

# **Persistent symptoms following SARS-CoV-2 infection in a random community sample of 508,707 people**

Matthew Whitaker<sup>a,b\*</sup>, Joshua Elliott<sup>a\*</sup>, Marc Chadeau-Hyam<sup>a,b</sup>, Steven Riley<sup>a,c,d</sup>, Ara Darzi<sup>e,f</sup>,  
Graham Cooke<sup>f,g,h†</sup>, Helen Ward<sup>c,f,g†</sup>, Paul Elliott<sup>a,b,f,g,i,j†</sup>

<sup>a</sup> School of Public Health, Imperial College London, UK

<sup>b</sup> MRC Centre for Environment and Health, Imperial College London, UK

<sup>c</sup> MRC Centre for Global infectious Disease Analysis, Imperial College London, UK

<sup>d</sup> Abdul Latif Jameel Institute for Disease & Emergency Analytics, Imperial College London, UK

<sup>e</sup> Institute of Global Health Innovation at Imperial College London, UK

<sup>f</sup> Imperial College Healthcare NHS Trust, UK

<sup>g</sup> National Institute for Health Research Imperial Biomedical Research Centre, UK

<sup>h</sup> Department of Infectious Disease, Imperial College London, London, UK

<sup>i</sup> Health Data Research (HDR) UK London at Imperial College, UK

<sup>j</sup> UK Dementia Research Institute at Imperial College, UK

\*Joint first author

†Equal contribution

## **Address for Correspondence:**

Professor Paul Elliott, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, St Mary's campus, London W2 1PG, UK

Email: [p.elliott@imperial.ac.uk](mailto:p.elliott@imperial.ac.uk)

## **Keywords**

COVID-19, symptoms, Long COVID

Link to data tables included in this report is [here](#).

The study materials and questionnaires used in this study can be found [here](#).

## **Research in context**

### **Evidence before this study**

Recent systematic reviews have documented the wide range of symptoms and reported prevalence of persistent symptoms following COVID-19. A dynamic review of Long COVID studies (NIHR Evidence) in March 2021 summarised the literature on the prevalence of persistent symptoms after acute COVID19, and reported that most studies (14) were of hospitalised patients, with higher prevalence of persistent symptoms compared with two community-based studies. There was limited evidence from community studies beyond 12 weeks. Another systematic review reported a median of over 70% of people with symptoms lasting at least 60 days. A review of risk factors for Long COVID found consistent evidence for an increased risk amongst women and those with high body mass index (BMI) but inconsistent findings on the role of age and little evidence concerning risks among different socioeconomic or ethnic groups which are often not well captured in routine healthcare records. Long COVID is increasingly recognised as heterogenous, likely underpinned by differing biological mechanisms, but there is not yet consensus on defining subtypes of the condition.

### **Added value of this study**

This community-based study of over half a million people was designed to be representative of the adult population of England. A random sample of adults ages 18 years and above registered with a GP were invited irrespective of previous access to services for COVID-19, providing an estimate of population prevalence that was representative of the whole population. The findings show substantial declines in symptom prevalence over the first 12 weeks following Covid-19, reported by nearly one fifth of respondents, of whom over a third remained symptomatic at 12 weeks and beyond, with little evidence for decline thereafter.

Risk factors identified for persistent symptoms (12 weeks or more) suggestive of Long COVID confirm some previous findings - an increased risk in women, obese and overweight individuals

and those hospitalised for COVID-19, with strong evidence for an increasing risk with age. Additional evidence was found for an increased risk in those with lower income, smoking or vaping and healthcare or care home workers. A lower risk was found in those of Asian ethnicity.

Clustering identified two distinct groups of individuals with different symptom profiles at 12 weeks, highlighting the heterogeneity of clinical presentation. The smaller cluster had higher prevalence of respiratory and related symptoms, while for those in the larger cluster tiredness was the dominant symptom, with lower prevalence of organ-specific symptoms.

### **Implications of available evidence**

There is a high prevalence of persistent symptoms beyond 12 weeks after acute COVID-19, with little evidence of decline thereafter. This highlights the needs for greater support for patients, both through specialised services and, for those from low-income settings, financial support. The understanding that there are distinct clusters of persistent symptoms, the most common of which is dominated by fatigue, is important for the recognition and clinical management of the condition outside of specialised services.

## **Summary**

### **Background**

Long COVID, describing the long-term sequelae after SARS-CoV-2 infection, remains a poorly defined syndrome. There is uncertainty about its predisposing factors and the extent of the resultant public health burden, with estimates of prevalence and duration varying widely.

### **Methods**

Within rounds 3–5 of the REACT-2 study, 508,707 people in the community in England were asked about a prior history of COVID-19 and the presence and duration of 29 different symptoms. We used uni- and multivariable models to identify predictors of persistence of symptoms (12 weeks or more). We estimated the prevalence of symptom persistence at 12 weeks, and used unsupervised learning to cluster individuals by symptoms experienced.

### **Findings**

Among the 508,707 participants, the weighted prevalence of self-reported COVID-19 was 19.2% (95% CI: 19.1,19.3). 37.7% of 76,155 symptomatic people post COVID-19 experienced at least one symptom, while 14.8% experienced three or more symptoms, lasting 12 weeks or more. This gives a weighted population prevalence of persistent symptoms of 5.75% (5.68, 5.81) for one and 2.22% (2.1, 2.26) for three or more symptoms. Almost a third of people (8,771/28,713 [30.5%]) with at least one symptom lasting 12 weeks or more reported having had severe COVID-19 symptoms (“significant effect on my daily life”) at the time of their illness, giving a weighted prevalence overall for this group of 1.72% (1.69,1.76). The prevalence of persistent symptoms was higher in women than men (OR: 1.51 [1.46,1.55]) and, conditional on reporting symptoms, risk of persistent symptoms increased linearly with age by 3.5 percentage points per decade of life. Obesity, smoking or vaping, hospitalisation, and deprivation were also associated with a higher probability of persistent symptoms, while Asian ethnicity was associated with a lower probability. Two stable clusters were identified based on symptoms that persisted for 12 weeks or more: in the largest cluster, tiredness predominated, while in the second there was a high prevalence of respiratory and related symptoms.

### **Interpretation**

A substantial proportion of people with symptomatic COVID-19 go on to have persistent symptoms for 12 weeks or more, which is age-dependent. Clinicians need to be aware of the differing manifestations of Long COVID which may require tailored therapeutic approaches. Managing the long-term sequelae of SARS-CoV-2 infection in the population will remain a major challenge for health services in the next stage of the pandemic.

**Funding**

The study was funded by the Department of Health and Social Care in England.

## Introduction

The UK has experienced one of the largest epidemics of COVID-19 in Europe. As a new disease, the natural history beyond the immediate illness and the possible long-term sequelae remain largely unknown. As well as the acute risk of hospitalisation and death from COVID-19, some people who develop symptoms have a prolonged and debilitating illness that may continue for weeks or months (so-called Long COVID or post-COVID syndrome).<sup>1-5</sup> Patients sharing their experience on social media and establishing support groups were key to raising awareness of persistent symptoms and coined the term Long COVID.

The frequency and duration of persistent symptoms from COVID-19 are poorly understood and represent a major knowledge gap if effective treatments and management strategies are to be developed. Reported symptoms/signs include severe fatigue, breathlessness, chest pain or heaviness, fever, palpitations, cognitive impairment ('brain fog'), anosmia, ageusia, skin rash, joint pain or swelling.<sup>1-5</sup> Estimates of symptom prevalence and persistence vary substantially, arguably due to heterogeneous study designs and syndrome definitions.<sup>6-9</sup> It has been suggested that Long COVID describes a group of disparate conditions, including post-viral syndromes, long-term tissue or organ damage and ongoing inflammation.<sup>3,7,10,11</sup>

Occurrence of Long COVID appears to be associated with the severity of acute COVID-19 symptoms; for example, high prevalence of persistent symptoms has been reported among people hospitalised with COVID-19.<sup>12,13</sup> The number of acute symptoms has also been associated with risk of long COVID, alongside older age and female sex.<sup>6</sup>

While many Long COVID studies so far have focused on hospitalised COVID-19 cases,<sup>12-16</sup> here we report data from random community-based samples of the population in England. These involved over half a million people who took part in rounds three to five of the Real-Time Assessment of Community Transmission-2 (REACT-2) study between September 2020 and February 2021. We estimate symptom prevalence and investigate co-occurrence of symptoms among participants reporting symptoms lasting 12 weeks or more following suspected or confirmed COVID-19.

## **Methods**

### **Participants**

The REACT-2 programme evaluates community prevalence of SARS-CoV-2 anti-spike protein antibody positivity in England. Random population samples of adults in England were invited to take part every 2–4 months using the National Health Service (NHS) patient list to achieve similar numbers of participants in each of 315 lower-tier local authority (LTLA) areas.<sup>17</sup> Participants registered via an online portal or by telephone. Those registered were sent a test kit by post that included a self-administered point-of-care lateral flow immunoassay (LFIA) test with instructions and a link to an online video. Participants completed a survey (online/telephone) upon completion of their self-test. Participants provided information on demographics, household composition, whether or not they thought that they had had COVID-19, whether or not they had had a PCR test, comorbidities, symptoms related to COVID-19, severity of symptoms, and duration of any of a list of 29 symptoms.<sup>18</sup> In addition, we asked participants to report any other symptoms in free text. Personalised invitations were sent to between 560,000 and 600,000 individuals aged 18 years and above in each of rounds three to five of the REACT-2 study, carried out from 15 to 28 September 2020 (round 3), 27 October to 10 November 2020 (round 4) and 25 January to 8 February 2021 (round 5). Registrations closed after ~190,000 people had signed up at each round.

### **Data analysis**

We obtained prevalence estimates for reporting of one or more of the 29 symptoms by sex, age and other characteristics, at time of suspected or confirmed COVID-19, and for persistence of symptoms at four and 12 weeks. Our main analyses focused on individual symptoms reported as lasting for 12 weeks (84 days) or more (this excluded a small number of participants with inconsistent or missing data, see Supplementary Figure S1). Prevalence estimates were weighted by sex, age, ethnicity, LTLA population and index of multiple deprivation, to take account of the sampling design that gave approximately equal numbers of participants in each LTLA, and

differential response rates, to obtain prevalence estimates that were representative of the population of England as a whole.

We used logistic regression (univariable, and sex, age adjusted) to investigate the associations of demographic and lifestyle factors with persistence of symptoms at 12 weeks or more, and gradient boosted tree models<sup>19</sup> to investigate predictive ability (area under the curve, AUC) changes from adding variables to the model for persistent symptoms at 12 weeks or more.

To identify a more specific set of persistent symptoms associated with history of COVID-19, in sensitivity analyses, we carried out variable selection in a 30% subset of symptomatic participants: in univariable models, we identified a subset of persistent symptoms (12 or more weeks) that were positively associated with a reported prior positive PCR test, and estimated the population prevalence of persistence of one or more of these symptoms. We also repeated the logistic and gradient boosted tree modeling with this subset of symptoms as outcome variables.

Generalised additive models (GAMs) were constructed with likelihood of symptom persistence at 12 weeks or more modelled as a smoothed function of sex and age. A default thin plate spline was used and the smoothed functions were plotted to visualise the relationship between risk of persistent symptoms and age.

We used free-text analysis to identify single and co-occurring words to indicate other symptoms recorded by participants, and plotted these in a network.

