

1 **Attributes and predictors of Long-COVID: analysis of COVID cases and their symptoms**  
2 **collected by the Covid Symptoms Study App**

3  
4 Carole H. Sudre<sup>1,2,3</sup>, Benjamin Murray<sup>3</sup>, Thomas Varsavsky<sup>3</sup>, Mark S. Graham<sup>3</sup>, Rose S. Penfold<sup>4</sup>,  
5 Ruth C. Bowyer<sup>5</sup>, Joan Capdevila Pujol<sup>5</sup>, Kerstin Klaser<sup>3</sup>, Michela Antonelli<sup>3</sup>, Liane S. Canas<sup>3</sup>,  
6 Erika Molteni<sup>3</sup>, Marc Modat<sup>3</sup>, M. Jorge Cardoso<sup>3</sup>, Anna May<sup>5</sup>, Sajaysurya Ganesh<sup>5</sup>, Richard  
7 Davies<sup>5</sup>, Long H Nguyen<sup>6</sup>, David A. Drew<sup>6</sup>, Christina M. Astley<sup>7</sup>, Amit D. Joshi<sup>6</sup>, Jordi Merino<sup>8,9,10</sup>,  
8 Neli Tsereteli<sup>11</sup>, Tove Fall<sup>12</sup>, Maria F. Gomez<sup>11</sup>, Emma L. Duncan<sup>4</sup>, Cristina Menni<sup>4</sup>, Frances M.K.  
9 Williams<sup>4</sup>, Paul W. Franks<sup>4,11</sup>, Andrew T. Chan<sup>6</sup>, Jonathan Wolf<sup>5</sup>, Sebastien Ourselin<sup>3\*</sup>, Tim  
10 Spector<sup>4\*</sup>, Claire J. Steves<sup>4\*\*</sup>

11  
12 1 MRC Unit for Lifelong Health and Ageing, Department of Population Health Sciences,  
13 University College London, WC1E 7HB London

14 2 Centre for Medical Image Computing, Department of Computer Science, University College  
15 London

16 3 School of Biomedical Engineering & Imaging Sciences, King's College London, SE1 7EU, London  
17 UK

18 4 Department of Twin Research and Genetic Epidemiology, King's College London,  
19 Westminster Bridge Road, SE17EH London, UK

20 5 Zoe Global Limited, 164 Westminster Bridge Road, London SE1 7RW, UK

21 6 Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, MA, USA

22 7 Division of Endocrinology & Computational Epidemiology, Boston Children's Hospital, MA,  
23 USA

24 8 Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston,  
25 MA, USA.

26 9 Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.

27 10 Department of Medicine, Harvard Medical School, Boston, MA, USA.

28 11 Lund University Diabetes Centre, Department of Clinical Sciences, Malmö, Sweden

29 12 Molecular epidemiology, Department of Medical Sciences, Uppsala University, SE-751 85  
30 Uppsala Sweden

31  
32 \* Contributed equally

33 Corresponding author: Claire J. Steves – [Claire.j.steves@kcl.ac.uk](mailto:Claire.j.steves@kcl.ac.uk)

34  
35  
36 **Reports of “Long-COVID”, are rising but little is known about prevalence, risk factors, or**

37 **whether it is possible to predict a protracted course early in the disease. We analysed data**

38 **from 4182 incident cases of COVID-19 who logged their symptoms prospectively in the COVID**

39 **Symptom Study app. 558 (13.3%) had symptoms lasting  $\geq 28$  days, 189 (4.5%) for  $\geq 8$  weeks**

40 and 95 (2.3%) for  $\geq 12$  weeks. Long-COVID was characterised by symptoms of fatigue,  
41 headache, dyspnoea and anosmia and was more likely with increasing age, BMI and female  
42 sex. Experiencing more than five symptoms during the first week of illness was associated  
43 with Long-COVID, OR=3.53 [2.76;4.50]. A simple model to distinguish between short and long-  
44 COVID at 7 days, which gained a ROC-AUC of 76%, was replicated in an independent sample  
45 of 2472 antibody positive individuals. This model could be used to identify individuals for  
46 clinical trials to reduce long-term symptoms and target education and rehabilitation services.

47  
48 COVID-19 can manifest a wide severity spectrum from asymptomatic to fatal forms<sup>1</sup>. A further  
49 source of heterogeneity is the duration of symptoms, which could have considerable impact  
50 due to the huge scale of the pandemic. Hospitalised patients are well recognised to have lasting  
51 dyspnoea and fatigue in particular<sup>2</sup>, yet such patients constitute the ‘tip of the iceberg’ of  
52 symptomatic SARS CoV2 disease<sup>3</sup>. Few studies capture symptoms prospectively in the general  
53 population to ascertain with accuracy the duration of illness and the prevalence of long-lasting  
54 symptoms.

