pone.0243342
4
5

6 546 26 Peto T, Affron D, Afrough B, *et al.* COVID-19: Rapid antigen detection for SARS-CoV-
7
8 547 2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for
9
10 548 mass-testing. *EClinicalMedicine* 2021;**36**:100924.
11
12 doi:10.1016/J.ECLINM.2021.100924
13 549
14

15
16 550 27 García-Fiñana M, Hughes DM, Cheyne CP, *et al.* Performance of the Innova SARS-
17
18 551 CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot:
19
20 552 population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637
21
22

23 553 28 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical
24
25 554 data for patients attending hospital: development and prospective validation of an
26
27 555 artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87.
28
29 556 doi:10.1016/S2589-7500(20)30274-0
30
31

32
33 557
34
35

36 558
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 559 **Figure legends**
4

5 560
6
7

8 561 **Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B)**
9
10 562 **dataset.**

11
12
13 563 *All patient admissions with routine venous blood sampling at the emergency department (ED)*
14
15 564 *were included. For the development dataset, completeness of the lab panel was assessed for*
16
17 565 *all the 29 laboratory tests (see Table 1), for the temporal validation dataset this was only*
18
19 566 *necessary for 10 laboratory tests (see Table 2). The major causes of missingness are*
20
21 567 *described in the text. In the development dataset, presentations with extreme values (>10 SD)*
22
23 568 *were excluded. The same limits were applied to the temporal validation dataset (see Table 2*
24
25 569 *for limits).*

26
27
28
29
30 570
31
32

33 571 **Figure 2: Probability density plot of the CoLab-linear predictor.**

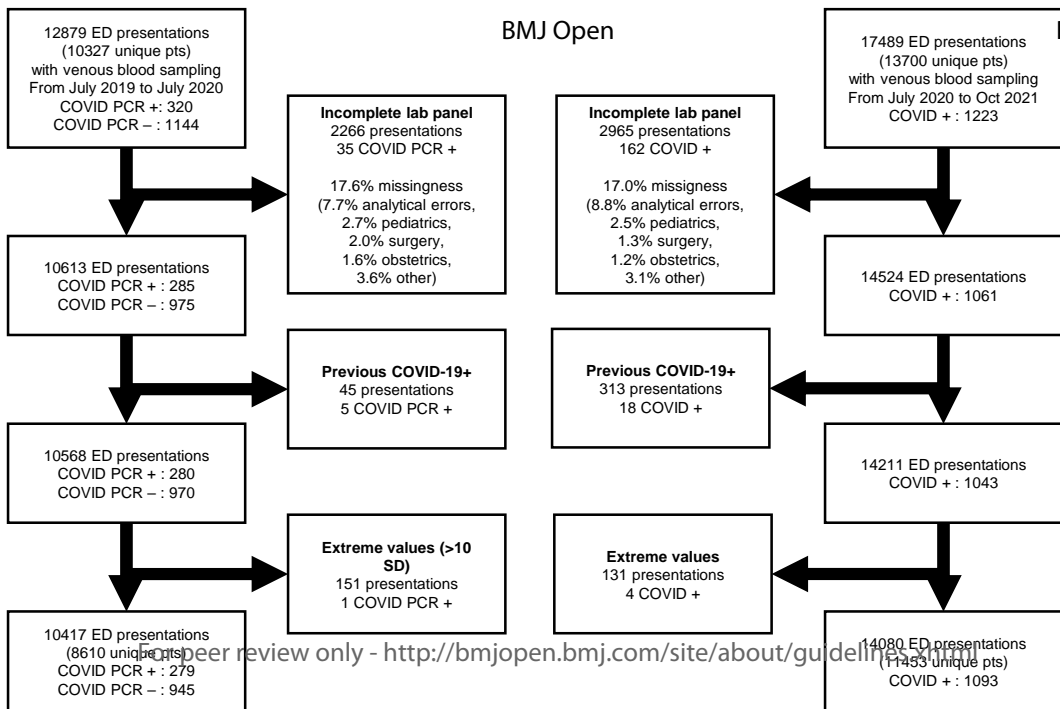
34
35
36 572 *The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are*
37
38 573 *plotted against the linear predictor (see table 2). The CoLab-score cut-offs (-5.83 , -4.02 , $-$*
39
40 574 *3.29 , -2.34 and -1.64) are depicted with vertical dashed lines. The white-boxed numbers*
41
42 575 *(between the cut-offs) represent the corresponding CoLab-score. Note that while the area*
43
44 576 *under both curves is identical (since these are probability density functions), in absolute*
45
46 577 *numbers the “negative or untested”-group is about 36 times larger than the PCR positive*
47
48 578 *group.*

49
50
51
52
53 579
54
55

56 580 **Figure 3: Inclusion flow of ED patients in three external centers.**
57
58
59
60

1
2
3 581 *All emergency department (ED) presentations with routine venous blood sampling were*
4
5 582 *included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see*
6
7 583 *Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were*
8
9 584 *excluded as “previous COVID-19+”. Presentations with any laboratory result above the*
10
11 585 *limits of the CoLab-score (see Table 2) were excluded.*
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only



CoLab-score

Threshold

SARS CoV-2 PCR

- Negative or untested
- Positive

BMJ Open

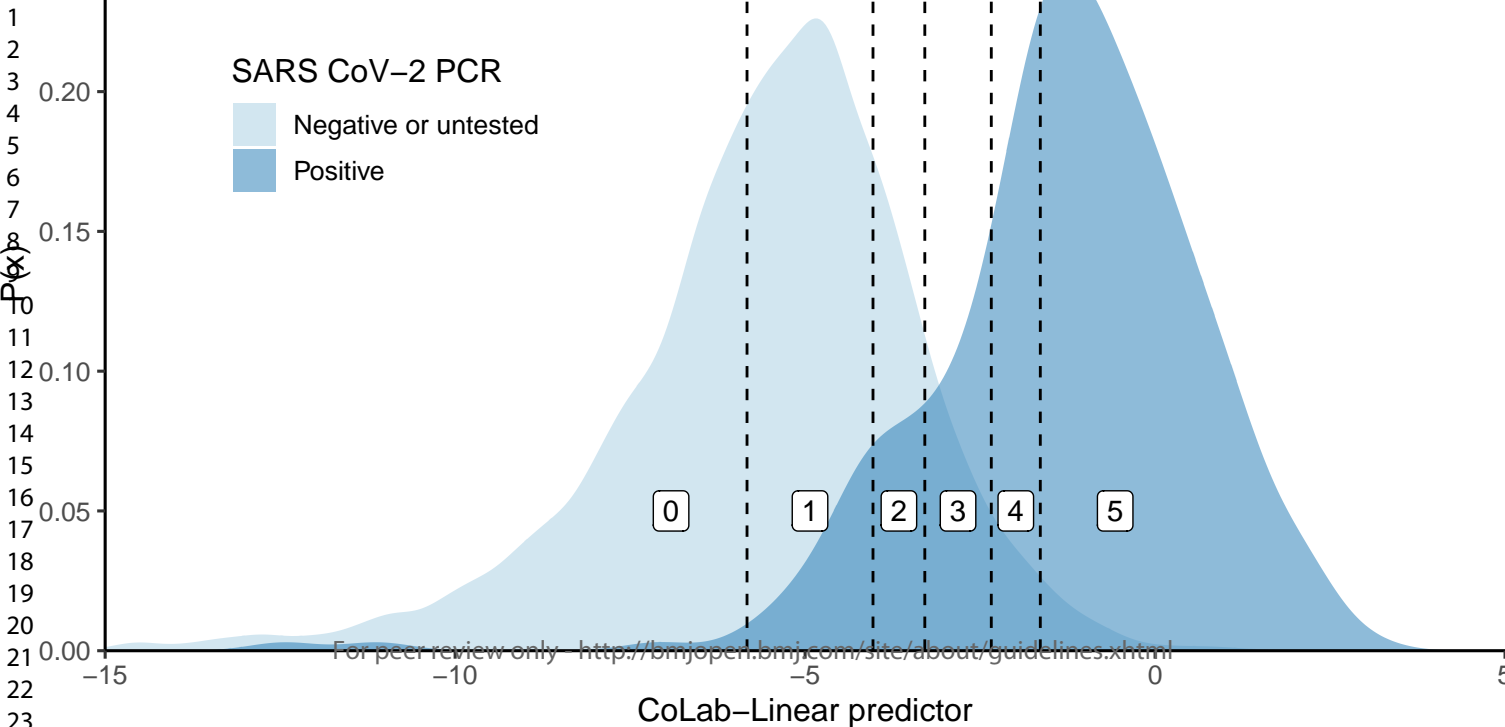
-3.29

-1.64

-4.03

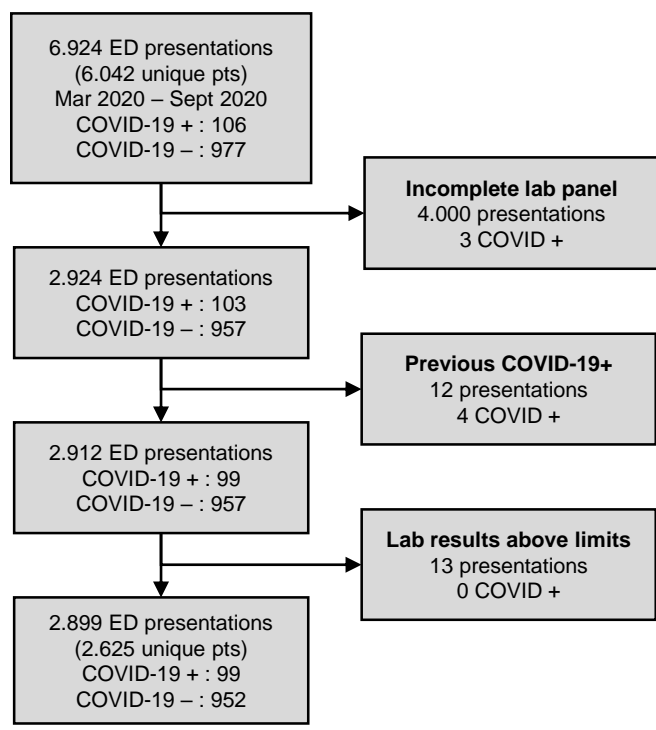
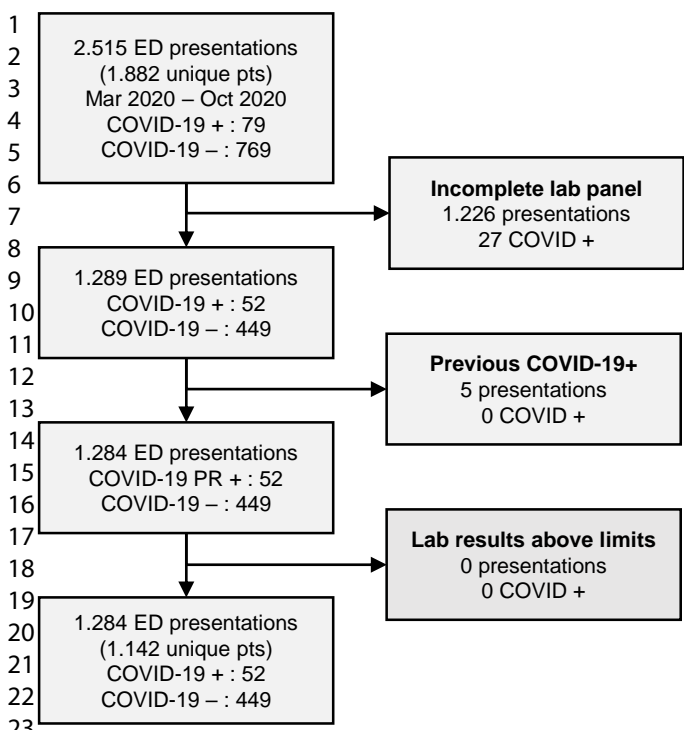
-2.34

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

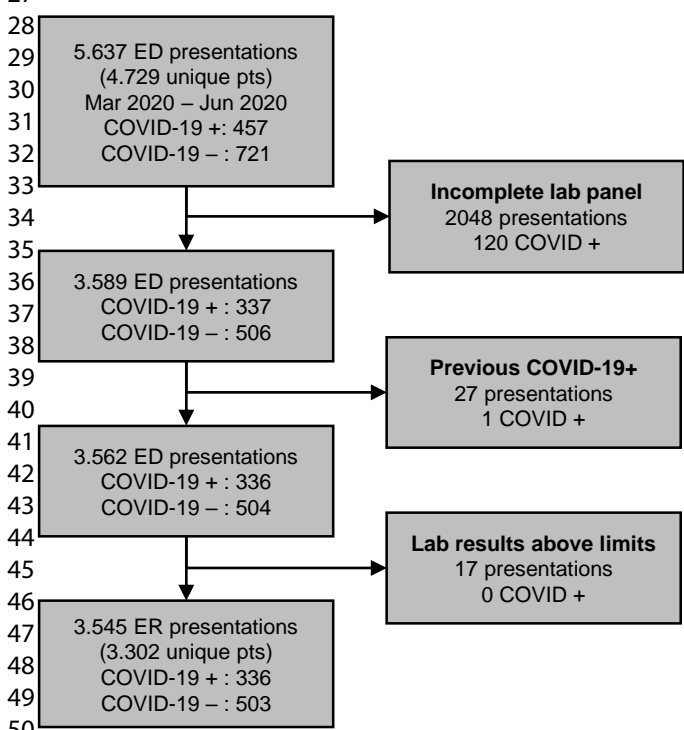


Center 1

Center 2



Center 3



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 (\hat{w}) was calculated by $\hat{w} = 1/|\hat{\beta}_{ridge}|^2$. 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: $lp_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in}$. Where n is the number of variables in the final model, x_{in} are the observed predictor variables for subject i and β_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:

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

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:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639$$

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$

$$lp_i(\tau) \geq \ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

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

$$lp_i(\tau = 0.4) \geq -5.83 \text{ (CoLab-score = 1)}$$

$$lp_i(\tau = 0.1) \geq -4.03 \text{ (CoLab-score = 2)}$$

$$lp_i(\tau = 0.05) \geq -3.29 \text{ (CoLab-score = 3)}$$

$$lp_i(\tau = 0.02) \geq -2.34 \text{ (CoLab-score = 4)}$$

$$lp_i(\tau = 0.01) \geq -1.64 \text{ (CoLab-score = 5)}$$

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

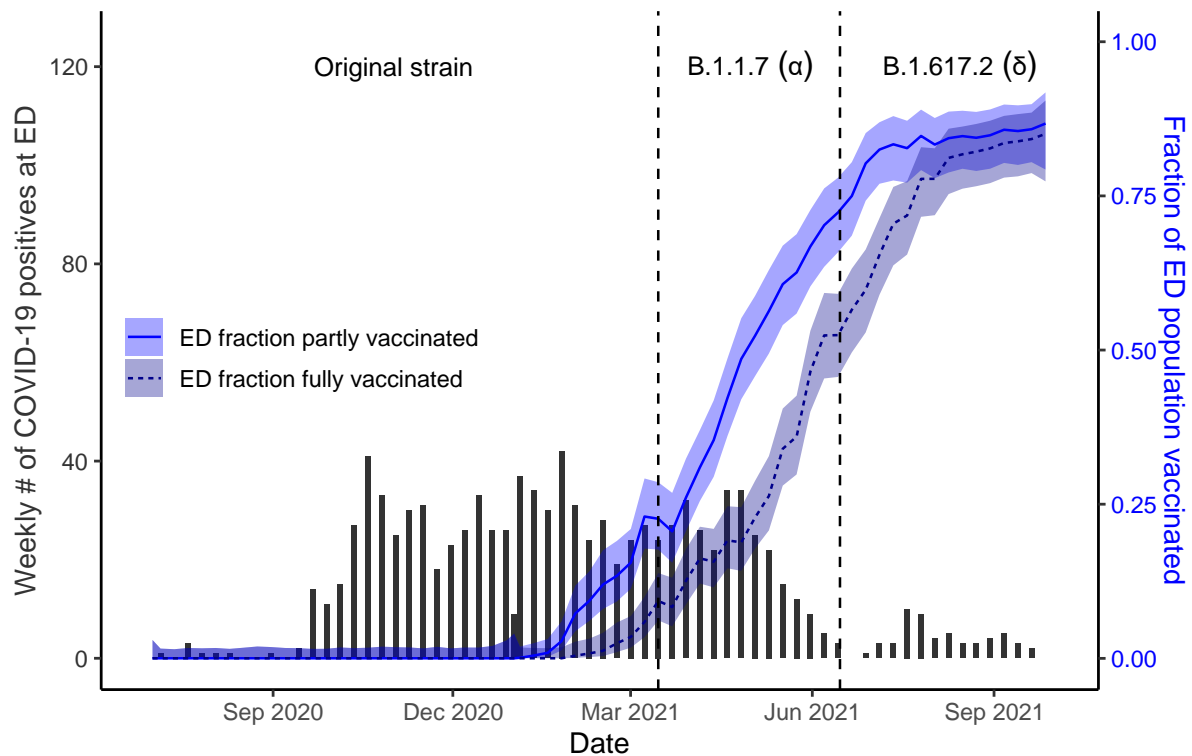


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

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.

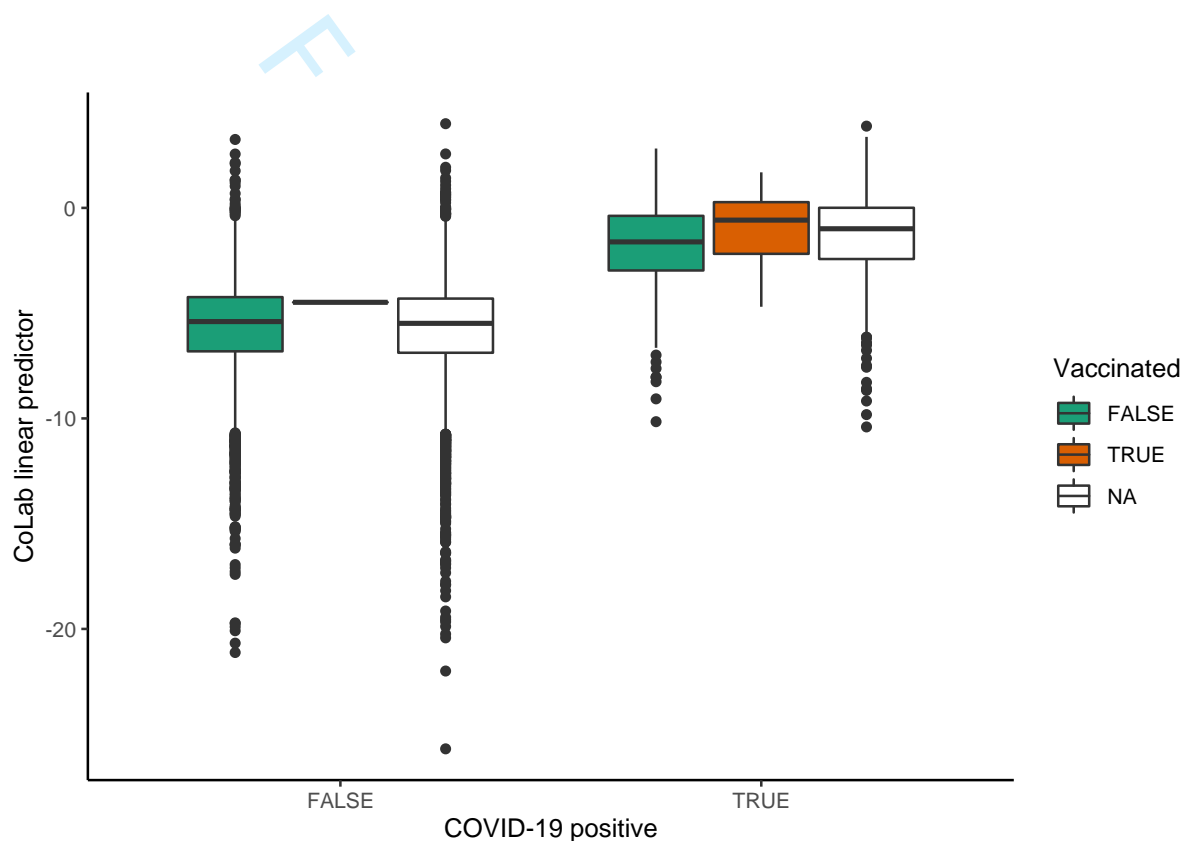


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

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.

1
2
3 *Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or*
4 *considered positive until proven otherwise, therefore there is only one COVID-19 negative*
5 *patient with a registered vaccination status.*
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Supplemental material 3

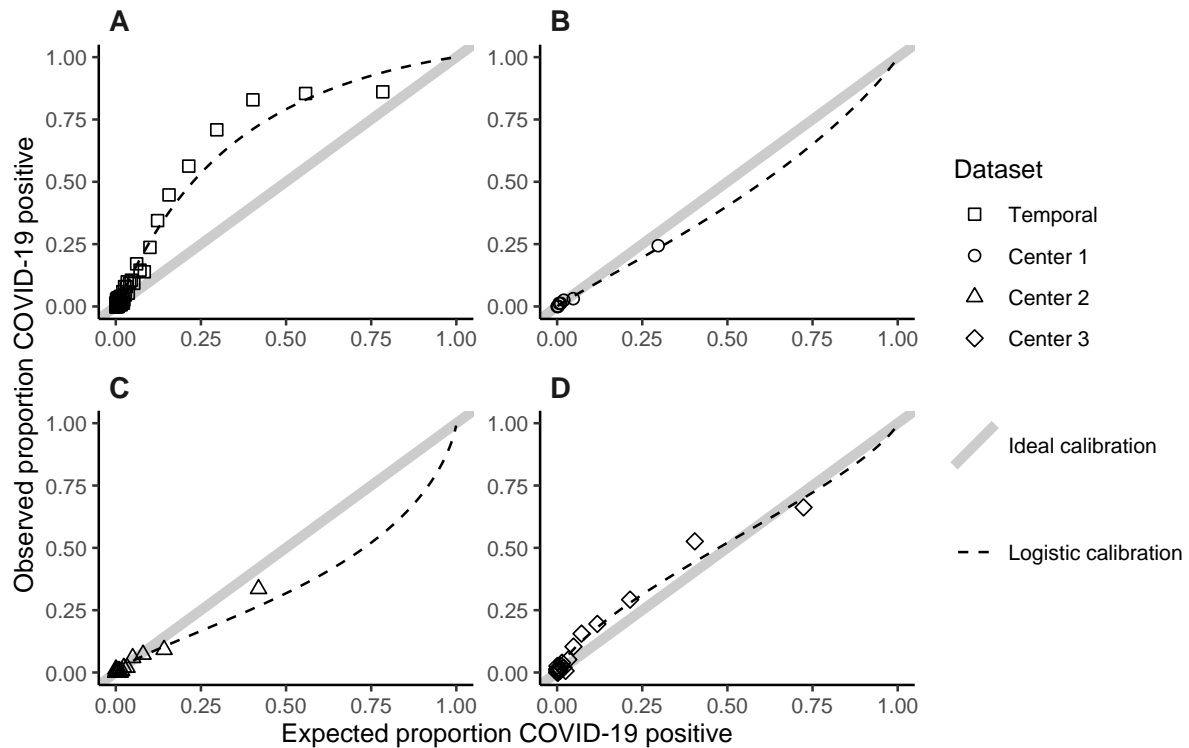


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

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.

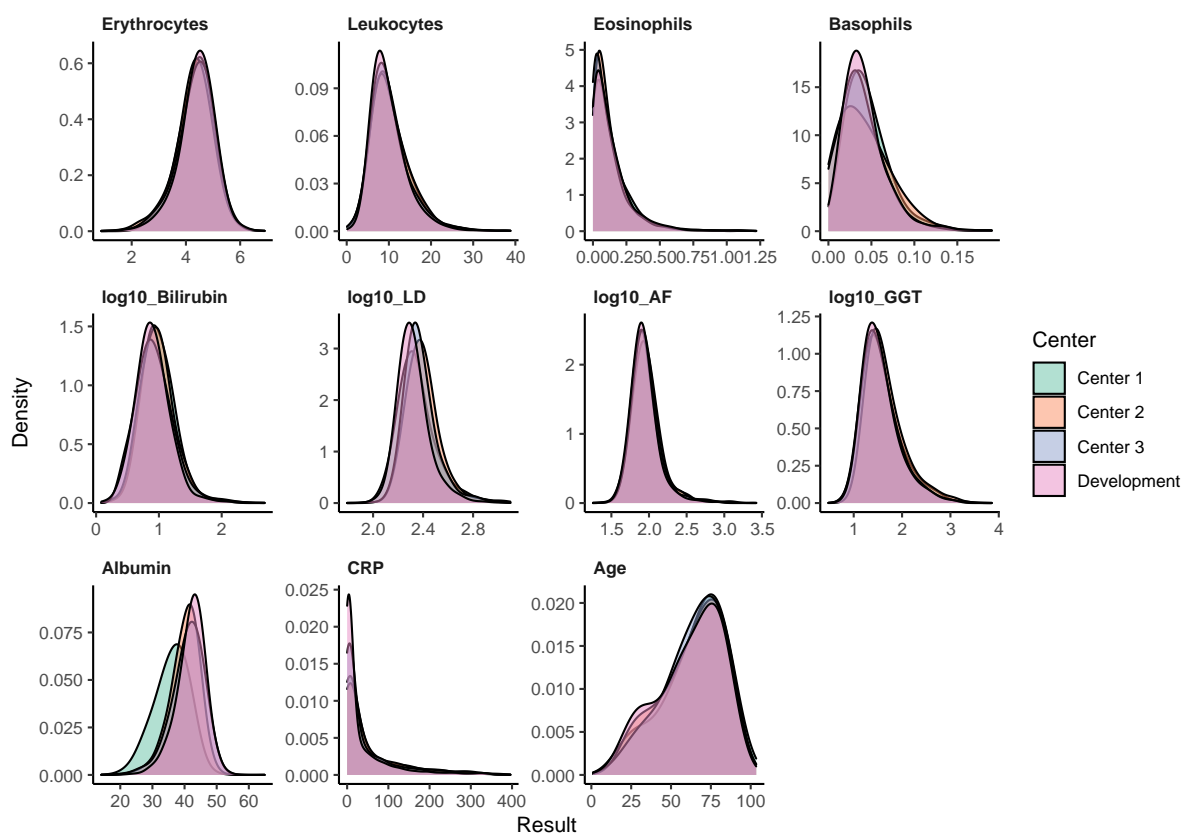


Figure 2: Probability density plots of laboratory parameters.

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

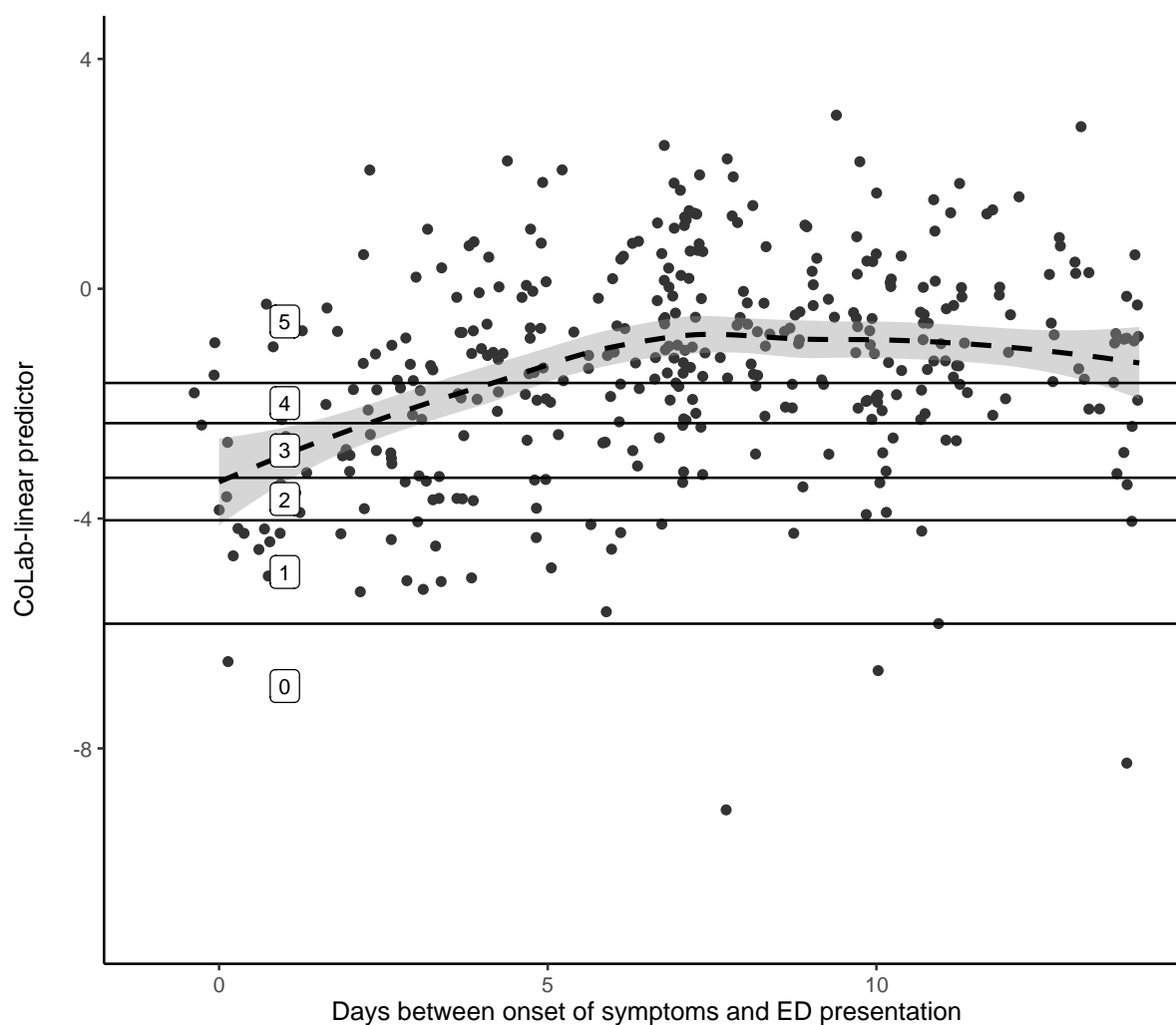


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predictor 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.