To identify symptom clusters segmenting participants, two binary matrices were constructed for presence or absence (1 or 0) of each of the 29 surveyed symptoms at (i) time of symptom onset and (ii) 12 weeks after, for each participant. Clustering was performed, separately, both row-wise (to identify groups of participants with similar symptoms) and column-wise (to group symptoms based on their co-occurrence) using the CLustering LARge Applications (CLARA) extension of the Partitioning Around Medoids (PAM) algorithm, implemented in the R package *fpc*.<sup>20</sup> Briefly, PAM searches for the most representative data points to become cluster centroids by minimising the sum of dissimilarities between data points



and their assigned centroids. CLARA uses a sampling approach to reduce the computational burden for large data sets. We used Hamming distance as a measure of dissimilarity between participants (row-wise clustering) and symptoms (column-wise clustering). We determined the optimal number of clusters using the average silhouette width. We used two methods to assess cluster stability. First, we bootstrapped and re-clustered 100 times, then quantified the difference between bootstrapped and non-bootstrapped clusters using the Jaccard coefficient, which can range from 0 (no overlap) to 1 (perfect overlap).<sup>21</sup> Second, we removed each symptom in turn, re-clustered, then calculated the average proportion of non-overlap (APN) between these and whole-dataset clusters as a proxy for the individual variable importance and contribution to the population segmentation.<sup>22</sup>

To further describe patterns of symptom co-occurrence, we took the cross-product of the symptom matrix at symptom onset and at 12 weeks to find pairwise symptom co-occurrence counts, and visualised them as heatmaps.

## **Results**

### **Prevalence of persistent symptoms**

A total of 508,707 people took part in REACT-2 rounds three to five and completed surveys. The weighted prevalence of self-reported COVID-19 was 19.2% [19.1,19.3] with 92,116 people reporting one or more of 29 symptoms, of whom 76,155 (82.7%) reported a valid date of symptom onset 12 weeks or more before their survey date (Supplementary Figure S1A). Of those self-reporting COVID-19, 28,713/76,155 (37.7%) experienced at least one symptom for 12 weeks or more and 11,241 (14.8%) experienced at least three symptoms for the same period. This gives a weighted population prevalence of persistent symptoms of 5.75% (5.68, 5.81) for one and 2.22% (2.18, 2.26) for three or more symptoms for England to early February 2021. Almost a third of people with at least one symptom lasting 12 weeks or more (8,771/28,713 [30.5%]) reported having had severe COVID-19 symptoms (“significant effect on my daily life”) at the time of their illness, giving a weighted prevalence overall of people with persistent

symptoms at 12 weeks who had reported severe symptoms of 1.72% (1.69,1.76) (Supplementary Table S1A).

Figure 1 shows how the proportion of people with one or multiple symptoms declined over time since infection. There was a rapid drop-off by four weeks, a further, smaller drop by 12 weeks, but then little evidence of further decline over time up to ~22 weeks for both men and women, with higher prevalence of symptoms at each time point among women.

### **Factors associated with persistent symptoms**

Among symptomatic people, the persistence of one or more symptoms for 12 weeks or more was higher in women than men (age-adjusted OR: 1.51 [1.46,1.55]), and increased with age, with a linear increase of 3.5 percentage points per decade of life (Table 1, Figure 2, Supplementary Figure S2, Supplementary Table S2). With adjustment for sex and age, persistent symptoms were associated with self-reported overweight (OR: 1.16 [1.12, 1.21]) and obesity (OR: 1.53 [1.47,1.59]) compared with normal weight individuals, smoking (OR: 1.35 [1.28,1.41]), vaping (OR: 1.26 [1.18,1.34]), healthcare or care home workers (OR: 1.33 [1.26, 1.41]) and hospitalisation with COVID-19 (OR: 3.46 [2.93,4.09]), while Asian ethnicity (OR: 0.80 [0.74,0.88]) was associated with lower risk of persistent symptoms compared to people of white ethnicity (Table 1, Figure 2, Supplementary Table S2).

There was a higher proportion with persistent symptoms among those with low incomes at 51.0% (49.5, 52.4) compared with high incomes at 28.7% (27.2, 30.4) and among people living in the most deprived areas at 42.6% (41.5, 43.6) compared with the most affluent areas at 34.7% (34.0, 35.3) (Table 1).

Prevalence of persistent symptoms at 12 or more weeks was around 50% or more among people reporting co-morbidities, ranging up to 67.9% (65.6,70.1) for “other lung condition” (Table 1).

In addition to the 29 symptoms enquired about on the questionnaire, 8,370 respondents gave free-text descriptions of other symptoms, of whom 1,860 reported symptoms that persisted for 12 weeks or more. Free-text analysis of co-occurring words indicated common additional

symptoms which were not in our survey, including brain-fog, hair-loss, blood-pressure, heart-palpitations, severe-joint-pain (Supplementary Figure S3).

### **Sensitivity analysis**

In sensitivity analyses, we assessed the extent by which the observed prevalence of, and factors associated with, persistent symptoms were altered by using a smaller set of symptoms. To improve specificity of the symptoms for COVID-19, we selected symptoms that were positively predictive of self-reported PCR positivity (among a 30% holdout set of people, Supplementary Figure S1), thus identifying a subset of 15 of the surveyed 29 symptoms (Supplementary Table S1B, Supplementary Figure S4). Based on this more limited set, the number of people with symptoms reporting one or more persistent symptoms at 12 weeks was 24,867/76,155 (32.7%), giving a weighted population prevalence of 4.95% (4.89, 5.01) (Supplementary Table S1B). Logistic regression modelling with persistence of any one of these 15 symptoms at 12 weeks as a binary outcome showed similar results to the full 29-symptom analysis (Figure 2, Supplementary Figure S5).

### **Clustering analysis**

In clustering analysis, two stable clusters of participants were identified based on symptom profiles at 12 weeks. Participants in Cluster L1 (“tiredness cluster”) experienced high prevalence of tiredness, which co-occurred with muscle aches, difficulty sleeping and shortness of breath (Supplementary Figure S6). Participants in Cluster L2 (“respiratory cluster”) experienced high prevalence of respiratory symptoms including shortness of breath and tight chest, as well as chest pain (Figure 3, Figure 4). A higher proportion of people in the respiratory cluster reported severe symptoms at the time of their COVID-19 illness (43.5%, [42.0,44.9]) than in the tiredness cluster (27.4%, [26.7,28.1]) (Supplementary Table S3).

We obtained research ethics approval from the South Central-Berkshire B Research Ethics Committee (IRAS ID: 283787). The REACT Public Advisory Panel provides regular review of the study processes and results.

## Discussion

In this large community-based study of symptoms following COVID-19 among adults aged 18 years and above in England, participants reported high prevalence of persistent symptoms lasting 12 weeks or more. Estimates ranged from 5.8% of the population experiencing one or more persistent symptoms post-COVID-19 (corresponding to over 2 million adults in England), to 2.2% for three or more persistent symptoms (just under a million adults in England), and 1.7% with one or more symptoms lasting at least 12 weeks in people who reported severe COVID-19 symptoms affecting their daily life at the time of their illness.

Our estimates of the proportion of people with persistent COVID-19 symptoms are higher than in many other studies, although previous estimates have varied widely. At the low end, one study found that 2.3% of people with COVID-19 still reported symptoms at 12 weeks;<sup>6</sup> other studies have reported 13.7% of people were symptomatic at 84 days,<sup>7</sup> 14.8% symptomatic at 90 days,<sup>8</sup> 27% at 60 days,<sup>23</sup> 34.7% at 7 months,<sup>24</sup> and as high as 46% at six months.<sup>9</sup> Our comparatively high estimate, at 37.7% of people with COVID-19 experiencing one or more symptoms at 12 weeks, may partly reflect the large list of symptoms we surveyed, many of which are common and not specific to COVID-19. However, we asked participants only about symptoms that they related to a confirmed or suspected episode of COVID-19, and not to symptoms more generally. In addition, to improve specificity, in a sensitivity analysis we restricted COVID-19 symptoms to a subset of 15 that were predictive of reported PCR positivity and which included among others, loss or change of sense of smell or taste, tiredness, shortness of breath, muscle aches, heavy arms/legs, severe fatigue, tight chest and chest pain. This gave a similar proportion of people with persistent symptoms at 12 weeks (32.7%), and a population prevalence (5%) similar to that obtained when all 29 surveyed symptoms were considered.

Increasing age, female sex, BMI, hospitalisation and co-morbidities have previously been identified as risk factors for Long COVID<sup>6,25</sup>. Our finding of a linear association between age and persistent symptoms in people with symptomatic COVID-19 contrasts with some other studies that suggest the highest prevalence is found in middle-aged groups.<sup>7</sup> Our finding is conditional on symptomatic COVID-19, reflecting the fact that older age groups in the community have

lower infection rates than younger people<sup>26</sup> and are more likely to be asymptomatic.<sup>27</sup> Our findings of an association of persistent symptoms with deprivation, and with smoking or vaping, are not well described in the literature, although smokers have been found to have a higher risk of severe COVID-19.<sup>28</sup>

Our identification of two stable and well-differentiated symptom clusters at 12 weeks supports the characterisation of Long COVID as a diverse set of overlapping conditions. Previous studies have taken a similar unsupervised approach to characterising subtypes of Long COVID, albeit at earlier time points: Sudre et al.<sup>6</sup> identified two symptom clusters at 28 days post-symptom-onset, which resembled the early clusters identified in our study but not the 12-week clusters. Huang et al.<sup>23</sup> identified five clusters at 61 days, two of which reported high prevalences of respiratory symptoms similar to cluster L2 in our study.

### *Strengths and Limitations*

This study uses a large random community sample with a high response rate (26–29% across the three rounds) to describe the persistence of COVID-19 symptoms. It is therefore more likely to be representative of the range of disease severity in the population compared to some others, especially those based on hospitalised cases alone<sup>5</sup>. The focus in our questionnaire on persistence of self-reported COVID-19 symptoms, without specific reference to Long COVID (in contrast to the questions in some other studies<sup>7</sup>) has allowed us to investigate the persistence of a wide range of specific symptoms that have been suggested as relevant to Long COVID.<sup>3,5</sup> However, it is clear that a wide spectrum of symptoms and clinical presentations post-COVID-19 may be involved; our open free-text question identified a number of symptoms not included in our questionnaire including “brain fog”, “palpitations” and “hair loss”.<sup>29</sup> However, as the study was based on self-reported data and because many of the symptoms are common and not specific to COVID-19, we may, as noted, have overestimated the prevalence of persistent symptoms.

A further limitation is the retrospective study design, which introduces the possibility of recall bias. Nonetheless, in earlier analyses we have shown that participant reports of date of onset of their symptoms produce an epidemic curve that very closely tracks the epidemic.<sup>27,30,31</sup> Respondents were restricted to reporting a single date of (initial) symptom onset which does not

allow for delayed onset of some symptoms, nor does it allow for the reporting of relapsing symptoms which appear to be a feature of Long COVID.<sup>6</sup> A further limitation, despite the high response rate for a community surveillance study, is the possibility of participation bias as the REACT-2 study included a home antibody self-test;<sup>27</sup> it is plausible that people with persistent symptoms may have been more likely to participate in order to ascertain their antibody status.

### *Implications*

The scale of morbidity identified in this study presents significant challenges for the affected individuals and their families, and for health services and society more broadly. After the initial decline in symptom prevalence between 4 and 12 weeks the prevalence of persistent symptoms plateaued indicating that large numbers of people may have chronic symptoms requiring investigation and intervention including rehabilitation. We show here that economically disadvantaged people and those in deprived areas appear to have a higher burden of persistent symptoms post COVID-19, compounding the excess burden of severe illness and mortality from COVID-19 experienced by these groups<sup>32</sup>. Inability to work due to disability from Long COVID could have a high impact on these populations. Given the high potential population health burden, investment is urgently needed to expand the network of services investigating and managing people with Long COVID, to link these services to national research studies and consistent data collection to improve our understanding, and to create the infrastructure for trials in the same way as has been achieved for acute COVID-19.