55  
56 Here we report a prospective observational cohort study of COVID-19 symptoms in a subset of  
57 4182 users of the COVID Symptom Study app meeting inclusion criteria (see online methods)<sup>4,5</sup>,  
58 compared to matched symptomatic test-negative controls. Briefly, the cases comprised  
59 individuals who reported testing positive for SARS-CoV2 by swab testing who started on the  
60 app “feeling physically normal” to be able to determine symptom onset. We compare cases  
61 with symptoms persisting over 28 days, LC28) and short duration (symptoms lasting less than

62 10 days, short-COVID). Our previous findings that clusters of symptoms predicted the need for  
63 acute care<sup>6</sup> led us to hypothesize that persistent symptomatology in COVID-19 (Long-COVID) is  
64 associated with early symptom patterns which could be used to predict who might be affected.  
65 Figure 1 shows the duration of symptoms reported in COVID+ cases (orange) over-laid on age,  
66 sex and BMI matched negative-testing symptomatic controls (blue). The overall median  
67 symptom duration was 11 days (IQR[6;19]).

68  
69 Of the 4182 COVID-19 swab positive users, 558 (13.3%) met the LC28 definition (Median 41,  
70 IQR[33,63]) of whom 189 (4.5%) met LC56, and 108 (2.6%) LC84. In contrast 1591 (38.0%) had  
71 short-COVID (median 6, IQR[4-8]). The proportions were comparable in three countries (LC28:  
72 GB 13.3%, USA 16.1%, Sweden 12.1%  $p=0.35$ ; LC56: GB 4.7%, USA 5.5%, Sweden 2.5%  $p=0.07$ ).

73  
74 Table 1 summarises the descriptive characteristics of the study population stratifying by  
75 symptom/disease duration. Age was significantly associated with LC28, rising from 9.9% in 18-  
76 49 year-olds to 21.9% in those aged  $\geq 70$  ( $p < 0.0005$ ), with escalating OR by age decile (Figure  
77 1b, Supplementary Table 2). LC28 disproportionately affected women (14.9%) compared to  
78 men (9.5%), although not in the older age-group. Long-COVID affected all socio-economic  
79 groups (assessed using Index of Multiple Deprivation), (Supplementary Figure 2). Individuals  
80 with Long-COVID were more likely to have required hospital assessment. Asthma was the  
81 only/unique pre-existing condition providing significant association with LC28 (OR=2.14 [1.55-  
82 2.96]).

83  
84  
85

86 Fatigue (97.7%) and headache (91.2%) were the most reported symptoms in those with LC28,  
87 followed by anosmia and lower respiratory symptoms, and headache was more often reported  
88 intermittently (Figure 2, supplementary Table s1). Free-text additional symptoms were more  
89 commonly reported in LC28 cases (81%) compared to Short-COVID (45%), with cardiac  
90 symptoms (palpitations, tachycardia) ( LC28,6.1%; short-COVID 0.5%  $p < 0.0005$ ), concentration  
91 or memory issues (4.1% vs 0.2%,  $p < 0.0005$ ), tinnitus and earache (3.6% vs 0.2%  $p < 0.0005$ ) and  
92 peripheral neuropathy symptoms (pins and needles and numbness) (2% vs 0.5%  $p = 0.004$ )  
93 disproportionately reported in LC28. Most of these symptoms were reported for the first time  
94 3-4 weeks post-symptom onset.

95  
96 We found two main patterns of symptomatology within LC28: those reporting exclusively  
97 fatigue, headache and upper respiratory complaints (shortness of breath, sore throat,  
98 persistent cough and loss of smell) and those with additional multi-system complaints, including  
99 ongoing fever and gastroenterological symptoms (Supplementary figure 3). In the individuals  
100 with long duration (LC28), ongoing fever OR 2.16 [1.50;3.13] and skipped meals OR 2.52  
101 [1.74;3.65] were strong predictors of a hospital visit. Details of the frequency of symptoms  
102 persisting beyond 28 and 56 days after disease onset are provided in Supplementary table 3.

103  
104  
105 Individuals with LC28 were more likely to report relapses (16.0% vs 8.4%) ( $p < 0.0005$ ). In  
106 comparison, in the matched group of SARS-CoV2 negative-tested individuals, relapse was  
107 reported in 11.5%, and relapse was longer in LC28 (median = 9 [5-18] vs 6 [4-10] days).