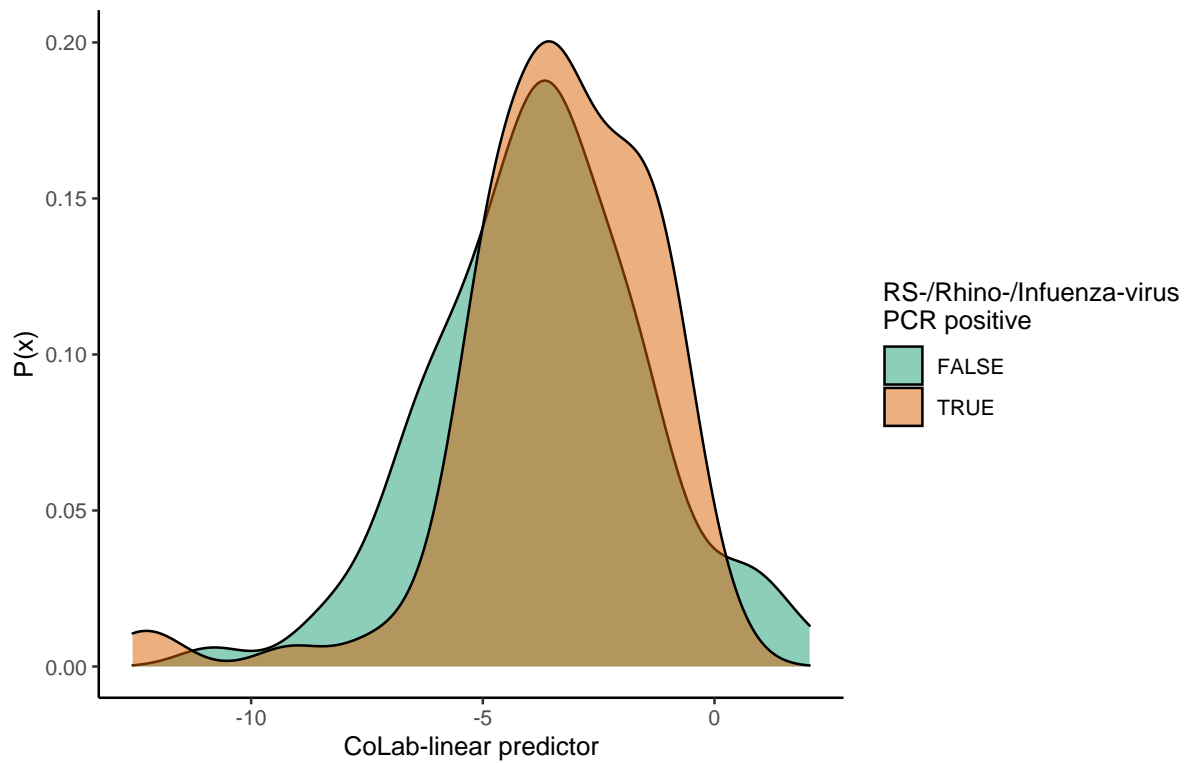


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

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: 0.4896-0.6563).

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 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:	<i>BMJ Open</i>
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, Maaïke; 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

SCHOLARONE™
Manuscripts



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.

1
2
3
4 1 **Development and validation of an early warning score to identify**
5
6 2 **COVID-19 in the emergency department based on routine laboratory**
7
8
9 3 **tests: a multicenter case-control study**
10
11
12 4

15 5 Arjen-Kars Boer^{1*}, Ruben Deneer^{1,2**†}, Maaïke Maas³, Heidi S. M. Ammerlaan⁴, Roland H. H.
16
17 6 van Balkom⁵, Wendy A.M.H. Thijssen³, Sophie Bennenbroek³, Mathie P. G. Leers⁶, Remy
18
19 7 J.H. Martens⁶, Madelon M. Buijs⁷, Jos J. Kerremans⁸, Muriël Messchaert⁹, Jeroen D.E. van
20
21 8 Suijlen⁹, Natal A.W. van Riel^{2,10} and Volkher Scharnhorst^{1,2}
22
23
24
25 9

26
27
28 10 * Both authors contributed equally
29

30
31 11 † Corresponding author
32
33
34 12

35
36
37 13 1 Department of Laboratory Medicine, Catharina Hospital Eindhoven,
38
39 14 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

40
41 15 2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene
42
43 16 Loper 3, 5612 AE, Eindhoven, the Netherlands

44
45
46 17 3 Department of Emergency Medicine, Catharina Hospital Eindhoven,
47
48 18 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

49
50 19 4 Department of Internal Medicine, Catharina Hospital Eindhoven,
51
52 20 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

53
54
55 21 5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven,
56
57 22 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands
58
59
60

- 1
2
3 23 6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der
4
5 24 Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands
6
7 25 7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the
8
9 Netherlands
10 26
11
12 27 8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital,
13
14 28 Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands
15
16
17 29 9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334
18
19 30 DZ Apeldoorn, the Netherlands
20
21 31 10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef
22
23 32 9, 1105 AZ Amsterdam, the Netherlands.
24
25
26
27 33
28
29

30 **Keywords**

31
32 35 COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction
33
34 36 model, routine laboratory tests
35
36
37
38 37
39

40 **Corresponding author**

41 38
42
43 39 Ruben Deneer
44
45
46 40 Michelangelolaan 2
47
48
49 41 5623 EJ Eindhoven, the Netherlands
50
51
52 42 ruben.deneer@catharinaziekenhuis.nl
53
54
55
56
57
58
59
60

43 **Abstract**

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

49 **Design:** Case-control study.

50 **Setting:** Secondary and tertiary hospitals in the Netherlands.

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

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

58 **Results:** The resulting “CoLab-score” consists of 10 routine laboratory measurements, and
59 age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945).
60 The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991,
61 specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity
62 (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were
63 confirmed in temporal and external validation.

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

1
2
3 66 ruled-out in over one third of ED presentations. Highly suspect cases can be identified
4
5 67 regardless of presenting symptoms. The CoLab-score is continuous, in contrast to the binary
6
7 68 outcome of lateral flow testing, and can guide PCR testing and triage ED patients.
8
9

10
11 69

12 13 14 70 **Article summary**

15 16 17 71 Strengths and limitations of this study

- 18
19 72 • A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency
20
21 73 department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- 22
23 74 • Using adaptive lasso regression analysis, the panel of 28 laboratory tests was reduced
24
25 75 to a single score consisting of a subset of 10 routine ED laboratory tests and age.
- 26
27 76 • The score was temporally validated from July 2020 to October 2021, in the presence of
28
29 77 vaccine roll-out and emergence of new SARS-CoV-2 variants.
- 30
31 78 • The score was externally validated in 3 other centers in the Netherlands.
- 32
33 79 • Missingness in the panel of laboratory tests varied between external centers, limiting
34
35 80 generalizability of the score to the ED population for which the complete panel of
36
37 81 laboratory tests was available.
38
39
40
41
42

43 82
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

83 Introduction

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

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

98 Many COVID-19 prediction models have already been developed, the living systematic
99 review by Wynants et. al [4] provides an extensive overview and critical appraisal.

100 Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early
101 models were based on relatively small sample sizes, hampered by selection bias or were over-
102 fitted by selecting too many features [4–6]. Aside from methodological shortcomings of early
103 models, most models are not developed as an early warning score for all ED patients. Firstly,
104 they require features from tests that are not routinely performed or logged for all ED patients
105 (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the
106 PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED
107 patient population. Secondly, the population on which models are commonly based, are PCR-

1
2
3 108 tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by
4
5 109 physicians.

7
8 110 Only two studies were identified that focus on patients presenting at the ED, include
9
10 111 unsuspected (and pre-pandemic) patients as controls, and rely solely on routine (laboratory)
11
12 112 tests [9,10].

15 113 In this study we report the development and validation of an early warning score that, based
16
17 114 on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in patients
18
19 115 who undergo routine laboratory testing at presentation. The score can assist ED physicians in
20
21 116 triaging patients and prevent further transmission of COVID-19 by quickly identifying
22
23
24 117 possibly infected patients or ruling out a possible infection when resources are scarce.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

118 **Methods**

119 *Study design*

120 This is a retrospective case-control study where routine laboratory test results, combined with
121 age and gender, from all patient presenting at the emergency department (ED) of the
122 Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-
123 2 PCR test results in a development dataset. A model that could predict the presence of a
124 COVID-19 infection was fit to this dataset. Performance of the model was assessed by i)
125 internal validation, ii) temporal validation and iii) external validation by using data from the
126 ED of three other centers. The study was reviewed by the Medical research Ethics
127 Committees United (MEC-U) under study number W20.071, which confirmed that the
128 Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this
129 study. The study was thereafter reviewed and approved by the internal hospital review board.

130

131 *Patient and Public Involvement*

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

133

134 *Development dataset*

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

1
2
3 142 quantitative results were not available due to analytical interference (hemolysis, lipemia or
4
5 143 icterus). The routine ED laboratory panel is requested for (adult) patients presenting with
6
7 144 abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific
8
9 145 complaints, or for patients (including non-adult patients) presenting with specific complaints
10
11 146 where a suspected diagnosis has to be ruled-in or ruled-out. Presentations with one or more
12
13 147 missing values in any of the 28 laboratory test in the routine ED panel, were excluded.
14
15 148 Presentations with one or more extreme lab results, > 10 times standard deviation from the
16
17 149 median, were also excluded to minimize the effect on the estimation of regression
18
19 150 coefficients. The median was chosen as a measure of central tendency due to its resistance for
20
21 151 outliers. After the first case of COVID-19 in the Netherlands, all patients with symptoms of
22
23 152 COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR
24
25 153 testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were
26
27 154 approved by the Dutch national institute of public health (RIVM). If a patient had a positive
28
29 155 PCR result in the past, subsequent presentations were excluded as re-presentations might be
30
31 156 clinically different from de novo presentations.
32
33
34
35
36
37
38 157 The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying
39
40 158 nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal
41
42 159 at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was
43
44 160 positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time
45
46 161 window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were
47
48 162 performed in the time window and the presentation occurred after the first case of COVID-19
49
50 163 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the
51
52 164 first case were labelled as "*Pre-COVID-19*".
53
54
55
56
57
58
59
60

166 *Laboratory tests*

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

179

180 *Modelling*

181 All data were processed and analyzed in R version 4.1.1 [11]. Laboratory results, combined
182 with age and gender were used as covariates in a regression model. Cases were defined as ED
183 presentations labelled as “*PCR-positive*”, controls were all other presentations (i.e. “*PCR-*
184 *negative*”, “*Untested*” or “*Pre-COVID-19*”). To achieve predictive accuracy, limit overfitting
185 and perform feature selection, penalized logistic regression with an adaptive lasso penalty was
186 chosen [12,13]. To minimize missing data, all non-numeric results at the extremes of the
187 measuring range, were converted to numeric results by removing the “<” and “>” signs. For
188 eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m² and
189 <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD,
190 CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model

1
2
3 191 predictions, these variables were transformed logarithmically. More details regarding model
4
5 192 fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the
6
7 193 glmnet-package [14].
8
9

10
11 194

12 13 195 *CoLab-score*

14
15
16 196 Since this is a retrospective case-control study, the sample prevalence may not reflect the
17
18 197 true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term
19
20 198 in the model should be adjusted according to the current prevalence (details can be found in
21
22 199 the document, **Supplemental Material 1**) [15]. However, adjusting the intercept term is not
23
24 200 straightforward to implement in clinical practice, therefore the linear predictor of the model
25
26 201 was categorized into a score, this score is hereafter referred to as the “CoLab-score”. The
27
28 202 categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15
29
30 203 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using
31
32 204 the intercept adjustment formula by King [15]. The intervals obtained through these breaks
33
34 205 correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and
35
36 206 score 5 reflects high-risk. More details regarding the rationale of the CoLab-score
37
38 207 categorization can be found in the document, **Supplemental Material 1**.
39
40
41
42
43

44 208

45 46 47 209 *Internal validation*

48
49 210 To assess model performance while taking overfitting into account, bootstrapping was
50
51 211 performed. 1000 bootstrap samples were generated from the original data. On each bootstrap
52
53 212 sample, the full model fitting procedure and CoLab-score conversion were performed.
54
55 213 Optimism adjusted performance measures of the CoLab-score were obtained by applying the
56
57 214 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [16]. Performance
58
59
60

1
2
3 215 measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative
4
5 216 predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate
6
7 217 performance measures [17]. Although the full inclusion period from July 2019 to July 2020
8
9 218 was used for model fitting, the performance was evaluated on the period starting from the first
10
11 219 COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain
12
13 220 performance measures that would reflect real world performance.
14
15
16
17
18 221

20 222 *Temporal validation*

23 223 For temporal validation, results from our center were prospectively analyzed from July 2020
24
25 224 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-
26
27 225 19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period
28
29 226 there was also more widespread external PCR testing by municipal health services. The
30
31 227 results of external conducted PCR tests were not available to our study. To overcome this
32
33 228 limitation, the outcome in the temporal validation cohort was chosen as a composite of the
34
35 229 hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test
36
37 230 result. This period also covers both the emergence of new SARS-CoV-2 variants as well as
38
39 231 vaccine rollout. However, neither vaccination status nor genomic sequencing was available to
40
41 232 determine whether a patient was vaccinated or which variant caused the infection. Therefore,
42
43 233 data from the Dutch national institute of public health (RIVM) was used, to divide the
44
45 234 temporal validation period into three phases: i) from July 2020 until March 2021, no
46
47 235 vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial
48
49 236 vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until
50
51 237 October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant.
52
53 238 See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted
54
55 239 of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold
56
57
58
59
60

1
2
3 240 for the entire period, as well as for each phase separately to determine a possible effect of
4
5 241 vaccination and new variants on performance (results in the **Supplemental Material 2**).
6
7 242 Model calibration was assessed graphically using the rms-package [18].
8
9

10 243

11
12
13 244 *External validation*

14
15 245 For the external validation, several centers in the Netherlands were approached and assessed
16
17 246 if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available.
18
19 247 Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals
20
21 248 (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and
22
23 249 Zuyderland Medical Center (center 3). The hematological parameters were measured with
24
25 250 Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and
26
27 251 Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with
28
29 252 Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott
30
31 253 Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external
32
33 254 validation was similar to the temporal validation and consisted of assessing the AUC
34
35 255 sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was
36
37 256 assessed graphically analogous to the temporal validation dataset.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

257 Results

258 Development dataset

259 12879 emergency department (ED) presentations of 10327 patients from July 2019 to July
 260 2020 were included. After excluding cases with an incomplete lab panel, patient presentations
 261 that occurred after a positive PCR test in the past (re-presentations) and presentations with
 262 extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of
 263 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/m ²	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]

264

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

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

275

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

281

282 **CoLab-score**

283 The model obtained through adaptive lasso regression contained eleven variables, which are
 284 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 μ mol/L	-1.142	Bilirubin >169 μ mol/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 %

285

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

287 *The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of*

288 *the 11 variables with their corresponding coefficients (β 's). CoLab-LP = - 6.885 +*

289 *[erythrocytes] \times 0.9379 - [leukocytes] \times 0.1298 - [eosinophils] \times 6.834 - [basophils] \times*

290 *47.7 - log₁₀([bilirubin]) \times 1.142 + log₁₀([LD]) \times 5.369 - log₁₀([ALP]) \times 3.114 +*

291 *log₁₀([gGT]) \times 0.3605 - [albumin] \times 0.1156 + [CRP] \times 0.02560 + [age] \times 0.002275. The*

292 *LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence*

293 *is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid*

294 *if any of the variables exceed the limits in the third column. The relative importance ranks the*

295 *importance of variables in predicting the outcome, relative to the most important variable (in*

296 *this case basophils).*

297

298 A larger β -coefficient does not imply that a variable is more important in predicting the odds

299 of testing positive for SARS-CoV-2, since variables are on different scales. The most

300 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)

309

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

1
2
3 317 *considered positive. Note that “0” lists the sensitivity and NPV of CoLab-score 0 and “ ≤ 4 ”*
4
5 318 *lists the specificity and PPV of CoLab-score 5.*

6
7
8 319
9
10
11 320 Diagnostic performance is shown in **Table 3**. A CoLab-score of 0 has a negative predictive
12
13 321 value (NPV) of 0.997 (95% CI: 0.994 to 0.999) and positive predictive value (PPV) of 0.115
14
15 322 (0.0932 - 0.141), one third (38.4%, 95% CI: 26.4 to 48.4%) of all ED presentations were
16
17 323 assigned this score and can therefore be safely excluded. Conversely, 6.2% (95% CI: 6.3 to
18
19 324 7.2%) of the ED patients had a CoLab-score = 5. Given the PPV of this score (0.682, 95% CI:
20
21 325 0.622 to 0.740, NPV: 0.970, 95% CI: 0.962 - 0.977), subsequent PCR testing is advised.
22
23
24
25
26 326

27 28 327 *Temporal validation*

29
30 328 As the CoLab-score was developed in our center after the first COVID-19-wave in the
31
32 329 Netherlands, the performance was evaluated in our center from July 2020 until October 2021.
33
34 330 Lab results from 17489 ED presentations were collected. After applying the inclusion flow as
35
36 331 shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a
37
38 332 COVID-19 infection.

39
40
41
42
43 333 The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal
44
45 334 validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the
46
47 335 development cohort, although sensitivity is slightly lower and specificity slightly higher (cf.
48
49 336 **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases
50
51 337 according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental**
52
53 338 **Material 2 Figure 1**). The discriminative ability was not lower in the second or third phase,
54
55 339 compared to the first phase. Diagnostic performance is preserved in terms of sensitivity and
56
57 340 specificity, except a moderately reduced sensitivity of scores ≥ 3 in the third phase as

1
2
3 341 compared to the first phase. PPV and NPV are incomparable due to different prevalence/pre-
4
5 342 test probabilities in each phase (see **Supplemental Material 2 Table 1**).

7
8 343 In terms of the predicted probabilities, model calibration shows that overall predicted
9
10 344 probabilities are too low (see **Supplemental Material 3** for the calibration plot), which is
11
12 345 expected since the prevalence differs and the intercept has to be adjusted to the prevalence.

13 346 In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that
14
15 347 initially did not present with COVID-specific symptoms. Most patients had neurological or
16
17 348 orthopedic presenting symptoms.

18
19
20
21
22
23 349

24 25 350 *External validation*

26
27 351 For external validation, data obtained from three other centers were used, center 1 (N = 1284,
28
29 352 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545,
30
31 353 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19
32
33 354 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was
34
35 355 lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of
36
37 356 the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891
38
39 357 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

40
41
42 358 Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers
43
44 359 is ≥ 0.96 . Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for
45
46 360 external centers are shown in **Supplemental Material 3**, the observed fraction of COVID-19
47
48 361 positives is slightly lower than expected in centers 1 and 2. For center 3, low probabilities
49
50 362 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)
	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)
≤ 1	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)
	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)
≤ 2	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)
	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	0.750 (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)

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)

363

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

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

371 Discussion

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

376 The main strength of this study is that this score can be used as an early-warning or triaging
377 tool for the ED population presenting with abdominal pain, chest pain, shortness of breath,
378 syncope, sepsis or other non-specific complaints where a routine blood panel is requested.
379 This is in contrast to the vast majority of COVID-19 diagnostic models that have been
380 developed on a pre-selected population of PCR-tested patients [9,20–26]. Moreover, the
381 CoLab-score requires only routine blood tests, instead of (features from) imaging such as CT-
382 scans or laboratory tests that are not routinely collected in the ED, e.g. interleukin-6 or 3-
383 hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous
384 result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous
385 score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable
386 to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific)
387 [27,28].

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

1
2
3 396 Since this is a retrospective case-control study, there is some unavoidable missing data. In our
4
5 397 cohort 17.6% of the ED presentations could not be used due to one or more missing
6
7 398 laboratory results. This is lower or equal to similar studies; 22% [23], 17% [21] and 11% [26].
9
10 399 Important to note is that 7.7% of missingness is due to analytical errors which can be assumed
11
12 400 to be missing completely at random. For the remaining 9.9% of missingness, the full lab panel
13
14 401 was most frequently missing for pediatric, obstetric and surgery patients. These patients are
15
16 402 presenting with specific complaints for which specific laboratory tests are requested, and
17
18 403 hence do not match the inclusion criteria for a routine blood panel. Overall the missingness
19
20 404 was significantly lower in the PCR-tested group versus the untested group (χ^2 -test p-value
21
22 405 <0.001).

23
24
25
26 406 In the external centers, there is a high level of missingness as a result of an incomplete
27
28 407 laboratory panel. In the case of centers 1 and 2, only internal medicine ED presentations were
29
30 408 tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED
31
32 409 lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain
33
34 410 the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine
35
36 411 ED presentations, which is reflected by the few PCR-positive patients excluded. Due to these
37
38 412 high levels of missingness, the results of the external centers cannot be used to show that the
39
40 413 CoLab-score generalizes to the entire ED population. Rather, the results show that for the
41
42 414 majority of COVID-19 positive patients presenting at the ED, a routine laboratory panel is
43
44 415 available from which the CoLab-score can be calculated, and that the performance of the
45
46 416 CoLab-score in this population is comparable to the development population.

47
48
49
50
51
52 417 The performance of the CoLab-score is affected by the time between the onset of symptoms
53
54 418 and ED presentations. The score increases with the duration of symptoms and gradually
55
56 419 decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of
57
58 420 COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some

1
2
3 421 COVID-19 patients with early or late presentation after onset of symptoms can be missed.
4
5 422 Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and
6
7 423 <10 days prior to ED presentation. Chemotherapy that causes myeloid suppression, will
8
9 424 decrease neutrophilic, basophilic and eosinophilic counts and thereby “falsely” increasing the
10
11 425 CoLab-score. Conversely, COVID-19 patients with severe anemia could have “falsely”
12
13 426 lowered CoLab-scores. To minimize false negatives, we have therefore advised to report
14
15 427 CoLab-scores only when the concentration of erythrocytes is ≥ 2.9 /pL.
16
17
18
19 428 It was chosen to exclude re-presentations after a previous presentation with COVID-19. Since
20
21 429 the median time between initial presentation and re-presentation was 12 days, these patients
22
23 430 were most likely not re-infected patients, but patients who deteriorated after initial
24
25 431 presentation/treatment. Given that the CoLab-score follows the host-immune response, the
26
27 432 score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients
28
29 433 would impact the performance of the CoLab-score as patients in a later phase of the disease
30
31 434 show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to
32
33 435 patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate
34
35 436 after treatment for COVID-19. Other re-presentations were not excluded, which results in
36
37 437 some patients appearing multiple times in a dataset. This was not corrected for in the
38
39 438 regression model since the assumption was made that ED presentations are independent
40
41 439 observations. The median time between re-presentations is 38 days, most likely resulting in
42
43 440 variations in laboratory results between presentations, and hence, little to no correlation
44
45 441 between presentations. A sensitivity analysis was performed whereby only the first
46
47 442 presentation was included for each patient (**Supplemental Material 4 Table 1**), but no
48
49 443 difference was found in performance in terms of sensitivity, specificity and AUC.
50
51
52
53
54 444 The CoLab-score does not serve as a replacement for PCR-testing or LFTs, and can be used to
55
56 445 guide PCR-testing when routine blood tests are available. Note the performance of the
57
58
59
60

1
2
3 446 CoLab-score in a suspected/PCR-tested cohort is not equal to the (see **Supplemental**
4
5 447 **Material 4 Table 1**).

6
7
8 448 Finally, the CoLab-score could lead to false positives by other viral infections. However, in an
9
10 449 historic patient cohort, the CoLab-score had only limited discriminative ability in separating
11
12 450 influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4**
13
14 451 **Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-
15
16 452 response to the virus, it is expected that the CoLab-score is also sensitive to future SARS-
17
18 453 CoV-2 variants. This is supported by the fact that the diagnostic performance is sustained in
19
20 454 periods with different dominant variants. Moreover, there is no evidence that the
21
22 455 discriminative ability of the CoLab-score is lowered by a change in the ED patient population
23
24 456 as a result of widespread vaccination. Although vaccination status is not registered for all
25
26 457 presenting patients, in a small subgroup of 12 patients for whom vaccination status was
27
28 458 registered, and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5)
29
30 459 (see **Supplemental Material 2 Figure 2**),

31
32
33
34
35
36 460 To conclude, the CoLab-score developed and validated in this study, based on 10 routine
37
38 461 laboratory results and age, is available within 1 hour for any patient presenting at the ED. The
39
40 462 score can be used by clinicians to guide PCR testing or triage patients and helps to identify
41
42 463 COVID-19 in patients presenting at the ED with abdominal pain, chest pain, shortness of
43
44 464 breath, syncope, sepsis or other non-specific complaints where a routine blood panel is
45
46 465 requested. The lowest CoLab-score can be used to effectively rule-out a possible SARS-CoV-
47
48 466 2 infection, the highest score to alert physicians to a possible infection. The CoLab-score is
49
50 467 therefore a valuable tool to rule out COVID-19, guide PCR testing and is available to any
51
52 468 center with access to routine laboratory tests.

53
54
55
56
57
58 469

470 **Funding statement**

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

472

473 **Competing interests**

474 A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no
475 conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT
476 reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of
477 interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no
478 conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR
479 reports no conflict of interest. VS reports no conflict of interest.