We identified two clusters of participants based on their symptoms. Individuals with predominantly respiratory symptoms comprised around one quarter of those with persistent symptoms -- the larger proportion comes from a cluster of less organ-specific symptoms, particularly fatigue. Clinicians will need to be aware of the range of presenting symptoms to best support their patients towards recovery and there will need to be education of healthcare professionals to recognise and respect the experiences of people reporting these symptoms. Societal recognition of Long COVID as an important outcome of the pandemic (including as an occupational illness for those acquiring it through work) is needed, including for those who may require benefits as well as rehabilitation. The Long COVID patient groups provide much needed

support to those involved, but will require resources if they are to be able to work effectively with health services, especially to reach people in more deprived and marginalised groups.

In summary we have identified significant ongoing morbidity among people post COVID-19, with a substantial proportion experiencing persistent symptoms lasting 12 weeks or more. Managing the long term sequelae of SARS-CoV-2 infection in the population will remain a major challenge for health services in the next stage of the pandemic.

## **Data availability**

Summary tabular data are provided with this paper.

## **Declaration of interests**

We declare no competing interests.

## **Author contributions**

MW - conceptualisation, formal analysis, visualisation, methodology, writing – original draft, writing– review & editing; JE - conceptualisation, formal analysis, methodology, writing – original draft, writing– review & editing; MCH - supervision, methodology, writing– review & editing; SR: supervision, methodology, writing– review & editing; AD - funding acquisition, supervision, writing– review & editing; GC - conceptualisation, supervision, methodology, writing – original draft, writing– review & editing; HW - conceptualisation, supervision, methodology, writing – original draft, writing– review & editing; PE - funding acquisition, conceptualisation, supervision, methodology, writing – original draft, writing– review & editing.

## **Acknowledgements**

GC is supported by an NIHR Professorship. WSB is the Action Medical Research Professor, AD is an NIHR senior investigator and DA and PE are Emeritus NIHR Senior Investigators. SR acknowledges support from MRC Centre for Global Infectious Disease Analysis, NIHR Health Protection Research Unit, Wellcome Trust (200861/Z/16/Z, 200187/Z/15/Z), and Centres for Disease Control and Prevention (US, U01CK0005-01-02). HW is a National Institute for Health Research (NIHR) Senior Investigator and acknowledges support from NIHR Biomedical Research Centre of Imperial College NHS Trust, NIHR School of Public Health Research, NIHR Applied Research Collaborative North West London, and Wellcome Trust (UNS32973). PE is Director of the MRC Centre for Environment and Health (MR/L01341X/1, MR/S019669/1). PE acknowledges support from the NIHR Imperial Biomedical Research Centre and the NIHR HPRUs in Chemical and Radiation Threats and Hazards and in Environmental Exposures and Health, the British Heart Foundation Centre for Research Excellence at Imperial College London (RE/18/4/34215), Health Data Research UK (HDR UK) and the UK Dementia Research Institute at Imperial (MC\_PC\_17114). We thank The Huo Family Foundation for their support of our work on COVID-19. Our work on Long COVID is also being supported by grants from NIHR and UK Research and Innovation (UKRI): REACT GE (MR/V030841/1) and COV-LT-0040 – REACT Long COVID (REACT-LC).

We thank key collaborators on this work -- Ipsos MORI: Stephen Finlay, John Kennedy, Kevin Pickering, Duncan Peskett, Sam Clemens and Kelly Beaver; Institute of Global Health Innovation at Imperial College London: Gianluca Fontana, Sutha Satkunarajah, Didi Thompson; the Imperial Patient Experience Research Centre and the REACT Public Advisory Panel; NHS Digital for access to the NHS Register; Dr Nisreen Alwan.



## References

1. Urgent need for more research to understand Long Covid.  
<https://royalsociety.org/news/2020/10/urgent-need-to-understand-long-covid/>.
2. Del Rio, C., Collins, L. F. & Malani, P. Long-term Health Consequences of COVID-19. *JAMA* **324**, 1723–1724 (2020).
3. *Living with Covid19 – Second review*.  
<https://evidence.nihr.ac.uk/themedreview/living-with-covid19-second-review/> (2021)  
doi:10.3310/themedreview\_45225.
4. Overview | COVID-19 rapid guideline: managing the long-term effects of COVID-19 | Guidance | NICE.
5. Lopez-Leon, S. *et al.* More than 50 Long-term effects of COVID-19: a systematic review and meta-analysis. *medRxiv* (2021) doi:10.1101/2021.01.27.21250617.
6. Sudre, C. H. *et al.* Attributes and predictors of long COVID. *Nat. Med.* **27**, 626–631 (2021).
7. Ayoubkhani, D. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics.  
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/1april2021> (2021).
8. Cirulli, E. T. *et al.* Long-term COVID-19 symptoms in a large unselected population. *bioRxiv* (2020) doi:10.1101/2020.10.07.20208702.
9. Klein, H. *et al.* Onset, duration and unresolved symptoms, including smell and taste changes, in mild COVID-19 infection: a cohort study in Israeli patients. *Clin. Microbiol. Infect.* (2021)  
doi:10.1016/j.cmi.2021.02.008.
10. Yong, S. J. Long-haul COVID-19: Putative pathophysiology, risk factors, and treatments. *Preprints*

- (2020) doi:10.20944/preprints202012.0242.v1.
11. *Living with Covid19*. <https://evidence.nihr.ac.uk/themedreview/living-with-covid19/> (2020) doi:10.3310/themedreview\_41169.
  12. Arnold, D. T. *et al.* Patient outcomes after hospitalisation with COVID-19 and implications for follow-up: results from a prospective UK cohort. *Thorax* (2020) doi:10.1136/thoraxjnl-2020-216086.
  13. Carfi, A., Bernabei, R., Landi, F. & Gemelli Against COVID-19 Post-Acute Care Study Group. Persistent Symptoms in Patients After Acute COVID-19. *JAMA* **324**, 603–605 (2020).
  14. Venturelli, S. *et al.* Surviving COVID-19 in Bergamo province: a post-acute outpatient re-evaluation. *Epidemiol. Infect.* **149**, e32 (2021).
  15. Tomasoni, D. *et al.* Anxiety and depression symptoms after virological clearance of COVID-19: A cross-sectional study in Milan, Italy. *J. Med. Virol.* **93**, 1175–1179 (2021).
  16. Moreno-Pérez, O. *et al.* Post-acute COVID-19 syndrome. Incidence and risk factors: A Mediterranean cohort study. *J. Infect.* **82**, 378–383 (2021).
  17. Riley, S. *et al.* REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: Study protocol. *Wellcome Open Res.* **5**, 200 (2020).
  18. Real-time Assessment of Community Transmission (REACT) study. <https://www.imperial.ac.uk/medicine/research-and-impact/groups/react-study/>.
  19. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. CatBoost: unbiased boosting with categorical features. *arXiv [cs.LG]* (2017).
  20. Hennig, C. & Imports, M. Package ‘fpc’. Available at: Available at: <https://cran.r-project.org/web/packages/fpc/index.html> ENT **91**, (2015).
  21. Hennig, C. Cluster-wise assessment of cluster stability. in *Computational Statistics and Data Analysis* (2007).
  22. Datta, S. & Datta, S. Comparisons and validation of statistical clustering techniques for microarray

- gene expression data. *Bioinformatics* **19**, 459–466 (2003).
23. Huang, Y. *et al.* COVID Symptoms, Symptom Clusters, and Predictors for Becoming a Long-Hauler: Looking for Clarity in the Haze of the Pandemic. *medRxiv* (2021) doi:10.1101/2021.03.03.21252086.
  24. Augustin, M. *et al.* Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study. *Lancet Reg Health Eur* **6**, 100122 (2021).
  25. Sykes, D. L. *et al.* Post-COVID-19 Symptom Burden: What is Long-COVID and How Should We Manage It? *Lung* **199**, 113–119 (2021).
  26. Riley, S. *et al.* Resurgence of SARS-CoV-2: Detection by community viral surveillance. *Science* **372**, 990–995 (2021).
  27. Ward, H. *et al.* SARS-CoV-2 antibody prevalence in England following the first peak of the pandemic. *Nat. Commun.* **12**, 905 (2021).
  28. Elliott, J. *et al.* COVID-19 mortality in the UK Biobank cohort: revisiting and evaluating risk factors. *Eur. J. Epidemiol.* **36**, 299–309 (2021).
  29. Fernández-de-Las-Peñas, C. *et al.* Long-term post-COVID symptoms and associated risk factors in previously hospitalized patients: A multicenter study. *J. Infect.* (2021) doi:10.1016/j.jinf.2021.04.036.
  30. Ward, H. *et al.* Prevalence of antibody positivity to SARS-CoV-2 following the first peak of infection in England: Serial cross-sectional studies of 365,000 adults. *Lancet Reg Health Eur* **4**, 100098 (2021).
  31. Ward, H. *et al.* REACT-2 Round 5: increasing prevalence of SARS-CoV-2 antibodies demonstrate impact of the second wave and of vaccine roll-out in England. (2021) doi:10.1101/2021.02.26.21252512.
  32. Davies, B. *et al.* Community factors and excess mortality in first wave of the COVID-19 pandemic in England. *Nat. Commun.* **12**, 3755 (2021).

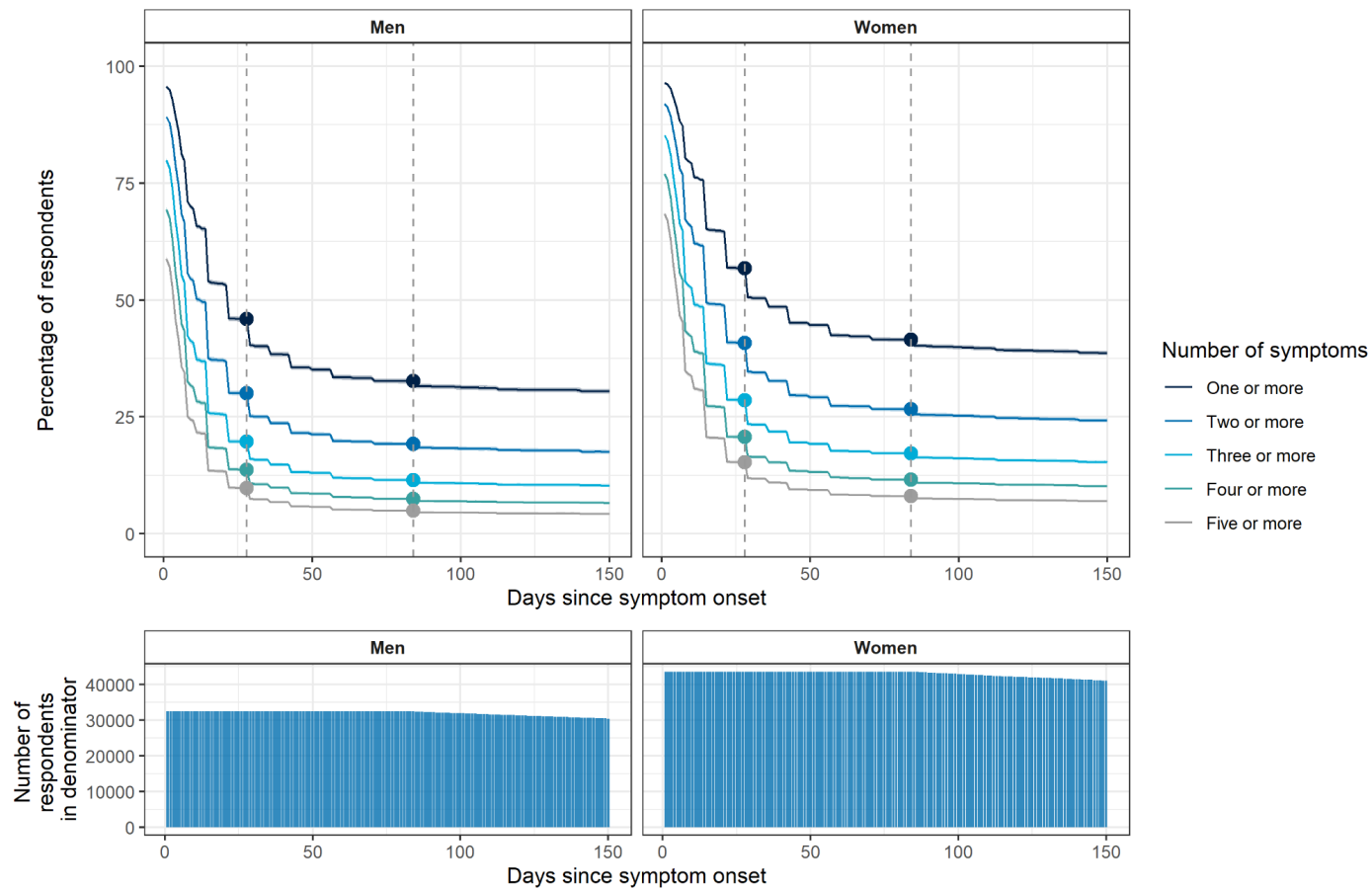
## Figures and Tables

**Table 1.** Numbers and proportions of participants who reported one or more symptoms (from a list of 29 surveyed symptoms) of COVID-19 at i) time of symptom onset, ii) 4 weeks post symptom onset, and iii) 12 weeks post symptom onset, among the 76,155 symptomatic participants for whom we have 12 weeks' follow-up and complete data.