108

109 We explored how to estimate risk of LC28 among positive individuals from data available early  
110 in the disease course. Individuals reporting more than 5 symptoms in the first week (the  
111 median number reported) were significantly more likely to go on to experience LC28, (OR=3.95  
112 [3.10;5.04]). This strongest risk factor was predictive in both sexes and all age groups  
113 (supplementary Figures 4, a-e).

114

115 The five symptoms experienced during the first week most predictive of LC28 in the positive  
116 individuals were: fatigue OR=2.83 [2.09;3.83], headache OR=2.62 [2.04;3.37], dyspnoea  
117 OR=2.36 [1.91;2.91], hoarse voice OR=2.33 [1.88;2.90] and myalgia OR=2.22 [1.80;2.73] (Figure  
118 3). Similar patterns were observed in both genders. In adults aged over 70, loss of smell (which  
119 is less common) was the most predictive of long-COVID OR=7.35 [1.58;34.22] before fever  
120 OR=5.51 [1.75;17.36] and hoarse voice OR=4.03 [1.21;13.42] (Supplementary figures 4). Co-  
121 occurrence plots of symptoms in short-COVID versus LC28 further illustrate the importance of  
122 early multi-symptom involvement (Figure 3c).

123

124 We created Random Forest Prediction models using a combination of the first week's symptom  
125 reporting, personal characteristics and comorbidities. Using all features, the average ROC AUC  
126 was 76.7% (SD=2.5) (Figure 3d) in the classification between short-COVID and LC28. The  
127 strongest predictor was age (29.2%) followed by the number of symptoms during the first week  
128 (16.3%). Feature importance was relatively similar across age-specific models. However, in the  
129 over 70s, early features such as fever, anosmia and comorbidities were important, and may be  
130 'red flags' in older adults (Supplementary figure 6).

131

132 To create a model usable in healthcare settings, we simplified the prediction model to include  
133 only symptom number in the first week with age, and sex in a logistic regression model  
134 obtaining ROC AUC of 76.7% (SD 2.5) (Figure 3d), for which the calibration slope had an average  
135 of 1.02 (0.15). When optimising the balance between false positives and false negatives, we  
136 obtained a specificity of 73.4% (SD 9.7) and a sensitivity of 68.7% (SD 9.9). Specificity,  
137 Sensitivity, PPV and NPV values at different thresholds are presented in Supplementary table 6.  
138  
139 Key predictive findings of our analysis were validated in an independent dataset of 2412  
140 individuals who reported testing antibody positive (but no positive PCR result) for SARS-CoV2  
141 from 2 weeks after symptom onset where, again, the number of symptoms in the first week of  
142 illness was the strongest predictor, OR=4.60 [95% CI 3.28; 6.46]. The simple prediction model,  
143 was similarly predictive of LC28 in the antibody group, with a ROC-AUC of 75.9% (SD=4.3%)  
144 (Figure 3-e).

145  
146 While this study provides important insights into the disease presentation, any generalisation  
147 should be considered carefully. Our study was limited by being confined to app users who were  
148 disproportionately female and under-represented those >70years which could increase or  
149 decrease our estimate of the extent of Long-COVID respectively and caution is needed in  
150 interpreting associations found in smaller population subgroups. Swab test results were self-  
151 reported and were all assumed to be RT-PCR, as antigen tests were not available at the time.  
152 Applying a weighting following the UK population (see Supplementary Methods), the estimated  
153 proportion of people experiencing symptomatic COVID-19 going on to suffer Long-COVID were

154 similar: 14.5%, 5.1% and 2.2% for 4, 8- and 12-weeks duration respectively. While estimates  
155 could be inflated because early PCR testing was restricted to those more severely unwell, or if  
156 regular logging or test results encouraged a systematic bias in symptom reporting, Long-COVID  
157 may here be underestimated if individuals with prolonged symptoms were more likely to stop  
158 logging symptoms on the app. Our participant selection criteria were chosen to confidently  
159 identify cases, and upper and lower bounds for estimates given each exclusion criteria are  
160 presented in Supplementary Table 4. Symptom reporting rates through the study period for all  
161 users are also presented in Supplementary Table 6. Taken together, these data suggest that our  
162 estimates may be conservative. We had insufficient numbers to explore risk factors for disease  
163 over 2 months and were unable to analyse the impact of ethnicity due to incomplete data. In  
164 addition, the list of symptoms on the app is necessarily non-exhaustive, however, analysis of  
165 the free-text responses allowed us to highlight other symptoms present in Long-COVID, such as  
166 cardiac and neurological manifestations. With emerging evidence of ongoing myocardial  
167 inflammation and change in<sup>8,9</sup> associated with COVID-19, this calls for specific studies of cardiac  
168 and neurological longer-term sequelae of COVID-19.

