

# 1           **Validation of a clinical and genetic model for predicting severe COVID-19**

2

3 Gillian S. Dite<sup>1\*</sup> (ORCID: 0000-0002-2448-2548)

4 Nicholas M. Murphy<sup>1</sup> (ORCID: 0000-0001-7881-1952)

5 Erika Spaeth<sup>2</sup> (ORCID: 0000-0002-5397-3493)

6 Richard Allman<sup>1</sup> (ORCID: 0000-0002-8635-5237)

7 Lifelines Corona Research Initiative<sup>3</sup>

8

9 1. Genetic Technologies Limited, Fitzroy, Victoria, Australia

10 2. Phenogen Sciences Inc, Charlotte, North Carolina, United States of America

11 3. Members of the Lifelines Corona Research Initiative are as follows:

12           Marieke H. Boezen,<sup>a</sup> Jochen O. Mierau,<sup>b,c</sup> H. Lude Franke,<sup>d</sup> Jackie Dekens,<sup>d,e</sup>

13           Patrick Deelen,<sup>d</sup> Pauline Lanting,<sup>d</sup> Judith M. Vonk,<sup>a</sup> Ilja Nolte,<sup>a</sup> Anil P.S. Ori,<sup>d,f</sup>

14           Annie Claringbould,<sup>d</sup> Floranne Boulogne,<sup>d</sup> Marjolein X.L. Dijkema,<sup>d</sup> Henry H.

15           Wiersma,<sup>d</sup> Robert Warmerdam,<sup>d</sup> Soesma A. Jankipersadsing,<sup>d</sup> Irene van

16           Blokland<sup>d,g</sup>

17           a. Department of Epidemiology, University of Groningen, University Medical  
18           Center Groningen, Groningen, The Netherlands

19           b. Faculty of Economics and Business, University of Groningen, Groningen, The  
20           Netherlands

21           c. Aletta Jacobs School of Public Health, Groningen, The Netherlands

- 22 d. Department of Genetics, University of Groningen, University Medical Center  
23 Groningen, Groningen, The Netherlands
- 24 e. Center of Development and Innovation, University of Groningen, University  
25 Medical Center Groningen, Groningen, The Netherlands
- 26 f. Department of Psychiatry, University of Groningen, University Medical  
27 Center Groningen, Groningen, The Netherlands
- 28 g. Department of Cardiology, University of Groningen, University Medical  
29 Center Groningen, Groningen, The Netherlands

30

31 \* Corresponding author. Genetic Technologies Limited, 60–66 Hanover St, Fitzroy VIC  
32 3065, Australia.

33 Email: [gillian.dite@gtglabs.com](mailto:gillian.dite@gtglabs.com)

34

35 Word count: 2495

36

37 Key words: severe COVID-19, risk prediction, validation, risk factors, single-nucleotide  
38 polymorphism

## 39 **Summary**

40 Using nested case–control data from the Lifelines COVID-19 cohort, we undertook a  
41 validation study of a clinical and genetic model to predict the risk of severe COVID-19 in  
42 people with confirmed COVID-19 and in people with confirmed or self-reported COVID-19.  
43 The model performed well in terms of discrimination of cases and controls for all ages (area  
44 under the receiver operating characteristic curve [AUC] = 0.680 for confirmed COVID-19  
45 and AUC = 0.689 for confirmed and self-reported COVID-19) and in the age group in which  
46 the model was developed (50 years and older; AUC = 0.658 for confirmed COVID-19 and  
47 AUC= 0.651 for confirmed and self-reported COVID-19). There was no evidence of over- or  
48 under-dispersion of risk scores but there was evidence of overall over-estimation of risk in all  
49 analyses (all  $P < 0.0001$ ). In the light of large numbers of people worldwide remaining  
50 unvaccinated and continuing uncertainty regarding vaccine efficacy over time and against  
51 variants of concern, identification of people at high risk of severe COVID-19 may encourage  
52 the uptake of vaccinations (including boosters) and the use of non-pharmaceutical inventions.

