Letter to the Editor

A model to identify individuals with a high probability of a SARS-CoV-2 infection

Dear Editor,

Persons with a high probability of a SARS-CoV-2 infection should be tested with priority if testing resources are limited. Recently, Clark et al. reported that the FebriDx point-of-care test which is based on the detection of the antiviral host response protein MxA had high accuracy to identify COVID-19 during the first wave.[1] They suggested that in hospitals, MxA positive patients should have a PCR test for confirmation or exclusion of COVID-19 with priority. Despite some limitations mentioned by the authors, FebriDx may be a valuable tool for triage in hospitals. In this study, we present a simple prediction model as another tool to identify persons with a high probability of a SARS-CoV-2 infection.

The City of Essen (Germany) established a Task Force during the SARS-CoV-2 pandemic. Data of all citizens who contacted the Task Force and either received a PCR test or were ordered a quarantine was recorded in a database. A self-administered questionnaire was sent to 4890 citizens who were entered in the database between February 27th and April 19th. We received 2234 questionnaires, 1937 (86.7%) of which could be assigned to a PCR test result. For 297 questionnaires, assignment was not possible because participants did not give their addresses and birth dates. For 1808 participants, there were complete data to develop and validate the model. The study was approved by the Ethical Committee of the Medical Faculty of the University Hospital Essen.

The self-administered questionnaire included questions on sociodemographic and anthropometric data, on reasons why the participants got tested for SARS-CoV-2, on contacts to persons with a confirmed SARS-CoV-2 infection, on the presence of symptoms typical for COVID-19 at the time of the test, on chronic comorbidities, on the course of their COVID-19 infection including hospitalization and referral to an intensive care unit. For temporal validation, we split the questionnaires according to the completion dates of the questionnaires.[2] We used 1089 (60%) questionnaires completed until April 29th to develop the model, and the remaining 719 (40%) questionnaires to validate the model. For model development, we used twelve candidate variables (sex, age, return from abroad, close contact (< 1.5 m, ≥ 15 min) to a person with a confirmed SARS-CoV-2 infection, the presence of any COVID-19 type symptoms, fever, cough, exhaustion, taste or smell disorder, current smoking (≥ 10 cigarettes per day), general health condition at the time of the test, and number of comorbidities). We followed established methods for model development.[3,4] For each candidate predictor, univariable logistic regression models were fitted, and variables with $p < 0.2$ were used for subsequent selection from a multivariable logistic regression model. Backward elimination was used to build the final model, and the significance level to enter

and retain variables was set at 0.1. To assess the goodness-of-fit of the final model, we used Hosmer–Lemeshow tests and calibration plots. To assess the discrimination of the final model, we estimated the area under the receiver operating curve (*c*-value), and the Tjur coefficient which is the difference between the mean predicted probability in the test positives and the mean predicted probability in the test negatives.

The proportion of participants reporting a taste or smell disorder was considerably larger in the group with a positive test (56% versus 11%), and there were fewer current smokers (≥ 10 cigarettes per day) among the positive tested than among the negative tested (6% versus 19%) (Table 1). From the final model, the probability of a positive test result can be estimated as follows:

p (positive test) $= 1 \,/\, (1 + \exp(-z))$,
with z=
– 2.7630
+ 0.4410 x sex (male=1, female=0)
– 1.3801 x smoking (current smoking of ≥ 10 cigarettes=1, else=0)
+ 2.0194 x taste or smell disorder (yes=1, no=0)
+ 0.7037 x close contact to infected person (yes=1, no=0)
+ 0.6915 x return from abroad (yes=1, no=0)
+ 0.4424 x exhaustion (yes=1, no=0).

Taste or smell disorder and non-smoking were the strongest predictors in the final model (Table 2). In the development data set, the *c*-statistic was 0.803, and the Tjur coefficient was 0.234. The model performed equally well in the validation data set ($c = 0.821$, Tjur coefficient=0.217). The calibration plot (not shown) and the Hosmer–Lemeshow test indicated good model fit in the development data set. In the validation data set, the probability of a positive test was lower than in the development data set (11.7% versus 18.5%), and, therefore, the model was poorly calibrated in this data set (Hosmer–Lemeshow $p < 0.05$).