169

170 At the population level, it is critical to quantify the burden of Long-COVID to better assess its  
171 impact on the healthcare system and appropriately distribute resources. In our study,  
172 prospective logging of a wide range of symptoms allowed us to conclude that the proportion of  
173 people with symptomatic COVID-19 who experience prolonged symptoms is considerable, and  
174 relatively stable across three countries with different cultures. Whether looking at a four-week  
175 or an eight-week threshold for defining long duration, those experiencing Long-COVID were

176 consistently older, more likely to be female and to require hospital assessment than in the  
177 group reporting symptoms for a short period of time. Those going on to experience LC28 had  
178 multi-system disease from the start, supporting the need for holistic support<sup>10</sup>. While asthma  
179 was not reported as a factor of risk for hospitalisation in<sup>11</sup>, its association with Long-COVID  
180 (LC28) warrants further investigation.

181  
182 We found early disease features were predictive of duration. With only three features - number  
183 of symptoms in the first week, age and sex, we were able to accurately distinguish individuals  
184 with LC28 from those with short duration. Importantly, the model generalised well to the  
185 population reporting antibody testing. This important information could feature in highly  
186 needed targeted education material for both patients and healthcare providers and we present  
187 typical nomograms for use in clinical settings in Supplementary Figure 7. Moreover, the method  
188 could help determine at-risk groups and could be used to target early intervention trials and  
189 clinical service developments to support rehabilitation in primary and specialist care<sup>14</sup> to  
190 alleviate Long-COVID and facilitate timely recovery.

191  
192 **Ethics:** In the UK, the App Ethics has been approved by KCL ethics Committee REMAS ID 18210,  
193 review reference LRS-19/20-18210 and all subscribers provided consent. In Sweden, ethics  
194 approval for the study was provided by the central ethics committee (DNR 2020-01803).  
195 **Funding:** Zoe provided in kind support for all aspects of building, running and supporting the  
196 app and service to all users worldwide. Support for this study was provided by the NIHR-funded  
197 Biomedical Research Centre based at GSTT NHS Foundation Trust. This work was supported by  
198 the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for  
199 Value Based Healthcare.  
200 Investigators also received support from the Wellcome Trust, the MRC/BHF, Alzheimer's  
201 Society, EU, NIHR, CDRF, and the NIHR-funded BioResource, Clinical Research Facility and BRC  
202 based at GSTT NHS Foundation Trust in partnership with KCL. ATC was supported in this work  
203 through a Stuart and Suzanne Steele MGH Research Scholar Award. CM is funded by the  
204 Chronic Disease Research Foundation and by the MRC AimHy project grant. LHN, DAD, ADJ,



205 ADS, CG, WL are supported by the Massachusetts Consortium on Pathogen Readiness  
206 (MassCPR) and Mark and Lisa Schwartz. JM was partially supported by the European  
207 Commission Horizon 2020 program (H2020-MSCA-IF-2015-703787). The work performed on the  
208 Swedish study is supported by grants from the Swedish Research Council, Swedish Heart-Lung  
209 Foundation and the Swedish Foundation for Strategic Research (LUDC-IRC 15-0067).

210 **Competing interests:**

211 Zoe Global Limited co-developed the app *pro bono* for non-commercial purposes. Investigators  
212 received support from the Wellcome Trust, the MRC/BHF, EU, NIHR, CDRF, and the NIHR-  
213 funded BioResource, Clinical Research Facility and BRC based at GSTT NHS Foundation Trust in  
214 partnership with KCL. RD, JW, JCP, AM and SG work for Zoe Global Limited and TDS and PWF  
215 are consultants to Zoe Global Limited. LHN, DAD, JM, PWF and ATC previously participated as  
216 investigators on a diet study unrelated to this work that was supported by Zoe Global Ltd.