## 53 Text

54 Severe coronavirus disease 2019 (COVID-19) disproportionately affects older adults,  
55 but can occur in people of all ages, especially those with comorbidities [1]. An abundance of  
56 research has identified clinical and genetic risk factors that are associated with developing  
57 severe disease if infected with severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-  
58 2) [1, 2]. In clinical practice, information on these risk factors can be useful when combined  
59 in a risk prediction model that provides a single estimate of absolute risk that enables health  
60 care providers to effectively communicate with their patients about risk.

61 We previously described a clinical and genetic model for predicting severe COVID-  
62 19 that was developed and validated using data from the UK Biobank [3]. We now report the  
63 results of a validation of the model using a case–control analysis of an external dataset from  
64 the Netherlands [4]. While the risk prediction model was developed in people aged 50 years  
65 and older, here we assess its performance in people aged 24 years and older.

66 We used data from participants in the Lifelines COVID-19 cohort [4] who were  
67 recruited from the Lifelines and Lifelines NEXT cohorts [5, 6]. Lifelines is a multi-  
68 disciplinary prospective population-based cohort study examining, in a unique three-  
69 generation design, the health and health-related behaviours of 167,729 people living in the  
70 north of the Netherlands. Lifelines employs a broad range of investigative procedures in  
71 assessing the biomedical, socio-demographic, behavioural, physical and psychological factors  
72 that contribute to the health and disease of the general population, with a special focus on  
73 multi-morbidity and complex genetics.

74 During 2020, questionnaire links were emailed to Lifelines and Lifelines NEXT  
75 participants, weekly from the end of March to mid-May and then every two weeks until July,  
76 after which the questionnaires were sent monthly through to April 2021 [4]. Lifelines  
77 COVID-19 cohort participants were aged 24 years or over and had completed at least one of

78 the regular online COVID-19 questionnaires via an emailed link during the first eight weeks  
79 of data collection [4].

80 The questionnaire response dates corresponded to the period from around one month  
81 after the beginning of the first wave of the COVID-19 pandemic in the Netherlands through  
82 to the peak of the fourth wave in May 2021. During this time, the original SARS-CoV-2 virus  
83 accounted for over 95% of infections in the Netherlands until early January 2021, after which  
84 the alpha variant became more prevalent and had accounted for over 95% of infections by the  
85 end of March 2021 [7]. The presence of the delta variant was negligible during the period of  
86 data collection for this study.

87 COVID-19 vaccinations became available in the Netherlands in mid-January 2021  
88 and were initially offered to high-risk groups and then progressively to other groups (such as  
89 care workers) and younger age groups until all adults became eligible in mid-June 2021 [8].  
90 From questionnaire 18 (March 2021) onwards, participants were asked about their  
91 vaccination status and we excluded questionnaires where (and after which) a participant  
92 reported having had one or two doses of a vaccine.

93 At the beginning of data collection, when testing for SARS-CoV-2 infection was not  
94 widely available in the Netherlands, Lifelines COVID-19 questionnaires 1–4 asked  
95 participants whether a doctor had told them they had COVID-19 [4]. From questionnaire 5  
96 (early May 2020) onwards, the questionnaires also asked about positive test results. From  
97 these questions we identified a group of participants with confirmed COVID-19. In addition,  
98 the questionnaires asked participants to self-report having had COVID-19. We used this  
99 question with the previous questions to identify a broader group of participants who had  
100 either confirmed or self-reported COVID-19.

101 Given the limited availability of testing early in the data collection period, the  
102 confirmed COVID-19 group is likely to miss some participants who had COVID-19.

103 Conversely, the broader group including participants with self-reported COVID-19 is likely  
104 to have some false positives. The true number of participants who had COVID-19 will be  
105 somewhere between the two. Therefore, we conducted two sets of analyses: (i) using  
106 participants with confirmed COVID-19 and (ii) using participants with confirmed and self-  
107 reported COVID-19.

