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

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

1 MRC Unit for Lifelong Health and Ageing, Department of Population Health Sciences, University College London, WC1E 7HB London
2 Centre for Medical Image Computing, Department of Computer Science, University College London
3 School of Biomedical Engineering & Imaging Sciences, King's College London, SE1 7EU, London UK
4 Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, SE17EH London, UK
5 Zoe Global Limited,164 Westminster Bridge Road, London SE1 7RW, UK
6 Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, MA, USA
7 Division of Endocrinology & Computational Epidemiology, Boston Children's Hospital, MA, USA
8Diabetes Unit and Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA.
9Program in Medical and Population Genetics, Broad Institute, Cambridge, MA, USA.
10Department of Medicine, Harvard Medical School, Boston, MA, USA.
11 Lund University Diabetes Centre, Department of Clinical Sciences, Malmö, Sweden
12 Molecular epidemiology, Department of Medical Sciences, Uppsala University, SE-751 85 Uppsala Sweden

Corresponding author: Claire J. Steves – Claire.j.steves@kcl.ac.uk

Reports of "Long-COVID", are rising but little is known about prevalence, risk factors, or whether it is possible to predict a protracted course early in the disease. We analysed data from 4182 incident cases of COVID-19 who logged their symptoms prospectively in the COVID Symptom Study app. 558 (13.3%) had symptoms lasting >28 days, 189 (4.5%) for >8 weeks and 95 (2.3%) for >12 weeks. Long-COVID was characterised by symptoms of fatigue,

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

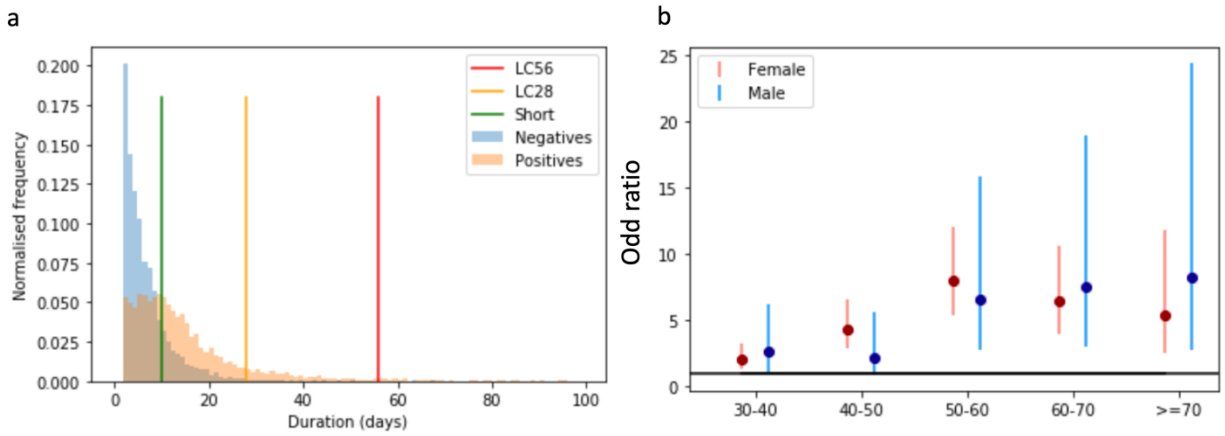
COVID-19 can manifest a wide severity spectrum from asymptomatic to fatal forms [1]. A further source of heterogeneity is the duration of symptoms after the acute stage which could have considerable impact due to the huge scale of the pandemic. Hospitalised patients are well recognised to have lasting dyspnoea and fatigue in particular [2], yet such patients constitute the 'tip of the iceberg' of symptomatic SARS CoV2 disease [3]. Few studies capture symptoms prospectively in the general population to ascertain with accuracy the duration of illness and the prevalence of long-lasting symptoms.

Here we report a prospective observational cohort study of COVID-19 symptoms in a subset of 4182 users of the COVID Symptom Study app meeting inclusion criteria (see online methods) [4,5]. Briefly, the subset comprised individuals who had tested positive for SARS-CoV2 by PCR swab testing who logged as "feeling physically normal" before the start of illness (up to 14 days before testing) so that we could determine onset. We compare cases of long (reporting symptoms lasting more than 28 days, LC28) and short duration (reporting symptoms lasting less than 10 days, short-COVID). Our previous findings that clusters of symptoms predicted the

need for acute care [6] led us to hypothesize that persistent symptomatology in COVID-19 (Long-COVID) is associated with a particular symptom pattern early in the disease which could be used to predict who might be affected. In particular, dyspnoea has been shown to be a significant predictor of long-term symptoms in an unselected population[7].

Figure 1 shows the duration of symptoms reported in COVID+ cases (orange) over-laid on age, sex and BMI matched negative testing symptomatic controls (blue), depicting lines for the definitions of short-COVID, LC28 and LC56 (symptoms for more than 56 days) used in this study. The duration of COVID-19 symptoms followed an approximately log-normal distribution (sigma = 0.97, location =0.78, scale = 10.07), with an overall median symptom duration of 11 days (IQR[6;19]).

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



Of the 4182 COVID-19 swab positive users, 558 (13.3%) met the LC28 definition with a median duration of 41 days (IQR[33,63] of whom 189 (4.5%) met LC-56, and 95 (2.3%) LC94. In contrast

1591 (38.0%) had short disease duration (median 6, IQR[4-8]). The proportion with LC28 were comparable in all three separate countries (GB 13.3%, USA 16.1%, Sweden 12.1% p=0.35) and for LC56 (GB 4.7%, USA 5.5%, Sweden 2.5% p=0.07).

Table 1 summarises the descriptive characteristics of the study population overall and stratifying by symptom/disease duration. Age was significantly associated with Long-COVID (LC28) rising from 9.9% in 18-49 year olds to 21.9% in those aged >=70 ($p < 0.0005$), with a clear escalation in OR by age decile (Figure 1b), although females aged 50-60 had the highest odds. (ST2). Individuals with Long-COVID were more likely to have required hospital assessment in the acute period. LC28 disproportionately affected women (14.9%) compared to men (9.5%), although this sex effect was not significant in the older age-group. Long-COVID affected all socio-economic groups (assessed using Index of Multiple Deprivation), (Supplementary Figure 2). Asthma was the only/unique pre-existing condition providing significant association with long-COVID-19 (OR = 2.14 [1.55-2.96]).

Table 1. Characteristics of COVID-19 by duration of symptoms.