217 **Data and materials availability:** Data used in this study is available to bona fide researchers  
218 through UK Health Data Research using the following link

219 <https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259>

220

221

222 **References**

223

- 224 1. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C.  
225 Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019  
226 (COVID-19): A Review. *JAMA - Journal of the American Medical Association* vol. 324 782–  
227 793 (2020).
- 228 2. Sheehy, L. M. Considerations for postacute rehabilitation for survivors of COVID-19.  
229 *Journal of Medical Internet Research* vol. 22 (2020).
- 230 3. Alwan, N. A. A negative COVID-19 test does not mean recovery. *Nat.* 2020 5847820  
231 (2020).
- 232 4. Menni, C. *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-  
233 19. *Nat. Med.* **26**, 1037–1040 (2020).
- 234 5. Drew, D. A. *et al.* Rapid implementation of mobile technology for real-time epidemiology  
235 of COVID-19. *Science (80-. )*. eabc0473 (2020) doi:10.1126/science.abc0473.
- 236 6. Sudre, C. H. *et al.* Symptom clusters in Covid19: A potential clinical prediction tool from  
237 the COVID Symptom study app. *medRxiv* 2020.06.12.20129056 (2020)  
238 doi:10.1101/2020.06.12.20129056.
- 239 7. Cirulli, E. T. *et al.* Long-term COVID-19 symptoms in a large unselected population.  
240 *medRxiv* (2020) doi:10.1101/2020.10.07.20208702.
- 241 8. Mitrani, R. D., Dabas, N. & Goldberger, J. J. COVID-19 cardiac injury: Implications for long-  
242 term surveillance and outcomes in survivors. *Hear. Rhythm* (2020)  
243 doi:10.1016/j.hrthm.2020.06.026.
- 244 9. Padda, I., Khehra, N., Jaferi, U. & Parmar, M. S. The Neurological Complexities and  
245 Prognosis of COVID-19. *SN Compr. Clin. Med.* (2020) doi:10.1007/s42399-020-00527-2.
- 246 10. *Living with Covid19*. <https://evidence.nihr.ac.uk/themedreview/living-with-covid19/>

- 247 (2020) doi:10.3310/themedreview\_41169.
- 248 11. Robinson, L. B. *et al.* Title: COVID-19 severity in asthma patients: A multi-center matched  
249 cohort study. *medRxiv* 2020.10.02.20205724 (2020) doi:10.1101/2020.10.02.20205724.
- 250 12. Mahase, E. Covid-19: Demand for dexamethasone surges as RECOVERY trial publishes  
251 preprint. *BMJ* **369**, m2512 (2020).
- 252 13. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Preliminary Report. *N.*  
253 *Engl. J. Med.* (2020) doi:10.1056/nejmoa2007764.
- 254 14. Greenhalgh, T., Knight, M., A’Court, C., Buxton, M. & Husain, L. Management of post-  
255 acute covid-19 in primary care. *BMJ* **370**, (2020).
- 256 15. Drew, D. A. *et al.* Rapid implementation of mobile technology for real-time epidemiology  
257 of COVID-19. *Science (80-. )*. (2020).
- 258 16. Spiel, C. *et al.* A Euclidean distance-based matching procedure for non randomized  
259 comparison studies. *Eur. Psychol.* **13**, 180–187 (2008).

260  
261  
262  
263  
264  
265  
266  
267  
268

Table 1. Characteristics of Individuals with COVID-19 by duration of symptoms, compared to matched sample testing negative for COVID-19. In statistical comparison, the short COVID group is the reference.

	Positive PCR test				Matched Negative sample
	Short (<10 days)	LC28 (>=28 days) (including LC56)	LC56 (>=56 days)	Overall	
<b>Number</b>	1591	558	189	4182	4182
<b>GB/SE/US [numbers, %]</b>	1365 / 139 / 87 85.8 / 8.7 / 5.5	466/57/35 83.5 / 10.2 / 6.3	165/12/12 87.3 / 6.3 / 6.3	3491 / 473 / 218 83.5 / 11.3 / 5.2	3882/131/169 92.8/3.1/4.1
<b>Male (%)</b>	32.7	20.3***	16.9*	28.5	28.5
<b>Age (years) [median, IQR]</b>	38 [29;49]	50 [39;57]***	52 [43;59]***	42 [32;53]	42 [32;53]
<b>Age group (18-49/ 50-69/ &gt;70)</b>	1122/331/38 75.3 / 22.2 / 2.5	259/262/24 47.5 / 48.1 / 4.4	69/96/11 39.2 / 54.5 / 6.3	2627/1195/96 62.8 / 28.6 / 2.3	2821 / 1264/97 67.5 / 30.2 / 2.3
<b>Obese (%)</b>	23.8	27.6*	26.5	26.3	26.4
<b>BMI (kg/m<sup>2</sup>) [median, IQR]</b>	25.5 [22.7;29.7]	26.1 [23.3;30.5]	25.9[23.3;30.5]	25.9[23.3;30.3]	25.9 [23.0;30.3]
<b>Asthma (%)</b>	7.7	15.8***	18.0***	10.0	13.7
<b>Lung disease (%)</b>	12.8	16.5**	15.9	13.6	13.7
<b>Diabetes (%)</b>	3.0	3.9	5.8*	2.9	2.8
<b>Heart (%)</b>	1.7	3.2**	4.8**	1.9	1.7