108 As we did previously, we used hospitalization as a proxy for severe COVID-19 [3].  
109 The Lifelines COVID-19 questionnaires specifically asked participants whether they had  
110 been hospitalized for COVID-19. The questionnaires also asked about being given  
111 supplemental oxygen, admission to an intensive care unit and being placed on a ventilator,  
112 but there were too few positive responses to these questions to allow separate analysis.

113 The risk factors included in the calculation of the risk of severe COVID-19 are age;  
114 sex; body mass index; a history of cerebrovascular disease, diabetes, haematological cancer,  
115 non-haematological cancer, hypertension, kidney disease or respiratory disease (excluding  
116 asthma); and the genotypes of seven single nucleotide polymorphisms (SNPs) –  
117 rs112641600, rs10755709, rs118072448, rs7027911, rs71481792, rs112317747 and  
118 rs2034831 [3]. The log odds of the risk of severe COVID-19 is the sum of the intercept and  
119 the product of the value and beta coefficient for each of the risk factors listed in  
120 Supplementary Table S1. The probability of severe COVID-19 is then the inverse logit of the  
121 log odds ( $x$ ), that is,  $\frac{1}{(1+e^{-x})}$ .

122 We used the age reported at the completion of the participant's first Lifelines COVID-  
123 19 questionnaire. The questionnaires asked about a history of cancer, cerebrovascular disease,  
124 diabetes, hypertension, kidney disease and respiratory disease on three occasions. If any of  
125 the participants' responses to the risk factor questions were missing for all answered  
126 questionnaires, we used responses from their Lifelines baseline questionnaire.

127 We were not able to identify the type of cancer reported by Lifelines COVID-19  
128 participants so we used the risk associated with having a non-haematological cancer for all  
129 reported cancers. In the Lifelines questionnaires, the respiratory disease question included  
130 asthma, whereas this is excluded in the model calculations. Because we were not able to  
131 distinguish respiratory disease solely due to asthma, we included all reports of respiratory  
132 disease in the model calculations. Gender, ethnicity, weight and height were taken from the  
133 Lifelines baseline questionnaire. If two weight or height measurements were available, we  
134 used the most recent weight measurement and the mean of the height measurements. Two of  
135 the SNPs in the risk model were not available on the Illumina CytoSNP-12v2 array used by  
136 Lifelines [6]. Instead, we used highly correlated proxy SNPs (rs10905502 was the proxy for  
137 rs71481792 [ $r^2 = 0.75$ ,  $D' = 1.0$ ] and rs78654835 was the proxy for rs112317747 [ $r^2 = 1.0$ ,  $D'$   
138  $= 1.0$ ]).

139 To extend the model to people aged less than 50 years, we estimated the risk  
140 associated with younger age groups using data from the Centers for Disease Control and  
141 Prevention [9] such that, compared with the 50–69 years baseline age group, people aged 18–  
142 29 years were at 0.27 times the risk, people aged 30–39 years were at 0.43 times the risk, and  
143 people aged 40–49 years were at 0.67 times the risk.

144 In each analysis – (i) using participants with confirmed COVID-19 and (ii) using  
145 participants with both confirmed and self-reported COVID-19 – the cases were those who  
146 reported having been hospitalized for COVID-19 and the controls were the remainder of the  
147 group. We also did analyses restricting the dataset to those aged 50 years or older (the ages in  
148 which the model was developed).

149 As we did previously [3], we assessed the association between quintile of risk score  
150 and severe COVID-19 using logistic regression. We used the area under the receiver  
151 operating characteristic curve (AUC) to assess discrimination. We used logistic regression of

152 the log odds of the risk score to assess calibration in terms of the overall estimation of risk  
153 (the intercept) and the dispersion of risk (the slope), and we drew calibration plots of deciles  
154 of expected and observed cases of severe COVID-19. We used Stata MP version 13.1  
155 (StataCorp LP, College Station, Texas, USA) for all analyses and all statistical tests were two  
156 sided.