480

481 **Data sharing statement**

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

485

486 **Author contributorship statement**

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

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

- 1
2
3 493 Maaïke Maas: Conceptualization (Supporting), Resources (Supporting), Supervision
4
5 494 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
6
7
8 495 Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision
9
10 496 (Supporting), Validation (Equal), Writing-review & editing (Equal).
11
12
13 497 Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision
14
15 498 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
16
17
18 499 Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision
19
20 500 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
21
22
23
24 501 Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision
25
26 502 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
27
28
29 503 Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
30
31
32 504 Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
33
34
35 505 Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
36
37
38 506 Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
39
40
41 507 Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
42
43
44 508 Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing
45
46 509 (Equal).
47
48
49 510 Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal),
50
51 511 Writing-review & editing (Equal).
52
53
54 512 Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project
55
56 513 administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing
57
58 514 (Equal).
59
60

515 **References**

- 516 1 Coronavirus Disease (COVID-19) Situation Reports.
517 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
518 (accessed 4 Feb 2021).
- 519 2 Guan W, Ni Z, Hu Y, *et al*. Clinical Characteristics of Coronavirus Disease 2019 in
520 China. <https://doi.org/10.1056/NEJMoa2002032> 2020;**382**:1708–20.
521 doi:10.1056/NEJMOA2002032
- 522 3 Vetter P, Vu DL, L’Huillier AG, *et al*. Clinical features of covid-19. *BMJ* 2020;**369**.
523 doi:10.1136/BMJ.M1470
- 524 4 Wynants L, Van Calster B, Collins GS, *et al*. Prediction models for diagnosis and
525 prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18.
526 doi:10.1136/bmj.m1328
- 527 5 Albahri AS, Hamid RA, Alwan J k., *et al*. Role of biological Data Mining and Machine
528 Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19):
529 A Systematic Review. *J. Med. Syst.* 2020;**44**:122. doi:10.1007/s10916-020-01582-x
- 530 6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical
531 Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417
- 532 7 Prokop M, Everdingen W van, Vellinga T van R, *et al*. CO-RADS: A Categorical CT
533 Assessment Scheme for Patients Suspected of Having COVID-19—Definition
534 and Evaluation. <https://doi.org/10.1148/radiol2020201473> 2020;**296**:E97–104.
535 doi:10.1148/RADIOL.2020201473
- 536 8 Goodacre S, Thomas B, Sutton L, *et al*. Derivation and validation of a clinical severity
537 score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort

- 1
2
3 538 study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840
4
5
6 539 9 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine
7
8 540 learning tool to rule out COVID-19 among adults in the emergency department using
9
10 541 routine blood tests: A large, multicenter, real-world study. *J Med Internet Res*
11
12 542 2020;**22**:e24048. doi:10.2196/24048
13
14
15
16 543 10 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical
17
18 544 data for patients attending hospital: development and prospective validation of an
19
20 545 artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87.
21
22 546 doi:10.1016/S2589-7500(20)30274-0
23
24
25
26 547 11 R Core Team. R: A Language and Environment for Statistical Computing.
27
28 548 2020.<https://www.r-project.org/>
29
30
31 549 12 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29.
32
33 550 doi:10.1198/016214506000000735
34
35
36 551 13 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B*
37
38 552 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x
39
40
41 553 14 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models
42
43 554 via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01
44
45
46
47 555 15 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63.
48
49 556 doi:10.1093/oxfordjournals.pan.a004868
50
51
52 557 16 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation.
53
54 558 *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973
55
56
57 559 17 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to
58
59 560 analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-

- 1
2
3 561 2105-12-77
4
5
6 562 18 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.[https://cran.r-](https://cran.r-project.org/package=rms)
7
8 563 [project.org/package=rms](https://cran.r-project.org/package=rms)
9
10
11 564 19 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons
12
13 and progress. *Scand. J. Clin. Lab. Invest.* 1999;**59**:491–500.
14 565
15 doi:10.1080/00365519950185229
16 566
17
18 567 20 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from
19
20 Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst*
21 568
22 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4
23 569
24
25
26 570 21 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of
27
28 SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J*
29 571
30 *Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502
31 572
32
33
34 573 22 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical
35
36 and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807.
37 574
38 doi:10.1007/s00330-020-07022-1
39 575
40
41 576 23 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-
42
43 infected patients at the emergency department using routine testing. *Clin Chem Lab Med*
44 577
45 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593
46 578
47
48
49 579 24 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a
50
51 diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int*
52 580
53 *J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305
54 581
55
56 582 25 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in
57
58 the emergency department: derivation and validation of the LDH-CRP-Lymphocyte
59 583
60

- 1
2
3 584 (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336
4
5
6 585 26 Tordjman M, Mekki A, Mali RD, *et al*. Pre-test probability for SARS-Cov-2-related
7
8 586 infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342.
9
10 587 doi:10.1371/journal.pone.0243342
11
12
13 588 27 Peto T, Affron D, Afrough B, *et al*. COVID-19: Rapid antigen detection for SARS-CoV-
14
15 589 2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for
16
17 mass-testing. *EClinicalMedicine* 2021;**36**:100924.
18 590
19
20 591 doi:10.1016/J.ECLINM.2021.100924
21
22
23 592 28 García-Fiñana M, Hughes DM, Cheyne CP, *et al*. Performance of the Innova SARS-
24
25 593 CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot:
26
27 594 population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637
28
29
30
31 595
32
33
34 596
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 597 **Figure legends**
4

5 598
6

7
8 599 **Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B)**
9
10 600 **dataset.**

11
12
13 601 *All patient admissions with routine venous blood sampling at the emergency department (ED)*
14
15 602 *were included. For the development dataset, completeness of the lab panel was assessed for*
16
17 603 *all 28 laboratory tests, for the temporal validation dataset this was only necessary for 10*
18
19 604 *laboratory tests. The major causes of missingness are described in the text. In the*
20
21 605 *development dataset, presentations with extreme values (>10 SD) were excluded. The same*
22
23 606 *limits were applied to the temporal validation dataset (see Table 2 for limits).*
24
25
26
27

28 607
29

30
31 608 **Figure 2: Probability density plot of the CoLab-linear predictor.**
32

33
34 609 *The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are*
35
36 610 *plotted against the linear predictor (see table 2). The CoLab-score cut-offs (-5.83 , -4.02 , $-$*
37
38 611 *3.29 , -2.34 and -1.64) are depicted with vertical dashed lines. The white-boxed numbers*
39
40 612 *(between the cut-offs) represent the corresponding CoLab-score. Note that while the area*
41
42 613 *under both curves is identical (since these are probability density functions), in absolute*
43
44 614 *numbers the “negative or untested”-group is about 36 times larger than the PCR positive*
45
46 615 *group.*
47
48
49

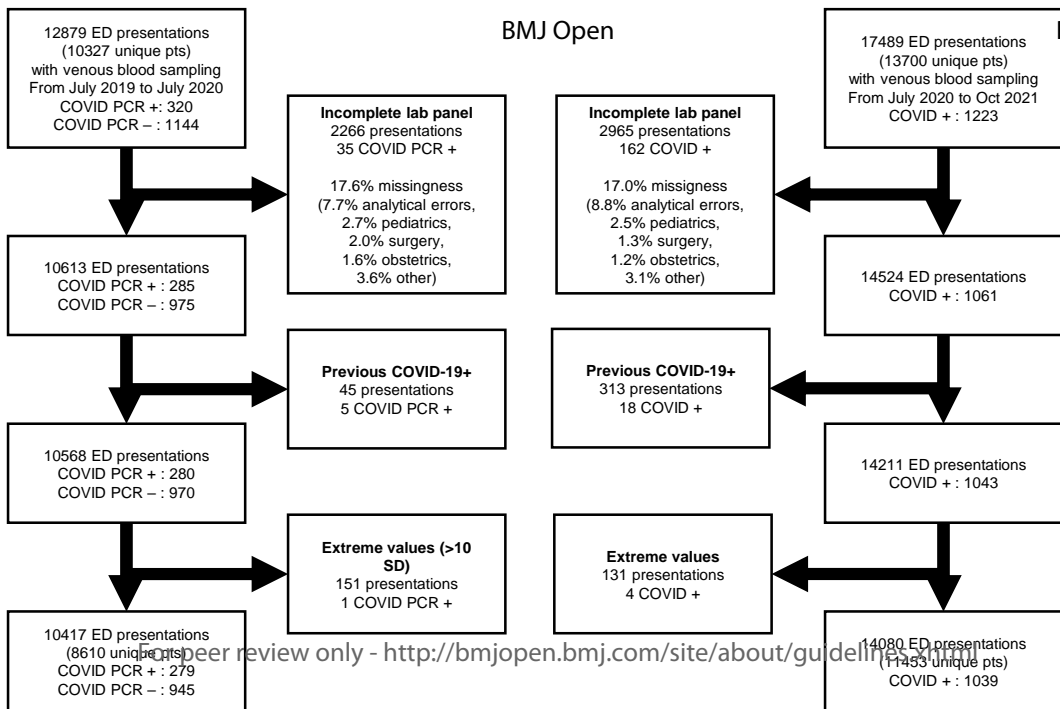
50 616
51

52
53 617 **Figure 3: Inclusion flow of ED patients in three external centers.**
54

55
56 618 *All emergency department (ED) presentations with routine venous blood sampling were*
57
58 619 *included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see*
59
60

1
2
3 620 *Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were*
4
5 621 *excluded as “previous COVID-19+”. Presentations with any laboratory result above the*
6
7 622 *limits of the CoLab-score (see Table 2) were excluded.*
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

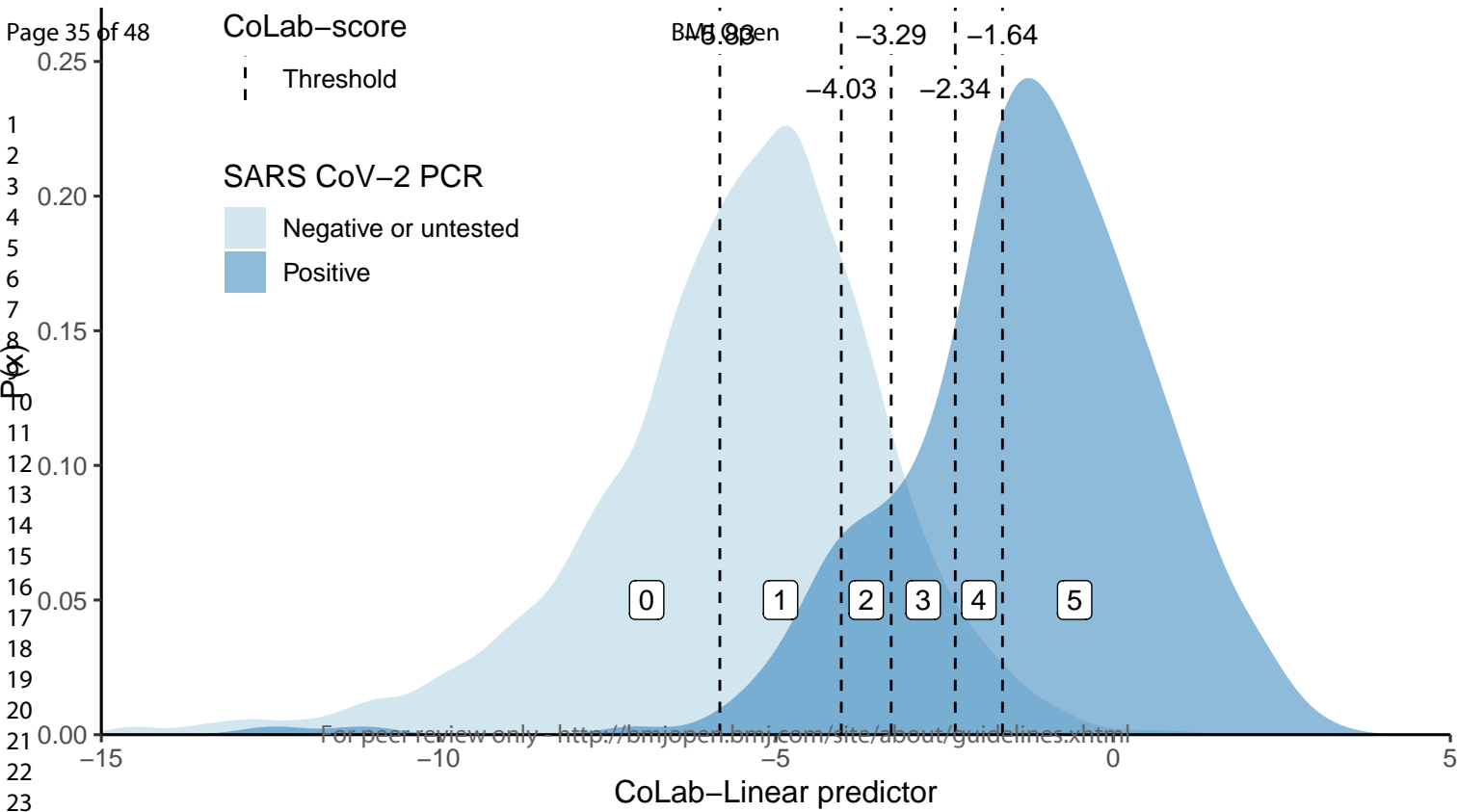


CoLab-score

Threshold

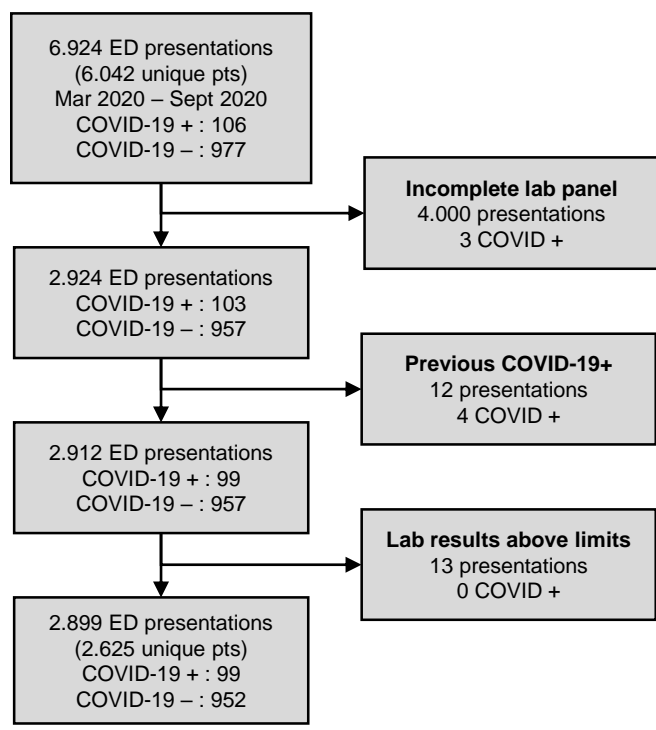
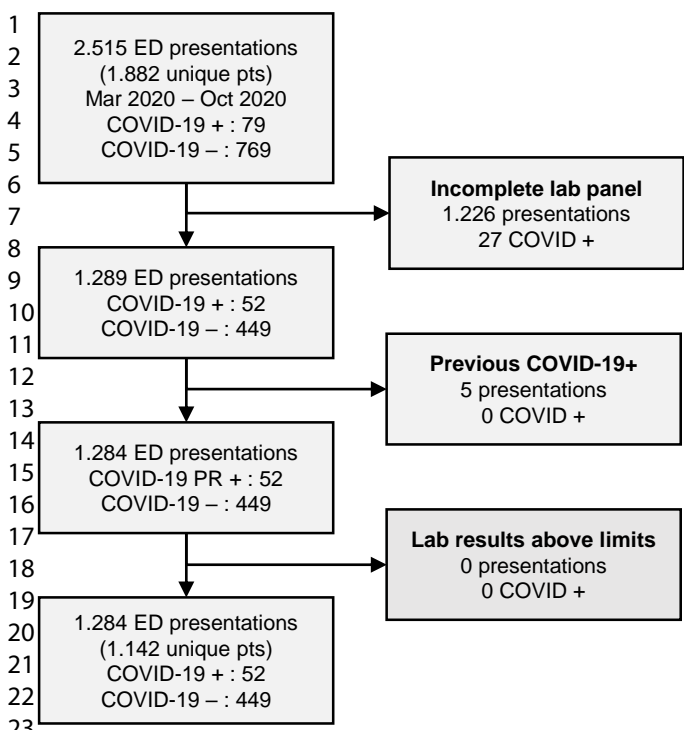
SARS CoV-2 PCR

- Negative or untested
- Positive

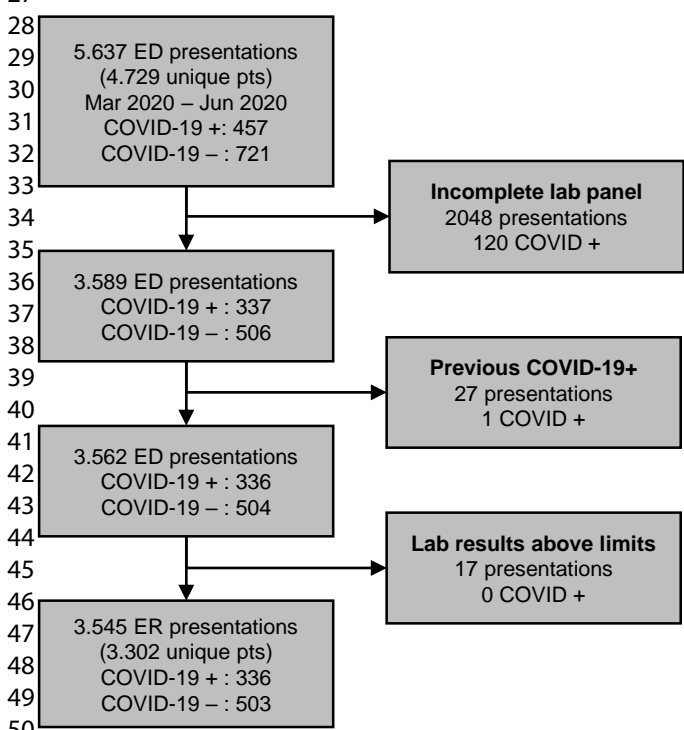


Center 1

Center 2



Center 3



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 (\hat{w}) was calculated by $\hat{w} = 1/|\hat{\beta}_{ridge}|^2$. 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: $lp_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in}$. Where n is the number of variables in the final model, x_{in} are the observed predictor variables for subject i and β_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:

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

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:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639$$

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$

$$lp_i(\tau) \geq \ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

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

$$lp_i(\tau = 0.4) \geq -5.83 \text{ (CoLab-score} = 1)$$

$$lp_i(\tau = 0.1) \geq -4.03 \text{ (CoLab-score} = 2)$$

$$lp_i(\tau = 0.05) \geq -3.29 \text{ (CoLab-score} = 3)$$

$$lp_i(\tau = 0.02) \geq -2.34 \text{ (CoLab-score} = 4)$$

$$lp_i(\tau = 0.01) \geq -1.64 \text{ (CoLab-score} = 5)$$

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

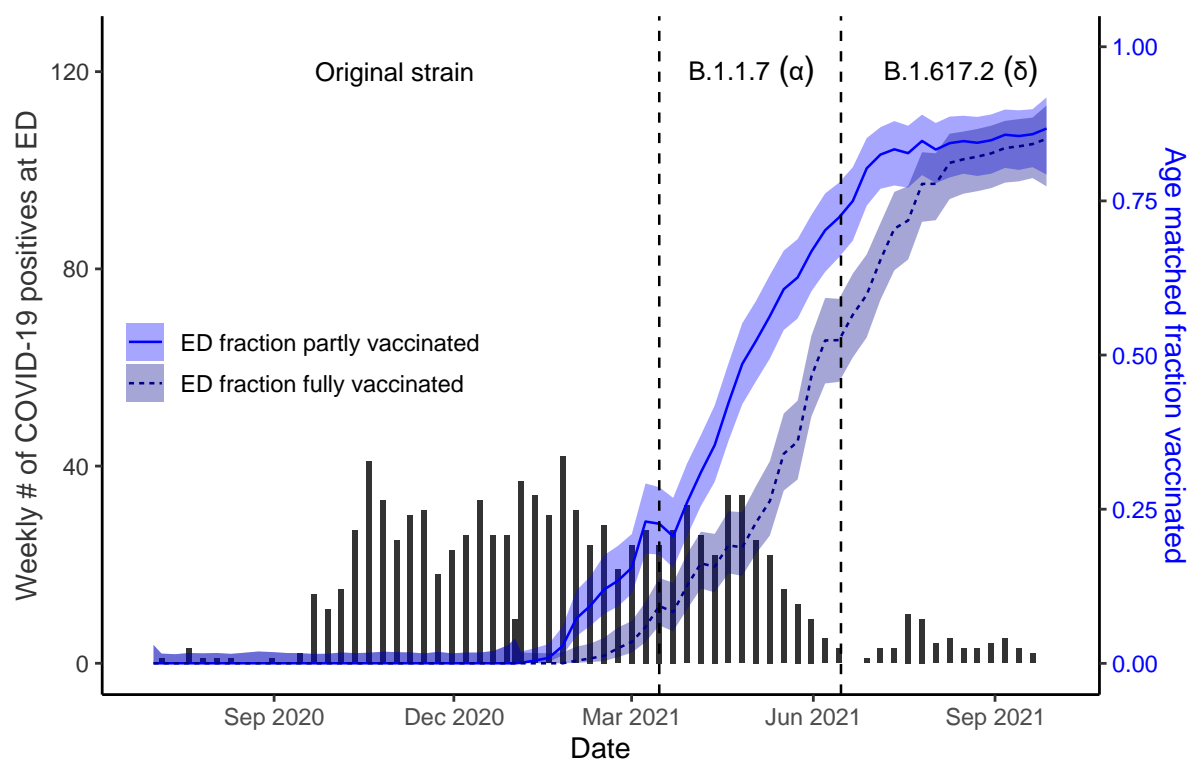


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

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).

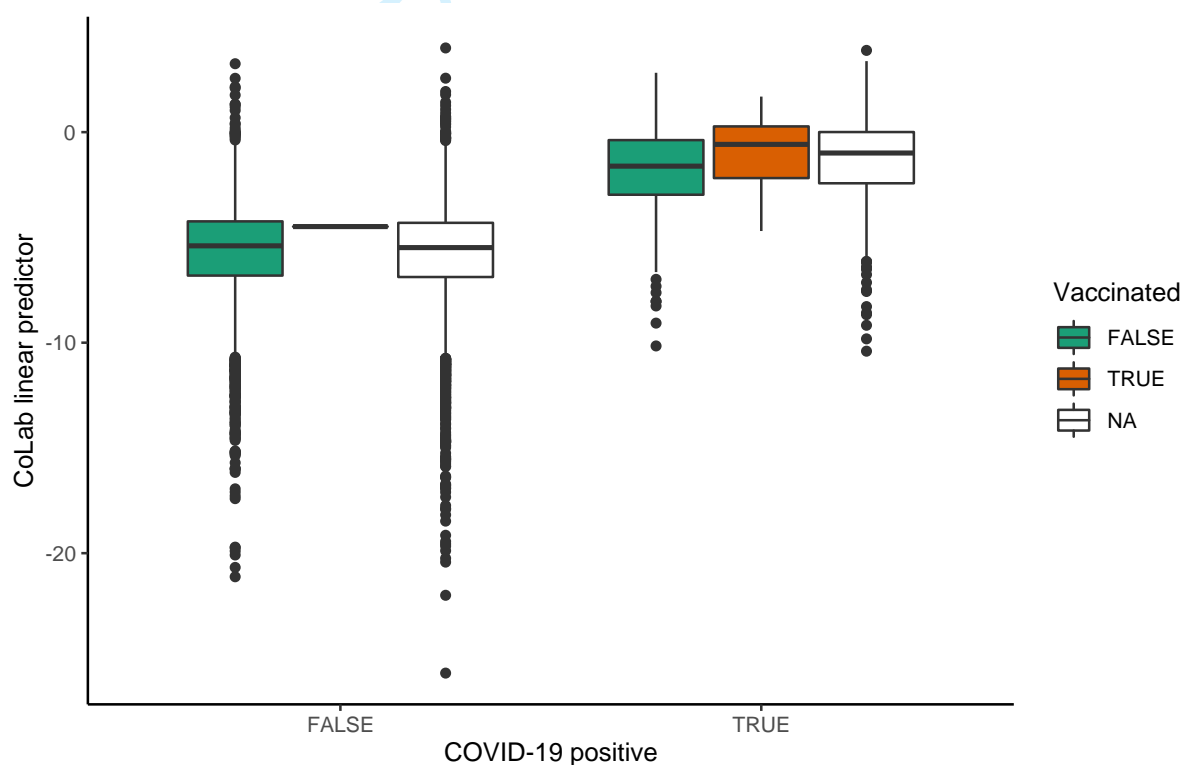


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

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

1
2
3 *vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13*
4
5 *presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative.*
6
7 *Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or*
8
9 *considered positive until proven otherwise, therefore there is only one COVID-19 negative*
10
11 *patient with a registered vaccination status.*
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Supplemental material 3

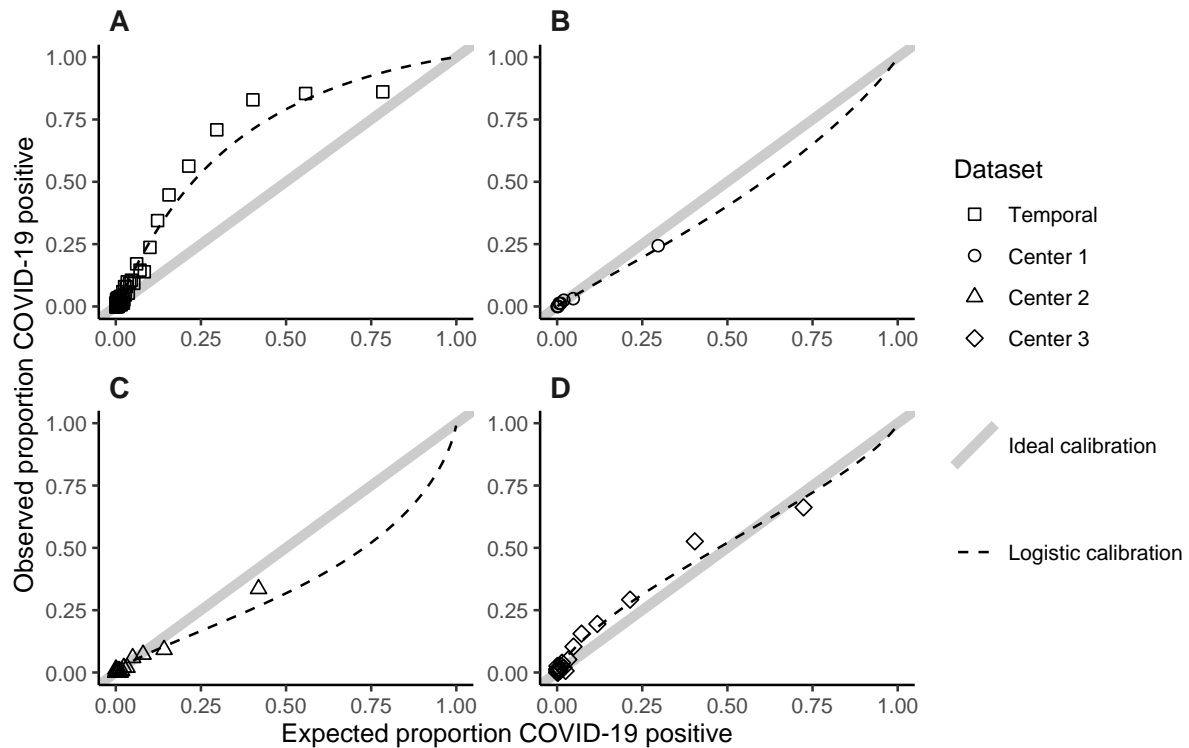


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

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.

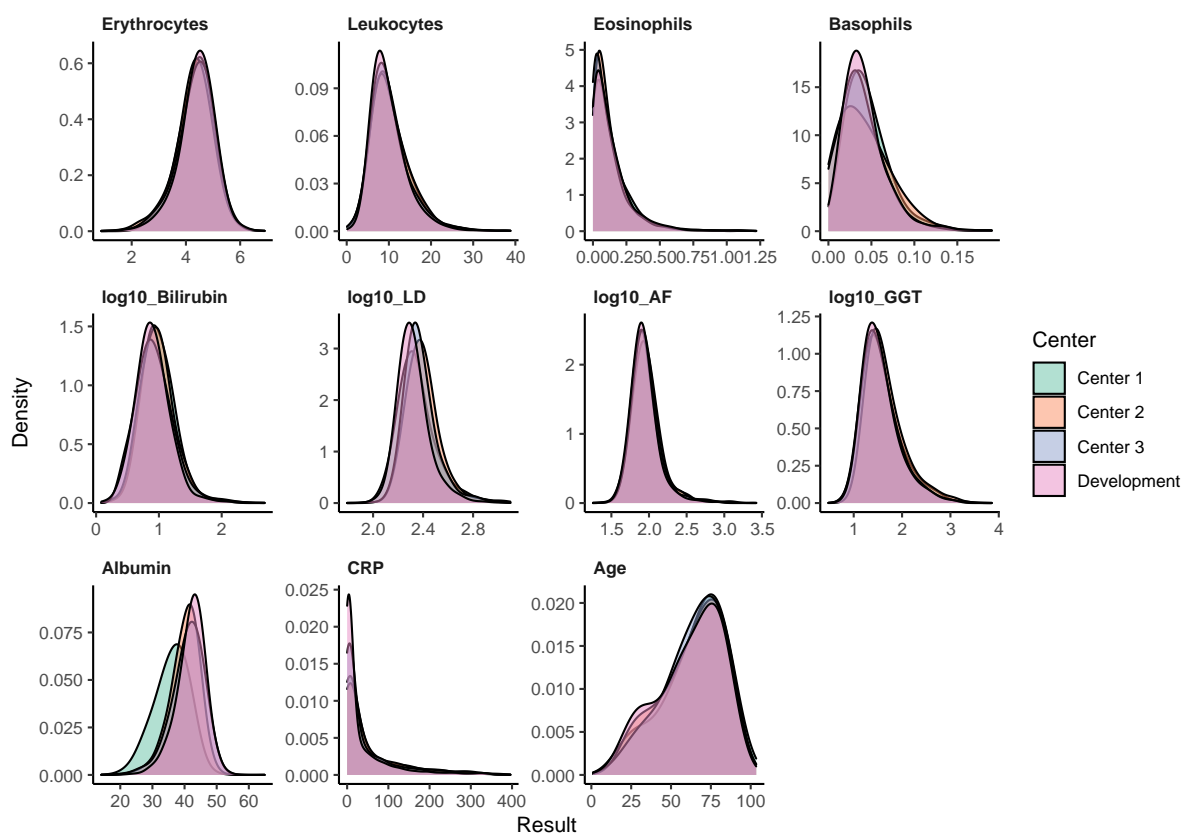


Figure 2: Probability density plots of laboratory parameters.