	Overall	**Short**	**LC28 (including LC56)**	**LC56**
Number	4182	1591	558	189
GB/SE/US	3491 / 473 / 218	1365 / 139 / 87	466/57/35	165/12/12
Male (%)	28.5	32.7	20.3***	16.9*
Age (years)	42.8 (13.4)	40.1 (13.6)	48.9 (12.7)***	50.9 (12.5)***
Age group	2627/1195/96	1122/331/38	259/262/24	69/96/11

Obese (%)	26.3	23.8	27.6*	26.5
BMI	27.3 (5.9)	26.8 (5.7)	27.5 (5.7)	27.5 (5.8)
Asthma (%)	10.0	7.7	15.8***	18.0***
Lung (%)	13.6	12.8	16.5**	15.9
Diabetes (%)	2.9	3.0	3.9	5.8*
Heart (%)	1.9	1.7	3.2**	4.8**
Kidney (%)	0.6	0.5	0.9	0.5
IMD (average decile)	6.4 (2.7)	6.3 (2.7)	6.7 (2.8)	6.6 (2.9)
IMD quintilesϰ	158 / 194 /830/363 /1653	64/75/334/132/634	23/23/86/49/240	10/9/26/18/88
Visit to hospital (%)	13.9	7.0	31.5***	43.9***
Number of symptoms	5.9 (2.83)	5.0(2.7)	6.9 (2.9)***	7.0 (3.1)***

* indicates p <0.1 ** <0.05 ***<0.01 when comparing to short covid. Comparison are performed with respect to the "short duration" group. Oneway ANOVA test are performed for continuous variables and chi square tests are performed when comparing proportions.
ϰ IMD information is only available for app users from the UK who have entered a complete post code
Acronyms: GB – Great Britain / SE – Sweden / US – United States / IMD – Index of multiple deprivation

Fatigue (97.7%) and headache (91.2%) were the most reported symptoms in those with Long-COVID, followed by anosmia and lower respiratory symptoms. Notably, while fatigue was reported continuously, other symptoms such as headache are reported intermittently (Figure 2, supplementary Table s1). To get better insight into the reported symptoms, we additionally analysed free text responses which were more common in Long-COVID cases (81%) than Short-COVID (45%). Cardiac symptoms (palpitations, tachycardia) were over-represented in the LC28 group (6.1%) compared to in short-COVID (0.5%) ($p<0.0005$) as were concentration or memory issues (4.1% vs 0.2%) ($p<0.0005$), tinnitus and earache (3.6% vs 0.2% $p<0.0005$) and peripheral

114 neuropathy symptoms (pins and needles and numbness) (2% vs 0.5%) (p=0.004). Most of these
115 symptoms were reported for the first time 3-4 weeks post symptom onset.
116
117 We examined whether there were different types of symptomatology within Long-COVID. We
118 found two main patterns: those reporting exclusively fatigue, headache and upper respiratory
119 complaints (shortness of breath, sore throat, persistent cough and loss of smell) and those with
120 multi-system complaints including ongoing fever and gastroenterological symptoms
121 (Supplementary figure 3). In the individuals with long duration (LC28), ongoing fever OR 2.16
122 [1.50 - 3.13] and skipped meals OR 2.52 [1.74; 3.65] were strong predictors of a subsequent
123 hospital visit. Details of the frequency of symptoms persisting beyond 28 and 56 days after
124 disease onset are provided in Supplementary table 3.
125
126

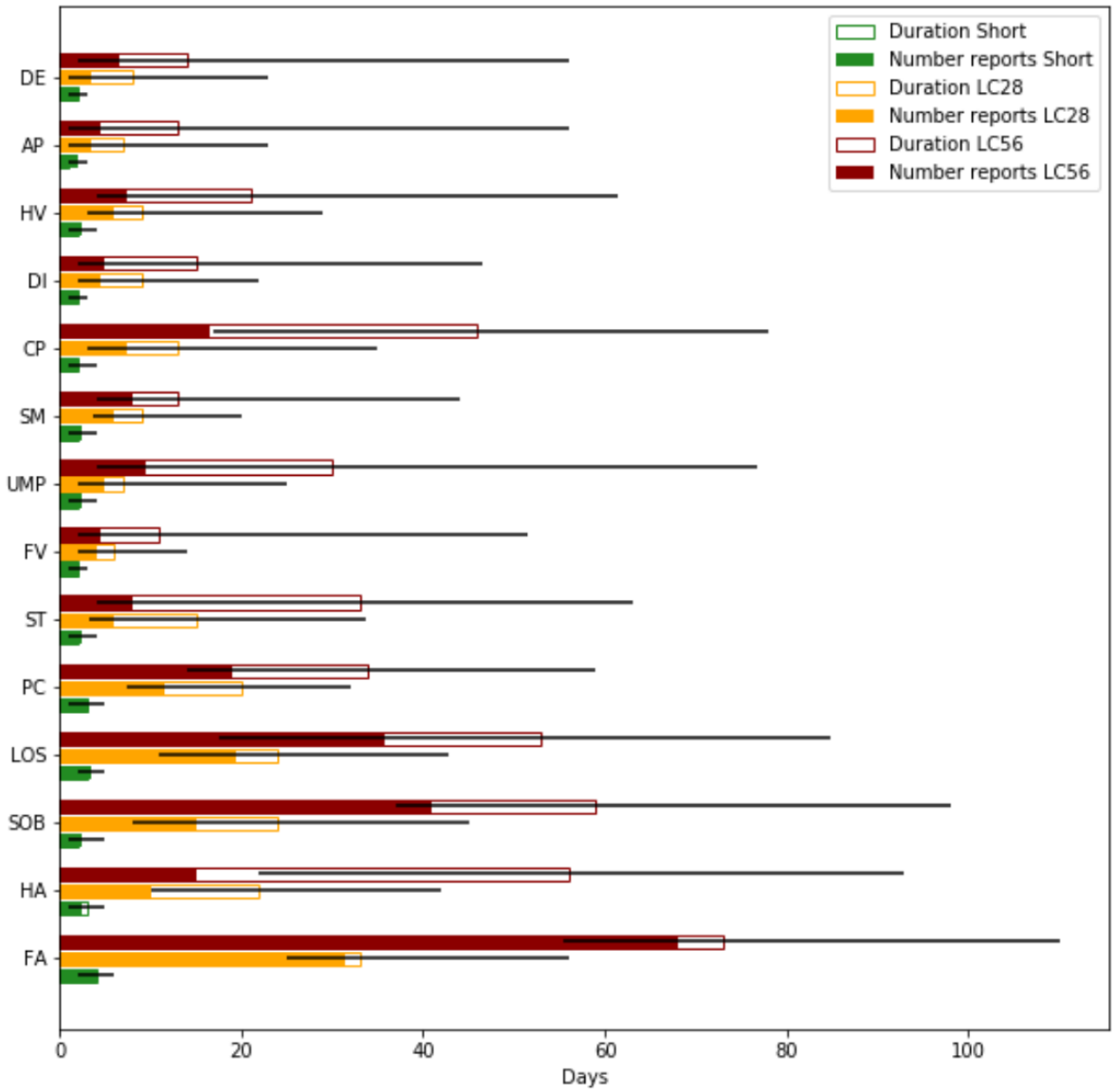


Figure 2 : Symptoms by duration. For each symptom (ordered from top to bottom by increasing frequency of occurrence) the median duration of report is presented by the total (hollowed) bar height, with associated interquartile range represented by the black line, for the short, LC28 and LC56 durations. The filled bars represent the number of times a report has been given. This highlights the differences in the symptoms in terms of their intermittence throughout the course of the disease. (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)

Individuals with long-COVID were more likely to report relapses (16.0%), compared to those not reporting long symptom duration (8.4%) ($p<0.0005$). In comparison, in the matched group of SARS-CoV2 negative tested individuals, a new bout of illness was reported in 11.5% of cases.