157 The Lifelines protocol has been approved by the Medical Ethical Committee of the  
158 University Medical Center Groningen, The Netherlands, under Approval Number 2007/152.  
159 All participants provided written informed consent to Lifelines before data collection began.  
160 This research was conducted using Lifelines data under Project Number OV20-00101.

161 The data used in this study was made available to us by Lifelines and is not publicly  
162 available. Researchers can apply to use the Lifelines data used in this study, and more  
163 information about how to request Lifelines data and the conditions of use can be found on  
164 their website (<https://www.lifelines.nl/researcher/how-to-apply>). Stata MP Version 13.1 code  
165 for the analysis is available for non-commercial purposes from the corresponding author on  
166 request.

167 Of the 26 845 Lifelines COVID-19 cohort participants who had genotyping data  
168 available and had completed at least one questionnaire, 3214 (12.0%) completed one  
169 questionnaire, 5742 (21.4%) completed 2–5 questionnaires, 4194 (15.6%) completed 6–10  
170 questionnaires, 3568 (13.3%) completed 11–15 questionnaires, 6106 (22.7%) completed 16–  
171 20 questionnaires and 4021 (15.0%) completed 21–23 questionnaires. We excluded 15 933  
172 questionnaires from 15 040 participants where (and after which) the participant reported  
173 being vaccinated.

174 In the final dataset, 55 participants were hospitalized for their COVID-19 infection  
175 and were considered cases in this study. We used two control groups: the first comprised the  
176 1355 participants who had confirmed COVID-19; the second comprised both the first control



177 group and the 2518 participants who self-reported having had COVID-19 (i.e. 3925  
178 participants). In the cases, there were 28 (50.9%) women and 27 (49.1%) men; their mean  
179 age was 57.6 years (standard deviation [SD] = 10.3) and the mean number of completed  
180 questionnaires was 17.0 (SD = 6.3). In the confirmed COVID-19 control group, there were  
181 905 (66.8%) women and 450 (33.2%) men; their mean age was 53.0 years (SD = 11.6) and  
182 the mean number of completed questionnaires was 14.1 (SD = 6.8). In the confirmed and  
183 self-reported COVID-19 control group, there were 2414 (62.3%) women and 1459 (37.7%)  
184 men; their mean age was 51.5 years (SD = 11.8) and the mean number of completed  
185 questionnaires was 12.2 (SD = 7.2).

186 In the cases, the mean probability of severe COVID-19 was 0.225 (SD = 0.019); in  
187 the confirmed COVID-19 controls, the mean was 0.165 (SD = 0.002); and in the confirmed  
188 and self-reported COVID-19 controls, the mean was 0.165 (SD = 0.001). The risk  
189 distribution for the cases, both control groups and the whole Lifelines COVID-19 cohort are  
190 show in in Supplementary Figure S1.

191 The top half of Table 1 shows the results of the analyses of the confirmed COVID-19  
192 group and the confirmed and self-reported COVID-19 group for all ages. Overall, the results  
193 were similar for the two groups. The odds ratios (OR) per quintile of risk (1.63 and 1.58,  
194 respectively) were a little lower than the OR of 1.77 seen in the validation group in the risk  
195 prediction model development paper [3]. Similarly, the AUCs (0.680 and 0.679, respectively)  
196 were a little lower than the AUC of 0.732 seen in the model development paper. In terms of  
197 calibration, there was no evidence of under- or over-dispersion in either group ( $\beta = 0.92$  and  
198  $0.90$ , respectively), as in the original paper ( $\beta = 0.90$ ). In both groups, the model  
199 overestimated risk ( $\alpha = -1.78$  and  $-2.86$ , respectively), whereas the validation group in the  
200 model development paper did not ( $\alpha = -0.08$ ).