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

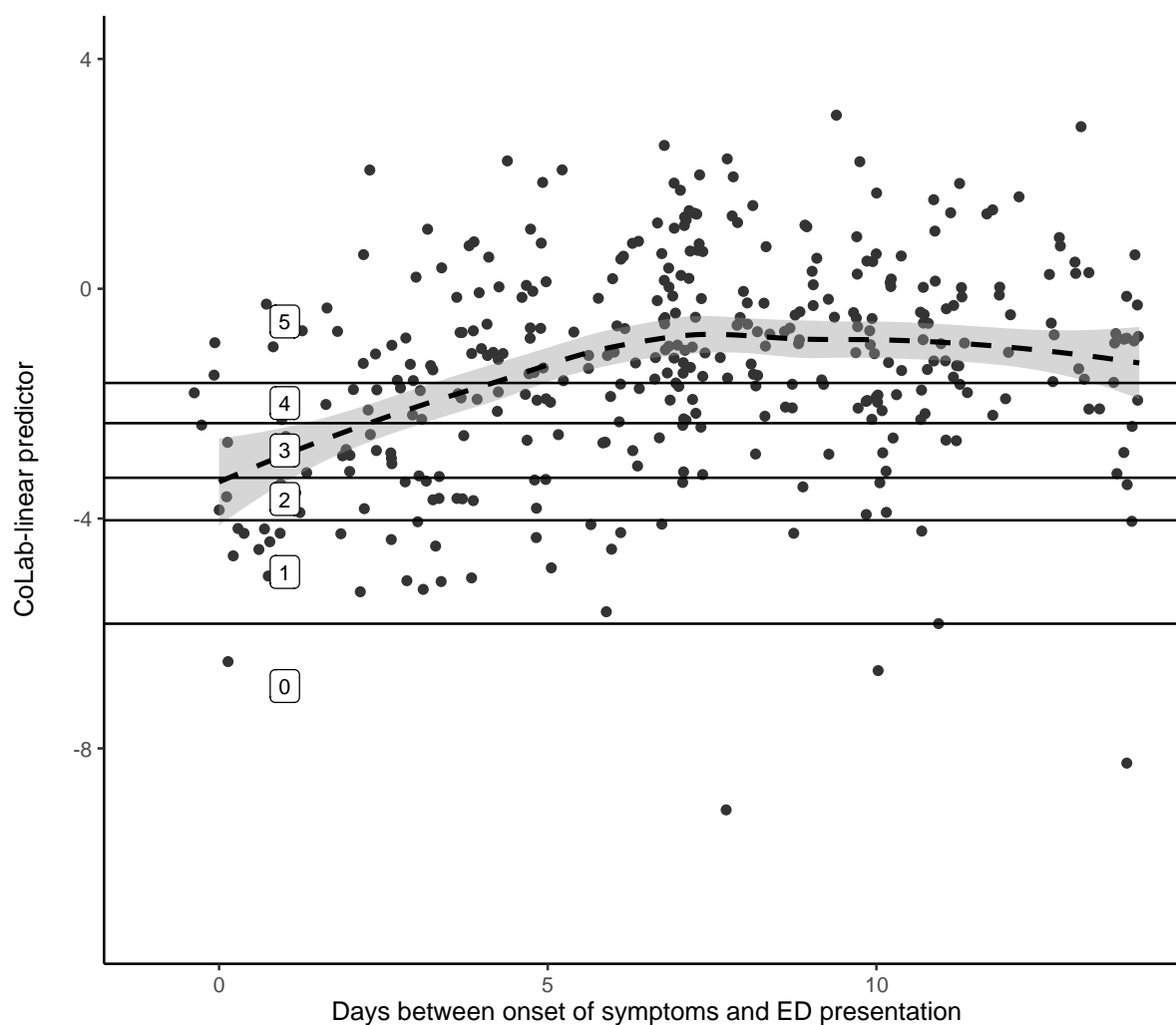


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predictor 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.

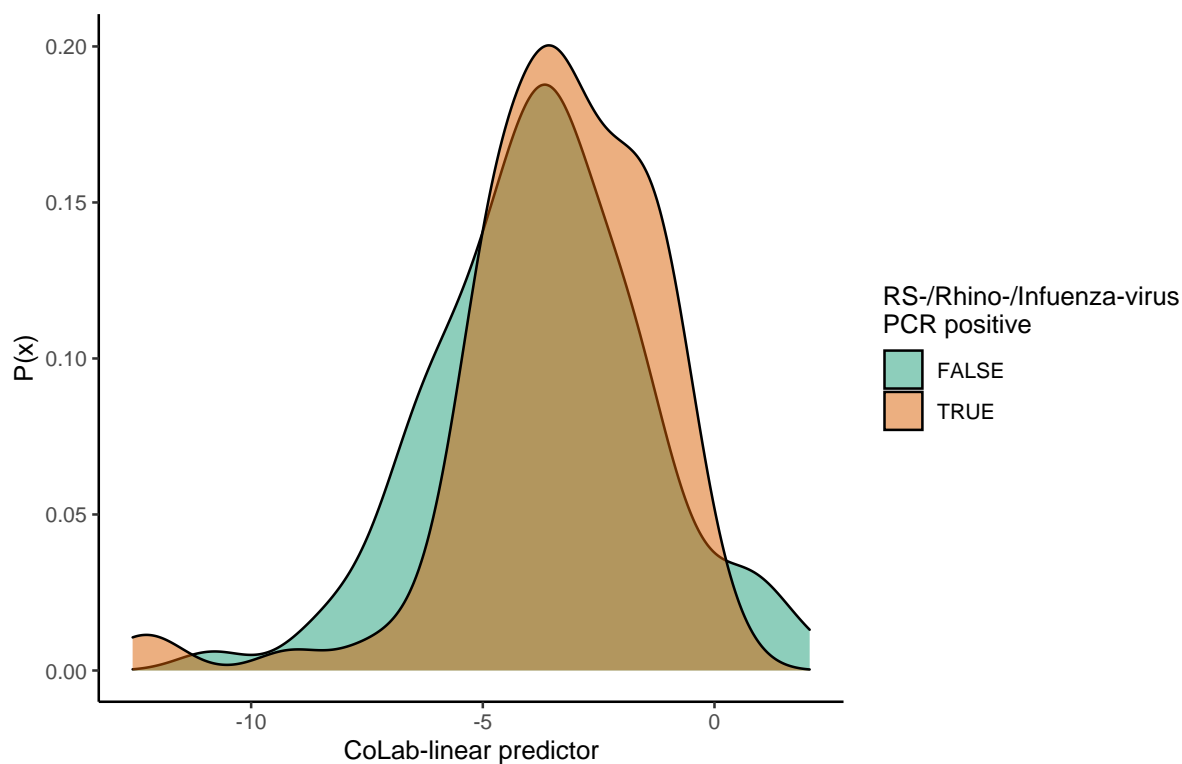


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

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: 0.4896-0.6563).

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)

1
2
3 **Table 1: Sensitivity analysis of the CoLab-score in the temporal validation dataset using**
4 **different inclusion criteria.**
5

6
7 *Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV),*
8 *true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are*
9 *shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals*
10 *in parentheses. The temporal validation dataset is used to compare the performance of the*
11 *CoLab-score with inclusion criteria that differ from the development dataset. The first line*
12 *shows the performance of the temporal validation dataset with the original inclusion criteria*
13 *as specified in Figure 1B. The second line shows the performance of the CoLab-score when*
14 *all re-presentations are excluded (i.e. no repeated presentations). The third line shows the*
15 *performance of the CoLab-score in the subgroup of patients that underwent PCR-testing.*
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

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 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:	<i>BMJ Open</i>
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, Maaïke; 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

SCHOLARONE™
Manuscripts



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.

1
2
3
4 1 **Development and validation of an early warning score to identify**
5
6 2 **COVID-19 in the emergency department based on routine laboratory**
7
8
9 3 **tests: a multicenter case-control study**
10
11
12 4

15 5 Arjen-Kars Boer^{1*}, Ruben Deneer^{1,2*†}, Maaïke Maas³, Heidi S. M. Ammerlaan⁴, Roland H. H.
16
17 6 van Balkom⁵, Wendy A.M.H. Thijssen³, Sophie Bennenbroek³, Mathie P. G. Leers⁶, Remy
18
19 7 J.H. Martens⁶, Madelon M. Buijs⁷, Jos J. Kerremans⁸, Muriël Messchaert⁹, Jeroen D.E. van
20
21
22 8 Suijlen⁹, Natal A.W. van Riel^{2,10} and Volkher Scharnhorst^{1,2}
23
24
25 9

26
27
28 10 * Both authors contributed equally
29

30
31 11 † Corresponding author
32
33
34 12

35
36
37 13 1 Department of Laboratory Medicine, Catharina Hospital Eindhoven,
38
39 14 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

40
41 15 2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene
42
43 16 Loper 3, 5612 AE, Eindhoven, the Netherlands

44
45
46 17 3 Department of Emergency Medicine, Catharina Hospital Eindhoven,
47
48 18 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

49
50 19 4 Department of Internal Medicine, Catharina Hospital Eindhoven,
51
52 20 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

53
54
55 21 5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven,
56
57 22 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands
58
59
60

- 1
2
3 23 6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der
4
5 24 Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands
6
7 25 7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the
8
9 26 Netherlands
10
11
12 27 8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital,
13
14 28 Simon Smitweg 1, 2353 GA Leiderdorp, the Netherlands
15
16
17 29 9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334
18
19 30 DZ Apeldoorn, the Netherlands
20
21 31 10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef
22
23 32 9, 1105 AZ Amsterdam, the Netherlands.
24
25
26
27 33
28
29

30 **Keywords**

31
32 35 COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction
33
34 36 model, routine laboratory tests
35
36
37
38 37
39

40 **Corresponding author**

41 38
42
43 39 Ruben Deneer
44
45
46 40 Michelangelolaan 2
47
48
49 41 5623 EJ Eindhoven, the Netherlands
50
51
52 42 ruben.deneer@catharinaziekenhuis.nl
53
54
55
56
57
58
59
60

43 Abstract

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

49 **Design:** Case-control study.

50 **Setting:** Secondary and tertiary hospitals in the Netherlands.

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

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

58 **Results:** The resulting “CoLab-score” consists of 10 routine laboratory measurements, and
59 age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945).
60 The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991,
61 specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity
62 (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were
63 confirmed in temporal and external validation.

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

1
2
3 66 ruled-out in over one third of ED presentations. Highly suspect cases can be identified
4
5 67 regardless of presenting symptoms. The CoLab-score is continuous, in contrast to the binary
6
7 68 outcome of lateral flow testing, and can guide PCR testing and triage ED patients.
8
9

10
11 69

12 13 14 70 **Article summary**

15 16 17 71 Strengths and limitations of this study

- 18
19 72 • A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency
20
21 73 department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- 22
23 74 • Using adaptive lasso regression analysis, the panel of 28 laboratory tests was reduced
24
25 75 to a single score consisting of a subset of 10 routine ED laboratory tests and age.
- 26
27 76 • The score was temporally validated from July 2020 to October 2021, in the presence of
28
29 77 vaccine roll-out and emergence of new SARS-CoV-2 variants.
- 30
31 78 • The score was externally validated in 3 other centers in the Netherlands.
- 32
33 79 • Missingness in the panel of laboratory tests varied between external centers, limiting
34
35 80 generalizability of the score to the ED population for which the complete panel of
36
37 81 laboratory tests was available.
- 38
39 82 • The score was not directly compared to lateral flow testing.
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

84 Introduction

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

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

99 Many COVID-19 prediction models have already been developed, the living systematic
100 review by Wynants et. al [4] provides an extensive overview and critical appraisal.

101 Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early
102 models were based on relatively small sample sizes, hampered by selection bias or were over-
103 fitted by selecting too many features [4–6]. Aside from methodological shortcomings of early
104 models, most models are not developed as an early warning score for all ED patients. Firstly,
105 they require features from tests that are not routinely performed or logged for all ED patients
106 (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the
107 PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED
108 patient population. Secondly, the population on which models are commonly based, are PCR-

1
2
3 109 tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by
4
5 110 physicians.

7 111 Only two studies were identified that focus on patients presenting at the ED, include
8
9
10 112 unsuspected (and pre-pandemic) patients as controls, and rely solely on routine (laboratory)
11
12 113 tests [9,10].

15 114 In this study we report the development and validation of an early warning score that, based
16
17 115 on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in patients
18
19 116 who undergo routine laboratory testing at presentation. The score can assist ED physicians in
20
21 117 triaging patients and prevent further transmission of COVID-19 by quickly identifying
22
23
24 118 possibly infected patients or ruling out a possible infection when resources are scarce.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

119 **Methods**

120 *Study design*

121 This is a retrospective case-control study where routine laboratory test results, combined with
122 age and gender, from all patient presenting at the emergency department (ED) of the
123 Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-
124 2 PCR test results in a development dataset. A model that could predict the presence of a
125 COVID-19 infection was fit to this dataset. Performance of the model was assessed by i)
126 internal validation, ii) temporal validation and iii) external validation by using data from the
127 ED of three other centers. The study was reviewed by the Medical research Ethics
128 Committees United (MEC-U) under study number W20.071, which confirmed that the
129 Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this
130 study. The study was thereafter reviewed and approved by the internal hospital review board.

131

132 *Patient and Public Involvement*

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

134

135 *Development dataset*

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

1
2
3 143 quantitative results were not available due to analytical interference (hemolysis, lipemia or
4
5 144 icterus). The routine ED laboratory panel is requested for (adult) patients presenting with
6
7 145 abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific
8
9 146 complaints, or for patients (including non-adult patients) presenting with specific complaints
10
11 147 where a suspected diagnosis has to be ruled-in or ruled-out. Presentations with one or more
12
13 148 missing values in any of the 28 laboratory test in the routine ED panel, were excluded.
14
15 149 Presentations with one or more extreme lab results, > 10 times standard deviation from the
16
17 150 median, were also excluded to minimize the effect on the estimation of regression
18
19 151 coefficients. The median was chosen as a measure of central tendency due to its resistance for
20
21 152 outliers. After the first case of COVID-19 in the Netherlands, all patients with symptoms of
22
23 153 COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR
24
25 154 testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were
26
27 155 approved by the Dutch national institute of public health (RIVM). If a patient had a positive
28
29 156 PCR result in the past, subsequent presentations were excluded as re-presentations might be
30
31 157 clinically different from de novo presentations.
32
33
34
35
36
37
38 158 The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying
39
40 159 nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal
41
42 160 at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was
43
44 161 positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time
45
46 162 window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were
47
48 163 performed in the time window and the presentation occurred after the first case of COVID-19
49
50 164 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the
51
52 165 first case were labelled as "*Pre-COVID-19*".
53
54
55
56
57
58
59
60

167 *Laboratory tests*

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

181 *Modelling*

182 All data were processed and analyzed in R version 4.1.1 [11]. Laboratory results, combined
183 with age and gender were used as covariates in a regression model. Cases were defined as ED
184 presentations labelled as “*PCR-positive*”, controls were all other presentations (i.e. “*PCR-*
185 *negative*”, “*Untested*” or “*Pre-COVID-19*”). To achieve predictive accuracy, limit overfitting
186 and perform feature selection, penalized logistic regression with an adaptive lasso penalty was
187 chosen [12,13]. To minimize missing data, all non-numeric results at the extremes of the
188 measuring range, were converted to numeric results by removing the “<” and “>” signs. For
189 eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m² and
190 <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD,
191 CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model

1
2
3 192 predictions, these variables were transformed logarithmically. More details regarding model
4
5 193 fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the
6
7 194 glmnet-package [14].
8
9

10
11 195

12 13 196 *CoLab-score*

14
15
16 197 Since this is a retrospective case-control study, the sample prevalence may not reflect the
17
18 198 true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term
19
20 199 in the model should be adjusted according to the current prevalence (details can be found in
21
22 200 the document, **Supplemental Material 1**) [15]. However, adjusting the intercept term is not
23
24 201 straightforward to implement in clinical practice, therefore the linear predictor of the model
25
26 202 was categorized into a score, this score is hereafter referred to as the “CoLab-score”. The
27
28 203 categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15
29
30 204 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using
31
32 205 the intercept adjustment formula by King [15]. The intervals obtained through these breaks
33
34 206 correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and
35
36 207 score 5 reflects high-risk. More details regarding the rationale of the CoLab-score
37
38 208 categorization can be found in the document, **Supplemental Material 1**.
39
40
41
42
43

44 209

45 46 47 210 *Internal validation*

48
49 211 To assess model performance while taking overfitting into account, bootstrapping was
50
51 212 performed. 1000 bootstrap samples were generated from the original data. On each bootstrap
52
53 213 sample, the full model fitting procedure and CoLab-score conversion were performed.
54
55 214 Optimism adjusted performance measures of the CoLab-score were obtained by applying the
56
57 215 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [16]. Performance
58
59
60

1
2
3 216 measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative
4
5 217 predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate
6
7 218 performance measures [17]. Although the full inclusion period from July 2019 to July 2020
8
9 219 was used for model fitting, the performance was evaluated on the period starting from the first
10
11 220 COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain
12
13 221 performance measures that would reflect real world performance.
14
15
16
17
18
19

20 223 *Temporal validation*

21
22
23 224 For temporal validation, results from our center were prospectively analyzed from July 2020
24
25 225 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-
26
27 226 19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period
28
29 227 there was also more widespread external PCR testing by municipal health services. The
30
31 228 results of external conducted PCR tests were not available to our study. To overcome this
32
33 229 limitation, the outcome in the temporal validation cohort was chosen as a composite of the
34
35 230 hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test
36
37 231 result. This period also covers both the emergence of new SARS-CoV-2 variants as well as
38
39 232 vaccine rollout. However, neither vaccination status nor genomic sequencing was available to
40
41 233 determine whether a patient was vaccinated or which variant caused the infection. Therefore,
42
43 234 data from the Dutch national institute of public health (RIVM) was used, to divide the
44
45 235 temporal validation period into three phases: i) from July 2020 until March 2021, no
46
47 236 vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial
48
49 237 vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until
50
51 238 October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant.
52
53 239 See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted
54
55 240 of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold
56
57
58
59
60

1
2
3 241 for the entire period, as well as for each phase separately to determine a possible effect of
4
5 242 vaccination and new variants on performance (results in the **Supplemental Material 2**).
6
7 243 Model calibration was assessed graphically using the rms-package [18].
8
9

10 244

11
12
13 245 *External validation*

14
15 246 For the external validation, several centers in the Netherlands were approached and assessed
16
17 247 if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available.
18
19 248 Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals
20
21 249 (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and
22
23 250 Zuyderland Medical Center (center 3). The hematological parameters were measured with
24
25 251 Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and
26
27 252 Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with
28
29 253 Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott
30
31 254 Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external
32
33 255 validation was similar to the temporal validation and consisted of assessing the AUC
34
35 256 sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was
36
37 257 assessed graphically analogous to the temporal validation dataset.
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

258 Results

259 Development dataset

260 12879 emergency department (ED) presentations of 10327 patients from July 2019 to July
 261 2020 were included. After excluding cases with an incomplete lab panel, patient presentations
 262 that occurred after a positive PCR test in the past (re-presentations) and presentations with
 263 extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of
 264 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/m ²	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]

265

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

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

276

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

282

283 **CoLab-score**

284 The model obtained through adaptive lasso regression contained eleven variables, which are
 285 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 μ mol/L	-1.142	Bilirubin >169 μ mol/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 %

286

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

288 *The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of*

289 *the 11 variables with their corresponding coefficients (β 's). CoLab-LP = - 6.885 +*

290 *[erythrocytes] \times 0.9379 - [leukocytes] \times 0.1298 - [eosinophils] \times 6.834 - [basophils] \times*

291 *47.7 - log₁₀([bilirubin]) \times 1.142 + log₁₀([LD]) \times 5.369 - log₁₀([ALP]) \times 3.114 +*

292 *log₁₀([gGT]) \times 0.3605 - [albumin] \times 0.1156 + [CRP] \times 0.02560 + [age] \times 0.002275. The*

293 *LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence*

294 *is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid*

295 *if any of the variables exceed the limits in the third column. The relative importance ranks the*

296 *importance of variables in predicting the outcome, relative to the most important variable (in*

297 *this case basophils).*

298

299 A larger β -coefficient does not imply that a variable is more important in predicting the odds

300 of testing positive for SARS-CoV-2, since variables are on different scales. The most

301 important variables are basophiles, eosinophils and lactate dehydrogenase (LD).

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

305

306 *Internal validation*

307 The model was validated in the period starting from the first COVID-19 infection to July
 308 2020, in this period the mean prevalence was 7.2%. The AUC of the CoLab-score is 0.930
 309 (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)

310

311 **Table 3: Bootstrapped diagnostic performance of the CoLab-score in the development** 312 **dataset.**

313 *The development dataset was internally validated for the period March 2020 – July 2020 (N*
 314 *= 3868). The optimism-adjusted bootstrapped sensitivities, specificities, positive predictive*
 315 *values (PPV), negative predictive values (NPV), true positives (TP), true negatives (TN), false*
 316 *positives (FP) and false negatives (FN) and fraction of presentations (%) are shown for fixed*
 317 *cut-offs (CoLab-score 0 till ≤ 4). The numbers in round brackets represent the 95% optimism-*

1
2
3 318 *adjusted bootstrapped confidence intervals. The first column defines the threshold above*
4
5 319 *which CoLab-score a patient is considered positive. Note that “0” lists the sensitivity and*
6
7 320 *NPV of CoLab-score 0 and “≤ 4” lists the specificity and PPV of CoLab-score 5. Also note*
8
9 321 *that TP, TN, FP and FN are not whole numbers, as these are obtained through bootstrapping*
10
11 322 *and each bootstrap replicate contains a different number of controls and cases.*
12
13
14

15 323

16
17
18 324 Diagnostic performance is shown in **Table 3**. A CoLab-score of 0 has a negative predictive
19
20 325 value (NPV) of 0.997 (95% CI: 0.993 to 0.999) and positive predictive value (PPV) of 0.115
21
22 326 (0.0934 - 0.147), one third (38%, 95% CI: 28 to 514%) of all ED presentations were assigned
23
24 327 this score and can therefore be safely excluded. Conversely, 6% (95% CI: 6 to 8%) of the ED
25
26 328 patients had a CoLab-score = 5. Given the PPV of this score (0.683, 95% CI: 0.628 to 0.746,
27
28 329 NPV: 0.970, 95% CI: 0.963 - 0.978), subsequent PCR testing is advised.
29
30
31

32 330

33 331 *Temporal validation*

34
35
36 332 As the CoLab-score was developed in our center after the first COVID-19-wave in the
37
38 333 Netherlands, the performance was evaluated in our center from July 2020 until October 2021.
39
40 334 Lab results from 17489 ED presentations were collected. After applying the inclusion flow as
41
42 335 shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a
43
44 336 COVID-19 infection.
45
46
47
48
49

50 337 The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal
51
52 338 validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the
53
54 339 development cohort, although sensitivity is slightly lower and specificity slightly higher (cf.
55
56 340 **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases
57
58 341 according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental**
59
60

1
2
3 342 **Material 2 Figure 1**). The discriminative ability was not lower in the second or third phase,
4
5 343 compared to the first phase. Diagnostic performance is preserved in terms of sensitivity and
6
7 344 specificity, except a moderately reduced sensitivity of scores ≥ 3 in the third phase as
8
9
10 345 compared to the first phase. PPV and NPV are incomparable due to different prevalence/pre-
11
12 346 test probabilities in each phase (see **Supplemental Material 2 Table 1**).

13
14
15 347 In terms of the predicted probabilities, model calibration shows that overall predicted
16
17 348 probabilities are too low (see **Supplemental Material 3** for the calibration plot), which is
18
19 349 expected since the prevalence differs and the intercept has to be adjusted to the prevalence.

20 350 In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that
21
22 351 initially did not present with COVID-specific symptoms. Most patients had neurological or
23
24 352 orthopedic presenting symptoms.
25
26
27
28
29

30 353

31 32 33 354 *External validation*

34
35 355 For external validation, data obtained from three other centers were used, center 1 (N = 1284,
36
37 356 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545,
38
39 357 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19
40
41 358 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was
42
43 359 lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of
44
45 360 the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891
46
47 361 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.
48
49
50

51 362 Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers
52
53 363 is ≥ 0.96 . Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for
54
55 364 external centers are shown in **Supplemental Material 3**, the observed fraction of COVID-19
56
57
58
59
60

365 positives is slightly lower than expected in centers 1 and 2. For center 3, low probabilities
 366 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)
	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)
≤ 1	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)
	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)
≤ 2	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)
	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)
≤ 3	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)

367

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

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

375 Discussion

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

380 The main strength of this study is that this score can be used as an early-warning or triaging
381 tool for the ED population presenting with abdominal pain, chest pain, shortness of breath,
382 syncope, sepsis or other non-specific complaints where a routine blood panel is requested.
383 This is in contrast to the vast majority of COVID-19 diagnostic models that have been
384 developed on a pre-selected population of PCR-tested patients [9,20–26]. Moreover, the
385 CoLab-score requires only routine blood tests, instead of (features from) imaging such as CT-
386 scans or laboratory tests that are not routinely collected in the ED, e.g. interleukin-6 or 3-
387 hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous
388 result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous
389 score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable
390 to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific)
391 [27,28].

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

1
2
3 400 Since this is a retrospective case-control study, there is some unavoidable missing data. In our
4
5 401 cohort 17.6% of the ED presentations could not be used due to one or more missing
6
7 402 laboratory results. This is lower or equal to similar studies; 22% [23], 17% [21] and 11% [26].
8
9 403 Important to note is that 7.7% of missingness is due to analytical errors which can be assumed
10
11 404 to be missing completely at random. For the remaining 9.9% of missingness, the full lab panel
12
13 405 was most frequently missing for pediatric, obstetric and surgery patients. These patients are
14
15 406 presenting with specific complaints for which specific laboratory tests are requested, and
16
17 407 hence do not match the inclusion criteria for a routine blood panel. Overall the missingness
18
19 408 was significantly lower in the PCR-tested group versus the untested group (χ^2 -test p-value
20
21 409 <0.001). It is assumed that all presentations in the untested group are COVID-19 negative.
22
23 410 However, some presentations with asymptomatic COVID-19 could be present in the untested
24
25 411 control group. The impact of these 'false controls' is most likely small as other studies
26
27 412 indicate that there is a very low positivity rate among asymptomatic ED presentations (only a
28
29 413 few in over a thousand tested asymptomatic cases) [29,30].
30
31
32
33
34
35 414 In the external centers, there is a high level of missingness as a result of an incomplete
36
37 415 laboratory panel. In the case of centers 1 and 2, only internal medicine ED presentations were
38
39 416 tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED
40
41 417 lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain
42
43 418 the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine
44
45 419 ED presentations, which is reflected by the few PCR-positive patients excluded. Due to these
46
47 420 high levels of missingness, the results of the external centers cannot be used to show that the
48
49 421 CoLab-score generalizes to the entire ED population. Rather, the results show that for the
50
51 422 majority of COVID-19 positive patients presenting at the ED, a routine laboratory panel is
52
53 423 available from which the CoLab-score can be calculated, and that the performance of the
54
55 424 CoLab-score in this population is comparable to the development population.
56
57
58
59
60

1
2
3 425 The performance of the CoLab-score is affected by the time between the onset of symptoms
4
5 426 and ED presentations. The score increases with the duration of symptoms and gradually
6
7 427 decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of
8
9 428 COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some
10
11 429 COVID-19 patients with early or late presentation after onset of symptoms can be missed.
12
13 430 Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and
14
15 431 <10 days prior to ED presentation. Chemotherapy that causes myeloid suppression, will
16
17 432 decrease neutrophilic, basophilic and eosinophilic counts and thereby “falsely” increasing the
18
19 433 CoLab-score. Conversely, COVID-19 patients with severe anemia could have “falsely”
20
21 434 lowered CoLab-scores. To minimize false negatives, we have therefore advised to report
22
23 435 CoLab-scores only when the concentration of erythrocytes is ≥ 2.9 /pL.
24
25
26
27
28
29 436 It was chosen to exclude re-presentations after a previous presentation with COVID-19. Since
30
31 437 the median time between initial presentation and re-presentation was 12 days, these patients
32
33 438 were most likely not re-infected patients, but patients who deteriorated after initial
34
35 439 presentation/treatment. Given that the CoLab-score follows the host-immune response, the
36
37 440 score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients
38
39 441 would impact the performance of the CoLab-score as patients in a later phase of the disease
40
41 442 show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to
42
43 443 patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate
44
45 444 after treatment for COVID-19. Other re-presentations were not excluded, which results in
46
47 445 some patients appearing multiple times in a dataset. This was not corrected for in the
48
49 446 regression model since the assumption was made that ED presentations are independent
50
51 447 observations. The median time between re-presentations is 38 days, most likely resulting in
52
53 448 variations in laboratory results between presentations, and hence, little to no correlation
54
55 449 between presentations. A sensitivity analysis was performed whereby only the first
56
57
58
59
60

1
2
3 450 presentation was included for each patient (**Supplemental Material 4 Table 1**), but no
4
5 451 difference was found in performance in terms of sensitivity, specificity and AUC.
6
7

8 452 The CoLab-score does not serve as a replacement for PCR-testing or LFTs, and can be used to
9
10 453 guide PCR-testing when routine blood tests are available. Important to note is that the CoLab-
11
12 454 score is only valid for ED presentations where routine blood testing is requested, and as a
13
14 455 consequence does not generalize to the ED population who is otherwise well and does not
15
16 456 undergo routine blood testing. Using the CoLab-score in a symptomatic/PCR-tested cohort
17
18 457 also results in different diagnostic performance characteristics, as compared to using the score
19
20 458 on the full ED cohort (see **Supplemental Material 4 Table 1**).
21
22
23

24
25 459 Finally, the CoLab-score could lead to false positives by other viral infections. However, in an
26
27 460 historic patient cohort, the CoLab-score had only limited discriminative ability in separating
28
29 461 influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4**
30
31 462 **Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-
32
33 463 response to the virus, it is expected that the CoLab-score is also sensitive to future SARS-
34
35 464 CoV-2 variants. This is supported by the fact that the discriminative ability is sustained in
36
37 465 periods with different dominant variants. Moreover, there is no evidence that the
38
39 466 discriminative ability of the CoLab-score is lowered by a change in the ED patient population
40
41 467 as a result of widespread vaccination. Although vaccination status is not registered for all
42
43 468 presenting patients, in a small subgroup of 12 patients for whom vaccination status was
44
45 469 registered, and were COVID-19 positive, 8 of 12 patients had the highest CoLab-score (= 5)
46
47 470 (see **Supplemental Material 2 Figure 2**),
48
49
50
51
52

53 471 To conclude, the CoLab-score developed and validated in this study, based on 10 routine
54
55 472 laboratory results and age, is available within 1 hour for any patient presenting at the ED
56
57 473 where routine blood testing is requested. The score can be used by clinicians to guide PCR
58
59 474 testing or triage patients and helps to identify COVID-19 in patients presenting at the ED with
60

1
2
3 475 abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific
4
5 476 complaints where a routine blood panel is requested. The lowest CoLab-score can be used to
6
7 477 effectively rule-out a possible SARS-CoV-2 infection, the highest score to alert physicians to
8
9 478 a possible infection. The CoLab-score is therefore a valuable tool to rule out COVID-19,
10
11 479 guide PCR testing and is available to any center with access to routine laboratory tests.
12
13
14

15 480
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

481 **Funding statement**

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

483

484 **Competing interests**

485 A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no
486 conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT
487 reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of
488 interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no
489 conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR
490 reports no conflict of interest. VS reports no conflict of interest.