<b>Kidney (%)</b>	0.5	0.9	0.5	0.6	0.6
<b>IMD (median decile - IQR)</b>	7 [4;9]	7 [5;9]	7 [5;9]	7 [4;9]	7[5;9]***
<b>IMD quintiles</b>	64/75/334/132/634 5.2/6.1/27.0/10.7/51.2	23/23/86/49/240 5.5/5.5/20.4/11.6/57.0	10/9/26/18/88 6.6/6.0/17.2/11.9/58.3	158 / 194 /830/363 /1653 4.9 /6.1/26.0/11.4/51.7	118/193/895/376/2057 3.2/5.3/24.6/10.3/56.5
<b>Visit to hospital (%)</b>	7.0	31.5***	43.9***	13.9	4.1
<b>Number of symptoms in the first week [median [IQR]]</b>	5 [3;7]	7 [5-9]***	7 [5;9]***	6 [4;8]	3 [2;4]***

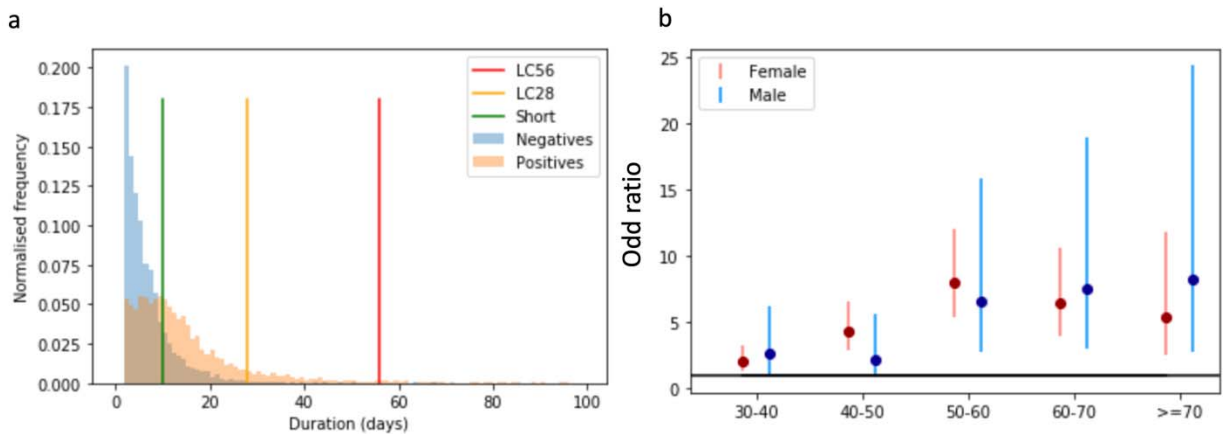
269 \* indicates  $p < 0.1$  \*\*  $< 0.05$  \*\*\*  $< 0.01$  when comparing to short covid. Comparison are performed with respect to  
 270 the “short duration” within the positive group. Matched Negatives are compared to the overall positive population  
 271 Mann Whitney U tests are performed for continuous variables and chi square tests are performed when  
 272 comparing proportions.

273 Index of Multiple Deprivation (IMD) information is only available for app users from the UK who have entered a  
 274 complete post code