201           The bottom half of Table 1 shows the analyses limited to participants aged 50 years  
202           and older. Compared with the analysis of all ages, there was a reduction in the ORs per  
203           quintile of risk for both the confirmed COVID-19 group and the confirmed and self-reported  
204           COVID-19 group (1.54 and 1.47, respectively), and a reduction in the AUCs (0.658 and  
205           0.651, respectively). The calibration slopes suggested over-dispersion of risk but were not  
206           statistically significant. The overestimations of risk (−1.84 and −2.87, respectively) were  
207           similar to those seen in the analysis of all ages.

208           The true number of people with COVID-19 is unknown but is likely to be somewhere  
209           between the number who test positive for SARS-CoV-2 infection and the number who self-  
210           report having had COVID-19. In this study we have addressed this uncertainty by conducting  
211           two sets of analyses: the first in individuals with confirmed COVID-19, and the second with  
212           individuals with confirmed and self-reported COVID-19. In terms of discrimination, the  
213           AUC of the risk prediction model was almost identical in the two analyses and only slightly  
214           lower than the AUC in validation group in the model development paper [3]. This and the  
215           similarity in the association per quintile of risk (Table 1) provide confidence in the model’s  
216           application across adult populations. Risk of COVID-19 severity was overestimated in this  
217           study, but in a clinical setting, overestimation of risk is preferred to an underestimation given  
218           that the risk-reduction options of vaccination, masking and social distancing are benign in  
219           nature. Our results were similar for the full dataset and when limiting analyses to people aged  
220           50 years and over.

221           It is possible that some severe cases of COVID-19 have not been ascertained in this  
222           dataset. Death registry linkage identified 77 deaths in the broader Lifelines COVID-19 cohort  
223           in people who did not have confirmed or self-reported COVID-19. While these deaths may  
224           have been unrelated to COVID-19, some will represent people who became infected and

225 were too unwell to complete a Lifelines COVID-19 questionnaire before they died. This  
226 limitation may have attenuated some of the results seen in this study.

227 As the pandemic continues to evolve, there are two major issues that can affect the  
228 utility of our risk model. First, we have to address the impact of viral variants on the  
229 performance of the risk model. The model development paper [3] and the present study used  
230 datasets in which the original and alpha SARS-CoV-2 variants were predominant. We have  
231 not been able to assess our model in datasets with known delta or omicron SARS-CoV-2  
232 variants. We hypothesize that the clinical and genetic risk factors have broad effects in terms  
233 of risk of severe disease because the delta and omicron SARS-CoV-2 variants appear to  
234 affect transmissibility rather than severity [10].

235 Second, our model does not incorporate the protection offered by vaccination. Thus in  
236 vaccinated adults, the model will overestimate their risk of developing severe disease.  
237 However, we know that vaccine immunity wanes over about six months through a steady  
238 reduction in antibody levels leading to greater number of breakthrough infections among the  
239 vaccinated [11]. The wide range of immunity across individuals makes it hard to predict the  
240 impact of waning vaccination in terms of risk. Thus, we believe that the model can be used to  
241 provide a baseline risk of developing severe disease, even in the context of vaccinated adults.

242 Herein, we have validated our model to predict risk of severe COVID-19 if infected  
243 with SARS-CoV-2 in a dataset unrelated to the one in which the model was originally  
244 developed and validated. Despite new SARS-CoV-2 variants of concern, the model may  
245 complement current public health efforts in vaccine (and booster) uptake and may enable  
246 healthcare providers to have more informed discussions with patients about their risk-  
247 mitigation options and early treatment awareness, if ever infected.