Most other models to predict a COVID-19 infection are based either on medical imaging or include invasive parameters.[5] Few published models include only non-invasive parameters.[6–9] The six parameters of the present model can easily be assessed in telephone interviews, and excel sheets are suitable to estimate the probability of being test positive quickly. Poor calibration is not a problem as long as the model is only used to rank-order people by probability of being infected. To improve agreement between observed and predicted probabilities of being test positive, recalibration techniques could be used.[10]

A limitation of our study is that the prediction model was developed in spring during the first wave, and that we do not know how its diagnostic accuracy will be in another season and in another wave of the pandemic after potential mutation of the virus. A strength of our study is that the model includes only non-invasive

Table 1

Characteristics of all participants with a clearly assigned test result.

		Positive test result	Negative test result
N		296	1641
Sex (male)[a]		137 (47.2%)	622 (38.7%)
Age (years)[b]		53.2 ± 19.0	53.9 ± 21.6
COVID-19 type symptoms		198 (66.9%)	849 (51.7%)
Belonging to risk group for severe course of disease		55 (18.6%)	542 (33.0%)
Return from risk region		60 (20.3%)	134 (8.2%)
Return from abroad		49 (16.6%)	105 (6.4%)
Close contact with confirmed COVID-19 infected person		111 (37.5%)	352 (21.5%)
General health condition[c]	Normal	65 (22.6%)	489 (34.1%)
	Slightly reduced	88 (30.6%)	353 (24.6%)
	Somewhat reduced	87 (30.2%)	329 (23.0%)
	Strongly reduced	48 (16.7%)	261 (18.2%)
Taste disorder		152 (51.4%)	149 (9.1%)
Smell disorder		129 (43.6%)	107 (6.5%)
Taste or smell disorder		166 (56.1%)	175 (10.7%)
Taste and smell disorder		115 (38.9%)	81 (4.9%)
Exhaustion		201 (67.9%)	668 (40.7%)
Fever		133 (44.9%)	459 (28.0%)
Cough		165 (55.7%)	678 (41.3%)
Exhaustion and fever and cough		62 (21.0%)	207 (12.6%)
Number of comorbidities		1 (0; 2)	1 (0; 3)
No comorbidities		115 (38.9%)	550 (33.5%)
Current smoking (≥ 10 cigarettes per day)[d]		17 (5.9%)	296 (19.0%)

Data are proportions (%), means (± standard deviation), or median (first quartile; third quartile).

[a] 6 missings in test positives, 34 in test negatives.

[b] 6 missings in test positives, 63 in test negatives.

[c] 8 missings in test positives, 209 in test negatives.

[d] 8 missings in test positives, 84 missings in test negatives.

Table 2

Multivariable prediction of positive test results for the development, validation and combined data set (OR, 95% CI).

	Development $N = 1089$	Validation $N = 719$	Combined $N = 1808$
Number of test positives (*N* (%))	202 (18.5%)	84 (11.7%)	286 (15.8%)
Predictors	OR (95% CI)	OR (95% CI)	OR (95% CI)
Sex (male vs female)	1.55 (1.09–2.21)	1.17 (0.68–2.00)	1.44 (1.07–1.93)
≥ 10 cigarettes per day (yes vs no)	0.25 (0.13–0.50)	0.35 (0.14–0.84)	0.28 (0.16–0.48)
Taste or smell disorder (yes vs no)	7.53 (5.14–11.04)	11.86 (6.82–20.65)	8.79 (6.42–12.04)
Close contact with confirmed COVID-19 infected person (yes vs no)	2.02 (1.41–2.91)	1.84 (1.00–3.40)	2.07 (1.52–2.82)
Return from abroad (yes vs no)	2.00 (1.22–3.27)	2.40 (0.89–6.50)	2.19 (1.41–3.41)
Exhaustion (yes vs no)	1.56 (1.06–2.30)	1.53 (0.88–2.67)	1.56 (1.13–2.14)
Model performance			
c-Statistic	0.803 (0.768–0.838)	0.821 (0.770–0.873)	0.814 (0.786–0.843)
Tjur coefficient[a]	0.234	0.217	0.232
Mean predicted probability in test positives	0.376	0.329	0.362
Mean predicted probability in test negatives	0.142	0.112	0.130
Mean observed versus predicted	18.5% versus 18.5%	11.7% versus 13.7%	15.8% versus 16.6%
Hosmer Lemeshow test (*p* value)	0.31	0.02	0.002

OR: odds ratio; CI: confidence interval.