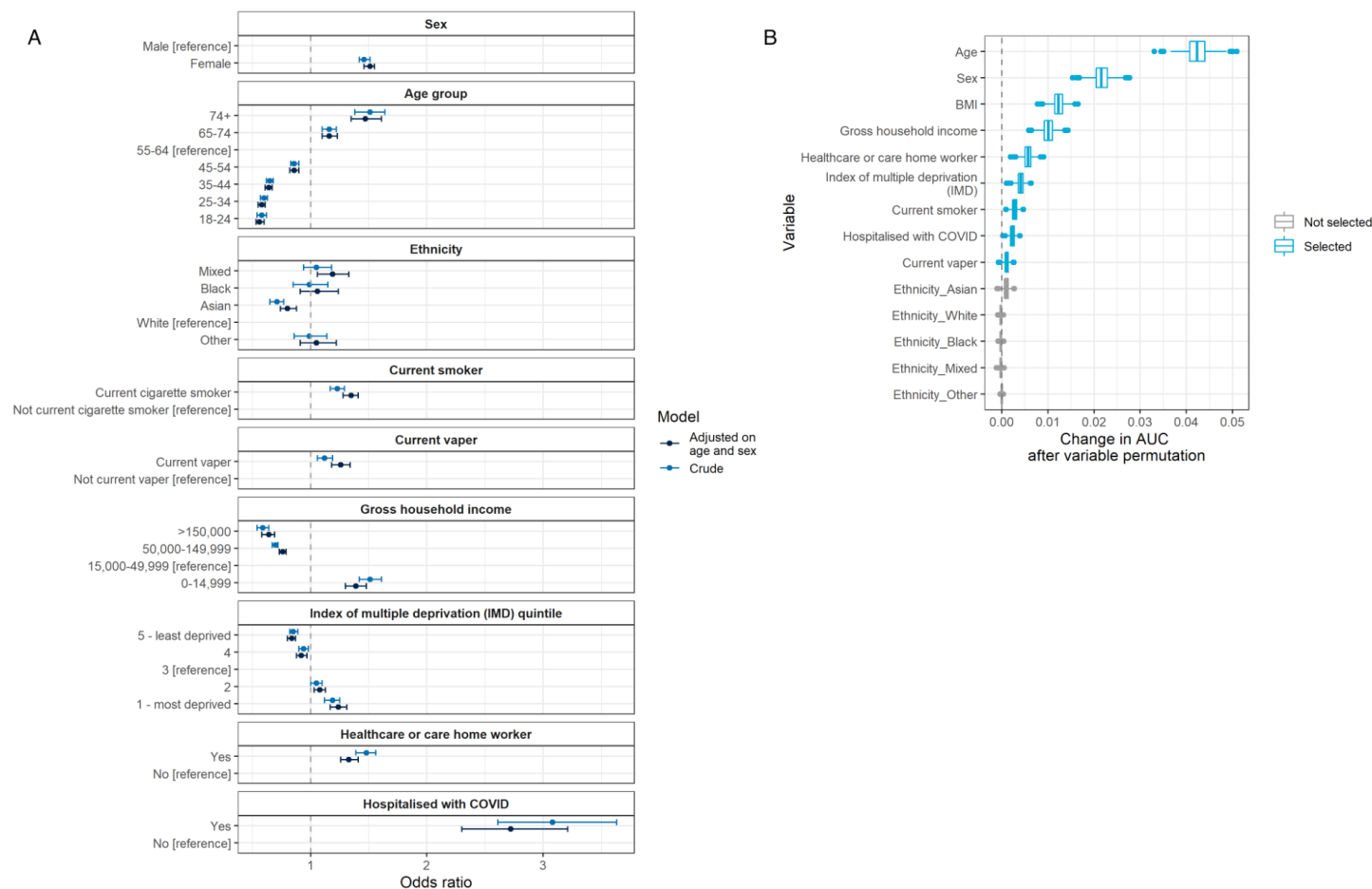
Category		Symptomatic at infection	4 weeks after initial symptom onset		12 weeks after initial symptom onset	
			Number symptomatic	Percentage symptomatic	Number symptomatic	Percentage symptomatic
All participants		76155	39737	52.2 [51.8-52.5]	28713	37.7 [37.4-38.0]
Sex	Women	43654	24812	56.8 [56.4-57.3]	18109	41.5 [41.0-41.9]
	Men	32500	14924	45.9 [45.4-46.5]	10604	32.6 [32.1-33.1]
Age group	18-24	5085	2133	41.9 [40.6-43.3]	1534	30.2 [28.9-31.4]
	25-34	12245	5325	43.5 [42.6-44.4]	3784	30.9 [30.1-31.7]
	35-44	15992	7646	47.8 [47.0-48.6]	5232	32.7 [32.0-33.4]
	45-54	17548	9490	54.1 [53.3-54.8]	6858	39.1 [38.4-39.8]
	55-64	15148	8787	58.0 [57.2-58.8]	6461	42.7 [41.9-43.4]
	65-74	7811	4823	61.7 [60.7-62.8]	3615	46.3 [45.2-47.4]
	74+	2326	1533	65.9 [64.0-67.8]	1229	52.8 [50.8-54.9]
Ethnicity	Asian	2676	1180	44.1 [42.2-46.0]	809	30.2 [28.5-32.0]
	Black	739	361	48.8 [45.3-52.5]	278	37.6 [34.2-41.2]
	Mixed	1244	631	50.7 [47.9-53.5]	487	39.1 [36.5-41.9]
	Other	793	413	52.1 [48.6-55.5]	299	37.7 [34.4-41.1]
	White	70103	36838	52.5 [52.2-52.9]	26603	37.9 [37.6-38.3]
BMI	Normal weight	26128	12803	49.0 [48.4-49.6]	8954	34.3 [33.7-34.8]
	Obese	16076	9416	58.6 [57.8-59.3]	7107	44.2 [43.4-45.0]
	Overweight	23958	12429	51.9 [51.2-52.5]	8916	37.2 [36.6-37.8]
	Underweight	886	450	50.8 [47.5-54.1]	333	37.6 [34.5-40.8]
Healthcare or care home worker	Yes	5243	3041	58.0 [56.7-59.3]	2283	43.5 [42.2-44.9]
Index of multiple deprivation (IMD) quintile	1 - most deprived	8096	4438	54.8 [53.7-55.9]	3445	42.6 [41.5-43.6]
	2	12789	6836	53.5 [52.6-54.3]	5066	39.6 [38.8-40.5]
	3	16491	8722	52.9 [52.1-53.7]	6336	38.4 [37.7-39.2]
	4	18202	9529	52.4 [51.6-53.1]	6728	37.0 [36.3-37.7]
	5 - least deprived	20577	10212	49.6 [48.9-50.3]	7138	34.7 [34.0-35.3]
Current smoker	Current cigarette smoker	8441	4536	53.7 [52.7-54.8]	3552	42.1 [41.0-43.1]
	Not current cigarette smoker	67021	34880	52.0 [51.7-52.4]	24916	37.2 [36.8-37.5]
	Prefer not to say	693	321	46.3 [42.6-50.0]	245	35.4 [31.9-39.0]
Current vaper	Current vaper	4933	2534	51.4 [50.0-52.8]	1986	40.3 [38.9-41.6]
	Not current vaper	70802	37015	52.3 [51.9-52.6]	26586	37.5 [37.2-37.9]
	Prefer not to say	420	188	44.8 [40.1-49.5]	141	33.6 [29.2-38.2]
Severity of COVID symptoms	Mild symptoms	21454	7519	35.0 [34.4-35.7]	5441	25.4 [24.8-25.9]
	Moderate symptoms	37976	20413	53.8 [53.3-54.3]	14501	38.2 [37.7-38.7]
	Severe symptoms	16725	11805	70.6 [69.9-71.3]	8771	52.4 [51.7-53.2]
Treatment sought/received for COVID	No medical attention sought	47642	21009	44.1 [43.7-44.5]	15098	31.7 [31.3-32.1]
	Sought medical attention from pharmacist / by phone (NHS 111/GP)	17794	11713	65.8 [65.1-66.5]	8717	49.0 [48.3-49.7]
	Visited GP/walk-in centre/A&E	6154	4269	69.4 [68.2-70.5]	2811	45.7 [44.4-46.9]
	Hospital admission	629	494	78.5 [75.2-81.6]	408	64.9 [61.1-68.5]
Gross household income	0-14,999	4543	2862	63.0 [61.6-64.4]	2316	51.0 [49.5-52.4]
	15,000-49,999	21468	11847	55.2 [54.5-55.8]	8746	40.7 [40.1-41.4]
	50,000-149,999	21923	10382	47.4 [46.7-48.0]	7100	32.4 [31.8-33.0]

	>150,000	3082	1330	43.2 [41.4-44.9]	886	28.7 [27.2-30.4]
Comorbidities	Organ transplant recipient	70	43	61.4 [49.7-72.0]	35	50.0 [38.6-61.4]
	Diabetes (type I or II)	3282	2045	62.3 [60.6-64.0]	1596	48.6 [46.9-50.3]
	Heart disease or heart problems	2822	1894	67.1 [65.4-68.8]	1521	53.9 [52.1-55.7]
	Hypertension (high blood pressure)	8269	5203	62.9 [61.9-64.0]	4009	48.5 [47.4-49.6]
	Stroke	415	261	62.9 [58.1-67.4]	211	50.8 [46.0-55.6]
	Kidney disease	547	405	74.0 [70.2-77.5]	325	59.4 [55.2-63.5]
	Liver disease	309	209	67.6 [62.2-72.6]	171	55.3 [49.8-60.8]
	Anaemia	1371	952	69.4 [66.9-71.8]	801	58.4 [55.8-61.0]
	Asthma	9119	5616	61.6 [60.6-62.6]	4419	48.5 [47.4-49.5]
	Other lung condition	1639	1252	76.4 [74.3-78.4]	1113	67.9 [65.6-70.1]
	Cancer	820	533	65.0 [61.7-68.2]	411	50.1 [46.7-53.5]
	Condition affecting the brain and nerves	686	457	66.6 [63.0-70.0]	378	55.1 [51.4-58.8]
	A weakened immune system/reduced ability to deal with infections	3080	2165	70.3 [68.7-71.9]	1815	58.9 [57.2-60.7]
	Depression	6911	4680	67.7 [66.6-68.8]	3931	56.9 [55.7-58.0]
	Anxiety	10516	6768	64.4 [63.4-65.3]	5553	52.8 [51.9-53.8]
	Psychiatric disorder	641	417	65.1 [61.3-68.6]	372	58.0 [54.2-61.8]

**Figure 1** Plots showing persistence of symptoms as a proportion of those who reported symptoms at any time. Plots account for ‘censoring’ – i.e. the denominator in the proportion at day X is the number of respondents who reported a symptomatic infection X or more days prior to responding to the REACT-2 survey (the denominator is shown in the bottom panel plots). Women have higher rates of persistent symptoms; a slower decline in symptom prevalence is observed after 12 weeks in both sexes. The vertical dashed lines show 4 and 12 weeks.