## 248 **References**

- 249 1. **Gao YD, et al.** (2021) Risk factors for severe and critically ill COVID-19 patients: A  
250 review. *Allergy*; **76**: 428-455.
- 251 2. **Velavan TP, et al.** (2021) Host genetic factors determining COVID-19 susceptibility  
252 and severity. *EBioMedicine*; **72**: 103629.
- 253 3. **Dite GS, Murphy NM, Allman R.** (2021) Development and validation of a clinical  
254 and genetic model for predicting risk of severe COVID-19. *Epidemiology and*  
255 *Infection*; **149**: e162.
- 256 4. **Mc Intyre K, et al.** (2021) Lifelines COVID-19 cohort: investigating COVID-19  
257 infection and its health and societal impacts in a Dutch population-based cohort. *BMJ*  
258 *Open*; **11**: e044474.
- 259 5. **Warmink-Perdijk WDB, et al.** (2020) Lifelines NEXT: a prospective birth cohort  
260 adding the next generation to the three-generation Lifelines cohort study. *European*  
261 *Journal of Epidemiology*; **35**: 157-168.
- 262 6. **Scholtens S, et al.** (2015) Cohort Profile: LifeLines, a three-generation cohort study  
263 and biobank. *International Journal of Epidemiology*; **44**: 1172-1180.
- 264 7. **National Institute for Public Health and the Environment.** Coronavirus variants.  
265 Available at <https://coronadashboard.government.nl/landelijk/varianten> (Accessed 10  
266 November 2021).
- 267 8. **National Institute for Public Health and the Environment.** COVID-19 vaccinations.  
268 Available at <https://coronadashboard.government.nl/landelijk/vaccinaties> (Accessed 16  
269 November 2021).

- 270 9. **Centers for Disease Control and Prevention.** COVIDView: a weekly surveillance  
271 summary of U.S. COVID-19 activity. Available at  
272 <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/pdf/covidview-08-07-2020.pdf>  
273 (Accessed 15 November 2021).
- 274 10. **Lewis T.** Omicron’s effect won’t be as mild as hoped. *Scientific American*; 23  
275 December. Available at [https://www.scientificamerican.com/article/omicrons-effect-](https://www.scientificamerican.com/article/omicrons-effect-wont-be-as-mild-as-hoped1/)  
276 [wont-be-as-mild-as-hoped1/](https://www.scientificamerican.com/article/omicrons-effect-wont-be-as-mild-as-hoped1/) (Accessed 13 January 2021).
- 277 11. **Cromer D, et al.** (2021) Prospects for durable immune control of SARS-CoV-2 and  
278 prevention of reinfection. *Nature Reviews: Immunology*; **21**: 395-404.

## 279 **Acknowledgements**

280 The authors wish to acknowledge the services of the Lifelines Cohort Study  
281 (especially the data management team), the contributing research centres delivering data to  
282 Lifelines and all the study participants.

## 283 **Financial support**

284 The Lifelines initiative has been made possible by subsidy from the Dutch Ministry of  
285 Health, Welfare and Sport, the Dutch Ministry of Economic Affairs, the University Medical  
286 Center Groningen, Groningen University and the Provinces in the North of the Netherlands  
287 (Drenthe, Friesland, Groningen).

288 The analyses undertaken in this study were fully funded by Genetic Technologies  
289 Limited.

## 290 **Conflict of interest**

291 GSD, NMM, and RA are employees of Genetic Technologies Limited. ES is an  
292 employee of Phenogen Sciences Inc (a subsidiary of Genetic Technologies Limited).  
293 Genetic Technologies Limited had no role in the conceptualization, design, data analysis,  
294 decision to publish or preparation of the manuscript.

295 Aspects of this manuscript are covered by Provisional Patent Application  
296 AU\_2021900392 (pending), Methods of assessing risk of developing a severe response to  
297 Coronavirus infection. GSD, NMM and RA are named inventors on the patent application,  
298 which is assigned to Genetic Technologies Limited.

299 **Table**

300 **Table 1.** Validation analysis of model to predict risk of severe COVID-19 for participants of  
 301 all ages and for participants aged 50 years and older