Relapse in the context of long-COVID was longer than in the matched controls (median = 9 [5-18] vs 6 [4-10] days).

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

When analysed individually after adjusting for age and sex, every symptom in isolation was positively predictive of longer illness duration. The five symptoms experienced during the first week most predictive of Long-COVID were: fatigue OR=2.83 [2.09; 3.83], headache OR=2.62 [2.04;3.37], dyspnoea 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 3). Similar patterns were observed in men and women. In adults aged over 70, loss of smell (which is less common) was the most predictive of long-COVID OR=7.35 [1.58 - 34.22] before fever OR=5.51 [1.75 - 17.36] and hoarse voice OR=4.03[1.21,13.42] (Supplementary figures 4). Plotting frequency of co-occurrence of symptoms in short-COVID versus long-COVID further illustrates the importance of early multi-symptom involvement (Figure 3c).

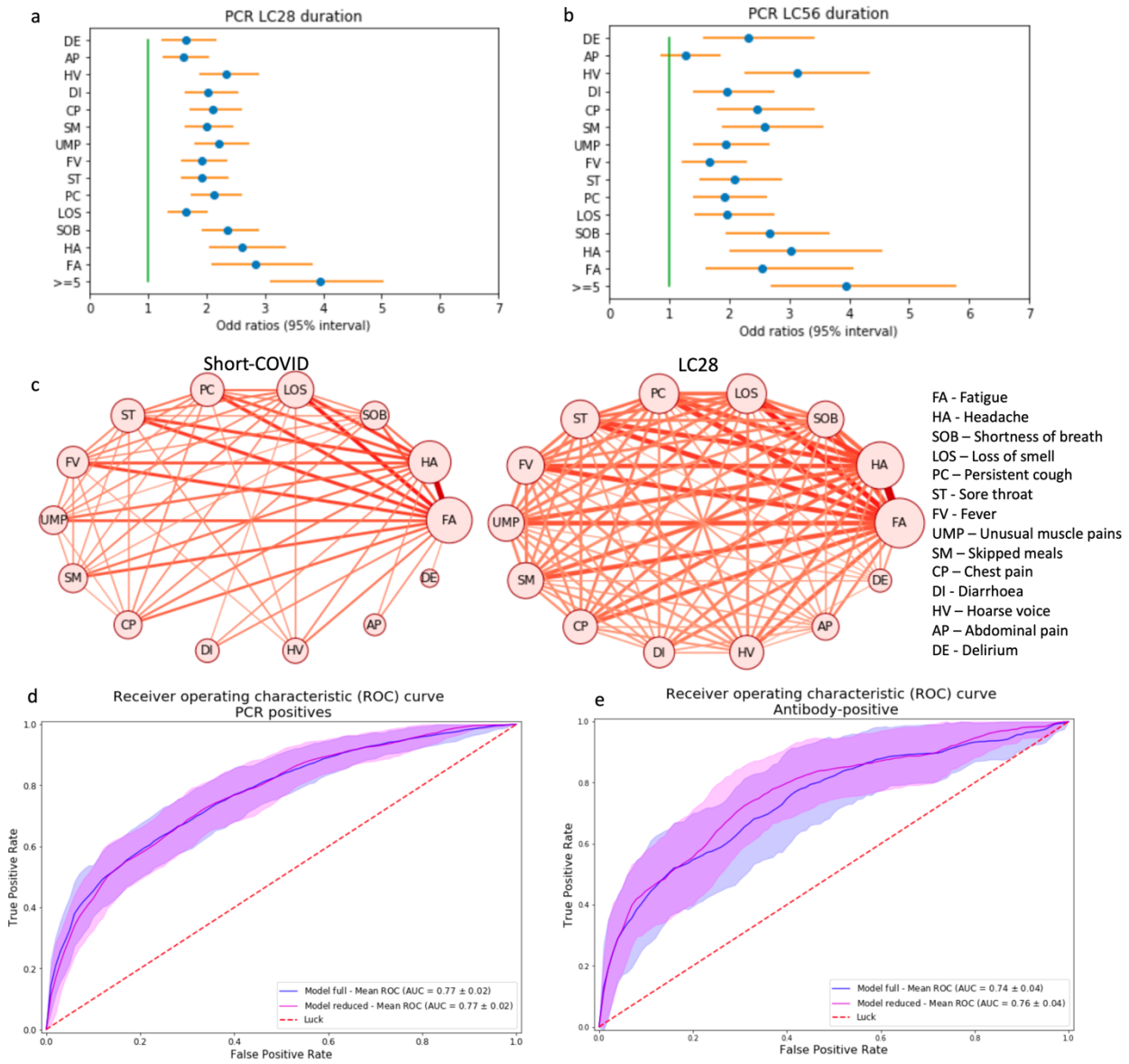


Figure 3: Symptom correlates of long-COVID for LC28 (a) and LC56 (b) with correction for age and gender. c) Co-occurrence network of symptom pairs with the frequency of symptom report as the size of the node and the likelihood of symptom pair co-occurrence by the weight of the edge linking them. Edges 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)

We further created Random Forest Prediction models using a combination of the first week's symptom reporting, personal characteristics and comorbidities. Using all features, the average

ROC AUC was 76.7% (SD=2.5) (Figure 3d) in the classification between short-COVID and LC28. The strongest predictor was age (29.2 %) followed by the number of symptoms during the first week (16.3%) and BMI (10.8%) while gender (3.7%) was ranked 6th shortly after hoarse voice (4.1) and shortness of breath (3.8). All individual symptoms, except abdominal pain and confusion, surpassed the comorbidity features. The ranking of feature importance was relatively similar across specific age group models. However, in the over 70s group it appeared that early features such as fever, loss of smell and comorbidities (especially heart and lung disease) were important, and thus could be considered 'red flags' in older adults (supplementary figure 6).

We simplified the prediction model to include only symptom number in the first week with age, and sex in a logistic regression model and we obtained ROC AUC of 76.7% (SD 2.5) (Figure 3d). When optimising the balance between false positives and false negatives, we obtained a specificity of 73.4% (SD 9.7) and a sensitivity of 68.7% (SD 9.9).