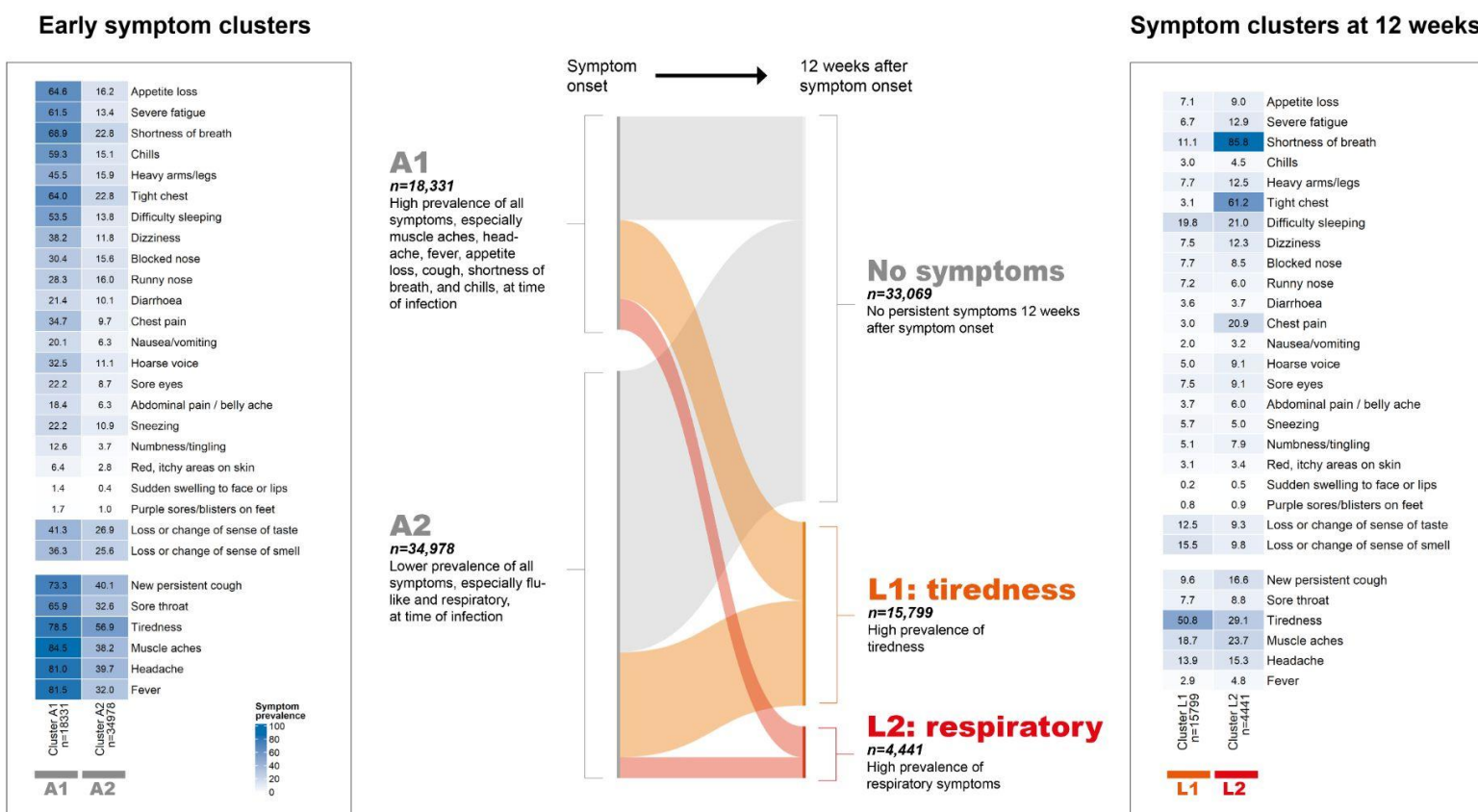


**Figure 2** (A) Logistic regression models with one or more symptoms at 12 weeks (y/n) as the binary outcome variable, both unadjusted and adjusted for sex and age; (B) Mean contribution to area under the curve (AUC) that each variable makes to a multivariable boosted tree model, derived by permuting each variable in turn (1000x to obtain a distribution) and quantifying the change in model performance. Age, sex, body mass index (BMI), household income, healthcare/care home worker, deprivation, smoking status and prior hospitalisation with COVID-19 are the strongest predictors of persistent symptoms in multivariable modelling, while Asian ethnicity is associated with a lower risk of persistent symptoms at 12 weeks.

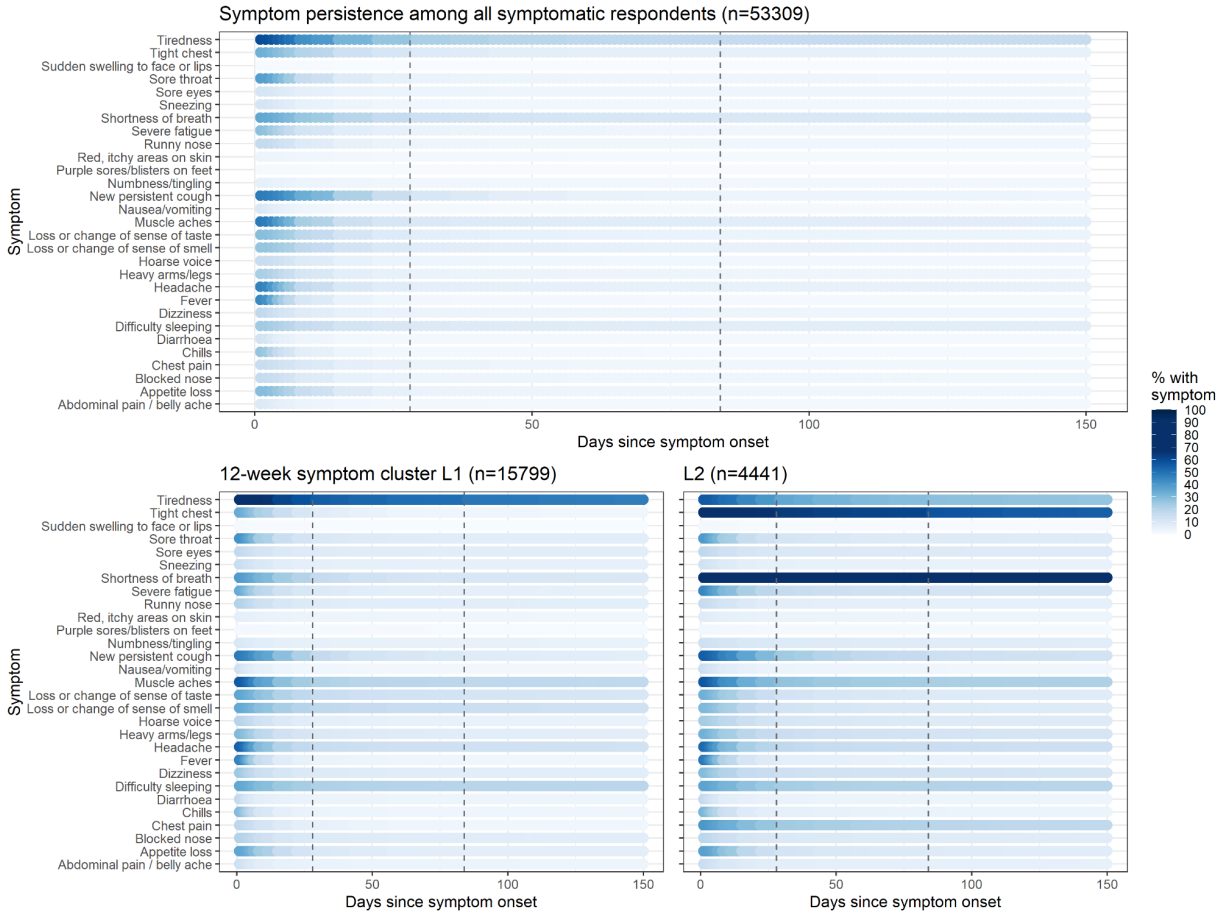




**Figure 3** Results of clustering on symptom profile at time of symptom onset and then reclustering at 12 weeks, using CLARA (partitioning around medoids) algorithm. Central Sankey plot shows transitions between early and 12-week clusters. Two stable clusters of symptomatic infections were identified at t0: cluster A1 was characterised primarily by higher prevalence of flu-like symptoms (muscle aches, headache, fever, appetite loss, chills) and respiratory symptoms (shortness of breath, tight chest, new persistent cough). Two stable clusters were identified at 12 weeks. Cluster L1 (“tiredness cluster”) had high prevalence of tiredness. Cluster L2 (“respiratory cluster”) was a smaller subset of 4,441 participants who had high prevalence of shortness of breath and tight chest as well as chest pain.



**Figure 4** Persistence of symptoms for all symptomatic participants (top panel) and in 12-week symptoms clusters L1 and L2 (bottom panels). Dashed lines show 4 and 12 weeks post symptom onset.



## **Supplementary material**

**Table S1** Numbers and weighted prevalence of currently or previously reported persistent symptoms in the full REACT-2 study population (rounds 3–5).

Table 1A shows prevalences of any of the 29 surveyed symptoms; Table 1B shows prevalences of any of a subset of 15 symptoms identified as predictive of PCR positivity. The final three columns show prevalences weighted to be representative of the adult population of England, by age, sex, ethnicity, lower-tier local authority population, and index of multiple deprivation (IMD).

(A) All 29 symptoms

Variable	Category	n	Counts at 12 weeks			Prevalence at 12 weeks		
			Number with 1 or more symptoms	Number with 2 or more symptoms	Number with 3 or more symptoms	Prevalence 1 or more symptoms (weighted) %	Prevalence 2 or more symptoms (weighted) %	Prevalence 3 or more symptoms (weighted) %
All participants		508707	28713	17860	11241	5.75 [5.68-5.81]	3.54 [3.49-3.59]	2.22 [2.18-2.26]
Sex	Women	284840	18109	11612	7513	6.55 [6.46-6.65]	4.2 [4.12-4.28]	2.73 [2.67-2.8]
	Men	223861	10604	6248	3728	4.90 [4.82-4.99]	2.85 [2.78-2.91]	1.68 [1.63-1.73]
Age group	18-24	28655	1534	885	514	5.65 [5.46-5.84]	3.19 [3.04-3.34]	1.84 [1.73-1.95]
	25-34	65159	3784	2105	1219	6.02 [5.86-6.18]	3.33 [3.21-3.45]	1.87 [1.78-1.96]
	35-44	85044	5232	3083	1815	6.19 [6.03-6.36]	3.62 [3.5-3.75]	2.14 [2.04-2.24]
	45-54	103478	6858	4311	2732	6.78 [6.62-6.95]	4.29 [4.15-4.42]	2.72 [2.62-2.83]
	55-64	105361	6461	4205	2713	6.31 [6.14-6.48]	4.16 [4.02-4.3]	2.73 [2.61-2.84]
	65-74	85297	3615	2406	1629	4.56 [4.4-4.72]	3.05 [2.92-3.19]	2.09 [1.98-2.2]
	74+	35713	1229	865	619	3.64 [3.48-3.8]	2.58 [2.45-2.72]	1.91 [1.8-2.03]
Ethnicity	Asian	17666	809	480	275	5.06 [4.84-5.29]	2.8 [2.63-2.97]	1.58 [1.46-1.71]
	Black	4091	278	138	84	6.56 [6.15-6.99]	3.48 [3.18-3.8]	2.02 [1.8-2.27]
	Mixed	6112	487	310	202	7.65 [6.99-8.37]	4.69 [4.17-5.27]	3.05 [2.63-3.53]
	Other	4434	299	180	118	6.71 [6.08-7.4]	4.1 [3.61-4.65]	2.76 [2.36-3.22]
	White	472750	26603	16600	10461	5.73 [5.67-5.8]	3.57 [3.52-3.63]	2.26 [2.21-2.3]
BMI	Normal weight	183305	8954	5325	3182	5.03 [4.93-5.13]	2.96 [2.88-3.03]	1.75 [1.69-1.81]
	Obese	97597	7107	4698	3127	7.29 [7.13-7.46]	4.69 [4.56-4.83]	3.15 [3.04-3.26]
	Overweight	161456	8916	5449	3398	5.72 [5.6-5.83]	3.48 [3.39-3.57]	2.18 [2.11-2.25]
	Underweight	6273	333	220	147	5.86 [5.36-6.41]	3.7 [3.3-4.14]	2.14 [1.84-2.49]
Healthcare or care home worker	Yes	27690	2283	1469	920	8.44 [8.12-8.78]	5.25 [4.99-5.52]	3.19 [2.98-3.4]