All ages	Estimate	95% CI	P value
<b>Association: OR per quintile of risk</b>			
Confirmed COVID-19	1.63	1.31, 2.04	<0.0001
Confirmed and self-reported COVID-19	1.58	1.27, 1.95	<0.0001
<b>Discrimination: AUC</b>			
Confirmed COVID-19	0.680	0.608, 0.752	
Confirmed and self-reported COVID-19	0.679	0.607, 0.751	
<b>Calibration slope: <math>\beta</math></b>			
Confirmed COVID-19	0.92	0.54, 1.30	0.7*
Confirmed and self-reported COVID-19	0.90	0.55, 1.25	0.6*
<b>Calibration intercept: <math>\alpha</math></b>			
Confirmed COVID-19	-1.78	-2.38, -1.19	<0.0001
Confirmed and self-reported COVID-19	-2.86	-3.41, -2.30	<0.0001
Aged 50 years and older	Estimate	95% CI	P value
<b>Association: OR per quintile of risk</b>			
Confirmed COVID-19	1.54	1.21, 1.96	<0.0001
Confirmed and self-reported COVID-19	1.47	1.16, 1.86	0.001
<b>Discrimination: AUC</b>			
Confirmed COVID-19	0.658	0.579, 0.737	
Confirmed and self-reported COVID-19	0.651	0.573, 0.730	
<b>Calibration slope: <math>\beta</math></b>			
Confirmed COVID-19	0.75	0.34, 1.16	0.2*
Confirmed and self-reported COVID-19	0.69	0.31, 1.08	0.1*
<b>Calibration intercept: <math>\alpha</math></b>			
Confirmed COVID-19	-1.85	-2.48, -1.21	<0.0001
Confirmed and self-reported COVID-19	-2.87	-3.46, -2.28	<0.0001

302 Note: AUC, area under the receiver operating characteristic curve; CI, confidence interval; OR, odds  
 303 ratio. \* P value for the null hypothesis that the calibration slope = 1.

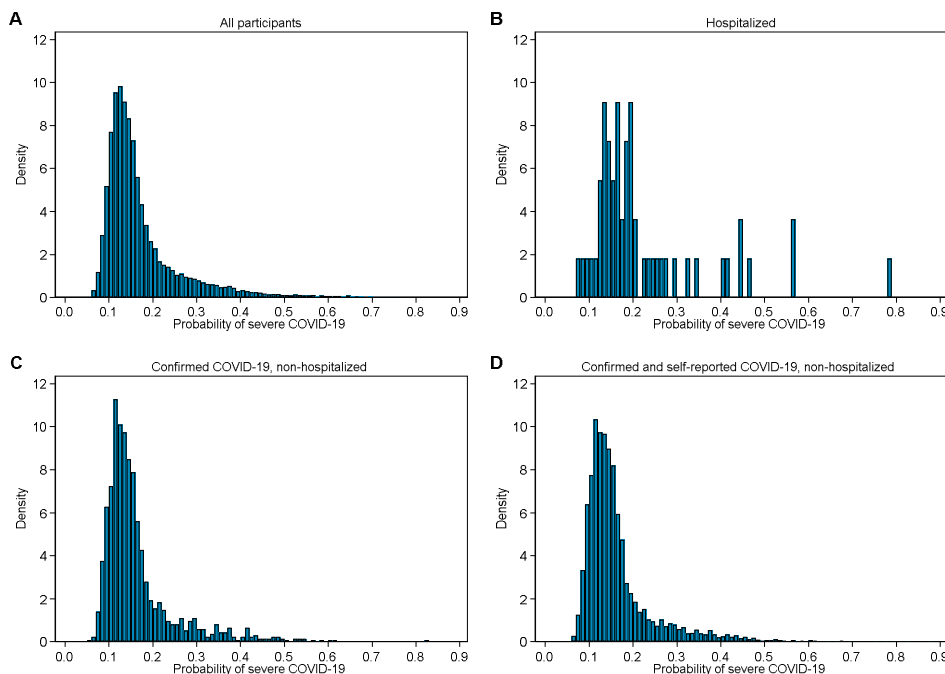
304 **Supplementary information**

305 **Supplementary Table S1.** Beta coefficients for calculation of risk of severe COVID-19