Key predictive findings of our analysis were validated in an independent dataset of 2472 individuals who reported testing antibody positive for SARS-CoV2 from 2 weeks after symptom onset where, again, the number of symptoms in the first week of illness was the strongest predictor of long-COVID, OR=5.12 [95% CI 3.65; 7.19]. The simple prediction model for Long-COVID, trained on the PCR positive cohort and including number of distinct symptoms experienced during the first week, age, and sex was similarly predictive of LC28, with a ROC AUC of 76.3% (SD=3.7%) (Figure 5 - b).

While this study provides important insights into the disease presentation, any generalisation should be considered carefully. Our study was limited by being confined to app users who were disproportionately female and under-represented those >70years which could increase or decrease our estimate of the extent of Long-COVID respectively. Applying a weighting following the UK population (see Supplementary Methods) the estimated proportion of people experiencing symptomatic COVID-19 going on to suffer long-COVID were similar: 14.5%, 5.1% and 2.2% for 4, 8- and 12-weeks duration respectively. These estimates may still be conservative: whilst estimates could be inflated due to PCR testing in the first wave being restricted to those more severely unwell, or if regular logging may have encouraged more symptoms to be noticed, Long-COVID may here be underestimated if individuals with prolonged symptoms were more likely to stop logging symptoms on the app. We had insufficient numbers to explore risk factors for disease over 2 months and were unable to analyse the impact of ethnicity due to incomplete data in this subset. In addition, while the list of symptoms on the app is necessarily non-exhaustive, the analysis of the free-text responses allowed us to highlight other symptoms present in long-COVID, such as cardiac and neurological manifestations starting generally a few weeks after the symptom onset. With emerging evidence of ongoing myocardial inflammation and change in neurological markers [8,9] associated with COVID-19, this calls for specific studies of cardiac and neurological longer-term sequelae of COVID-19.

217 At the population level, it is critical to quantify the burden of Long-COVID to better assess its
218 impact on the healthcare system and appropriately distribute resources. In our study,
219 prospective logging of a wide range of symptoms allowed us to conclude that the proportion of
220 people with symptomatic COVID-19 who experience prolonged symptoms is considerable, and
221 relatively stable across three countries with different cultures. Whether looking at a four-week
222 or an eight-week threshold for defining long duration, those experiencing Long-COVID were
223 consistently older, more female and were more likely to require hospital assessment than in the
224 group reporting symptoms for a short period of time. The multi-system nature of the initial
225 disease in Long-COVID was illustrated by the importance of the number of symptoms, and co-
226 occurrence networks showing that those going on to experience long-COVID had greater
227 number of concurrent symptoms, therefore supporting the need for holistic support [10]. While
228 asthma was not reported as a factor of risk for hospitalisation in [11], its association with Long-
229 COVID warrants further investigation.
230
231 We found early disease features were predictive of duration. With only three features - number
232 of symptoms in the first week, age and sex, a simple scoring derived from a logistic regression
233 was able to accurately distinguish individuals with Long-COVID from those with short duration.
234 Importantly, the model generalised well to the population reporting antibody testing. This
235 important information could feature in highly needed targeted education material for both
236 patients and healthcare providers. Moreover, the method could help determine at-risk groups
237 and could be used to target early intervention trials of treatment (for example, of

238 dexamethasone[12] and remdesivir [13]) and clinical service developments to support rehabilitation

239 in primary and specialist care [14] to alleviate Long-COVID and facilitate timely recovery.

240

241 **Ethics:** In the UK, the App Ethics has been approved by KCL ethics Committee REMAS ID 18210,
242 review reference LRS-19/20-18210 and all subscribers provided consent. In Sweden, ethics
243 approval for the study was provided by the central ethics committee (DNR 2020-01803).
244 **Funding:** Zoe provided in kind support for all aspects of building, running and supporting the
245 app and service to all users worldwide. Support for this study was provided by the NIHR-funded
246 Biomedical Research Centre based at GSTT NHS Foundation Trust. This work was supported by
247 the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for
248 Value Based Healthcare.
249 Investigators also received support from the Wellcome Trust, the MRC/BHF, Alzheimer's
250 Society, EU, NIHR, CDRF, and the NIHR-funded BioResource, Clinical Research Facility and BRC
251 based at GSTT NHS Foundation Trust in partnership with KCL. ATC was supported in this work
252 through a Stuart and Suzanne Steele MGH Research Scholar Award. CM is funded by the
253 Chronic Disease Research Foundation and by the MRC AimHy project grant. LHN, DAD, ADJ,
254 ADS, CG, WL are supported by the Massachusetts Consortium on Pathogen Readiness
255 (MassCPR) and Mark and Lisa Schwartz. JM was partially supported by the European
256 Commission Horizon 2020 program (H2020-MSCA-IF-2015-703787). The work performed on the
257 Swedish study is supported by grants from the Swedish Research Council, Swedish Heart-Lung
258 Foundation and the Swedish Foundation for Strategic Research (LUDC-IRC 15-0067).
259 **Competing interests:**
260 Zoe Global Limited co-developed the app *pro bono* for non-commercial purposes. Investigators
261 received support from the Wellcome Trust, the MRC/BHF, EU, NIHR, CDRF, and the NIHR-
262 funded BioResource, Clinical Research Facility and BRC based at GSTT NHS Foundation Trust in
263 partnership with KCL. RD, JW, JCP, AM and SG work for Zoe Global Limited and TDS and PWF
264 are consultants to Zoe Global Limited. LHN, DAD,JM, PWF and ATC previously participated as
265 investigators on a diet study unrelated to this work that was supported by Zoe Global Ltd.
266 **Data and materials availability:** Data used in this study is available to bona fide researchers
267 through UK Health Data Research using the following link

268 https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259

269
270
271 **References**
272
273 1. Wiersinga, W. J., Rhodes, A., Cheng, A. C., Peacock, S. J. & Prescott, H. C.
274 Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019
275 (COVID-19): A Review. *JAMA - Journal of the American Medical Association* vol. 324 782–
276 793 (2020).
277 2. Sheehy, L. M. Considerations for postacute rehabilitation for survivors of COVID-19.

Journal of Medical Internet Research vol. 22 (2020).