491

492 **Data sharing statement**

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

496

497 **Author contributorship statement**

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

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

- 1
2
3 504 Maaïke Maas: Conceptualization (Supporting), Resources (Supporting), Supervision
4
5 505 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
6
7
8 506 Heidi Ammerlaan: Conceptualization (Supporting), Resources (Supporting), Supervision
9
10 507 (Supporting), Validation (Equal), Writing-review & editing (Equal).
11
12
13 508 Roland van Balkom: Conceptualization (Supporting), Resources (Supporting), Supervision
14
15 509 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
16
17
18 510 Wendy Thijssen: Conceptualization (Supporting), Resources (Supporting), Supervision
19
20 511 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
21
22
23
24 512 Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision
25
26 513 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
27
28
29 514 Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
30
31
32 515 Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
33
34
35 516 Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
36
37
38 517 Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
39
40
41 518 Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
42
43
44 519 Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review & editing
45
46 520 (Equal).
47
48
49 521 Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision (Equal),
50
51 522 Writing-review & editing (Equal).
52
53
54 523 Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project
55
56 524 administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing
57
58 525 (Equal).
59
60

526 **References**

- 527
- 528 1 Coronavirus Disease (COVID-19) Situation Reports.
529 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
530 (accessed 4 Feb 2021).
- 531 2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in
532 China. <https://doi.org/10.1056/NEJMoa2002032> 2020;**382**:1708–20.
533 doi:10.1056/NEJMOA2002032
- 534 3 Vetter P, Vu DL, L’Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**.
535 doi:10.1136/BMJ.M1470
- 536 4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and
537 prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18.
538 doi:10.1136/bmj.m1328
- 539 5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine
540 Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-19):
541 A Systematic Review. *J. Med. Syst.* 2020;**44**:122. doi:10.1007/s10916-020-01582-x
- 542 6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical
543 Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417
- 544 7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT
545 Assessment Scheme for Patients Suspected of Having COVID-19—Definition
546 and Evaluation. <https://doi.org/10.1148/radiol2020201473> 2020;**296**:E97–104.
547 doi:10.1148/RADIOL.2020201473
- 548 8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity

- 1
2
3 549 score for acutely ill adults with suspected COVID-19: The PRIEST observational cohort
4
5 550 study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840
6
7
8 551 9 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a machine
9
10 552 learning tool to rule out COVID-19 among adults in the emergency department using
11
12 553 routine blood tests: A large, multicenter, real-world study. *J Med Internet Res*
13
14 554 2020;**22**:e24048. doi:10.2196/24048
15
16
17
18 555 10 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine clinical
19
20 556 data for patients attending hospital: development and prospective validation of an
21
22 557 artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87.
23
24 558 doi:10.1016/S2589-7500(20)30274-0
25
26
27
28 559 11 R Core Team. R: A Language and Environment for Statistical Computing.
29
30 560 2020.<https://www.r-project.org/>
31
32
33 561 12 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–29.
34
35 562 doi:10.1198/016214506000000735
36
37
38 563 13 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B*
39
40 564 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x
41
42
43
44 565 14 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models
45
46 566 via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01
47
48
49 567 15 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63.
50
51 568 doi:10.1093/oxfordjournals.pan.a004868
52
53
54 569 16 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-validation.
55
56 570 *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973
57
58
59 571 17 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to
60

- 1
2
3 572 analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-
4
5 573 2105-12-77
6
7
8 574 18 Harrell Jr FE. rms: Regression Modeling Strategies. 2021.https://cran.r-
9
10 575 project.org/package=rms
11
12
13 576 19 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons
14
15 and progress. *Scand. J. Clin. Lab. Invest.* 1999;**59**:491–500.
16 577 doi:10.1080/00365519950185229
17
18 578
19
20
21 579 20 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from
22
23 580 Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst*
24
25 581 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4
26
27
28
29 582 21 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of
30
31 583 SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J*
32
33 584 *Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502
34
35
36 585 22 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining clinical
37
38 586 and CT characteristics for the diagnosis of COVID-19. *Eur Radiol* 2020;**30**:6797–807.
39
40 587 doi:10.1007/s00330-020-07022-1
41
42
43
44 588 23 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-
45
46 589 infected patients at the emergency department using routine testing. *Clin Chem Lab Med*
47
48 590 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593
49
50
51 591 24 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a
52
53 592 diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An Int*
54
55 593 *J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305
56
57
58
59 594 25 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19 in
60

- 1
2
3 595 the emergency department: derivation and validation of the LDH-CRP-Lymphocyte
4
5 596 (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336
6
7
8 597 26 Tordjman M, Mekki A, Mali RD, *et al*. Pre-test probability for SARS-Cov-2-related
9
10 598 infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342.
11
12 599 doi:10.1371/journal.pone.0243342
13
14
15
16 600 27 Peto T, Affron D, Afrough B, *et al*. COVID-19: Rapid antigen detection for SARS-CoV-
17
18 601 2 by lateral flow assay: A national systematic evaluation of sensitivity and specificity for
19
20 602 mass-testing. *EClinicalMedicine* 2021;**36**:100924.
21
22 603 doi:10.1016/J.ECLINM.2021.100924
23
24
25
26 604 28 García-Fiñana M, Hughes DM, Cheyne CP, *et al*. Performance of the Innova SARS-
27
28 605 CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot:
29
30 606 population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637
31
32
33 607 29 Ford JS, Parikh A, Sandhu R, *et al*. Testing Asymptomatic Emergency Department
34
35 608 Patients for Coronavirus Disease 2019 (COVID-19) in a Low-prevalence Region. *Acad.*
36
37 609 *Emerg. Med.* 2020;**27**:771–4. doi:10.1111/acem.14044
38
39
40
41 610 30 Ravani P, Saxinger L, Chandran U, *et al*. COVID-19 screening of asymptomatic patients
42
43 611 admitted through emergency departments in Alberta: a prospective quality-improvement
44
45 612 study. *C Open* 2020;**8**:E887–94. doi:10.9778/cmajo.20200191
46
47
48
49 613
50
51 614
52
53
54
55
56
57
58
59
60

1
2
3 615 **Figure legends**
4

5 616
6
7

8 617 **Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B)**
9
10 618 **dataset.**

11
12
13 619 *All patient admissions with routine venous blood sampling at the emergency department (ED)*
14
15 620 *were included. For the development dataset, completeness of the lab panel was assessed for*
16
17 621 *all 28 laboratory tests, for the temporal validation dataset this was only necessary for 10*
18
19 622 *laboratory tests. The major causes of missingness are described in the text. In the*
20
21 623 *development dataset, presentations with extreme values (>10 SD) were excluded. The same*
22
23 624 *limits were applied to the temporal validation dataset (see Table 2 for limits).*
24
25
26
27

28 625

29
30
31 626 **Figure 2: Probability density plot of the CoLab-linear predictor.**

32
33
34 627 *The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are*
35
36 628 *plotted against the linear predictor (see table 2). The CoLab-score cut-offs (-5.83 , -4.02 , $-$*
37
38 629 *3.29 , -2.34 and -1.64) are depicted with vertical dashed lines. The white-boxed numbers*
39
40 630 *(between the cut-offs) represent the corresponding CoLab-score. Note that while the area*
41
42 631 *under both curves is identical (since these are probability density functions), in absolute*
43
44 632 *numbers the “negative or untested”-group is about 36 times larger than the PCR positive*
45
46 633 *group.*
47
48
49

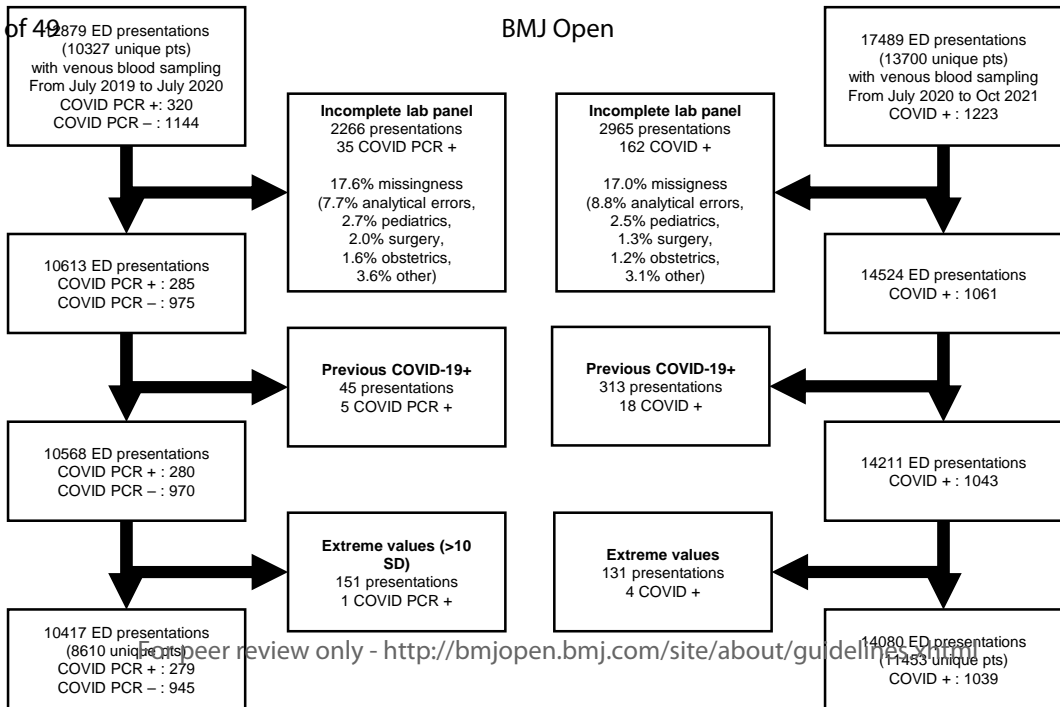
50
51 634

52
53
54 635 **Figure 3: Inclusion flow of ED patients in three external centers.**

55
56 636 *All emergency department (ED) presentations with routine venous blood sampling were*
57
58 637 *included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see*
59
60

1
2
3 638 *Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were*
4
5 639 *excluded as “previous COVID-19+”. Presentations with any laboratory result above the*
6
7 640 *limits of the CoLab-score (see Table 2) were excluded.*
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only



1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

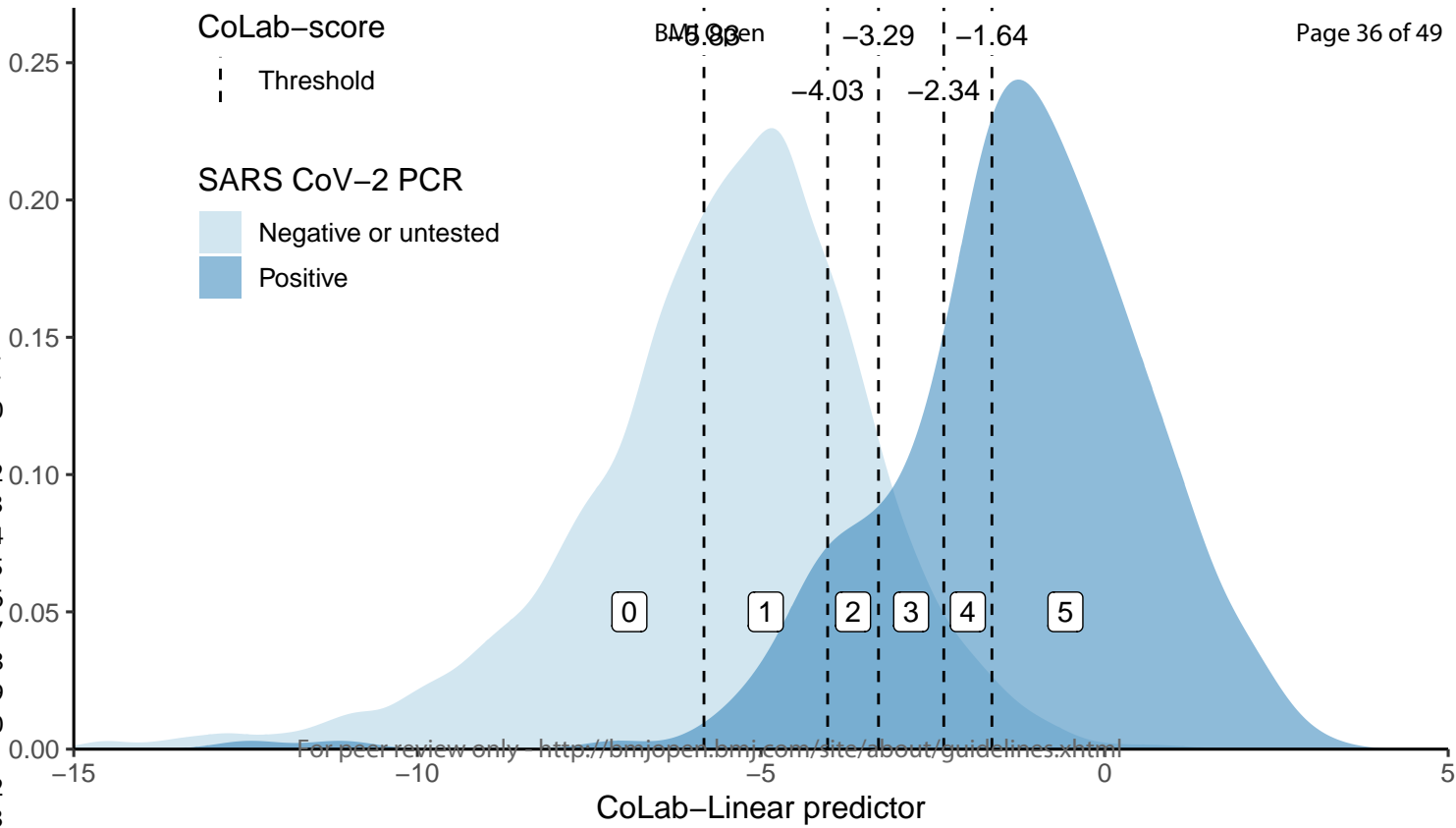
CoLab-score

Threshold

SARS CoV-2 PCR

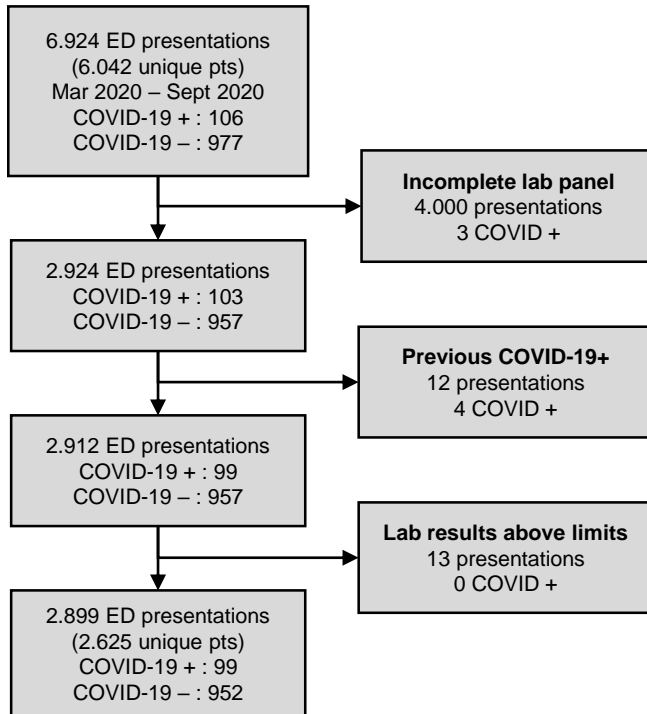
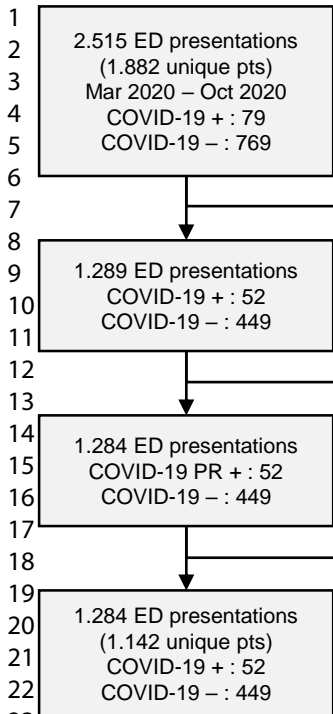
Negative or untested
Positive

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23

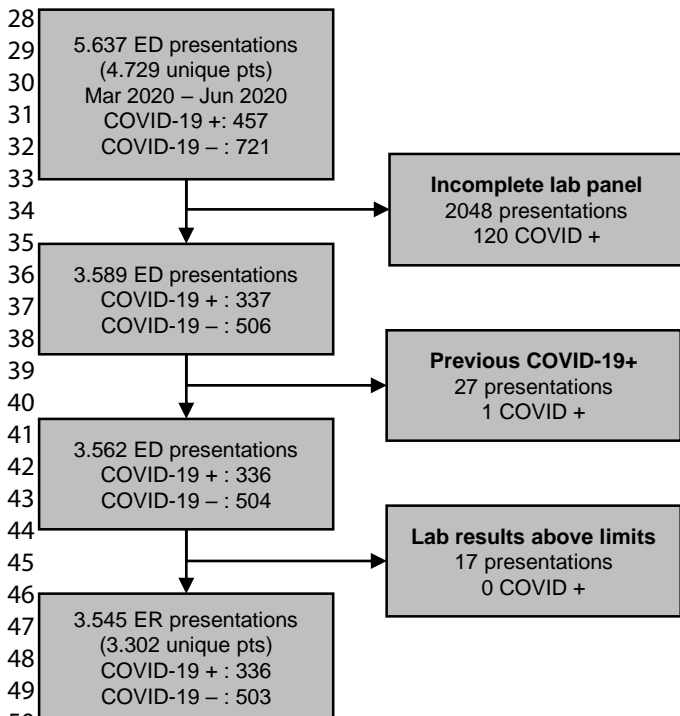


Center 1

Center 2



Center 3



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 (\hat{w}) was calculated by $\hat{w} = 1/|\hat{\beta}_{ridge}|^2$. 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: $lp_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in}$. Where n is the number of variables in the final model, x_{in} are the observed predictor variables for subject i and β_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:

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

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:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639$$

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$

$$lp_i(\tau) \geq \ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

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

$$lp_i(\tau = 0.4) \geq -5.83 \text{ (CoLab-score} = 1)$$

$$lp_i(\tau = 0.1) \geq -4.03 \text{ (CoLab-score} = 2)$$

$$lp_i(\tau = 0.05) \geq -3.29 \text{ (CoLab-score} = 3)$$

$$lp_i(\tau = 0.02) \geq -2.34 \text{ (CoLab-score} = 4)$$

$$lp_i(\tau = 0.01) \geq -1.64 \text{ (CoLab-score} = 5)$$

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

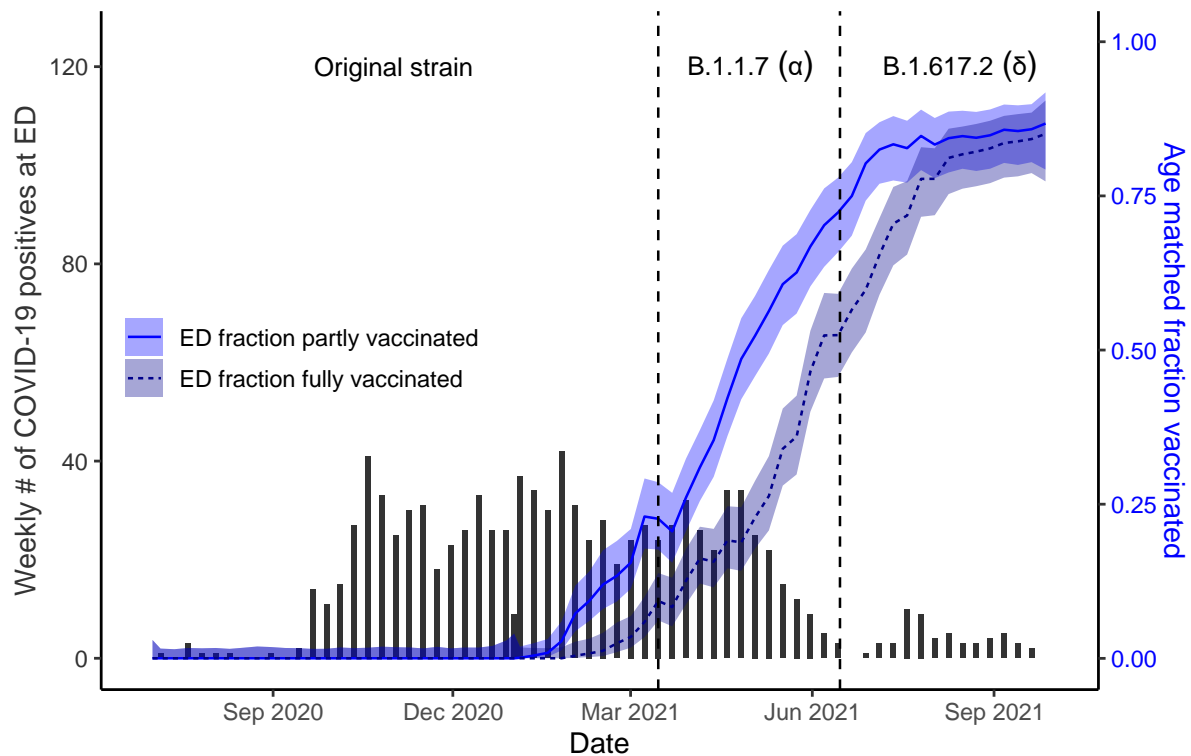


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

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).

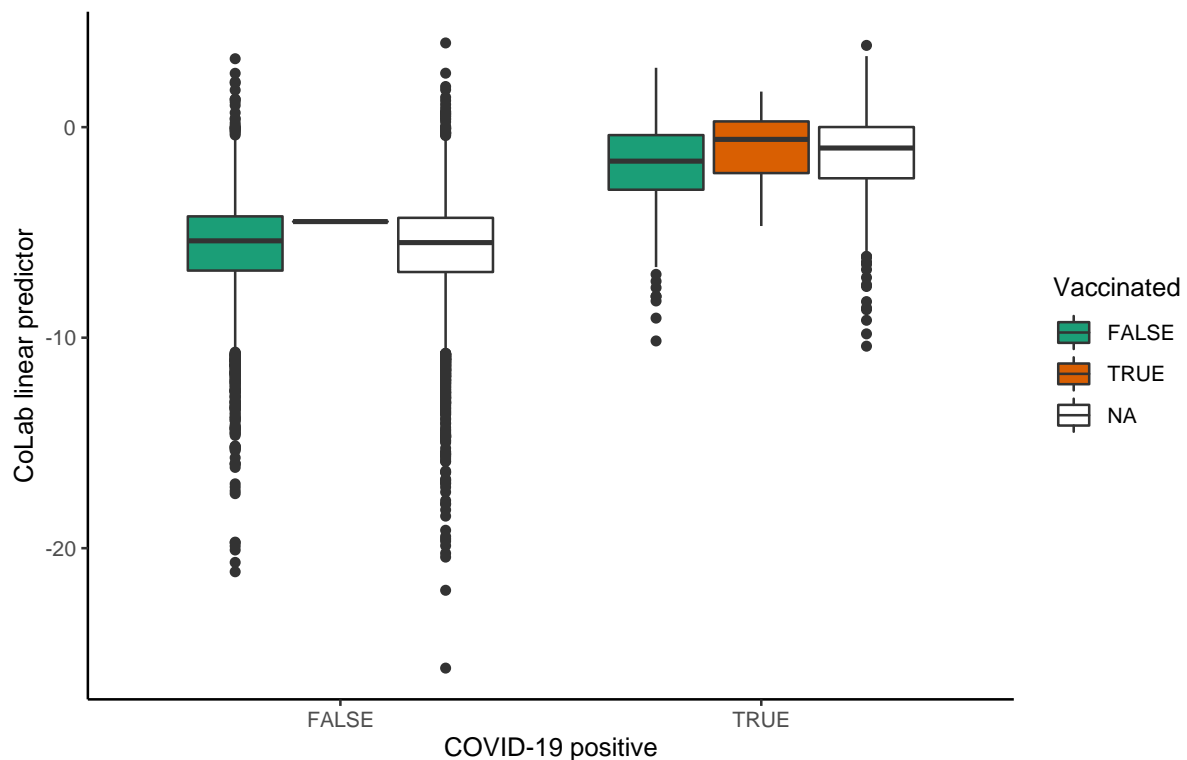


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

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

1
2
3 *vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13*
4
5 *presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative.*
6
7 *Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or*
8
9 *considered positive until proven otherwise, therefore there is only one COVID-19 negative*
10
11 *patient with a registered vaccination status.*
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Supplemental material 3

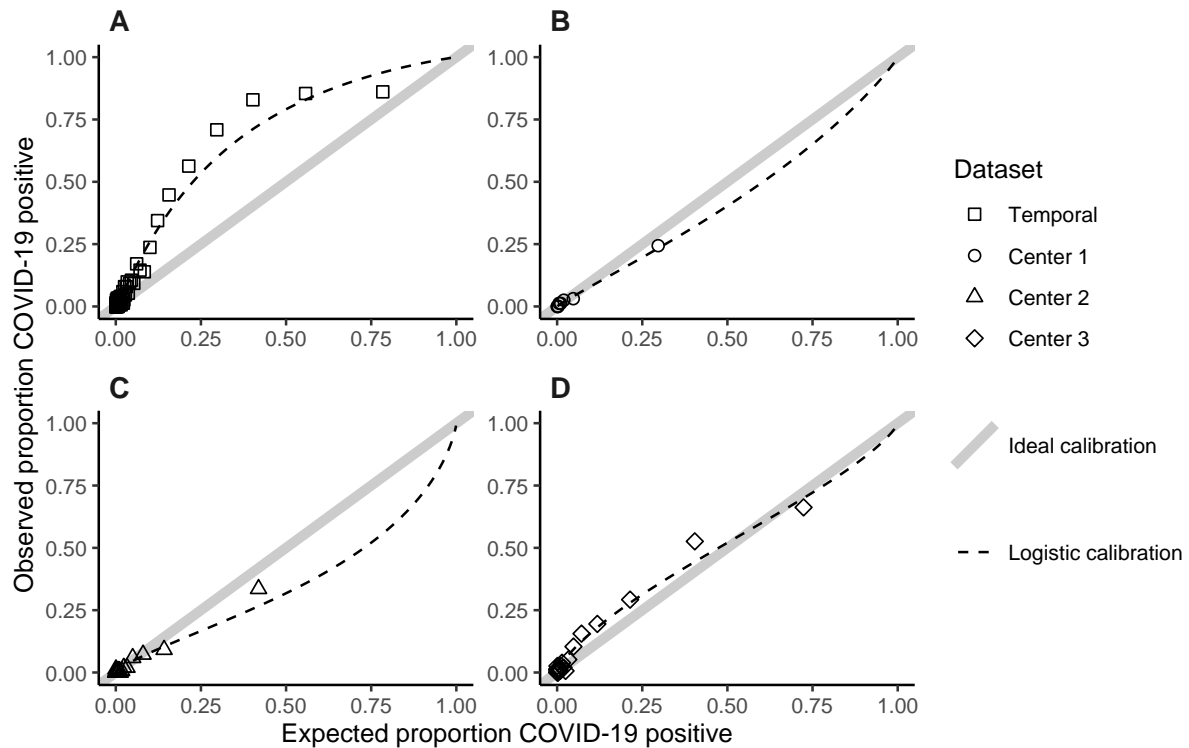


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

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.