Variable	Value	Beta coefficient
Intercept		-1.37
Age 18–29 years	0 = no, 1 = yes	-1.31
Age 30–39 years	0 = no, 1 = yes	-0.83
Age 40–49 years	0 = no, 1 = yes	-0.40
Age 50–69 years	0 = no, 1 = yes	0.00
Age 70–74 years	0 = no, 1 = yes	0.57
Age 75–79 years	0 = no, 1 = yes	0.82
Age 80+ years	0 = no, 1 = yes	1.01
Male	0 = no, 1 = yes	0.24
Inverse of BMI	10/BMI	-1.60
Cancer, haematological	0 = no, 1 = yes	1.00
Cancer, non-haematological	0 = no, 1 = yes	0.26
Cerebrovascular disease	0 = no, 1 = yes	0.40
Diabetes	0 = no, 1 = yes	0.43
Hypertension	0 = no, 1 = yes	0.29
Kidney disease	0 = no, 1 = yes	0.69
Respiratory disease (excluding asthma)	0 = no, 1 = yes	1.17
rs112317747	0 = T/T, 1 = C/T, 2 = C/C	0.27
rs2034831	0 = A/A, 1 = C/A, 2 = C/C	0.24
rs112641600	0 = C/C, 1 = T/C, 2 = T/T	-0.24
rs10755709	0 = A/A, 1 = G/A, 2 = G/G	0.12
rs118072448	0 = T/T, 1 = C/T, 2 = C/C	-0.20
rs7027911	0 = G/G, 1 = A/G, 2 = A/A	0.10
rs71481792	0 = A/A, 1 = T/A, 2 = T/T	-0.11

306 Note: Body mass index (BMI) is calculated as kg/m<sup>2</sup>; the inverse of BMI is calculated as 10 divided by  
 307 BMI. In the current analysis, we used rs10905502 as a proxy for rs71481792 and rs78654835 as a  
 308 proxy for rs112317747.

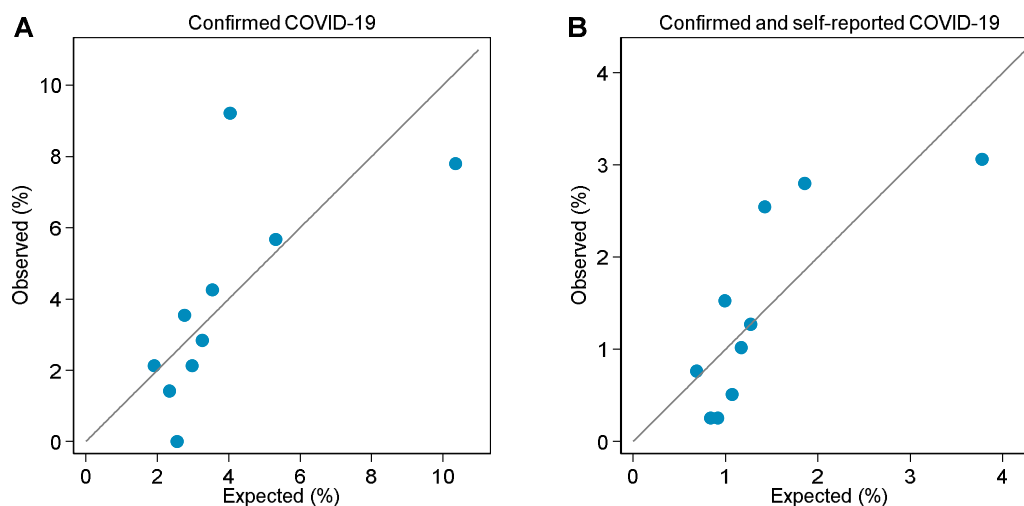




309

310 **Supplementary Figure S1.** Distribution of probability of severe COVID-19 in (A) all  
311 Lifelines COVID-19 cohort participants, (B) hospitalized (cases), (C) non-hospitalized  
312 confirmed COVID-19 (controls) and (D) non-hospitalized confirmed and self-reported  
313 COVID-19 (controls).

314



315

316 **Supplementary Figure S2.** Calibration plots for (A) the confirmed COVID-19 group and  
317 (B) the confirmed and self-reported COVID-19 group.