3. Alwan, N. A. A negative COVID-19 test does not mean recovery. *Nat. 2020 5847820* (2020).
4. Menni, C. *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* **26**, 1037–1040 (2020).
5. Drew, D. A. *et al.* Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science (80-.).* eabc0473 (2020) doi:10.1126/science.abc0473.
6. Sudre, C. H. *et al.* Symptom clusters in Covid19: A potential clinical prediction tool from the COVID Symptom study app. *medRxiv* 2020.06.12.20129056 (2020) doi:10.1101/2020.06.12.20129056.
7. Cirulli, E. T. *et al.* Long-term COVID-19 symptoms in a large unselected population. *medRxiv* (2020) doi:10.1101/2020.10.07.20208702.
8. Mitrani, R. D., Dabas, N. & Goldberger, J. J. COVID-19 cardiac injury: Implications for long-term surveillance and outcomes in survivors. *Hear. Rhythm* (2020) doi:10.1016/j.hrthm.2020.06.026.
9. Padda, I., Khehra, N., Jaferi, U. & Parmar, M. S. The Neurological Complexities and Prognosis of COVID-19. *SN Compr. Clin. Med.* (2020) doi:10.1007/s42399-020-00527-2.
10. *Living with Covid19*. https://evidence.nihr.ac.uk/themedreview/living-with-covid19/ (2020) doi:10.3310/themedreview_41169.
11. Robinson, L. B. *et al.* Title: COVID-19 severity in asthma patients: A multi-center matched cohort study. *medRxiv* 2020.10.02.20205724 (2020) doi:10.1101/2020.10.02.20205724.
12. Mahase, E. Covid-19: Demand for dexamethasone surges as RECOVERY trial publishes preprint. *BMJ* **369**, m2512 (2020).
13. Beigel, J. H. *et al.* Remdesivir for the Treatment of Covid-19 — Preliminary Report. *N. Engl. J. Med.* (2020) doi:10.1056/nejmoa2007764.
14. Greenhalgh, T., Knight, M., A'Court, C., Buxton, M. & Husain, L. Management of post-acute covid-19 in primary care. *BMJ* **370**, (2020).
15. Drew, D. A. *et al.* Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science (80-.).* (2020).
16. Spiel, C. *et al.* A Euclidean distance-based matching procedure for non randomized comparison studies. *Eur. Psychol.* **13**, 180–187 (2008).

322
323 **Online methods**
324
325 ## Methods

326 **Dataset:**

327 Data used in this study were acquired through the COVID 19 Symptom Study app, a mobile
328 health application developed by Zoe Global Limited with input from physicians and scientists at
329 King's College London, Massachusetts General Hospital, Lund and Uppsala Universities[15]. The
330 app, which collects data on personal characteristics and through prospective logging of
331 symptoms, was launched in the UK, the US and Sweden between 24th March (UK) and 30th April
332 (Sweden), and rapidly reached over 4 million users. This study focuses on 4182 users who
333 reported testing positive to SARS-CoV2 by PCR swab test and had a disease onset between 25th
334 March 2020 and 30th June 2020, for whom onset date matched with date of test and duration
335 of symptoms could be estimated (Supplementary figure 1 presents a flowchart of study
336 inclusion). We repeated analyses in an independent subgroup of 2472 app users who reported
337 positive testing for antibodies against SARS-CoV2 more than 2 weeks after symptom onset, but
338 without swab test results (Supplementary Figure 1).
339
340 To understand how the relapse rate compared to a comparable population not suffering from
341 COVID-19, we selected an additional matched sample from all app users meeting study
342 inclusion criteria but testing negative by PCR swab test, choosing for each COVID+ case the
343 individual from the negative group with the smallest Euclidean distance based on sex, age and
344 BMI [16].

345

346

347 **Definitions**

348 Onset of disease was defined as the first day of reporting at least one symptom lasting more
349 than one day.

350 Disease end was defined as the last day of unhealthy reporting before reporting healthy for
351 more than one week or the last day of reporting with less than 5 symptoms before ceasing
352 using the app. For the participants included who ceased using the app with a cumulative
353 number of symptoms of less than 5, disease end was considered as the last log.

354 Relapse was defined as 2 or more days of symptoms within a 7-day window after one week of
355 healthy logging, given initial symptoms close to a positive swab test.

356

357 Long-COVID was defined as symptoms persisting for a period of more than 4 weeks (28 days
358 LC28), more than 8 weeks (56 days, LC56) or more than 12 weeks (LC94) between symptom
359 onset and end, while short duration was defined as the interval between symptom onset and
360 end of less than 10 days, without a subsequent relapse (Short-COVID).

361

362 **Inclusion/Exclusion criteria**

363 To be included in the subsequent analysis, users of the COVID Symptoms Study app were
364 selected based on the following criteria:

365 *Inclusion criteria*: Age >=18 yrs; reporting a positive SARS-CoV-2 swab test (PCR) confirming the
366 diagnosis of COVID-19; disease onset between 14 days before and 7 days after the test date,
367 and before the 30th June 2020 (to limit right censoring).
368 *Exclusion criteria*: individuals who started app reporting when already unwell; users reporting
369 exclusively healthy throughout the study period; users with gaps of more than 7 days after an
370 unhealthy report and not reporting any hospital visit (to account for gaps due to
371 hospitalisation). In addition, individuals reporting for less than 28 days but reporting more than
372 5 symptoms at their last log were excluded, as duration could not be ascertained.
373
374
375 **Statistical testing and modelling**
376 Data collected prospectively until 02 September were included to allow sufficient time to
377 ascertain duration. Univariate and multivariate logistic regression was used to assess symptoms
378 associated with short- and long-COVID respectively, adjusting for sex and age, using
379 Statsmodels v0.11.1 Python3.7. Separate models were fitted to subgroups stratified by sex and
380 age (18-49; 50-69; >70). For analysis of relapse, results were compared for the long-COVID
381 group to the whole sample using Mann Whitney U test.
382
383 We used a K-mode clustering analysis to investigate whether there was evidence of different
384 sub-types of long-COVID, using the kmode package v0.10.2. Number of ideal symptom clusters
385 was obtained via a silhouette analysis with dice distance metrics. Differences between long-

386 COVID and short-COVID were visualised using a co-occurrence network (networkx for
387 visualisation), applying a 10% threshold to remove rare edges to aid visualisation.
388
389 Finally, to create a predictive model for long-COVID, we used sklearn v0.22.2.post1 package,
390 training random forest classifiers with stratified repeated cross-validation (10 times, 5 folds)
391 with hyperparameter grid search including, as features, information available during the first
392 week of illness, reported comorbidities (asthma, lung disease, heart disease, kidney disease and
393 diabetes) and personal characteristics (BMI, age, sex). In addition to a global consideration of
394 the studied sample population, separate models stratified by age were also entrained using a
395 similar cross-validation setting (hyperparameter search and stratified sampling). After running
396 the cross-validation for each model structure (50 times), the feature importance was averaged
397 across the different repeated folds. A final simplistic model using the key personal
398 characteristics and number of first week symptoms was further tested.
399 Using only 3 features, a logistic regression model was then assessed using the same
400 stratification and cross-validation.
401
402 To assess performance on the test dataset (antibody positive), cross-validation was also
403 performed to obtain an indication of the variability in performance using models that were
404 trained on the whole PCR positive sample.
405 For the reduced logistic regression model, the score was given by the following formula:
406

S = 0.259503 * NumberSymptoms + 0.055457 * Age -0.633310 * Sex where sex is encoded as 1 – Female / 2 – Male

Matching with negative sample:

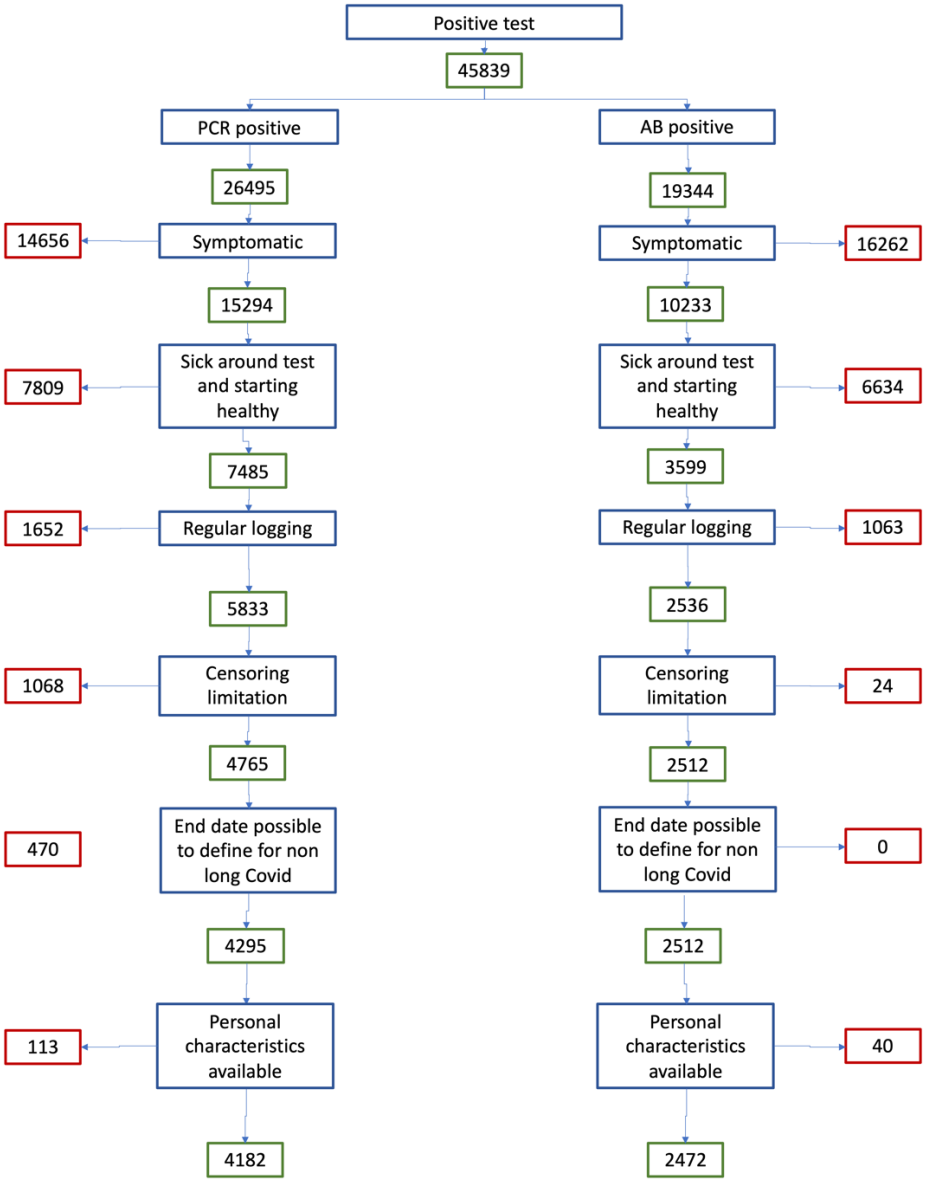
The negative cases selected for matching followed the same inclusion rules and were matched to the positive samples using the minimum Euclidean distance between the vectors of features created by age, BMI and sex. Sex feature was multiplied by 100 to ensure balance between feature strength.

Rebalancing to UK population demographics

Lastly, the rebalancing with respect to the UK population was performed by reweighting the age/sex proportions of LC28 in the studied sample by the one of the UK population based on census data from 2018. The weighting per age group is described in the table below

	Female	Male
18-49	0.263	0.264
50-69	0.156	0.150
>=70	0.093	0.075

455 **Supplementary tables and figures**
456
457 Supplementary Figure 1 presents a flowchart of the study inclusion. Individuals reporting
458 symptoms for at most 1 day are considered for the purpose of this analysis to be asymptomatic.
459 We further excluded users who joined the app already unhealthy, for which the onset of
460 disease was not calculable. Of the remainder, we further excluded those who only reported
461 intermittent unhealthy report and restricted to individuals reporting prospective symptoms at
462 least once a week over the course of the disease. The left part of the diagram represents the
463 inclusion flowchart for the individuals reporting a positive swab test while the right side reflects
464 the inclusion pathway for individuals with antibody positive test only.
465



466
467
468

Supplementary Table 1 Compares the frequency (over the whole disease course), existence in the first week, duration (difference between first and last day at which a symptom is reported) and number of reports of a given symptom for Short-COVID, LC28 and LC56. Symptoms are ordered by overall frequency in the LC28 group. Duration and number of reports are presented as median [IQR] while the frequencies are presented as percentages