275 Acronyms: GB – Great Britain / SE – Sweden / US – United States / IMD – Index of multiple deprivation

276

277



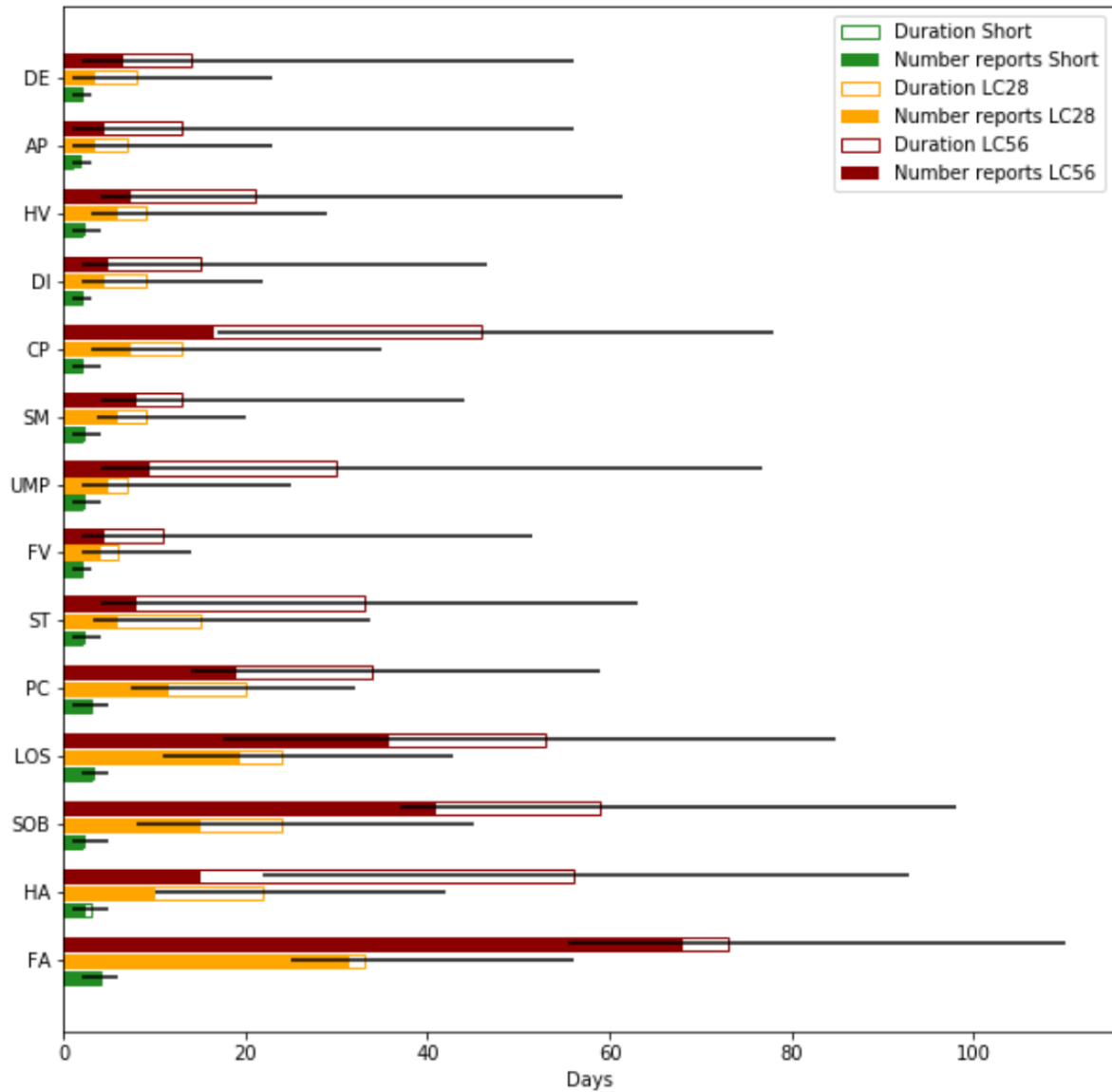
278

279

280 Figure 1. a) Distribution of duration of symptoms in COVID-19 – The coloured bars indicate the limits to define  
 281 short, LC28 and LC56. The y-axis reports the normalised frequency of duration of symptoms. b) OR and 95% CI of  
 282 LC28 with each successive decile compared to 20-30-year-olds