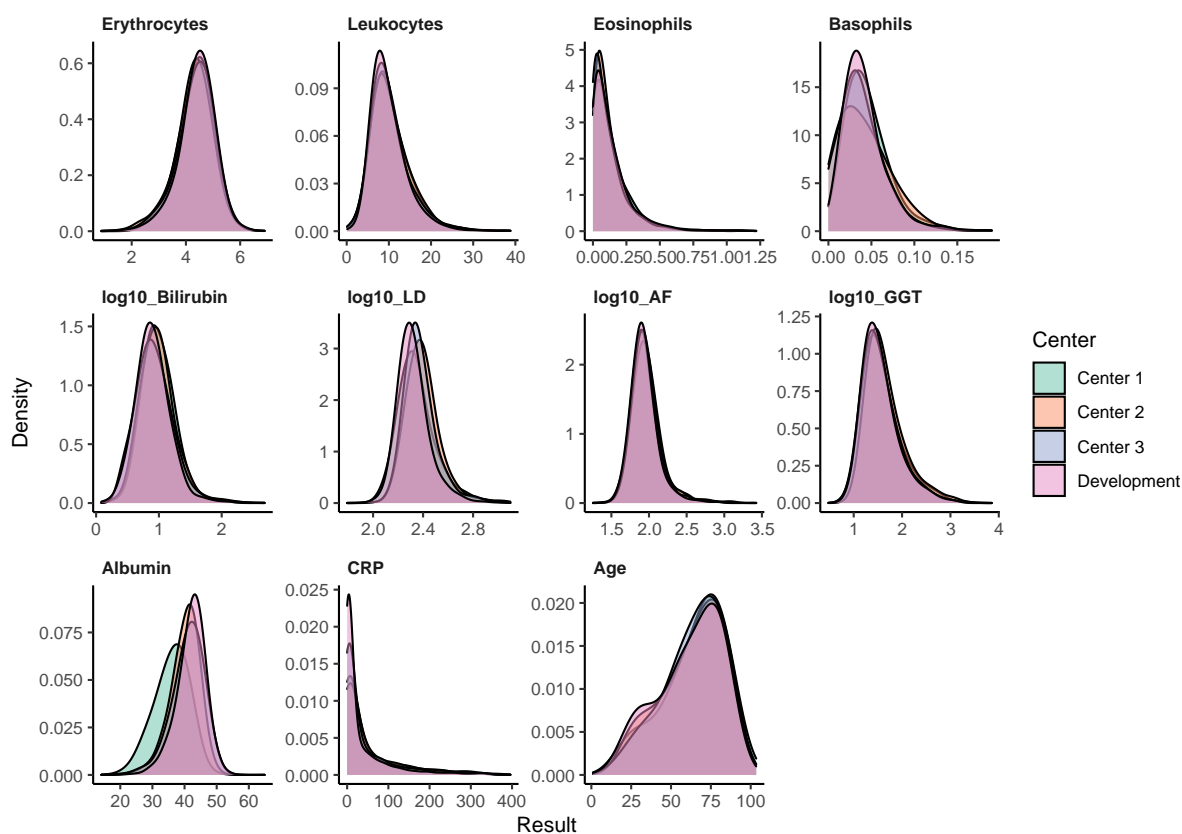


Figure 2: Probability density plots of laboratory parameters.

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

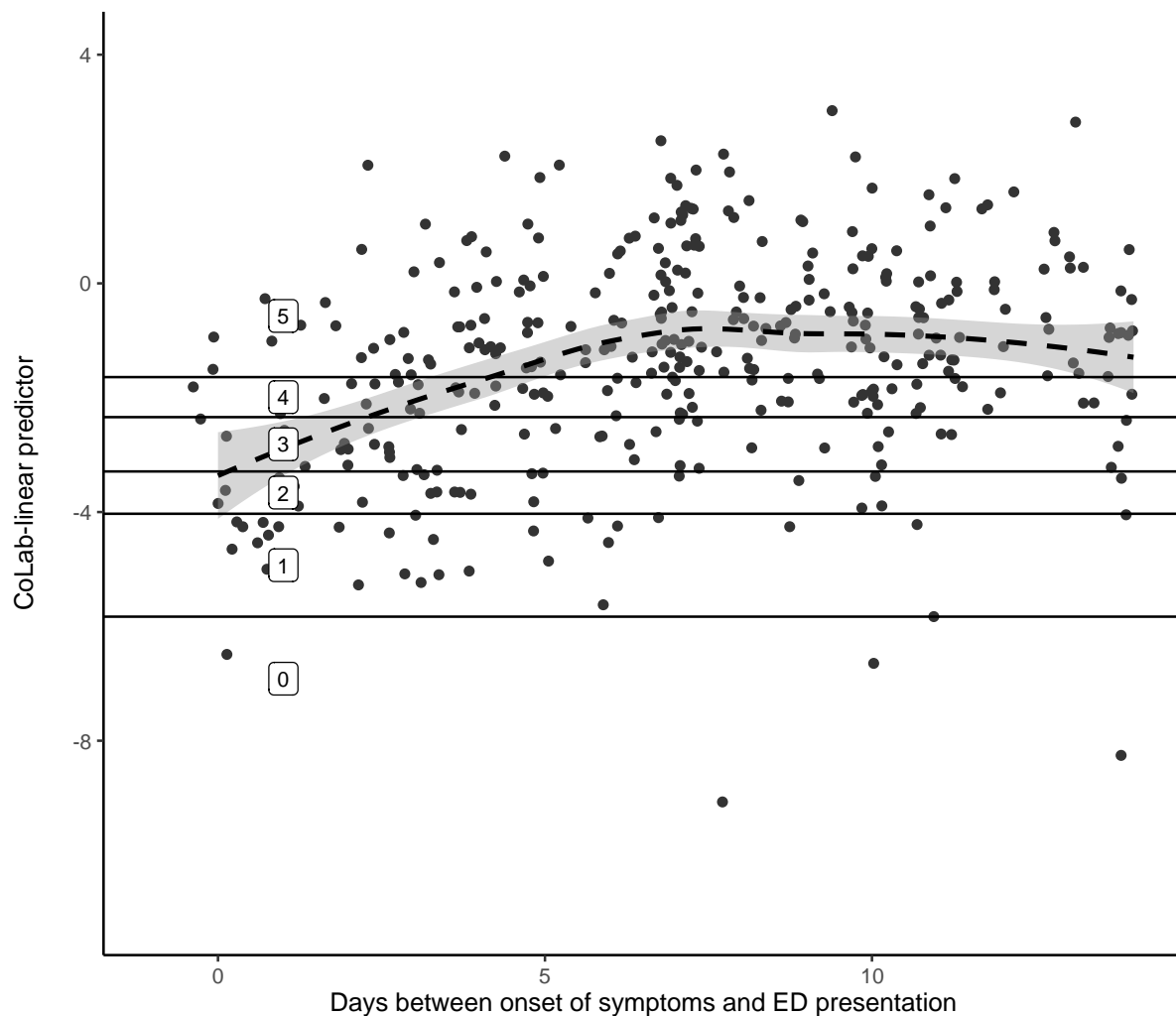


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predictor 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.

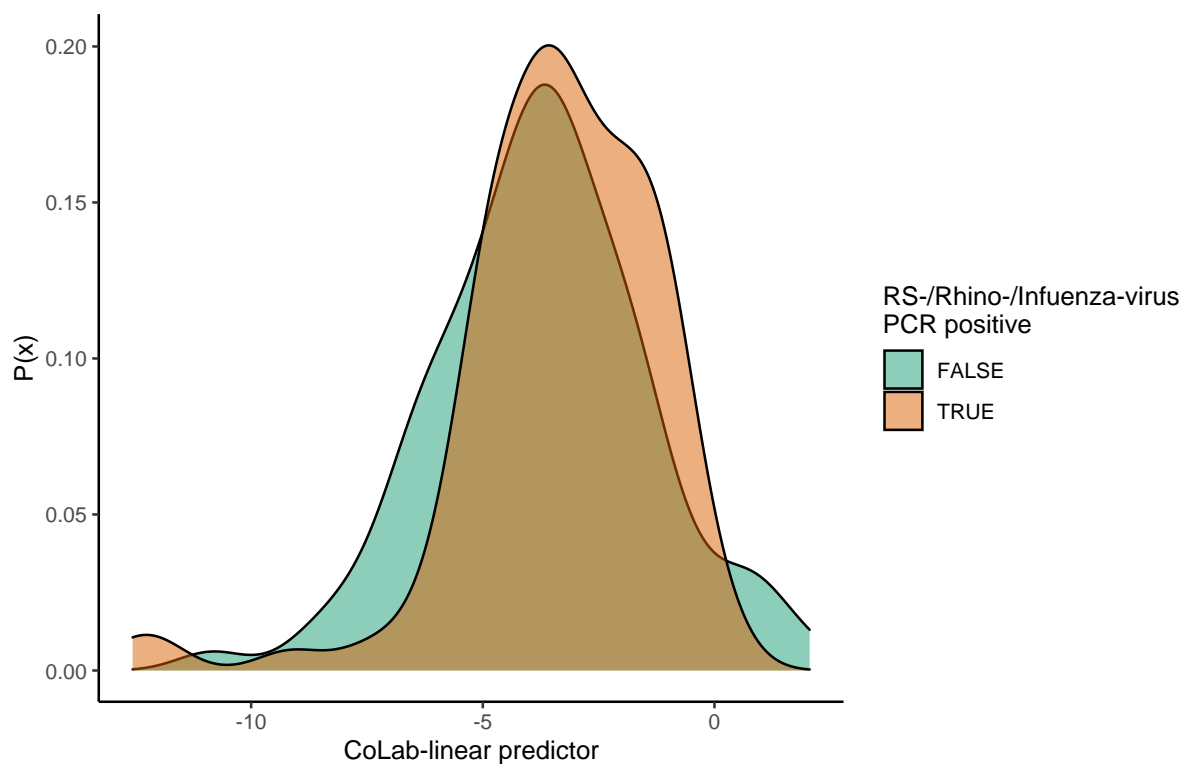


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

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: 0.4896-0.6563).

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)

1
2
3 **Table 1: Sensitivity analysis of the CoLab-score in the temporal validation dataset using**
4 **different inclusion criteria.**
5

6
7 *Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV),*
8 *true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are*
9 *shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals*
10 *in parentheses. The temporal validation dataset is used to compare the performance of the*
11 *CoLab-score with inclusion criteria that differ from the development dataset. The first line*
12 *shows the performance of the temporal validation dataset with the original inclusion criteria*
13 *as specified in Figure 1B. The second line shows the performance of the CoLab-score when*
14 *all re-presentations are excluded (i.e. no repeated presentations). The third line shows the*
15 *performance of the CoLab-score in the subgroup of patients that underwent PCR-testing.*
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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 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:	<i>BMJ Open</i>
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, Maaïke; 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

SCHOLARONE™
Manuscripts



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.

1
2
3
4 1 **Development and validation of an early warning score to identify**
5
6 2 **COVID-19 in the emergency department based on routine laboratory**
7
8
9 3 **tests: a multicenter case-control study**
10
11
12 4

15 5 Arjen-Kars Boer^{1*}, Ruben Deneer^{1,2**†}, Maaïke Maas³, Heidi S. M. Ammerlaan⁴, Roland H. H.
16
17 6 van Balkom⁵, Wendy A.M.H. Thijssen³, Sophie Bennenbroek³, Mathie P. G. Leers⁶, Remy
18
19 7 J.H. Martens⁶, Madelon M. Buijs⁷, Jos J. Kerremans⁸, Muriël Messchaert⁹, Jeroen D.E. van
20
21 8 Suijlen⁹, Natal A.W. van Riel^{2,10} and Volkher Scharnhorst^{1,2}
22
23
24
25 9

26
27
28 10 * Both authors contributed equally
29

30
31 11 † Corresponding author
32
33
34 12

35
36
37 13 1 Department of Laboratory Medicine, Catharina Hospital Eindhoven, Michelangelolaan
38
39 14 2, 5623 EJ, Eindhoven, the Netherlands

40
41 15 2 Faculty of Biomedical Engineering, Technical University Eindhoven, Groene Loper 3,
42
43 16 5612 AE, Eindhoven, the Netherlands

44
45
46 17 3 Department of Emergency Medicine, Catharina Hospital Eindhoven,
47
48 18 Michelangelolaan 2, 5623 EJ, Eindhoven, the Netherlands

49
50 19 4 Department of Internal Medicine, Catharina Hospital Eindhoven, Michelangelolaan 2,
51
52 20 5623 EJ, Eindhoven, the Netherlands

53
54
55 21 5 Department of Pulmonary Diseases, Catharina Hospital Eindhoven, Michelangelolaan
56
57 22 2, 5623 EJ, Eindhoven, the Netherlands
58
59
60

- 1
2
3 23 6 Department of Laboratory Medicine, Zuyderland Medical Center, Dr. H. van der
4
5 24 Hoffplein 1, 6162 BG, Sittard-Geleen, the Netherlands
6
7 25 7 Atalmedial Diagnostic Centers, Jan Tooropstraat 138, 1061 AD Amsterdam, the
8
9 26 Netherlands
10
11
12 27 8 Department of Medical Microbiology & Infection Prevention, Alrijne Hospital, Simon
13
14 28 Smitweg 1, 2353 GA Leiderdorp, the Netherlands
15
16
17 29 9 Department of Clinical Chemistry, Gelre Hospital, Albert Schweitzerlaan 31, 7334 DZ
18
19 30 Apeldoorn, the Netherlands
20
21 31 10 Amsterdam University Medical Centers, University of Amsterdam, Meibergdreef 9,
22
23 32 1105 AZ Amsterdam, the Netherlands.
24
25
26
27 33
28
29

30 **Keywords**

31
32 35 COVID-19, SARS-CoV-2, emergency department, triage, early warning score, prediction
33
34 36 model, routine laboratory tests
35
36
37
38
39
40

41 **Corresponding author**

42
43 39 Ruben Deneer

44
45
46 40 Michelangelolaan 2

47
48
49 41 5623 EJ Eindhoven, the Netherlands

50
51
52 42 ruben.deneer@catharinaziekenhuis.nl
53
54
55
56
57
58
59
60

44 **Abstract**

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

50 **Design:** Case-control study.

51 **Setting:** Secondary and tertiary hospitals in the Netherlands.

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

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

59 **Results:** The resulting “CoLab-score” consists of 10 routine laboratory measurements, and
60 age. The score showed good discriminative ability (AUC: 0.930, 95% CI: 0.909 to 0.945).
61 The lowest CoLab-score had a high sensitivity for COVID-19 (0.984, 95% CI: 0.970 to 0.991,
62 specificity: 0.411, 95% CI: 0.285 to 0.520). Conversely, the highest score had high specificity
63 (0.978, 95% CI: 0.973 to 0.983, sensitivity: 0.608, 95% CI: 0.522 to 0.685). Results were
64 confirmed in temporal and external validation.

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

1
2
3 67 ruled-out in over one third of ED presentations. Highly suspect cases can be identified
4
5 68 regardless of presenting symptoms. The CoLab-score is continuous, in contrast to the binary
6
7
8 69 outcome of lateral flow testing, and can guide PCR testing and triage ED patients.
9

10
11 70

13 71 **Article summary**

17 72 **Strengths and limitations of this study**

- 19 73 • A comprehensive panel of 28 laboratory tests was measured for 10.417 emergency
20
21 74 department (ED) presentations and combined with SARS-CoV-2 PCR test results.
- 23 75 • Using adaptive lasso regression analysis, the panel of 28 laboratory tests was reduced
24
25
26 76 to a single score consisting of a subset of 10 routine ED laboratory tests and age.
- 28 77 • The score was temporally validated from July 2020 to October 2021, in the presence
29
30
31 78 of vaccine roll-out and emergence of new SARS-CoV-2 variants.
- 33 79 • The score was externally validated in 3 other centers in the Netherlands.
- 35 80 • Missingness in the panel of laboratory tests varied between external centers, limiting
36
37
38 81 generalizability of the score to the ED population for which the complete panel of
39
40 82 laboratory tests was available.
- 42 83 • The score was not directly compared to lateral flow testing.

85 Introduction

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

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

100 Many COVID-19 prediction models have already been developed, the living systematic
101 review by Wynants et. al [4] provides an extensive overview and critical appraisal.

102 Unfortunately, only few models have found their way into routine care at the ED [5,6]. Early
103 models were based on relatively small sample sizes, hampered by selection bias or were over-
104 fitted by selecting too many features [4–6]. Aside from methodological shortcomings of early
105 models, most models are not developed as an early warning score for all ED patients. Firstly,
106 they require features from tests that are not routinely performed or logged for all ED patients
107 (e.g. the CO-RADS score from a CT-scan [7] or non-lab based clinical variables in the
108 PRIEST EWS [8]) and are therefore not straightforward to implement or scale to a large ED
109 patient population. Secondly, the population on which models are commonly based, are PCR-

1
2
3 110 tested patients, i.e. a pre-selection of a possible COVID-19 infection has already been done by
4
5 111 physicians.

6
7 112 Only two studies were identified that focus on patients presenting at the ED, include
8
9
10 113 unsuspected (and pre-pandemic) patients as controls, and rely solely on routine (laboratory)
11
12 114 tests [9,10].

13
14
15 115 In this study we report the development and validation of an early warning score that, based
16
17 116 on routine ED laboratory tests, estimates the risk of a possible COVID-19 infection in patients
18
19 117 who undergo routine laboratory testing at presentation. The score can assist ED physicians in
20
21 118 triaging patients and prevent further transmission of COVID-19 by quickly identifying
22
23
24 119 possibly infected patients or ruling out a possible infection when resources are scarce.
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

121 **Methods**

122 *Study design*

123 This is a retrospective case-control study where routine laboratory test results, combined with
124 age and gender, from all patient presenting at the emergency department (ED) of the
125 Catharina Hospital Eindhoven from July 2019 to July 2020 were combined with SARS-CoV-
126 2 PCR test results in a development dataset. A model that could predict the presence of a
127 COVID-19 infection was fit to this dataset. Performance of the model was assessed by i)
128 internal validation, ii) temporal validation and iii) external validation by using data from the
129 ED of three other centers. The study was reviewed by the Medical research Ethics
130 Committees United (MEC-U) under study number W20.071, which confirmed that the
131 Medical Research Involving Human Subjects Act (In Dutch: WMO) does not apply to this
132 study. The study was thereafter reviewed and approved by the internal hospital review board.

133

134 *Patient and Public Involvement*

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

136

137 *Development dataset*

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

1
2
3 145 quantitative results were not available due to analytical interference (hemolysis, lipemia or
4
5 146 icterus). The routine ED laboratory panel is requested for (adult) patients presenting with
6
7 147 abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific
8
9 148 complaints, or for patients (including non-adult patients) presenting with specific complaints
10
11 149 where a suspected diagnosis has to be ruled-in or ruled-out. Presentations with one or more
12
13 150 missing values in any of the 28 laboratory test in the routine ED panel, were excluded.
14
15 151 Presentations with one or more extreme lab results, > 10 times standard deviation from the
16
17 152 median, were also excluded to minimize the effect on the estimation of regression
18
19 153 coefficients. The median was chosen as a measure of central tendency due to its resistance for
20
21 154 outliers. After the first case of COVID-19 in the Netherlands, all patients with symptoms of
22
23 155 COVID-19 (either fever and/or respiratory symptoms) were subjected to nasopharyngeal PCR
24
25 156 testing for SARS-CoV-2 RNA. PCR testing was performed by commercial tests that were
26
27 157 approved by the Dutch national institute of public health (RIVM). If a patient had a positive
28
29 158 PCR result in the past, subsequent presentations were excluded as re-presentations might be
30
31 159 clinically different from de novo presentations.
32
33
34
35
36
37
38 160 The ED lab panel results were matched to SARS-CoV-2 PCR results if the underlying
39
40 161 nasopharyngeal swab had been taken ≤ 1 day prior, or ≤ 1 week after initial blood withdrawal
41
42 162 at the ED. If multiple PCR tests were performed in this window, and at least one PCR test was
43
44 163 positive, the presentation was labelled "*PCR-positive*". If all PCR test results in the time
45
46 164 window were negative, the presentation was labelled as "*PCR-negative*". If no PCR tests were
47
48 165 performed in the time window and the presentation occurred after the first case of COVID-19
49
50 166 in the Netherlands, the presentation was labelled as "*Untested*". All presentations before the
51
52 167 first case were labelled as "*Pre-COVID-19*".
53
54
55
56
57
58
59
60

169 *Laboratory tests*

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

183 *Modelling*

184 All data were processed and analyzed in R version 4.1.1 [11]. Laboratory results, combined
185 with age and gender were used as covariates in a regression model. Cases were defined as ED
186 presentations labelled as “*PCR-positive*”, controls were all other presentations (i.e. “*PCR-*
187 *negative*”, “*Untested*” or “*Pre-COVID-19*”). To achieve predictive accuracy, limit overfitting
188 and perform feature selection, penalized logistic regression with an adaptive lasso penalty was
189 chosen [12,13]. To minimize missing data, all non-numeric results at the extremes of the
190 measuring range, were converted to numeric results by removing the “<” and “>” signs. For
191 eGFR (CKD-epi) and CRP the raw precursor value was used instead of >90 ml/min/m² and
192 <6 mg/L, respectively. Considering that laboratory results of bilirubin, ASAT, ALAT, LD,
193 CK, ALP and gGT can have heavy (right) tailed distributions, which in turn impacts model

1
2
3 194 predictions, these variables were transformed logarithmically. More details regarding model
4
5 195 fitting can be found in the document, **Supplemental Material 1**. Models were fitted using the
6
7 196 glmnet-package [14].
8
9

10
11 197

12 13 198 *CoLab-score*

14
15
16 199 Since this is a retrospective case-control study, the sample prevalence may not reflect the
17
18 200 true/current COVID-19 prevalence. To obtain well-calibrated probabilities the intercept term
19
20 201 in the model should be adjusted according to the current prevalence (details can be found in
21
22 202 the document, **Supplemental Material 1**) [15]. However, adjusting the intercept term is not
23
24 203 straightforward to implement in clinical practice, therefore the linear predictor of the model
25
26 204 was categorized into a score, this score is hereafter referred to as the “CoLab-score”. The
27
28 205 categorization is based on a number needed to test of 15 (i.e. one is willing to PCR test 15
29
30 206 patients to find one positive) and prevalence cut-points of 1%, 2%, 5%, 10% and 40% using
31
32 207 the intercept adjustment formula by King [15]. The intervals obtained through these breaks
33
34 208 correspond to CoLab-scores 5 to 0, respectively. Score 0 reflects low-risk for COVID-19 and
35
36 209 score 5 reflects high-risk. More details regarding the rationale of the CoLab-score
37
38 210 categorization can be found in the document, **Supplemental Material 1**.
39
40
41
42
43
44
45

46
47 211

48 212 *Internal validation*

49 213 To assess model performance while taking overfitting into account, bootstrapping was
50
51 214 performed. 1000 bootstrap samples were generated from the original data. On each bootstrap
52
53 215 sample, the full model fitting procedure and CoLab-score conversion were performed.
54
55 216 Optimism adjusted performance measures of the CoLab-score were obtained by applying the
56
57 217 0.632 bootstrap rule to the in-sample and out-of-bag-sample performance [16]. Performance
58
59
60

1
2
3 218 measures included, AUC, sensitivity, specificity, positive predictive value (PPV) and negative
4
5 219 predictive value (NPV) of each CoLab-score. The pROC-package was used to calculate
6
7 220 performance measures [17]. Although the full inclusion period from July 2019 to July 2020
8
9 221 was used for model fitting, the performance was evaluated on the period starting from the first
10
11 222 COVID-19 infection (24th of February 2020) to July 2020. This was done to obtain
12
13 223 performance measures that would reflect real world performance.
14
15
16
17
18 224

20 225 *Temporal validation*

22 226 For temporal validation, results from our center were prospectively analyzed from July 2020
23
24 227 to October 2021. During this period, the Netherlands was struck by a second wave of COVID-
25
26 228 19 infections, starting in the fall of 2020 and subsiding in the summer of 2021. In this period
27
28 229 there was also more widespread external PCR testing by municipal health services. The
29
30 230 results of external conducted PCR tests were not available to our study. To overcome this
31
32 231 limitation, the outcome in the temporal validation cohort was chosen as a composite of the
33
34 232 hospital registration of a confirmed COVID-19 infection and/or at least one positive PCR test
35
36 233 result. This period also covers both the emergence of new SARS-CoV-2 variants as well as
37
38 234 vaccine rollout. However, neither vaccination status nor genomic sequencing was available to
39
40 235 determine whether a patient was vaccinated or which variant caused the infection. Therefore,
41
42 236 data from the Dutch national institute of public health (RIVM) was used, to divide the
43
44 237 temporal validation period into three phases: i) from July 2020 until March 2021, no
45
46 238 vaccination and no variants of concern identified ii) from March 2021 until June 2021, partial
47
48 239 vaccination and B.1.1.7 (Alpha) variant identified as dominant iii) from June 2021 until
49
50 240 October 2021, widespread vaccination and B.1.617.2 (Delta) variant identified as dominant.
51
52 241 See **Supplemental Material 2 Figure 1** for more details. The temporal validation consisted
53
54 242 of assessing the AUC, sensitivity, specificity, PPV and NPV of each CoLab-score threshold
55
56
57
58
59
60

1
2
3 243 for the entire period, as well as for each phase separately to determine a possible effect of
4
5 244 vaccination and new variants on performance (results in the **Supplemental Material 2**).
6
7 245 Model calibration was assessed graphically using the rms-package [18].
8
9

10
11 246

12
13
14 247 *External validation*

15
16 248 For the external validation, several centers in the Netherlands were approached and assessed
17
18 249 if the required panel of laboratory tests and SARS-CoV-2 PCR test results were available.
19
20 250 Seven centers responded and three centers fulfilled the inclusion criteria: Gelre Hospitals
21
22 251 (center 1), Atalmedial Diagnostic Centers, location Alrijne Hospital Leiderdorp (center 2) and
23
24 252 Zuyderland Medical Center (center 3). The hematological parameters were measured with
25
26 253 Sysmex XN10/XN20 (center 1), CELL-DYN-Sapphire (Abbott Laboratories) (center 2) and
27
28 254 Sysmex XN10 instruments (center 3). The clinical chemistry parameters were measured with
29
30 255 Architect c14100/c160000 (Abbott Laboratories) (center 1), Architect ci4100 (Abbott
31
32 256 Laboratories) (center 2) and Cobas 8000 instruments (Roche Dx) (center 3). The external
33
34 257 validation was similar to the temporal validation and consisted of assessing the AUC
35
36 258 sensitivity, specificity, PPV and NPV of each CoLab-score threshold. Calibration was
37
38 259 assessed graphically analogous to the temporal validation dataset.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

261 Results

262 Development dataset

263 12879 emergency department (ED) presentations of 10327 patients from July 2019 to July
 264 2020 were included. After excluding cases with an incomplete lab panel, patient presentations
 265 that occurred after a positive PCR test in the past (re-presentations) and presentations with
 266 extreme values (>10 times standard deviation) in any of the lab results, 10417 presentations of
 267 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/m ²	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]

268

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

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

279

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

285

286 **CoLab-score**

287 The model obtained through adaptive lasso regression contained eleven variables, which are
 288 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 μ mol/L	-1.142	Bilirubin >169 μ mol/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 %

289

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

291 *The CoLab-linear predictor (LP) is calculated by summing the intercept and the products of*
 292 *the 11 variables with their corresponding coefficients (β 's). CoLab-LP = - 6.885 +*
 293 *[erythrocytes] \times 0.9379 - [leukocytes] \times 0.1298 - [eosinophils] \times 6.834 - [basophils] \times*
 294 *47.7 - log₁₀([bilirubin]) \times 1.142 + log₁₀([LD]) \times 5.369 - log₁₀([ALP]) \times 3.114 +*
 295 *log₁₀([gGT]) \times 0.3605 - [albumin] \times 0.1156 + [CRP] \times 0.02560 + [age] \times 0.002275. The*
 296 *LP can be converted into a CoLab-score (see Figure 2) or into a probability if the prevalence*
 297 *is known or estimated (see details in Supplemental Material 1). The CoLab-score is not valid*
 298 *if any of the variables exceed the limits in the third column. The relative importance ranks the*
 299 *importance of variables in predicting the outcome, relative to the most important variable (in*
 300 *this case basophils).*

301

302 A larger β -coefficient does not imply that a variable is more important in predicting the odds
 303 of testing positive for SARS-CoV-2, since variables are on different scales. The most
 304 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**.