Index of multiple deprivation (IMD) quintile	1 - most deprived	50260	3445	2288	1535	6.8 [6.65-6.96]	4.39 [4.27-4.52]	2.89 [2.79-3]
	2	80894	5066	3223	2118	6.22 [6.08-6.37]	3.81 [3.69-3.93]	2.46 [2.37-2.56]
	3	109928	6336	3937	2500	5.55 [5.41-5.69]	3.39 [3.28-3.5]	2.1 [2.02-2.19]
	4	126237	6728	4122	2553	5.23 [5.1-5.37]	3.16 [3.05-3.27]	1.92 [1.84-2.01]
	5 - least deprived	141388	7138	4290	2535	4.92 [4.79-5.06]	2.94 [2.84-3.05]	1.72 [1.64-1.8]
Current smoker	Current cigarette smoker	50983	3552	2367	1610	7.02 [6.81-7.23]	4.59 [4.42-4.77]	3.1 [2.96-3.25]
	Not current cigarette smoker	453681	24916	15344	9533	5.58 [5.51-5.65]	3.4 [3.35-3.45]	2.11 [2.06-2.15]
	Prefer not to say	4043	245	149	98	6.08 [5.42-6.81]	3.97 [3.44-4.58]	2.35 [1.95-2.84]
Current vaper	Current vaper	28436	1986	1273	834	7.21 [6.93-7.5]	4.59 [4.36-4.83]	2.93 [2.75-3.13]
	Not current vaper	477729	26586	16493	10340	5.65 [5.59-5.72]	3.47 [3.41-3.52]	2.17 [2.13-2.21]
	Prefer not to say	2542	141	94	67	5.58 [4.8-6.48]	3.95 [3.3-4.72]	2.43 [1.92-3.05]
Severity of COVID symptoms	Mild symptoms	27572	5441	2746	1351	18.59 [18.16-19.02]	9.26 [8.94-9.59]	4.47 [4.24-4.71]
	Moderate symptoms	45357	14501	8841	5423	31.17 [30.76-31.59]	18.75 [18.4-19.11]	11.53 [11.25-11.83]
	Severe symptoms	19187	8771	6273	4467	44.37 [43.68-45.06]	31.78 [31.14-32.44]	22.57 [22-23.16]
Treatment sought/ received for COVID	No medical attention sought	58330	15098	8544	4912	24.72 [24.38-25.06]	13.82 [13.55-14.1]	7.93 [7.72-8.14]
	Sought medical attention from pharmacist / by phone (NHS 111/GP)	21388	8717	5991	4120	39.12 [38.5-39.75]	26.44 [25.87-27.01]	18.26 [17.77-18.76]
	Visited GP/walk-in centre/A&E	6792	2811	1911	1263	39.85 [38.69-41.02]	27.63 [26.58-28.71]	17.93 [17.04-18.86]
	Hospital admission	739	408	319	241	53.04 [49.66-56.39]	41.24 [37.96-44.6]	30.87 [27.84-34.08]
Gross household income	0-14,999	32819	2316	1673	1209	7.11 [6.86-7.38]	5.06 [4.84-5.29]	3.71 [3.52-3.91]
	15,000-49,999	148039	8746	5548	3521	6.06 [5.94-6.18]	3.82 [3.72-3.92]	2.38 [2.3-2.46]
	50,000-149,999	128957	7100	4088	2325	5.65 [5.52-5.78]	3.15 [3.05-3.25]	1.77 [1.7-1.84]
	>150,000	32819	2316	1673	1209	7.11 [6.86-7.38]	5.06 [4.84-5.29]	3.71 [3.52-3.91]
Comorbidities	Organ transplant	596	35	27	20	5.39 [3.84-7.51]	4.2 [2.86-6.13]	3.53 [2.32-5.34]

recipient							
Diabetes (type I or II)	27358	1596	1118	837	5.88 [5.61-6.16]	4.01 [3.79-4.25]	3.11 [2.91-3.32]
Heart disease or heart problems	21430	1521	1090	816	7.08 [6.75-7.43]	5.09 [4.8-5.39]	3.85 [3.6-4.12]
Hypertension (high blood pressure)	65792	4009	2744	1905	6.16 [5.98-6.35]	4.25 [4.09-4.41]	3.01 [2.88-3.15]
Stroke	3352	211	149	110	6.17 [5.42-7.02]	4.15 [3.53-4.86]	3.14 [2.62-3.78]
Kidney disease	3867	325	238	166	8.33 [7.51-9.23]	6.03 [5.33-6.82]	4.11 [3.54-4.78]
Liver disease	2107	171	123	103	8.28 [7.2-9.51]	5.58 [4.69-6.62]	4.67 [3.86-5.63]
Anemia	7575	801	586	439	10.99 [10.34-11.67]	7.8 [7.25-8.38]	5.78 [5.3-6.29]
Asthma	51337	4419	3123	2192	8.68 [8.44-8.93]	6.05 [5.85-6.26]	4.21 [4.04-4.39]
Other lung condition	12500	1113	840	643	9.02 [8.54-9.52]	6.73 [6.31-7.17]	5.32 [4.95-5.72]
Cancer	8042	411	270	196	5.2 [4.73-5.72]	3.49 [3.1-3.92]	2.65 [2.32-3.03]
Condition affecting the brain and nerves	5415	378	295	229	7.1 [6.45-7.82]	5.37 [4.8-6]	4.07 [3.58-4.63]
A weakened immune system/reduced ability to deal with infections	19302	1815	1400	1055	9.69 [9.27-10.12]	7.47 [7.1-7.86]	5.61 [5.28-5.95]
Depression	38125	3931	2873	2146	10.33 [10.04-10.63]	7.54 [7.29-7.8]	5.57 [5.35-5.79]
Anxiety	57140	5553	3923	2809	10.02 [9.79-10.27]	7.04 [6.84-7.24]	5.01 [4.84-5.19]
Psychiatric disorder	3385	372	284	226	10.78 [9.85-11.79]	8.08 [7.27-8.98]	6.13 [5.42-6.92]

(B) 15 symptoms (loss or change of sense of smell, loss or change of sense of taste, tiredness, shortness of breath, heavy arms/legs, severe fatigue, dizziness, numbness/tingling, appetite loss, red, itchy areas on skin, difficulty sleeping, tight chest, headache, chest pain).

Variable	Category	n	Counts at 12 weeks			Prevalence at 12 weeks		
			Number with 1 or more symptoms	Number with 2 or more symptoms	Number with 3 or more symptoms	Prevalence 1 or more symptoms (weighted) %	Prevalence 2 or more symptoms (weighted) %	Prevalence 3 or more symptoms (weighted) %
All participants		508707	24867	14810	8555	4.95 [4.89-5.01]	2.94 [2.9-2.99]	1.69 [1.65-1.72]
Sex	Women	284840	15923	9714	5761	5.75 [5.66-5.84]	3.53 [3.46-3.6]	2.08 [2.03-2.14]
	Men	223861	8944	5096	2794	4.12 [4.04-4.2]	2.33 [2.27-2.39]	1.27 [1.23-1.32]
Age group	18-24	28655	1295	700	357	4.64 [4.47-4.82]	2.56 [2.43-2.69]	1.33 [1.24-1.43]
	25-34	65159	3209	1728	923	5.1 [4.96-5.25]	2.74 [2.63-2.85]	1.39 [1.32-1.47]
	35-44	85044	4566	2600	1392	5.34 [5.19-5.5]	3.07 [2.96-3.19]	1.65 [1.57-1.74]
	45-54	103478	6050	3702	2201	6 [5.84-6.15]	3.68 [3.56-3.81]	2.2 [2.1-2.3]
	55-64	105361	5604	3470	2077	5.48 [5.32-5.65]	3.43 [3.31-3.56]	2.09 [1.99-2.2]
	65-74	85297	3087	1919	1175	3.91 [3.76-4.06]	2.46 [2.34-2.58]	1.52 [1.43-1.62]
Ethnicity	74+	35713	1056	691	430	3.17 [3.03-3.33]	2.13 [2.01-2.25]	1.36 [1.26-1.46]
	Asian	17666	687	390	214	4.15 [3.95-4.37]	2.34 [2.19-2.5]	1.23 [1.12-1.35]
	Black	4091	235	116	65	5.42 [5.04-5.81]	2.94 [2.66-3.24]	1.6 [1.4-1.83]
	Mixed	6112	426	263	169	6.58 [5.97-7.26]	3.85 [3.38-4.38]	2.35 [1.98-2.77]
	Other	4434	251	141	89	5.66 [5.08-6.3]	3.21 [2.78-3.71]	2.05 [1.71-2.46]
BMI	White	472750	23060	13769	7937	4.96 [4.9-5.03]	2.97 [2.92-3.02]	1.71 [1.67-1.75]
	Normal weight	183305	7631	4298	2341	4.3 [4.2-4.39]	2.41 [2.34-2.48]	1.3 [1.25-1.35]
	Obese	97597	6304	4014	2432	6.41 [6.25-6.56]	4.02 [3.89-4.14]	2.44 [2.34-2.54]
	Overweight	161456	7672	4506	2593	4.89 [4.78-5]	2.89 [2.81-2.97]	1.67 [1.6-1.73]
Healthcare or care home worker	Underweight	6273	293	180	122	4.99 [4.53-5.5]	2.96 [2.6-3.36]	1.69 [1.42-2]
	Yes	27690	2078	1292	758	7.45 [7.14-7.76]	4.65 [4.4-4.9]	2.61 [2.42-2.8]
Index of multiple deprivation (IMD) quintile	1 - most deprived	50260	3055	1950	1201	5.92 [5.77-6.07]	3.74 [3.62-3.86]	2.26 [2.17-2.35]
	2	80894	4424	2688	1661	5.42 [5.29-5.56]	3.18 [3.08-3.29]	1.89 [1.81-1.98]
	3	109928	5472	3294	1946	4.73 [4.6-4.86]	2.83 [2.73-2.93]	1.62 [1.55-1.7]
	4	126237	5805	3395	1900	4.47 [4.35-4.6]	2.59 [2.49-2.69]	1.41 [1.34-1.49]