[a] Mean predicted probability in test positives minus mean predicted probability in test negatives.

parameters so that probabilities of being test positive can quickly be estimated.

In conclusion, we developed a model to identify persons with a high probability of an infection with SARS-CoV-2 which can be used for primary triage in case test capacities are too limited to test all individuals who need or want a SARS-CoV-2 test.

Declaration of Competing Interest

None.

References

1. Clark TW, Brendish NJ, Poole S, Naidu VV, Mansbridge C, Norton N, et al. Diagnostic accuracy of the FebriDx host response point-of-care test in patients hospitalised with suspected COVID-19. *J Infect* 2020;**81**:607–13.
2. Steyerberg EW. *Clinical Prediction Models: A Practical Approach to Development, Validation and Updating*. New York: Springer; 2009.
3. Moons KGM, Kengne AP, Woodward M, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart* 2012;**98**:683–90.
4. Moons KGM, Kengne AP, Grobbee DE, Royston P, Vergouwe Y, Altman DG, et al. Risk prediction models II: external validation, model updating, and impact assessment. *Heart* 2012;**98**:691–8.
5. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of COVID-19: systematic review and critical appraisal. *BMJ* 2020;**369**:m1328.
6. Menni C, Valdes AM, Freidin MB, Ganesh S, El-Sayed Moustafa JS, Visconti A, et al. Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection. 2020. medRxiv preprint doi: https://doi.org/10.1101/2020.04.05.20048421.

7. Menni C, Valdes AM, Freidin MB. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;**26**:1037–40.
8. Sun Y, Koh V, Marimuthu K, Ng OT, Young B, Vasoo S, et al. Epidemiological and clinical predictors of COVID-19. *Clin Infect Dis* 2020;**71**:786–92.
9. Diaz-Quijano FA, JMNd Silva, Ganem F, Oliveira S, Vesga-Varela AL, Croda J. A model to predict SARS-CoV-2 infection based on the first three-month surveillance data in Brazil. *Trop Med Int Health* 2020 Aug 13:10.1111/tmi.13476Online ahead of print. doi:10.1111/tmi.13476.
10. Janssen KJ, Moons KG, Kalkman CJ, Grobbee DE, Vergouwe Y. Updating methods improved the performance of a clinical prediction model in new patients. *J Clin Epidemiol* 2008;**61**:76–86 .

B. Kowall*, M. Nonnemacher
Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Hufelandstraße 55, 45147 Essen, Germany

B. Brune
Medical Emergency Service of the City of Essen, Germany
Department for Trauma, Hand and Reconstructive Surgery, University Hospital of Essen, Germany

M. Brinkmann
Center for Clinical Trials, University Hospital Essen, Germany

M. Dudda
Medical Emergency Service of the City of Essen, Germany
Department for Trauma, Hand and Reconstructive Surgery, University Hospital of Essen, Germany

J. Böttcher
Public Health Office of the City of Essen, Germany

B. Schmidt, F. Standl, S. Stolpe
Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Hufelandstraße 55, 45147 Essen, Germany

U. Dittmer
Institute for Virology, University Hospital Essen, Germany

KH Jöckel
Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Hufelandstraße 55, 45147 Essen, Germany

A. Stang
Institute of Medical Informatics, Biometry and Epidemiology, University Hospital Essen, Hufelandstraße 55, 45147 Essen, Germany
School of Public Health, Department of Epidemiology, Boston University, Boston, United States

*Corresponding author.
E-mail address:* bernd.kowall@uk-essen.de (B. Kowall)