308

309 *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)

313

314 **Table 3: Bootstrapped diagnostic performance of the CoLab-score in the development** 315 **dataset.**

316 *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*
319 *cut-offs (CoLab-score 0 till ≤ 4). The numbers in round brackets represent the 95% optimism-*
320 *adjusted*

1
2
3 321 *adjusted bootstrapped confidence intervals. The first column defines the threshold above*
4
5 322 *which CoLab-score a patient is considered positive. Note that “0” lists the sensitivity and*
6
7 323 *NPV of CoLab-score 0 and “ ≤ 4 ” lists the specificity and PPV of CoLab-score 5. Also note*
8
9 324 *that TP, TN, FP and FN are not whole numbers, as these are obtained through bootstrapping*
10
11 325 *and each bootstrap replicate contains a different number of controls and cases.*
12
13
14
15 326

16
17
18 327 Diagnostic performance is shown in **Table 3**. A CoLab-score of 0 has a negative predictive
19
20 328 value (NPV) of 0.997 (95% CI: 0.993 to 0.999) and positive predictive value (PPV) of 0.115
21
22 329 (0.0934 - 0.147), one third (38%, 95% CI: 28 to 514%) of all ED presentations were assigned
23
24 330 this score and can therefore be safely excluded. Conversely, 6% (95% CI: 6 to 8%) of the ED
25
26 331 patients had a CoLab-score = 5. Given the PPV of this score (0.683, 95% CI: 0.628 to 0.746,
27
28 332 NPV: 0.970, 95% CI: 0.963 - 0.978), subsequent PCR testing is advised.
29
30
31
32
33
34

35 334 *Temporal validation*

36
37 335 As the CoLab-score was developed in our center after the first COVID-19-wave in the
38
39 336 Netherlands, the performance was evaluated in our center from July 2020 until October 2021.
40
41 337 Lab results from 17489 ED presentations were collected. After applying the inclusion flow as
42
43 338 shown in **Figure 1 B**, 14080 presentations remained, of which 1039 were associated with a
44
45 339 COVID-19 infection.
46
47
48
49

50 340 The mean prevalence in this period was 7.4%. The AUC of the CoLab-score in the temporal
51
52 341 validation set is 0.916 (95% CI: 0.906 to 0.927). The performance is comparable to the
53
54 342 development cohort, although sensitivity is slightly lower and specificity slightly higher (cf.
55
56 343 **Table 3** and **Table 4**). The temporal validation dataset was also split into three phases
57
58 344 according to dominant SARS-CoV-2 variants and vaccine roll-out (see **Supplemental**
59
60

1
2
3 345 **Material 2 Figure 1**). The discriminative ability was not lower in the second or third phase,
4
5 346 compared to the first phase. Diagnostic performance is preserved in terms of sensitivity and
6
7 347 specificity, except a moderately reduced sensitivity of scores ≥ 3 in the third phase as
8
9 348 compared to the first phase. PPV and NPV are incomparable due to different prevalence/pre-
10
11 349 test probabilities in each phase (see **Supplemental Material 2 Table 1**).

12
13
14
15 350 In terms of the predicted probabilities, model calibration shows that overall predicted
16
17 351 probabilities are too low (see **Supplemental Material 3 Figure 1** for the calibration plot),
18
19 352 which is expected since the prevalence differs and the intercept has to be adjusted to the
20
21 353 prevalence.

22
23
24 354 In this period at least 22 COVID-19 positive patients were identified by the CoLab-score, that
25
26 355 initially did not present with COVID-specific symptoms. Most patients had neurological or
27
28 356 orthopedic presenting symptoms.

29
30
31
32 357

33 34 35 358 *External validation*

36
37 359 For external validation, data obtained from three other centers were used, center 1 (N = 1284,
38
39 360 52 COVID-19 positive), center 2 (N = 2899, 99 COVID-19 positive) and center 3 (N = 3545,
40
41 361 336 COVID-19 positive). The inclusion flow is summarized in **Figure 3**. COVID-19
42
43 362 prevalence differed between the three centers (4.0%, 3.4% and 9.5% respectively) and was
44
45 363 lower in centers 1 and 2, and higher in center 3 than in the development dataset. The AUCs of
46
47 364 the CoLab-score are 0.904 (95% CI: 0.866 to 0.942), 0.886 (95% CI: 0.851 - 0.922) and 0.891
48
49 365 (95% CI: 0.872 - 0.909), for centers 1, 2, and 3 respectively.

50
51
52
53
54 366 Diagnostic performance is shown in **Table 4**. The sensitivity of CoLab-score 0 in all centers
55
56 367 is ≥ 0.96 . Therefore, the NPV of CoLab-score 0 was more than 99%. Calibration plots for
57
58 368 external centers are shown in **Supplemental Material 3 Figure 1**, the observed fraction of
59
60

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)
	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)
≤ 1	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)
	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)
≤ 2	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)

371
 372 **Table 4: Diagnostic performance of the CoLab-score in the validation dataset (temporal)**
 373 **and three external hospitals.**
 374 *Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV),*
 375 *true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are*
 376 *shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals*
 377 *in parentheses. Note that “0” lists the sensitivity and NPV of CoLab-score 0 and “≤ 4” lists*
 378 *the specificity and PPV of CoLab-score 5.*

380 Discussion

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

385 The main strength of this study is that this score can be used as an early-warning or triaging
386 tool for the ED population presenting with abdominal pain, chest pain, shortness of breath,
387 syncope, sepsis or other non-specific complaints where a routine blood panel is requested.
388 This is in contrast to the vast majority of COVID-19 diagnostic models that have been
389 developed on a pre-selected population of PCR-tested patients [9,20–26]. Moreover, the
390 CoLab-score requires only routine blood tests, instead of (features from) imaging such as CT-
391 scans or laboratory tests that are not routinely collected in the ED, e.g. interleukin-6 or 3-
392 hydroxybuteric acid [4]. Compared to lateral flow tests (LFTs), which provide a dichotomous
393 result within 30 minutes and are widely adopted in EDs, the CoLab-score is a continuous
394 score. The lowest CoLab-scores (0 - 1) offer higher sensitivity and are therefore more suitable
395 to rule-out COVID-19 than a LFT, which are only moderately sensitive (albeit more specific)
396 [27,28].

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

1
2
3 405 Since this is a retrospective case-control study, there is some unavoidable missing data. In our
4
5 406 cohort 17.6% of the ED presentations could not be used due to one or more missing
6
7 407 laboratory results. This is lower or equal to similar studies; 22% [23], 17% [21] and 11% [26].
8
9 408 Important to note is that 7.7% of missingness is due to analytical errors which can be assumed
10
11 409 to be missing completely at random. For the remaining 9.9% of missingness, the full lab panel
12
13 410 was most frequently missing for pediatric, obstetric and surgery patients. These patients are
14
15 411 presenting with specific complaints for which specific laboratory tests are requested, and
16
17 412 hence do not match the inclusion criteria for a routine blood panel. Overall the missingness
18
19 413 was significantly lower in the PCR-tested group versus the untested group (χ^2 -test p-value
20
21 414 <0.001). It is assumed that all presentations in the untested group are COVID-19 negative.
22
23 415 However, some presentations with asymptomatic COVID-19 could be present in the untested
24
25 416 control group. The impact of these 'false controls' is most likely small as other studies
26
27 417 indicate that there is a very low positivity rate among asymptomatic ED presentations (only a
28
29 418 few in over a thousand tested asymptomatic cases) [29,30]. The vast majority of controls were
30
31 419 not tested for COVID-19, because they were either pre-pandemic or untested patients (89% in
32
33 420 the development dataset). Clinical data always contains some unavoidable 'noise' in the form
34
35 421 of misregistrations, misdiagnoses or patients who were missed. We have tried to mitigate this
36
37 422 by including a large pre-pandemic control group and including all PCR tests within 1 week
38
39 423 after discharge.

40
41 424 In the external centers, there is a high level of missingness as a result of an incomplete
42
43 425 laboratory panel. In the case of centers 1 and 2, only internal medicine ED presentations were
44
45 426 tested with a laboratory panel containing the 10 tests required for the CoLab-score. The ED
46
47 427 lab panel of other disciplines (e.g. urology, surgery or pediatrics) differed and did not contain
48
49 428 the required tests. Nevertheless, the majority of COVID-19 patients were internal medicine
50
51 429 ED presentations, which is reflected by the few PCR-positive patients excluded. Due to these
52
53
54
55
56
57
58
59
60

1
2
3 430 high levels of missingness, the results of the external centers cannot be used to show that the
4
5 431 CoLab-score generalizes to the entire ED population. Rather, the results show that for the
6
7 432 majority of COVID-19 positive patients presenting at the ED, a routine laboratory panel is
8
9 433 available from which the CoLab-score can be calculated, and that the performance of the
10
11 434 CoLab-score in this population is comparable to the development population. Differences in
12
13 435 the distribution of CoLab variables between centers are shown in **Supplemental Material 3**
14
15 436 **Figure 2.**

16
17
18
19
20 437 The performance of the CoLab-score is affected by the time between the onset of symptoms
21
22 438 and ED presentations. The score increases with the duration of symptoms and gradually
23
24 439 decreases after day 7 (see **Supplemental Material 4 Figure 1** for a plot of the duration of
25
26 440 COVID-19 related symptoms and the CoLab-linear predictor). As a consequence, some
27
28 441 COVID-19 patients with early or late presentation after onset of symptoms can be missed.
29
30 442 Optimal performance of the CoLab-score is achieved when the onset of symptoms is >1 and
31
32 443 <10 days prior to ED presentation. Chemotherapy that causes myeloid suppression, will
33
34 444 decrease neutrophilic, basophilic and eosinophilic counts and thereby “falsely” increasing the
35
36 445 CoLab-score. Conversely, COVID-19 patients with severe anemia could have “falsely”
37
38 446 lowered CoLab-scores. To minimize false negatives, we have therefore advised to report
39
40 447 CoLab-scores only when the concentration of erythrocytes is ≥ 2.9 /pL.

41
42
43
44
45 448 It was chosen to exclude re-presentations after a previous presentation with COVID-19. Since
46
47 449 the median time between initial presentation and re-presentation was 12 days, these patients
48
49 450 were most likely not re-infected patients, but patients who deteriorated after initial
50
51 451 presentation/treatment. Given that the CoLab-score follows the host-immune response, the
52
53 452 score is time sensitive (see **Supplemental Material 4 Figure 1**). Including these patients
54
55 453 would impact the performance of the CoLab-score as patients in a later phase of the disease
56
57 454 show different biomarker profiles. The CoLab-score is aimed towards alerting clinicians to

1
2
3 455 patients presenting with a novel SARS-CoV-2 infection, rather than patients who deteriorate
4
5 456 after treatment for COVID-19. Other re-presentations were not excluded, which results in
6
7 457 some patients appearing multiple times in a dataset. This was not adjusted for in the
8
9
10 458 regression model since the assumption was made that ED presentations are independent
11
12 459 observations. The median time between re-presentations is 38 days, most likely resulting in
13
14 460 variations in laboratory results between presentations, and hence, little to no correlation
15
16
17 461 between presentations. A sensitivity analysis was performed whereby only the first
18
19 462 presentation was included for each patient (**Supplemental Material 4 Table 1**), but no
20
21 463 difference was found in performance in terms of sensitivity, specificity and AUC.
22
23
24 464 The CoLab-score does not serve as a replacement for PCR-testing or LFTs, and can be used to
25
26 465 guide PCR-testing when routine blood tests are available. Important to note is that the CoLab-
27
28 466 score is only valid for ED presentations where routine blood testing is requested, and as a
29
30 467 consequence does not generalize to the ED population who is otherwise well and does not
31
32 468 undergo routine blood testing. Using the CoLab-score in a symptomatic/PCR-tested cohort
33
34 469 also results in different diagnostic performance characteristics, as compared to using the score
35
36 470 on the full ED cohort (see **Supplemental Material 4 Table 1**).
37
38
39
40
41 471 Finally, the CoLab-score could lead to false positives by other viral infections. However, in an
42
43 472 historic patient cohort, the CoLab-score had only limited discriminative ability in separating
44
45 473 influenza-PCR-negative from influenza-PCR-positive patients (see **Supplemental Material 4**
46
47 474 **Figure 2**) implying specificity for SARS-CoV-2. Since the CoLab-score reflects the host-
48
49 475 response to the virus, it is hypothesized that the CoLab-score could also be sensitive to future
50
51 476 SARS-CoV-2 variants. This is supported by the fact that the discriminative ability is sustained
52
53 477 in periods with different dominant variants, although the sensitivity of scores ≥ 3 is somewhat
54
55 478 lower in the third phase (see **Supplemental Material 2 Table 1**). Although vaccination status
56
57 479 is not registered for all presenting patients, in a small subgroup of 12 patients for whom

1
2
3 480 vaccination status was registered, and were COVID-19 positive, 8 of 12 patients had the
4
5 481 highest CoLab-score (= 5) (see **Supplemental Material 2 Figure 2**). Continuous assessment
6
7 482 of the performance of the CoLab-score is required due to the emergence of new variants and
8
9 483 changes in the host's immune response.

10
11
12
13 484 To conclude, the CoLab-score developed and validated in this study, based on 10 routine
14
15 485 laboratory results and age, is available within 1 hour for any patient presenting at the ED
16
17 486 where routine blood testing is requested. The score can be used by clinicians to guide PCR
18
19 487 testing or triage patients and helps to identify COVID-19 in patients presenting at the ED with
20
21 488 abdominal pain, chest pain, shortness of breath, syncope, sepsis or other non-specific
22
23 489 complaints where a routine blood panel is requested. The lowest CoLab-score can be used to
24
25 490 effectively rule-out a possible SARS-CoV-2 infection, the highest score to alert physicians to
26
27 491 a possible infection. The CoLab-score is therefore a valuable tool to rule out COVID-19,
28
29 492 guide PCR testing and is available to any center with access to routine laboratory tests.
30
31
32
33
34
35
36

37 494 **Data Availability Statement**

38
39
40 495 Datasets with source data for Table 1, Figure 2 and Table 4, as well the R-code to fit the
41
42 496 model is available from a Dryad repository [31].
43
44

45 497 46 47 48 498 **Funding statement**

49
50 499 This was an investigator-initiated study and no funding was received for this study.
51
52
53

54 500
55
56
57
58
59
60

501 **Competing interests**

502 A-KB reports no conflict of interest. RD reports no conflict of interest. MM reports no
503 conflict of interest. HA reports no conflict of interest. RvB reports no conflict of interest. WT
504 reports no conflict of interest. SB reports not conflict of interest. ML reports no conflict of
505 interest. RM reports no conflict of interest. MB reports no conflict of interest. JK reports no
506 conflict of interest. MM reports no conflict of interest. JvS reports no conflict of interest. NvR
507 reports no conflict of interest. VS reports no conflict of interest.

508

509 **Author contributorship statement**

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

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

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

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

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

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

- 1
2
3 524 Sophie Bennenbroek: Conceptualization (Supporting), Resources (Supporting), Supervision
4
5 525 (Supporting), Validation (Supporting), Writing-review & editing (Equal).
6
7
8 526 Mathie Leers: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
9
10
11 527 Remy Martens: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
12
13
14 528 Madelon M. Buijs: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
15
16
17 529 Jos Kerremans: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
18
19
20 530 Muriël Messchaert: Resources (Equal), Validation (Equal), Writing-review & editing (Equal).
21
22
23 531 Jeroen van Suijlen: Resources (Supporting), Validation (Supporting), Writing-review &
24
25 532 editing (Equal).
26
27
28 533 Natal A.W. van Riel: Methodology (Supporting), Resources (Supporting), Supervision
29
30 534 (Equal), Writing-review & editing (Equal).
31
32
33 535 Volkher Scharnhorst: Conceptualization (Equal), Funding acquisition (Equal), Project
34
35 536 administration (Lead), Resources (Equal), Supervision (Lead), Writing-review & editing
36
37 537 (Equal).
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

539 **References**

- 540
- 541 1 Coronavirus Disease (COVID-19) Situation Reports.
542 <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/>
543 (accessed 4 Feb 2021).
- 544 2 Guan W, Ni Z, Hu Y, *et al.* Clinical Characteristics of Coronavirus Disease 2019 in
545 China. <https://doi.org/10.1056/NEJMoa2002032> 2020;**382**:1708–20.
546 doi:10.1056/NEJMOA2002032
- 547 3 Vetter P, Vu DL, L’Huillier AG, *et al.* Clinical features of covid-19. *BMJ* 2020;**369**.
548 doi:10.1136/BMJ.M1470
- 549 4 Wynants L, Van Calster B, Collins GS, *et al.* Prediction models for diagnosis and
550 prognosis of covid-19: Systematic review and critical appraisal. *BMJ* 2020;**369**:18.
551 doi:10.1136/bmj.m1328
- 552 5 Albahri AS, Hamid RA, Alwan J k., *et al.* Role of biological Data Mining and Machine
553 Learning Techniques in Detecting and Diagnosing the Novel Coronavirus (COVID-
554 19): A Systematic Review. *J. Med. Syst.* 2020;**44**:122. doi:10.1007/s10916-020-01582-
555 x
- 556 6 Hooli S, King C. Generalizability of Coronavirus Disease 2019 (COVID-19) Clinical
557 Prediction Models. *Clin Infect Dis* 2020;**71**:897–897. doi:10.1093/cid/ciaa417
- 558 7 Prokop M, Everdingen W van, Vellinga T van R, *et al.* CO-RADS: A Categorical CT
559 Assessment Scheme for Patients Suspected of Having COVID-19—
560 Definition and Evaluation. <https://doi.org/10.1148/radiol2020201473> 2020;**296**:E97–
561 104. doi:10.1148/RADIOL.2020201473

- 1
2
3 562 8 Goodacre S, Thomas B, Sutton L, *et al.* Derivation and validation of a clinical severity
4
5 563 score for acutely ill adults with suspected COVID-19: The PRIEST observational
6
7 564 cohort study. *PLoS One* 2021;**16**:e0245840. doi:10.1371/JOURNAL.PONE.0245840
9
- 10 565 9 Plante TB, Blau AM, Berg AN, *et al.* Development and external validation of a
11
12 566 machine learning tool to rule out COVID-19 among adults in the emergency
13
14 567 department using routine blood tests: A large, multicenter, real-world study. *J Med*
15
16 568 *Internet Res* 2020;**22**:e24048. doi:10.2196/24048
17
- 18 569 10 Soltan AAS, Kouchaki S, Zhu T, *et al.* Rapid triage for COVID-19 using routine
19
20 570 clinical data for patients attending hospital: development and prospective validation of
21
22 571 an artificial intelligence screening test. *Lancet Digit Heal* 2021;**3**:e78–87.
23
24 572 doi:10.1016/S2589-7500(20)30274-0
25
- 26 573 11 R Core Team. R: A Language and Environment for Statistical Computing.
27
28 574 2020. <https://www.r-project.org/>
29
- 30 575 12 Zou H. The adaptive lasso and its oracle properties. *J Am Stat Assoc* 2006;**101**:1418–
31
32 576 29. doi:10.1198/016214506000000735
33
- 34 577 13 Tibshirani R. Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B*
35
36 578 1996;**58**:267–88. doi:10.1111/j.2517-6161.1996.tb02080.x
37
- 38 579 14 Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models
39
40 580 via coordinate descent. *J Stat Softw* 2010;**33**:1–22. doi:10.18637/jss.v033.i01
41
- 42 581 15 King G, Zeng L. Logistic Regression in Rare Events Data. *Polit Anal* 2001;**9**:137–63.
43
44 582 doi:10.1093/oxfordjournals.pan.a004868
45
- 46 583 16 Efron B. Estimating the error rate of a prediction rule: Improvement on cross-
47
48 584 validation. *J Am Stat Assoc* 1983;**78**:316–31. doi:10.1080/01621459.1983.10477973
49
50
51
52
53
54
55
56
57
58
59
60

- 1
2
3 585 17 Robin X, Turck N, Hainard A, *et al.* pROC: An open-source package for R and S+ to
4
5 586 analyze and compare ROC curves. *BMC Bioinformatics* 2011;**12**:77. doi:10.1186/1471-
6
7 587 2105-12-77
8
9
10 588 18 Harrell Jr FE. rms: Regression Modeling Strategies. 2021. [https://cran.r-](https://cran.r-project.org/package=rms)
11
12 589 [project.org/package=rms](https://cran.r-project.org/package=rms)
13
14
15 590 19 Ricós C, Alvarez V, Cava F, *et al.* Current databases on biological variation: Pros, cons
16
17 591 and progress. *Scand. J. Clin. Lab. Invest.* 1999;**59**:491–500.
18
19 592 doi:10.1080/00365519950185229
20
21
22
23 593 20 Brinati D, Campagner A, Ferrari D, *et al.* Detection of COVID-19 Infection from
24
25 594 Routine Blood Exams with Machine Learning: A Feasibility Study. *J Med Syst*
26
27 595 2020;**44**:1–12. doi:10.1007/s10916-020-01597-4
28
29
30
31 596 21 Joshi RP, Pejaver V, Hammarlund NE, *et al.* A predictive tool for identification of
32
33 597 SARS-CoV-2 PCR-negative emergency department patients using routine test results. *J*
34
35 598 *Clin Virol* 2020;**129**:104502. doi:10.1016/j.jcv.2020.104502
36
37
38
39 599 22 Qin L, Yang Y, Cao Q, *et al.* A predictive model and scoring system combining
40
41 600 clinical and CT characteristics for the diagnosis of COVID-19. *Eur Radiol*
42
43 601 2020;**30**:6797–807. doi:10.1007/s00330-020-07022-1
44
45
46 602 23 Kurstjens S, van der Horst A, Herpers R, *et al.* Rapid identification of SARS-CoV-2-
47
48 603 infected patients at the emergency department using routine testing. *Clin Chem Lab*
49
50 604 *Med* 2020;**58**:1587–93. doi:10.1515/cclm-2020-0593
51
52
53
54 605 24 Fink DL, Khan PY, Goldman N, *et al.* Development and internal validation of a
55
56 606 diagnostic prediction model for COVID-19 at time of admission to hospital. *QJM An*
57
58 607 *Int J Med* Published Online First: 9 November 2020. doi:10.1093/qjmed/hcaa305
59
60

- 1
2
3 608 25 Giamello JD, Paglietta G, Cavalot G, *et al.* A simple tool to help ruling-out Covid-19
4
5 609 in the emergency department: derivation and validation of the LDH-CRP-Lymphocyte
6
7 610 (LCL) score. *Emerg Care J* 2020;**16**. doi:10.4081/ecj.2020.9336
8
9
10 611 26 Tordjman M, Mekki A, Mali RD, *et al.* Pre-test probability for SARS-Cov-2-related
11
12 612 infection score: The PARIS score. *PLoS One* 2020;**15**:e0243342.
13
14 613 doi:10.1371/journal.pone.0243342
15
16
17
18 614 27 Peto T, Affron D, Afrough B, *et al.* COVID-19: Rapid antigen detection for SARS-
19
20 615 CoV-2 by lateral flow assay: A national systematic evaluation of sensitivity and
21
22 616 specificity for mass-testing. *EClinicalMedicine* 2021;**36**:100924.
23
24 617 doi:10.1016/J.ECLINM.2021.100924
25
26
27
28 618 28 García-Fiñana M, Hughes DM, Cheyne CP, *et al.* Performance of the Innova SARS-
29
30 619 CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot:
31
32 620 population based cohort study. *BMJ* 2021;**374**:1637. doi:10.1136/BMJ.N1637
33
34
35
36 621 29 Ford JS, Parikh A, Sandhu R, *et al.* Testing Asymptomatic Emergency Department
37
38 622 Patients for Coronavirus Disease 2019 (COVID-19) in a Low-prevalence Region.
39
40 623 *Acad. Emerg. Med.* 2020;**27**:771–4. doi:10.1111/acem.14044
41
42
43 624 30 Ravani P, Saxinger L, Chandran U, *et al.* COVID-19 screening of asymptomatic
44
45 625 patients admitted through emergency departments in Alberta: a prospective quality-
46
47 626 improvement study. *C Open* 2020;**8**:E887–94. doi:10.9778/cmajo.20200191
48
49
50
51 627 [dataset] 31 Boer A-K, Deneer R. Source data for: Development and validation of an early
52
53 628 warning score to identify COVID-19 in the emergency department based on routine
54
55 629 laboratory tests: a multicenter case-control study. *Dryad Digit Repos* Published Online
56
57 630 First: 2021. doi:https://doi.org/10.5061/dryad.5hqbzkh6x
58
59
60

1
2
3 632 **Figure legends**

4
5
6 633
7
8
9 634 **Figure 1: Inclusion flow of patients in the development (A) and temporal validation (B)**
10
11 635 **dataset.**

12
13
14 636 *All patient admissions with routine venous blood sampling at the emergency department (ED)*
15
16 637 *were included. For the development dataset, completeness of the lab panel was assessed for*
17
18 638 *all 28 laboratory tests, for the temporal validation dataset this was only necessary for 10*
19
20 639 *laboratory tests. The major causes of missingness are described in the text. In the*
21
22 640 *development dataset, presentations with extreme values (>10 SD) were excluded. The same*
23
24 641 *limits were applied to the temporal validation dataset (see Table 2 for limits).*

25
26
27
28 642
29
30
31 643 **Figure 2: Probability density plot of the CoLab-linear predictor.**

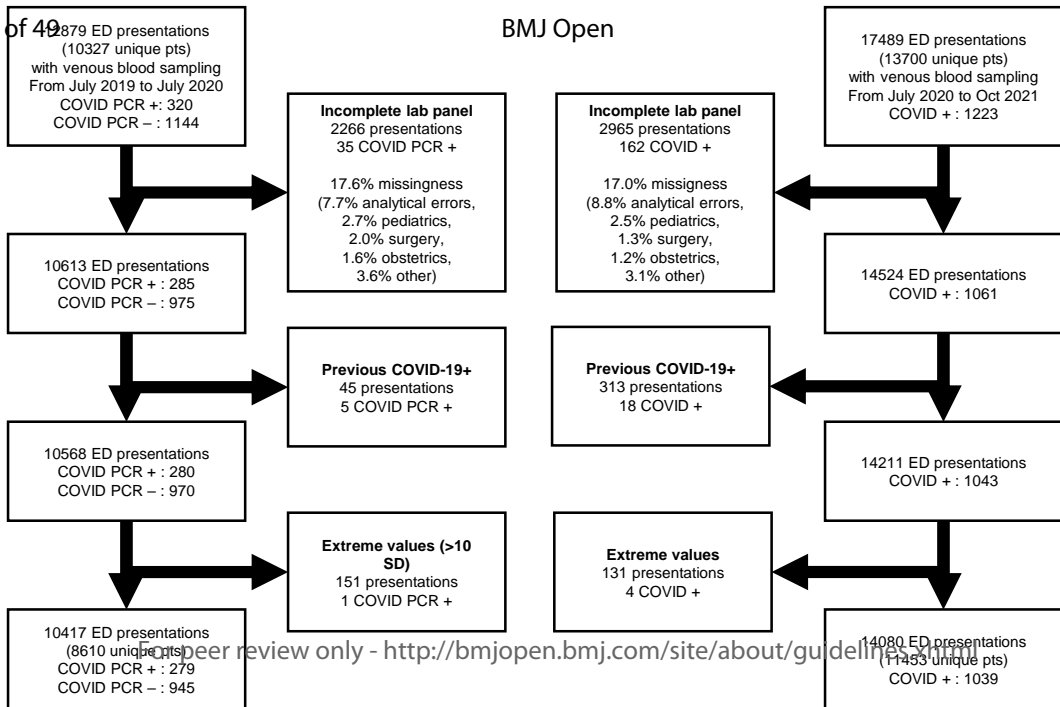
32
33
34 644 *The probability density plots for COVID (dark grey) and non-COVID patients (light grey) are*
35
36 645 *plotted against the linear predictor (see table 2). The CoLab-score cut-offs (-5.83 , -4.02 , $-$*
37
38 646 *3.29 , -2.34 and -1.64) are depicted with vertical dashed lines. The white-boxed numbers*
39
40 647 *(between the cut-offs) represent the corresponding CoLab-score. Note that while the area*
41
42 648 *under both curves is identical (since these are probability density functions), in absolute*
43
44 649 *numbers the “negative or untested”-group is about 36 times larger than the PCR positive*
45
46 650 *group.*

47
48
49
50
51 651
52
53
54 652 **Figure 3: Inclusion flow of ED patients in three external centers.**