	5 - least deprived	141388	6111	3483	1847	4.21 [4.08-4.34]	2.38 [2.28-2.47]	1.25 [1.19-1.32]
Current smoker	Current cigarette smoker	50983	3100	2001	1241	6.1 [5.91-6.3]	3.85 [3.69-4.01]	2.38 [2.25-2.5]
	Not current cigarette smoker	453681	21551	12680	7237	4.8 [4.74-4.86]	2.82 [2.77-2.87]	1.6 [1.56-1.64]
	Prefer not to say	4043	216	129	77	5.46 [4.83-6.16]	3.48 [2.99-4.06]	1.82 [1.47-2.25]
Current vaper	Current vaper	28436	1751	1075	651	6.34 [6.08-6.62]	3.95 [3.74-4.17]	2.25 [2.09-2.42]
	Not current vaper	477729	22988	13652	7851	4.86 [4.8-4.92]	2.87 [2.83-2.92]	1.65 [1.61-1.69]
	Prefer not to say	2542	128	83	53	5.13 [4.38-5.99]	3.33 [2.73-4.05]	1.77 [1.35-2.32]
Severity of COVID symptoms	Mild symptoms	27572	4316	2060	813	14.66 [14.27-15.06]	6.94 [6.66-7.22]	2.7 [2.52-2.89]
	Moderate symptoms	45357	12539	7228	4059	26.79 [26.39-27.19]	15.37 [15.04-15.69]	8.69 [8.44-8.95]
	Severe symptoms	19187	8012	5522	3683	40.51 [39.83-41.2]	28.15 [27.53-28.78]	18.43 [17.9-18.98]
Treatment sought/ received for COVID	No medical attention sought	58330	12756	6898	3525	20.74 [20.42-21.06]	11.14 [10.9-11.39]	5.72 [5.54-5.91]
	Sought medical attention from pharmacist / by phone (NHS 111/GP)	21388	7785	5126	3330	34.81 [34.2-35.42]	22.77 [22.23-23.31]	14.69 [14.24-15.15]
	Visited GP/walk-in centre/A&E	6792	2458	1590	958	34.97 [33.85-36.12]	23.12 [22.13-24.14]	13.31 [12.52-14.14]
	Hospital admission	739	393	292	208	51.01 [47.63-54.38]	38.14 [34.92-41.47]	26.7 [23.82-29.79]
Gross household income	0-14,999	32819	2061	1387	937	6.23 [5.99-6.48]	4.18 [3.98-4.38]	2.87 [2.71-3.05]
	15,000-49,999	148039	7601	4635	2692	5.27 [5.16-5.38]	3.21 [3.12-3.3]	1.84 [1.77-1.91]
	50,000-149,999	128957	6092	3359	1746	4.81 [4.69-4.93]	2.6 [2.51-2.69]	1.32 [1.26-1.38]
		32819	2061	1387	937	6.23 [5.99-6.48]	4.18 [3.98-4.38]	2.87 [2.71-3.05]
	>150,000	16279	751	385	180	4.61 [4.28-4.96]	2.25 [2.02-2.5]	1.01 [0.86-1.19]
Comorbidities	Organ transplant recipient	596	30	23	16	4.37 [3-6.33]	3.7 [2.45-5.53]	2.86 [1.79-4.53]
	Diabetes (type I or II)	27358	1417	957	656	5.18 [4.93-5.44]	3.45 [3.24-3.67]	2.49 [2.31-2.67]
	Heart disease or	21430	1349	926	646	6.36 [6.04-6.69]	4.3 [4.04-4.58]	2.99 [2.77-3.23]



heart problems							
Hypertension (high blood pressure)	65792	3525	2304	1468	5.46 [5.28-5.64]	3.61 [3.47-3.76]	2.37 [2.25-2.49]
Stroke	3352	189	131	96	5.55 [4.84-6.36]	3.57 [3.01-4.24]	2.72 [2.23-3.31]
Kidney disease	3867	294	209	133	7.7 [6.91-8.57]	5.02 [4.38-5.75]	3.31 [2.79-3.91]
Liver disease	2107	153	110	80	7.5 [6.47-8.68]	4.94 [4.11-5.93]	3.48 [2.79-4.33]
Anemia	7575	737	521	368	9.99 [9.37-10.64]	6.79 [6.28-7.34]	4.75 [4.32-5.21]
Asthma	51337	4021	2714	1757	7.81 [7.58-8.04]	5.26 [5.07-5.45]	3.34 [3.19-3.5]
Other lung condition	12500	1020	736	551	8.25 [7.79-8.73]	5.93 [5.54-6.35]	4.54 [4.2-4.91]
Cancer	8042	349	227	157	4.46 [4.02-4.94]	2.92 [2.57-3.32]	2.19 [1.89-2.54]
Condition affecting the brain and nerves	5415	348	271	202	6.55 [5.92-7.24]	4.94 [4.39-5.55]	3.57 [3.11-4.1]
A weakened immune system/reduced ability to deal with infections	19302	1651	1225	875	8.85 [8.45-9.27]	6.62 [6.27-6.99]	4.63 [4.34-4.95]
Depression	38125	3630	2533	1756	9.51 [9.23-9.8]	6.72 [6.48-6.97]	4.57 [4.37-4.77]
Anxiety	57140	5074	3433	2246	9.12 [8.89-9.35]	6.2 [6.01-6.4]	3.99 [3.84-4.15]
Psychiatric disorder	3385	347	249	189	9.86 [8.97-10.83]	6.89 [6.14-7.72]	5.21 [4.56-5.95]

**Table S2** Odds ratios for persistent symptoms at 12 weeks among symptomatic respondents, derived from logistic regression models (forest plot shown in Figure 2)

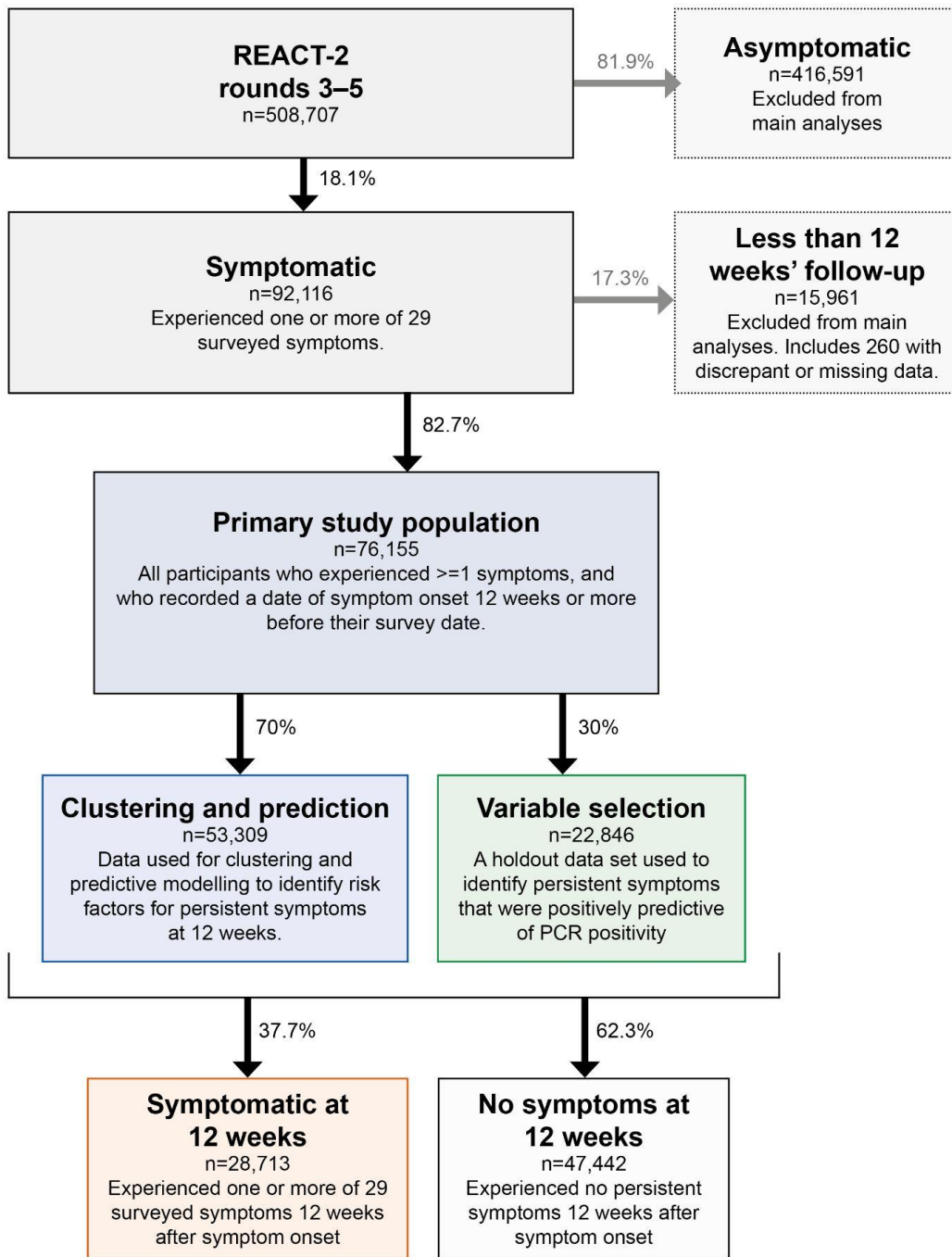
Predictor	Level	Crude (univariate) OR	Adjusted on age and sex
Sex	Male [reference]	-	-
	Female	1.46 [1.42-1.51]	1.51 [1.46-1.55]
Age group	18-24	0.58 [0.54-0.62]	0.56 [0.53-0.6]
	25-34	0.6 [0.57-0.63]	0.58 [0.55-0.61]
	35-44	0.65 [0.62-0.68]	0.64 [0.61-0.67]
	45-54	0.86 [0.83-0.9]	0.86 [0.82-0.9]
	55-64 [reference]	-	-
	65-74	1.16 [1.1-1.22]	1.16 [1.1-1.23]
	74+	1.51 [1.38-1.64]	1.47 [1.35-1.61]
Ethnicity	White [reference]	-	-
	Asian	0.71 [0.65-0.77]	0.8 [0.74-0.88]
	Black	0.99 [0.85-1.15]	1.06 [0.91-1.24]
	Mixed	1.05 [0.94-1.18]	1.19 [1.06-1.33]
	Other	0.99 [0.86-1.14]	1.05 [0.91-1.22]
BMI	Obese	1.52 [1.46-1.58]	1.53 [1.47-1.59]
	Overweight	1.14 [1.1-1.18]	1.16 [1.12-1.21]
	Normal weight [reference]	-	-
	Underweight	1.15 [1.01-1.33]	1.17 [1.02-1.35]
Healthcare or care home worker	No [reference]	-	-
	Yes	1.48 [1.39-1.56]	1.33 [1.26-1.41]
Index of multiple deprivation 1 - most deprived (IMD) quintile	1 - most deprived	1.19 [1.12-1.25]	1.24 [1.17-1.31]
	2	1.05 [1-1.1]	1.08 [1.03-1.13]
	3 [reference]	-	-
	4	0.94 [0.9-0.98]	0.92 [0.88-0.97]
	5 - least deprived	0.85 [0.82-0.89]	0.84 [0.8-0.87]
Current smoker	Not current cigarette smoker [reference]	-	-
	Current cigarette smoker	1.23 [1.17-1.29]	1.35 [1.28-1.41]
	Prefer not to say	0.92 [0.79-1.08]	1.03 [0.88-1.21]
Current vaper	Not current vaper [reference]	-	-
	Current vaper	1.12 [1.06-1.19]	1.26 [1.18-1.34]
	Prefer not to say	0.84 [0.69-1.03]	0.94 [0.77-1.16]
Treatment sought/ received for COVID	No medical attention sought [reference]	-	-
	Hospital admission	3.98 [3.37-4.69]	3.46 [2.93-4.09]
	Sought medical attention from pharmacist / by phone (NHS 111/GP)	2.07 [2-2.14]	2.05 [1.98-2.12]
	Visited GP/walk-in centre/A&E	1.81 [1.72-1.91]	1.67 [1.58-1.76]
Gross household income	0-14,999	1.51 [1.42-1.61]	1.39 [1.3-1.48]
	15,000-49,999 [reference]	-	-
	50,000-149,999	0.7 [0.67-0.72]	0.76 [0.73-0.79]
	>150,000	0.59 [0.54-0.64]	0.64 [0.58-0.69]

**Table S3** Characteristics of 12-week symptom clusters among 70% sub-sample (n=53,309) of people with symptoms and date of symptom onset 12 weeks or more before survey date -- compared with those who were not experiencing any persistent symptoms at 12 weeks. For categorical variables, column-wise within-group percentages shown in parentheses, with 95% confidence intervals in square brackets. Cluster L2 (respiratory) contained proportionately more people who were obese (31.6%, vs 27.0% in Cluster L1), more smokers (12.9% vs 12.2%) and more people who rated their symptoms as ‘severe’ (43.5% vs 27.4%). Prevalence of all surveyed comorbidities was higher in Cluster L2 than Cluster L1, and the prevalence of comorbidities was higher in the symptomatic clusters than in those who did not experience persistent symptoms at 12 weeks.