	Occurrence overall			Occurrence first week			Duration			Number of reports		
	Short	L28	L56	Short	L28	L56	Short	L28	L56	Short	L28	L56
FA	76.1%	97.7%	96.8%	75.9%	89.1%	87.8%	4 [2 ; 6]	33 [25 ; 56]	73 [55.5 ; 110]	4 [2 ; 6]	31.5 [19.5 ; 54.5]	68 [39 ; 118.5]
HA	65.6%	91.2%	93.7%	65.4%	80.6%	81.5%	3 [1 ; 5]	22 [10 ; 42]	56 [22 ; 93]	2.5 [1.5 ; 4]	10 [4.5 ; 19.5]	15 [7 ; 34.5]
SOB	29.4%	70.8%	75.7%	29.2%	48.4%	51.3%	2 [1 ; 5]	24 [8 ; 45]	59 [37 ; 98]	2.5 [1.5 ; 4.5]	15 [5.5 ; 34.5]	41 [15.75 ; 80.25]
LOS	49.2%	72.0%	75.1%	49.0%	56.8%	58.2%	3 [2 ; 5]	24 [11 ; 42.75]	53 [17.5 ; 84.75]	3.5 [2 ; 5]	19.5 [9 ; 34]	35.75 [11.63 ; 68.88]
PC	40.5%	68.6%	62.4%	40.4%	57.0%	52.9%	3 [1 ; 5]	20 [7.5 ; 32]	34 [14 ; 59]	3 [1.5 ; 4.5]	11.5 [4.5 ; 22.5]	19 [7 ; 40.5]
ST	41.2%	67.0%	72.5%	41.2%	53.6%	54.0%	2 [1 ; 4]	15 [3.25 ; 33.75]	33 [4 ; 63]	2.5 [1.5 ; 4]	6 [3 ; 12]	8 [3 ; 17]
FV	36.1%	62.9%	58.7%	36.1%	50.5%	45.5%	2 [1 ; 3]	6 [2 ; 14]	11 [2 ; 51.5]	2 [1.5 ; 3]	4 [2 ; 7.75]	4.5 [2 ; 11.5]
UMP	29.2%	64.0%	64.6%	29.2%	47.5%	43.4%	2 [1 ; 4]	7 [2 ; 25]	30 [4 ; 76.75]	2.5 [1.5 ; 3.5]	5 [2 ; 11]	9.5 [3 ; 30.63]
SM	29.9%	59.5%	66.7%	29.9%	46.6%	52.4%	2 [1 ; 4]	9 [3.75 ; 20]	13 [4 ; 44]	2.5 [1.5 ; 4]	6 [3 ; 14]	8 [3.5 ; 18.88]
CP	28.2%	60.0%	63.0%	28.1%	42.5%	45.0%	2 [1 ; 4]	13 [3 ; 35]	46 [17 ; 78]	2 [1.5 ; 4]	7.5 [2.5 ; 16.5]	16.5 [7.25 ; 42]
DI	20.4%	51.1%	54.5%	20.1%	34.6%	33.3%	2 [1 ; 3]	9 [2 ; 22]	15 [2 ; 46.5]	2 [1 ; 3]	4.5 [2 ; 9]	5 [2 ; 12.25]
HV	23.0%	53.0%	61.4%	22.9%	41.4%	47.6%	2 [1 ; 4]	9 [3 ; 29]	21 [4 ; 61.5]	2.5 [1.5 ; 4]	6 [3 ; 14.5]	7.5 [3.5 ; 20.625]
AP	17.2%	44.1%	49.2%	17.0%	26.0%	22.2%	1 [1 ; 3]	7 [1 ; 23]	13 [1 ; 56]	2 [1 ; 3]	3.5 [1.5 ; 8.5]	4.5 [2 ; 9.5]
DE	11.8%	30.3%	38.6%	11.7%	19.0%	24.3%	2 [1 ; 3]	8 [1 ; 23]	14 [2 ; 56]	2 [1 ; 3.5]	3.5 [1.5 ; 9]	6.5 [2 ; 16.5]

Supplementary Table 2 – Odd ratio of LC28 and LC56 per decade when compared to 18-30 category separating male and females

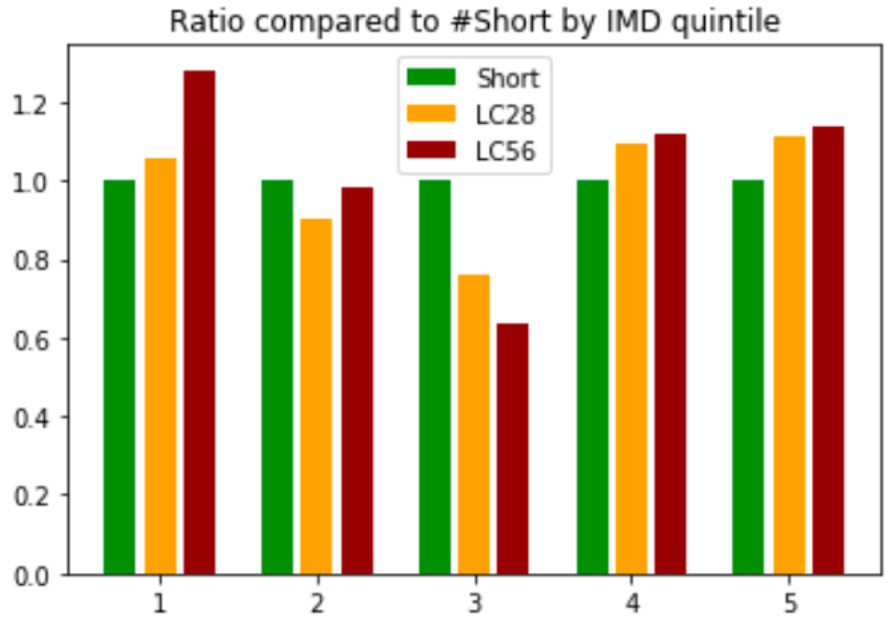
	Female		Male	
	LC28	LC56	LC28	LC56
30-40	2.11 [1.38 ; 3.23]	2.19 [1.02 ; 4.73]	2.62 [1.09 ; 6.27]	4.12 [0.49 ; 34.67]
40-50	4.35 [2.9 ; 6.53]	4.14 [2.02 ; 8.52]	2.24 [0.9 ; 5.62]	3.52 [0.39 ; 31.91]

50-60	8.03 [5.37 ; 12]	8.61 [4.33 ; 17.1]	6.65 [2.8 ; 15.8]	11.49 [1.44 ; 91.29]
60-70	6.53 [4.02 ; 10.6]	7.2 [3.31 ; 15.69]	7.5 [2.98 ; 18.9]	14 [1.68 ; 116.51]
>=70	5.46 [2.51 ; 11.85]	7.71 [2.6 ; 22.83]	8.27 [2.8 ; 24.39]	18.56 [1.99 ; 173.3]

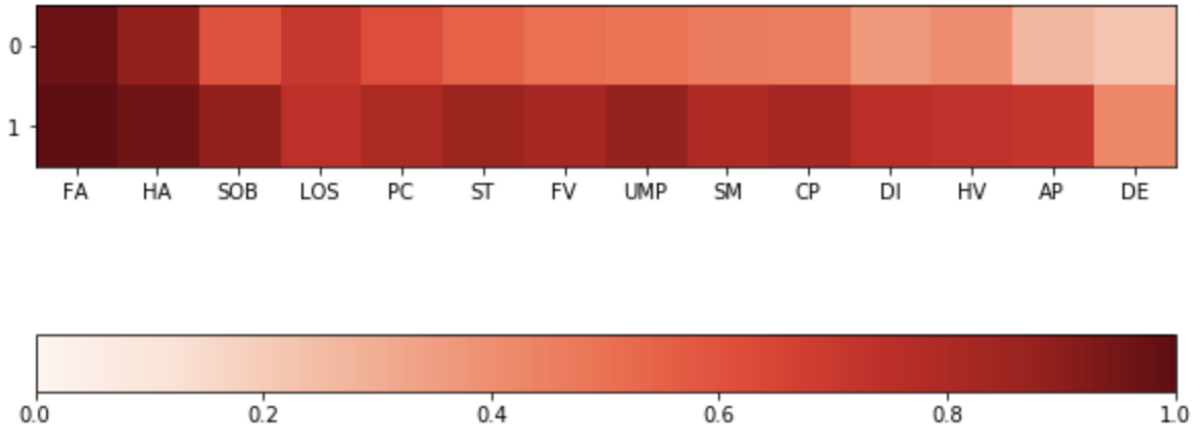
Supplementary Table 3 – Indication of the frequency of report of symptoms beyond 28 days and beyond 56 days in the LC28 and the LC56 groups.