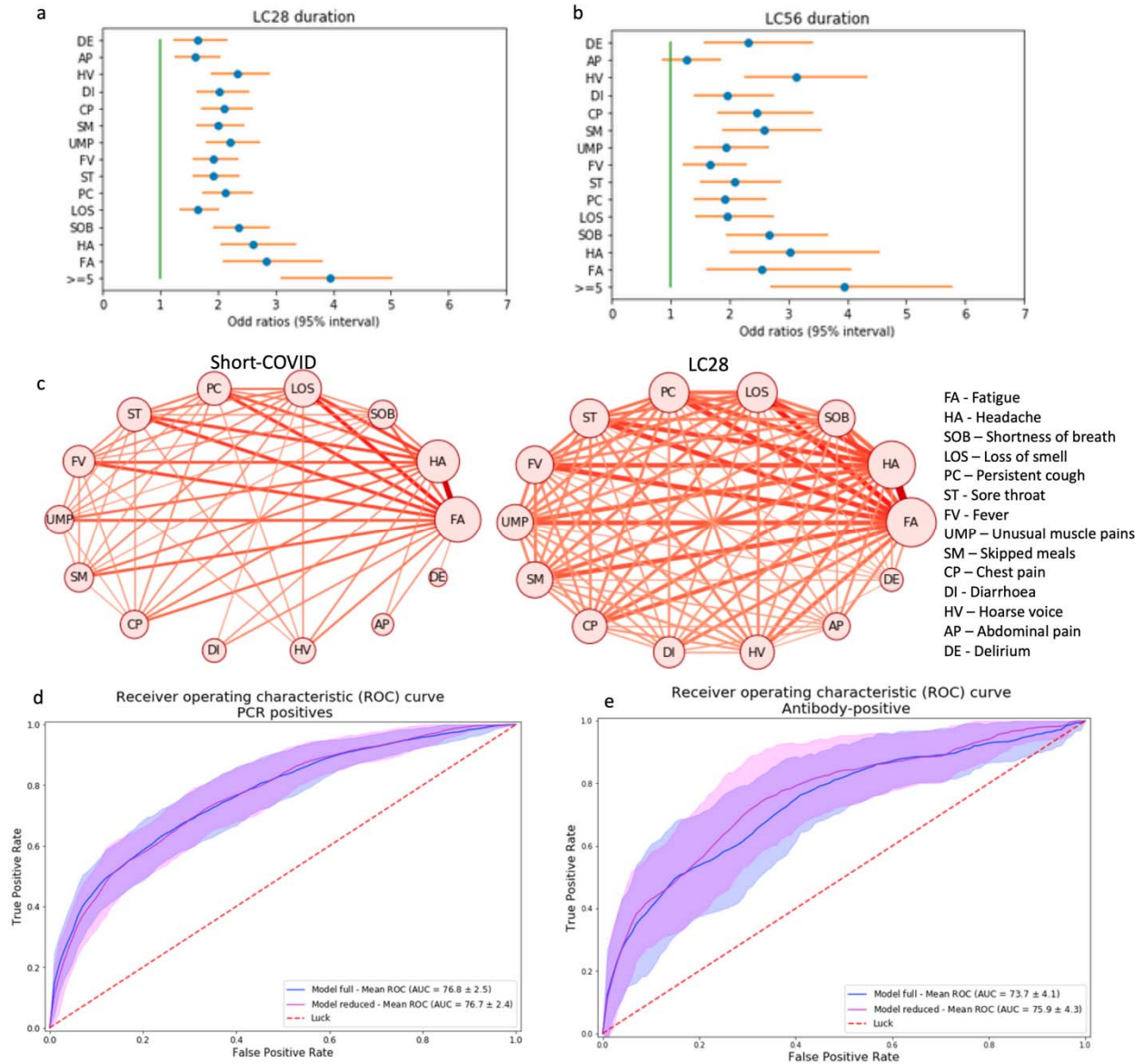
283

284



285  
286 **Figure 2** : Symptoms by duration. For each symptom (ordered from top to bottom by increasing frequency of  
287 occurrence) the median duration of report is presented by the total (hollowed) bar height, with associated  
288 interquartile range represented by the black line, for the short, LC28 and LC56 durations. The filled bars represent  
289 the number of times a report has been given. For both duration and number of reported days of symptoms, the x  
290 axis reflects the number of days. This highlights the differences in the symptoms in terms of their intermittence  
291 throughout the course of the disease. (Abbreviations DE – delirium, AP – Abdominal Pain, HV – Hoarse Voice, DI –  
292 Diarrhoea, CP – Chest Pain, SM – skipped meals, UMP – Unusual Muscle pains, FV – Fever, ST – Sore Throat, PC –  
293 Persistent Cough, LOS – Loss of smell, SOB – Shortness of breath, HA – Headache, FA – Fatigue)

294  
295  
296



297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312

**Figure 3:** Symptom correlates of long-COVID for LC28 (a) and LC56 (b) with correction for age and gender. c) Co-occurrence network of symptom pairs with the frequency of symptom report as the size of the node and the likelihood of symptom pair co-occurrence by the weight of the edge linking them. Edges representing a co-occurrence of less than 10% were removed. d) – Receiver Operating Characteristic (ROC) curve of the cross-validated full and reduced models on the PCR cohort. e)– ROC curve when training on the whole PCR cohort and testing on the antibody-positive cohort for the full (blue) and reduced (magenta) model. Random predictive probability is indicated in both panels as a dashed red line. (Abbreviations DE – delirium, AP – Abdominal Pain, HV – Hoarse Voice, DI – Diarrhoea, CP – Chest Pain, SM – skipped meals, UMP – Unusual Muscle pains, FV – Fever, ST – Sore Throat, PC – Persistent Cough, LOS – Loss of smell, SOB – Shortness of breath, HA – Headache, FA – Fatigue)

313

314