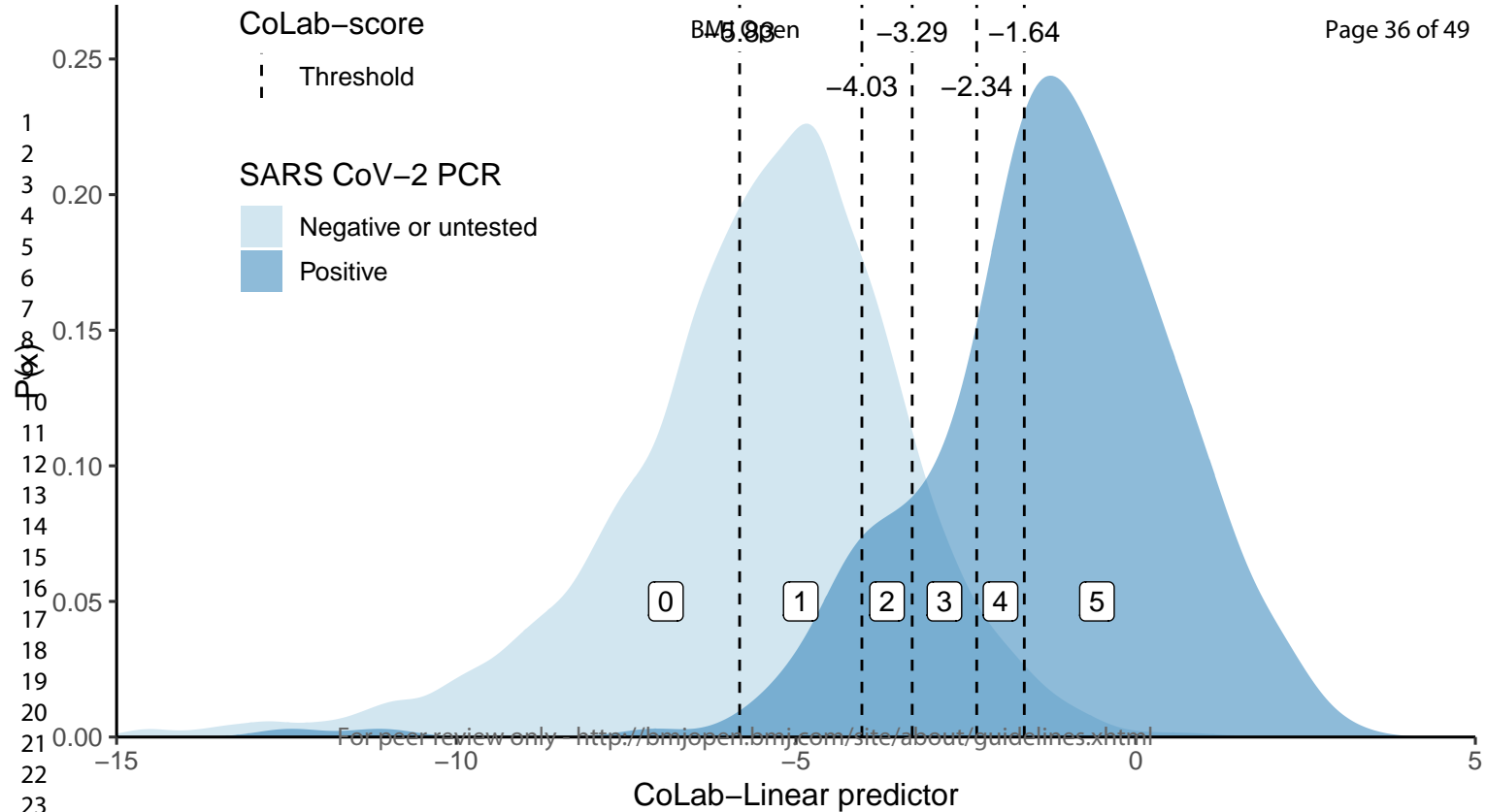
55
56
57 653 *All emergency department (ED) presentations with routine venous blood sampling were*
58
59 654 *included. Missingness of lab panels was assessed for the 11 variables in the CoLab-score (see*
60

1
2
3 655 *Table 2). Re-presentations after a positive PCR result or clinical COVID-19 registration were*
4
5 656 *excluded as “previous COVID-19+”. Presentations with any laboratory result above the*
6
7 657 *limits of the CoLab-score (see Table 2) were excluded.*
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

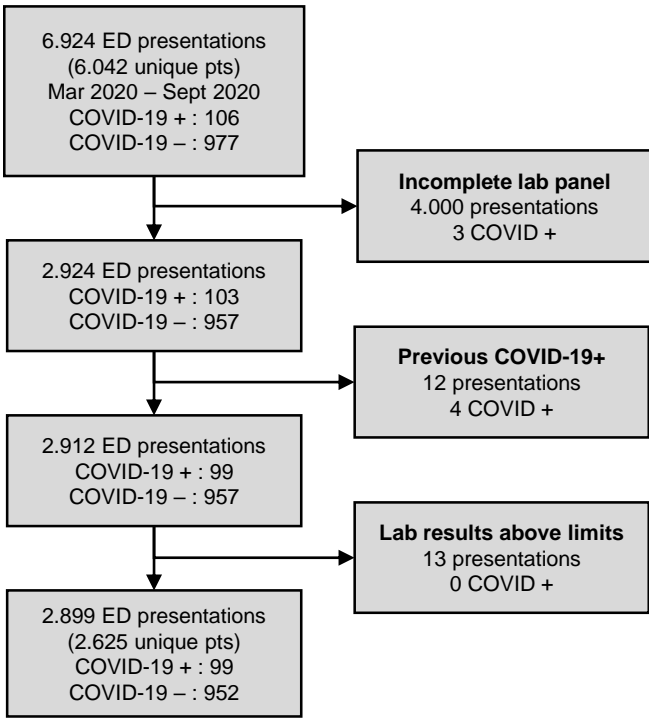
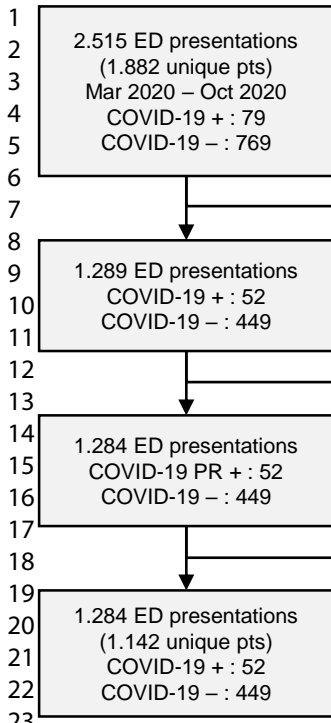


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20

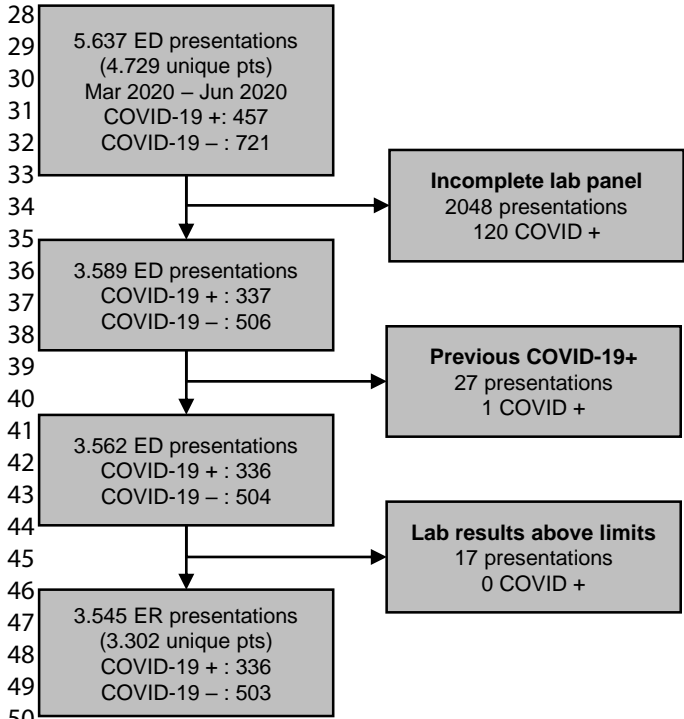


Center 1

Center 2



Center 3



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 (\hat{w}) was calculated by $\hat{w} = 1/|\hat{\beta}_{ridge}|^2$. 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: $lp_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in}$. Where n is the number of variables in the final model, x_{in} are the observed predictor variables for subject i and β_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:

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

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:

$$\beta_0 + \beta_{adj} + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639$$

$$\beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in} \geq -2.639 - \beta_{adj}$$

$$lp_i(\tau) \geq \ln\left(\frac{1-\tau}{\tau}\right) - 6.233$$

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

$$lp_i(\tau = 0.4) \geq -5.83 \text{ (CoLab-score} = 1)$$

$$lp_i(\tau = 0.1) \geq -4.03 \text{ (CoLab-score} = 2)$$

$$lp_i(\tau = 0.05) \geq -3.29 \text{ (CoLab-score} = 3)$$

$$lp_i(\tau = 0.02) \geq -2.34 \text{ (CoLab-score} = 4)$$

$$lp_i(\tau = 0.01) \geq -1.64 \text{ (CoLab-score} = 5)$$

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

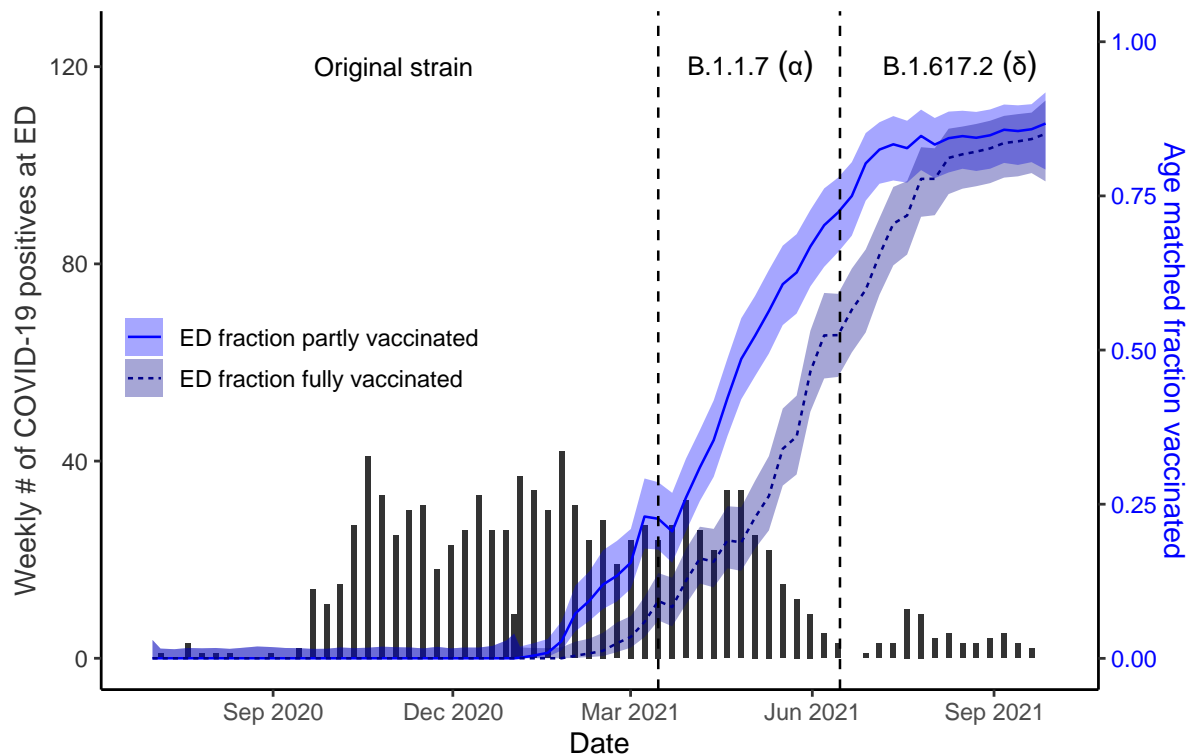


Figure 1: Temporal validation period split into three phases characterized by weekly number of new COVID-19 cases at the emergency department (ED) and estimated fraction of ED patients vaccinated.

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).

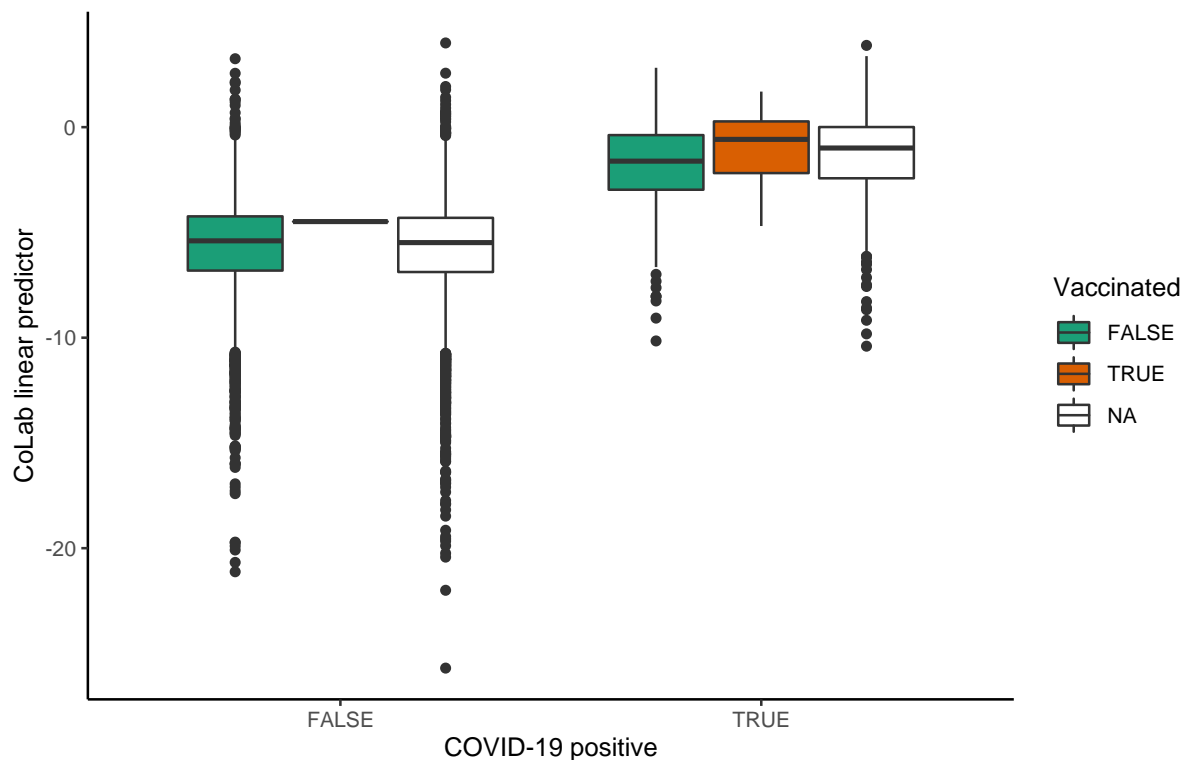


Figure 2: Boxplots of CoLab linear predictor versus COVID-19 positive, split by registered vaccination status.

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

1
2
3 *vaccine roll-out but where no status is registered are labeled NA (N = 8212). Of the 13*
4
5 *presentations who were registered as vaccinated, 12 were COVID-19 positive and 1 negative.*
6
7 *Note that vaccination status is only registered if a patient is SARS-CoV-2 PCR positive or*
8
9 *considered positive until proven otherwise, therefore there is only one COVID-19 negative*
10
11 *patient with a registered vaccination status.*
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For peer review only

Supplemental material 3

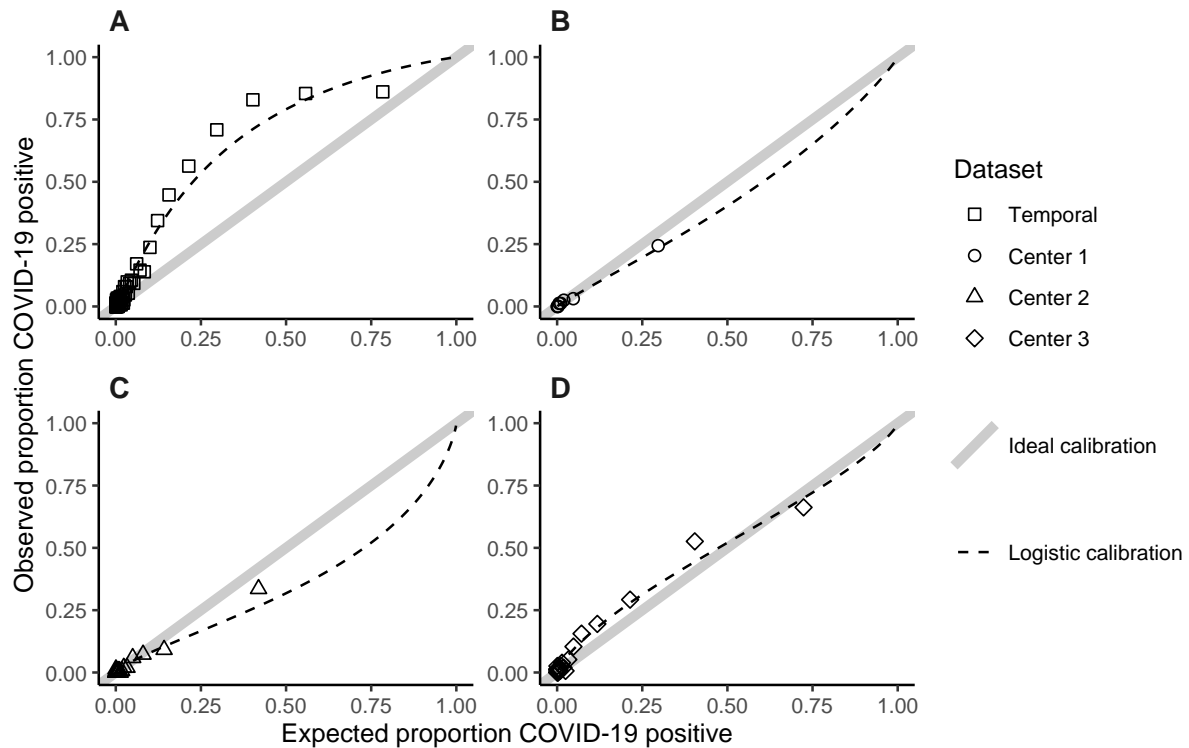


Figure 1: CoLab-score calibration plots of the temporal validation (A), external validation center 1 (B), external validation center 2 (C) and external validation center 3 (D).

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.

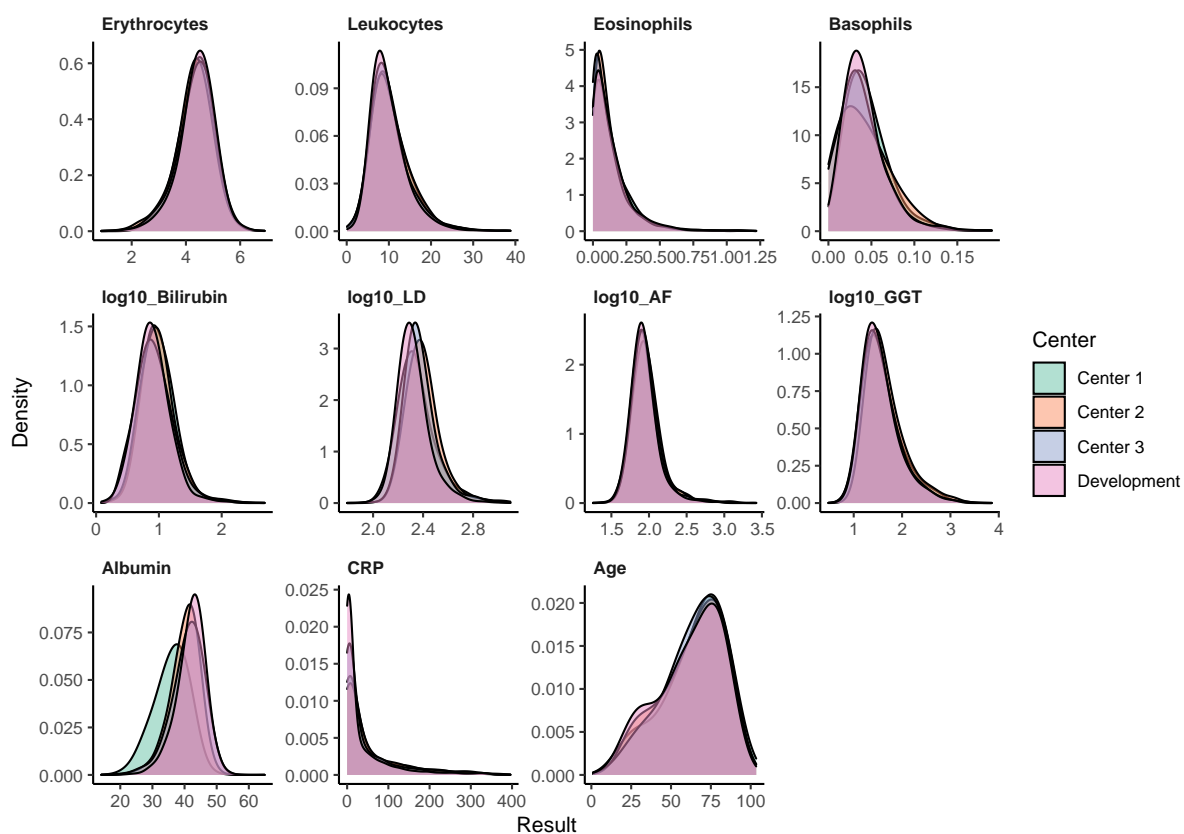


Figure 2: Probability density plots of laboratory parameters.

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

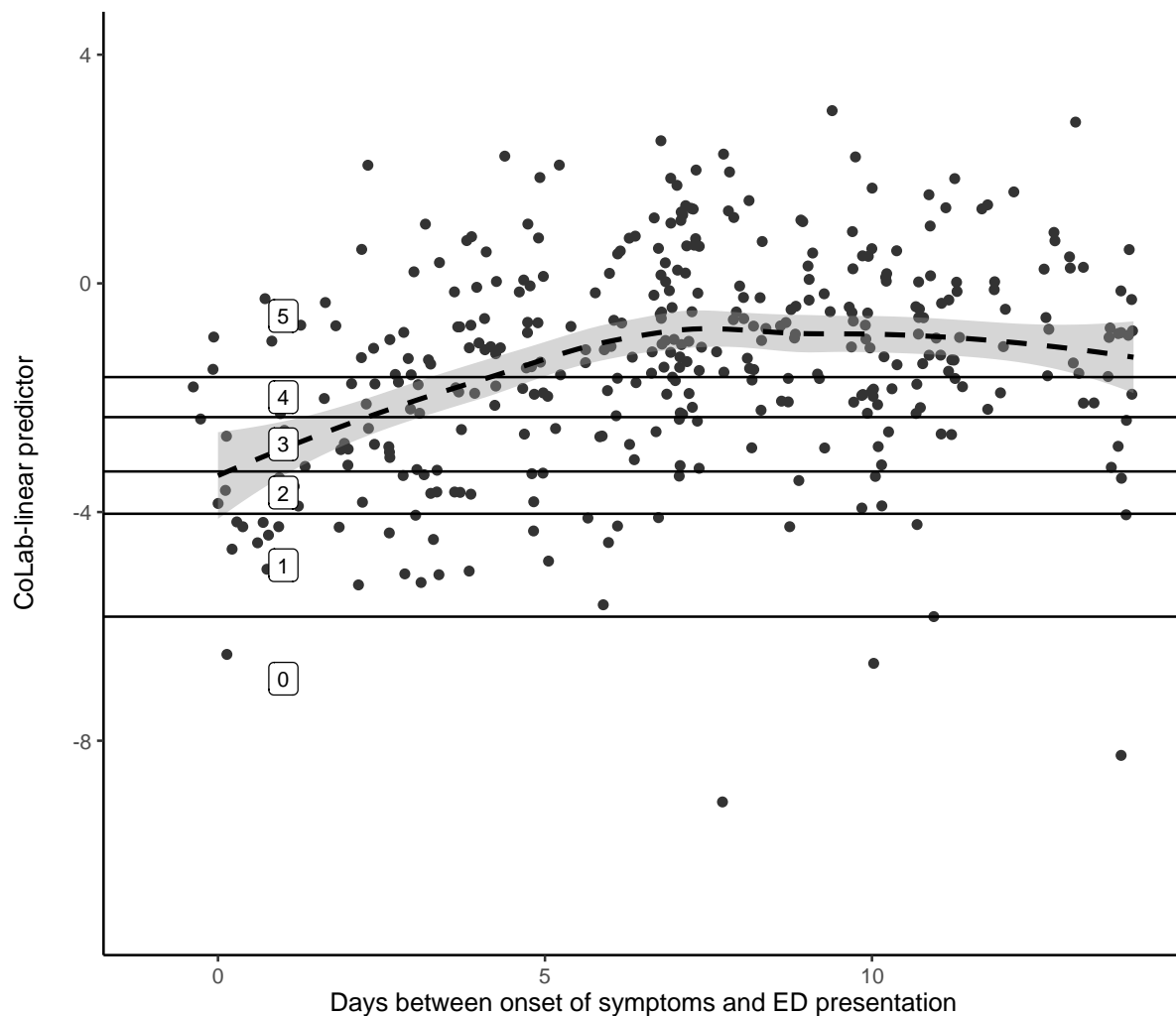


Figure 1: Association between the CoLab-linear predictor and the duration of COVID-19-related symptoms.

For all PCR-positive ED presentations in the development and temporal validation dataset, the CoLab-linear predictor 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.

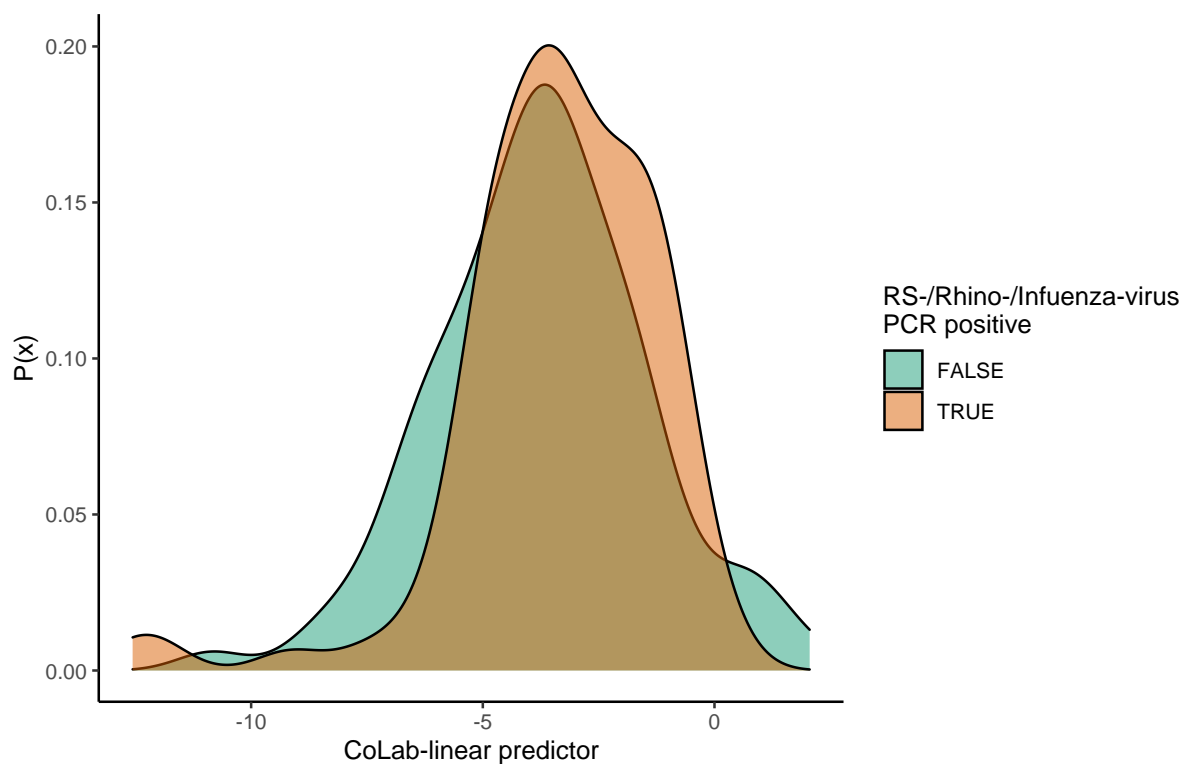


Figure 2: Probability density plot of CoLab-score for RS-, Rhino- and Influenza-virus PCR tested ED patients.

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: 0.4896-0.6563).

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)

1
2
3 **Table 1: Sensitivity analysis of the CoLab-score in the temporal validation dataset using**
4 **different inclusion criteria.**
5

6
7 *Sensitivities, specificities, positive predictive values (PPV), negative predictive values (NPV),*
8 *true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN) are*
9 *shown for fixed cut-offs (CoLab-score 0 till ≤ 4) with bootstrapped 95% confidence intervals*
10 *in parentheses. The temporal validation dataset is used to compare the performance of the*
11 *CoLab-score with inclusion criteria that differ from the development dataset. The first line*
12 *shows the performance of the temporal validation dataset with the original inclusion criteria*
13 *as specified in Figure 1B. The second line shows the performance of the CoLab-score when*
14 *all re-presentations are excluded (i.e. no repeated presentations). The third line shows the*
15 *performance of the CoLab-score in the subgroup of patients that underwent PCR-testing.*
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60



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 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.