	Reported beyond 28 days in LC28	Reported beyond 56 days in LC56
FA	0.68	0.73
HA	0.40	0.53
SOB	0.37	0.48
LOS	0.39	0.39
PC	0.27	0.22
ST	0.27	0.27
FV	0.12	0.16
UMP	0.20	0.30
SM	0.13	0.19
CP	0.23	0.31
DI	0.15	0.20
HV	0.17	0.22
AP	0.15	0.23
DE	0.11	0.15

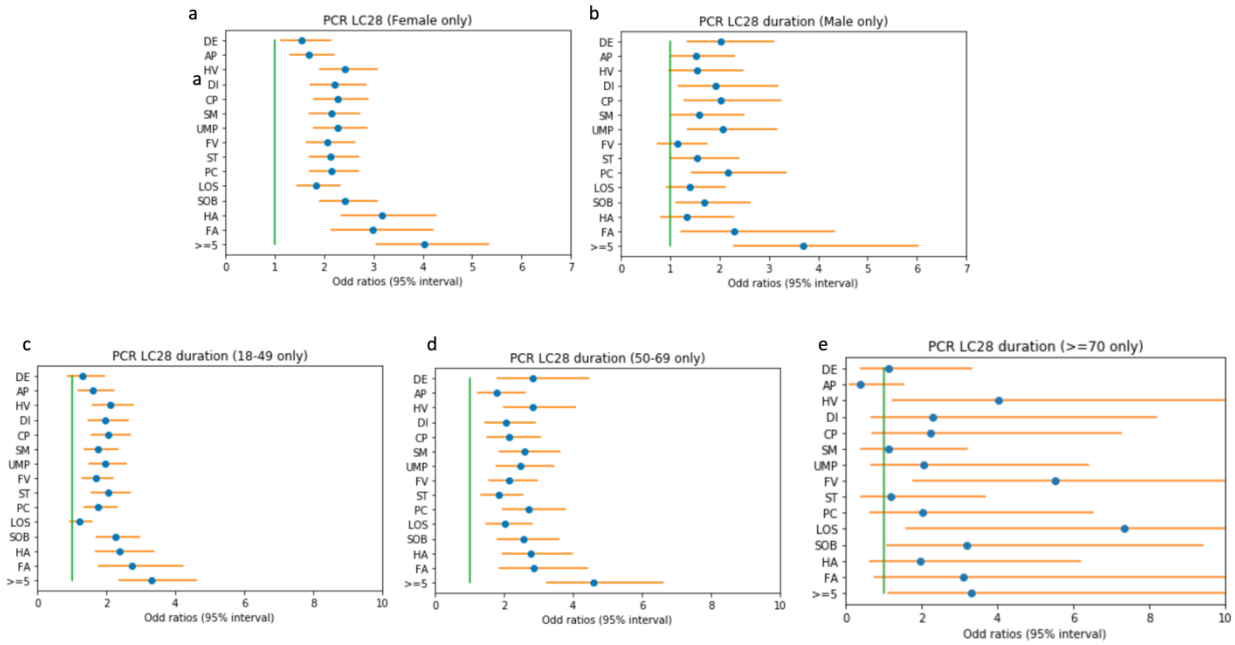
Supplementary Figure 2: Ratio of LC28 and LC56 vs short-COVID by IMD quintile



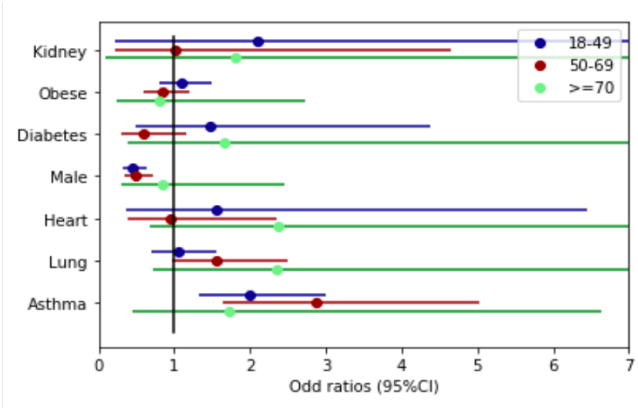
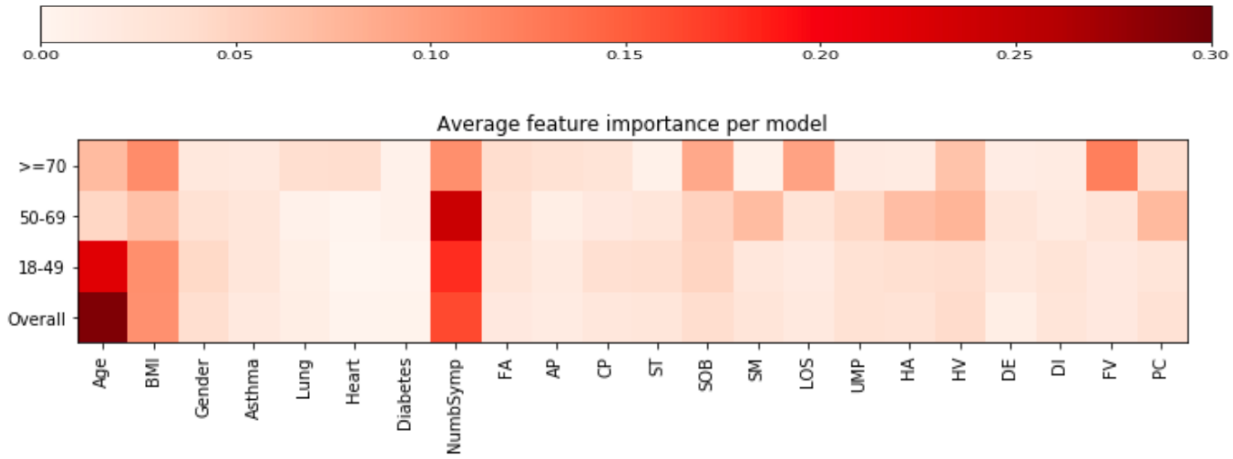
Supplementary Figure 3 – clustering of symptoms in the Long-COVID group indicating a common strong higher airways component with fatigue, headache and loss of smell for both groups and a more multi system presentation for the second group. Colouring presents the frequency of reporting of a given symptom. 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



Supplementary Figure 4. Odd ratios of continuing into long Covid 28 when presenting a given symptom during the first week correcting for age and gender (if necessary) in different subgroups female(a), male (b), 18-49 (c), 50-69 (d), >=70 (e). 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



Supplementary Figure 5 – Odd ratio for the risk of developing Long Covid 28 for each comorbidity or risk factors correcting for age and gender in each age group,



Supplementary Figure 6: Comparison of mean feature importance (proportion ranging from 0 to 1) for the cross-validated random forest models across the different age groups when considering personal characteristics and presented symptoms during the first week of the disease. Abbreviations - (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)