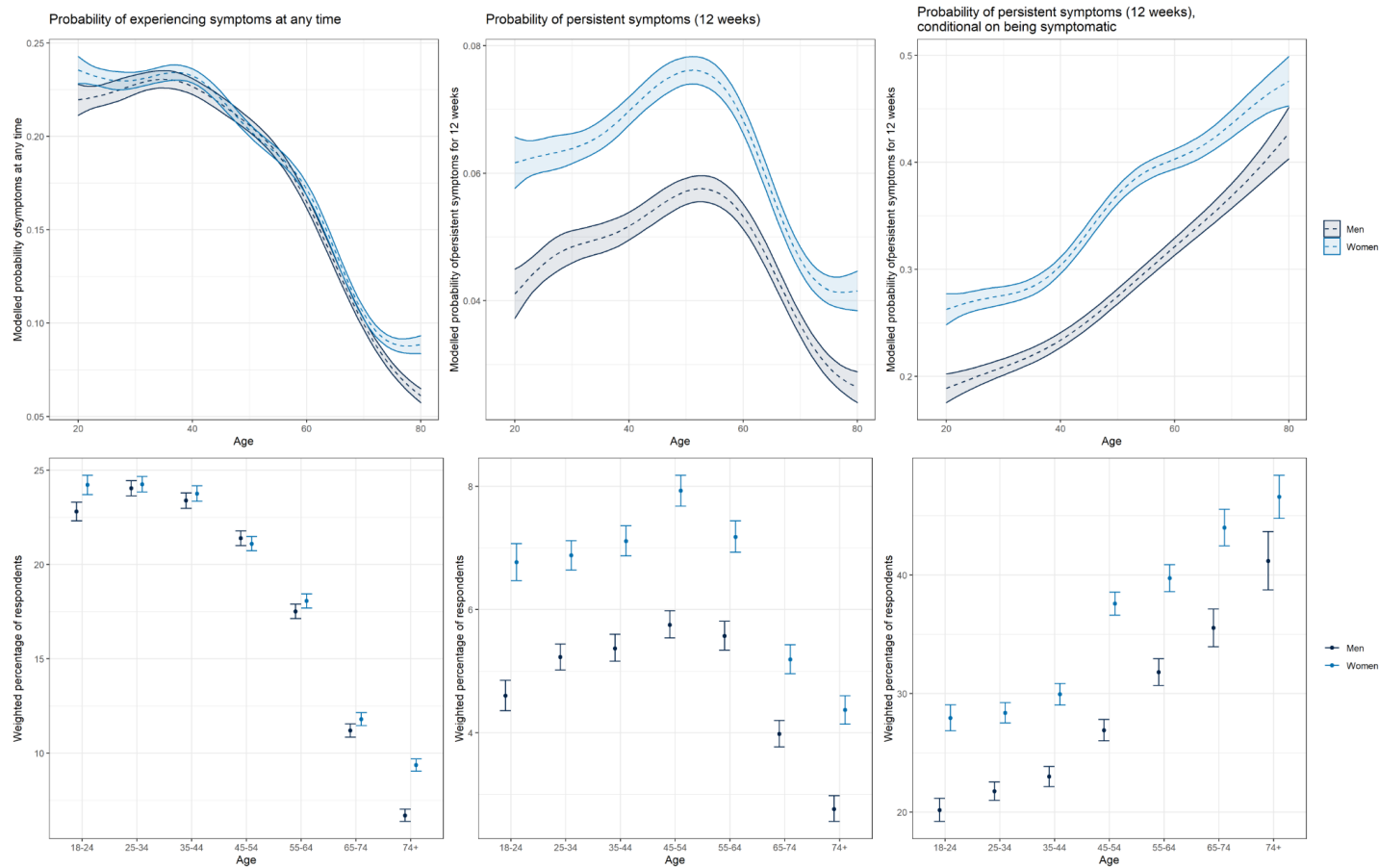
Variable	Category	Asymptomatic at 12 weeks	Cluster L1 (“tiredness”)	Cluster L2 (“respiratory”)	Total /mean (SD)
All participants	All participants	33069 (100%, [100-100])	15799 (100%, [100-100])	4441 (100%, [99.9-100])	53309
Sex	Women	17834 (53.9%, [53.4-54.5])	10041 (63.6%, [62.8-64.3])	2794 (62.9%, [61.5-64.3])	30669
	Men	15234 (46.1%, [45.5-46.6])	5758 (36.4%, [35.7-37.2])	1647 (37.1%, [35.7-38.5])	22639
Age group	18-24	2440 (7.4%, [7.1-7.7])	917 (5.8%, [5.5-6.2])	185 (4.2%, [3.6-4.8])	3542
	25-34	5911 (17.9%, [17.5-18.3])	2108 (13.3%, [12.8-13.9])	547 (12.3%, [11.4-13.3])	8566
	35-44	7478 (22.6%, [22.2-23.1])	2891 (18.3%, [17.7-18.9])	836 (18.8%, [17.7-20])	11205
	45-54	7476 (22.6%, [22.2-23.1])	3687 (23.3%, [22.7-24])	1108 (24.9%, [23.7-26.2])	12271
	55-64	6021 (18.2%, [17.8-18.6])	3492 (22.1%, [21.5-22.8])	1035 (23.3%, [22.1-24.6])	10548
	65-74	2971 (9%, [8.7-9.3])	2020 (12.8%, [12.3-13.3])	544 (12.2%, [11.3-13.2])	5535
Ethnicity	74+	772 (2.3%, [2.2-2.5])	684 (4.3%, [4-4.7])	186 (4.2%, [3.6-4.8])	1642
	Asian	1337 (4.1%, [3.9-4.3])	464 (3%, [2.7-3.2])	98 (2.2%, [1.8-2.7])	1899
	Black	319 (1%, [0.9-1.1])	168 (1.1%, [0.9-1.2])	34 (0.8%, [0.6-1.1])	521
	Mixed	524 (1.6%, [1.5-1.7])	262 (1.7%, [1.5-1.9])	92 (2.1%, [1.7-2.6])	878
	Other	354 (1.1%, [1-1.2])	164 (1%, [0.9-1.2])	42 (1%, [0.7-1.3])	560
	White	30275 (92.3%, [92-92.6])	14614 (93.2%, [92.8-93.6])	4146 (94%, [93.2-94.6])	49035
BMI	Normal weight	11981 (41.2%, [40.6-41.8])	5071 (36.5%, [35.7-37.3])	1241 (31.8%, [30.4-33.3])	18293
	Obese	6245 (21.5%, [21-21.9])	3756 (27%, [26.3-27.8])	1234 (31.6%, [30.2-33.1])	11235
	Overweight	10476 (36%, [35.5-36.6])	4898 (35.3%, [34.5-36.1])	1365 (35%, [33.5-36.5])	16739
	Underweight	381 (1.3%, [1.2-1.4])	166 (1.2%, [1-1.4])	63 (1.6%, [1.3-2.1])	610
Healthcare or care home worker	Yes	2109 (8.5%, [8.2-8.9])	1260 (11.8%, [11.2-12.4])	358 (12.2%, [11.1-13.4])	3727
Index of multiple deprivation (IMD) quintile	1 - most deprived	3244 (9.8%, [9.5-10.1])	1845 (11.7%, [11.2-12.2])	586 (13.2%, [12.2-14.2])	5675
	2	5463 (16.5%, [16.1-16.9])	2739 (17.3%, [16.8-17.9])	794 (17.9%, [16.8-19])	8996
	3	7123 (21.5%, [21.1-22])	3504 (22.2%, [21.5-22.8])	1008 (22.7%, [21.5-24])	11635
	4	8004 (24.2%, [23.7-24.7])	3790 (24%, [23.3-24.7])	1000 (22.5%, [21.3-23.8])	12794
	5 - least deprived	9235 (27.9%, [27.4-28.4])	3921 (24.8%, [24.2-25.5])	1053 (23.7%, [22.5-25])	14209
Current smoker	Current cigarette smoker	3430 (10.4%, [10-10.7])	1935 (12.2%, [11.7-12.8])	573 (12.9%, [11.9-13.9])	5938

	Not current cigarette smoker	29321 (88.7%, [88.3-89])	13731 (86.9%, [86.4-87.4])	3826 (86.2%, [85.1-87.1])	46878
	Prefer not to say	318 (1%, [0.9-1.1])	133 (0.8%, [0.7-1])	42 (0.9%, [0.7-1.3])	493
Current vaper	Current vaper	2042 (6.2%, [5.9-6.4])	1083 (6.9%, [6.5-7.3])	328 (7.4%, [6.7-8.2])	3453
	Not current vaper	30831 (93.2%, [93-93.5])	14639 (92.7%, [92.2-93.1])	4086 (92%, [91.2-92.8])	49556
	Prefer not to say	196 (0.6%, [0.5-0.7])	77 (0.5%, [0.4-0.6])	27 (0.6%, [0.4-0.9])	300
Severity of COVID symptoms	Mild symptoms	11133 (33.7%, [33.2-34.2])	3345 (21.2%, [20.5-21.8])	476 (10.7%, [9.8-11.7])	14954
	Moderate symptoms	16378 (49.5%, [49-50.1])	8130 (51.5%, [50.7-52.2])	2034 (45.8%, [44.3-47.3])	26542
	Severe symptoms	5558 (16.8%, [16.4-17.2])	4324 (27.4%, [26.7-28.1])	1931 (43.5%, [42.0-44.9])	11813
Treatment sought/ received for COVID	Hospital admission	149 (0.5%, [0.4-0.6])	163 (1.1%, [0.9-1.3])	122 (2.9%, [2.5-3.5])	434
	No medical attention sought	22687 (72%, [71.5-72.5])	8811 (59.2%, [58.4-59.9])	1758 (42.3%, [40.8-43.8])	33256
	Sought medical attention from pharmacist / by phone (NHS 111/GP)	6355 (20.2%, [19.7-20.6])	4525 (30.4%, [29.6-31.1])	1674 (40.3%, [38.8-41.7])	12554
	Visited GP/walk-in centre/A&E	2324 (7.4%, [7.1-7.7])	1397 (9.4%, [8.9-9.9])	605 (14.5%, [13.5-15.7])	4326
Gross household income	>150,000	1516 (6.8%, [6.5-7.2])	509 (4.9%, [4.5-5.3])	109 (3.7%, [3.1-4.4])	2134
	0-14,999	1545 (7%, [6.6-7.3])	1222 (11.7%, [11.1-12.3])	405 (13.8%, [12.6-15.1])	3172
	15,000-49,999	8858 (39.9%, [39.2-40.5])	4743 (45.4%, [44.5-46.4])	1380 (46.9%, [45.1-48.7])	14981
	50,000-149,999	10296 (46.3%, [45.7-47])	3968 (38%, [37.1-38.9])	1049 (35.6%, [33.9-37.4])	15313
Comorbidities	Organ transplant recipient	26 (0.1%, [0.1-0.1])	14 (0.1%, [0.1-0.1])	9 (0.2%, [0.1-0.4])	49
	Diabetes (type I or II)	1171 (3.5%, [3.3-3.7])	843 (5.3%, [5-5.7])	258 (5.8%, [5.2-6.5])	2272
	Heart disease or heart problems	914 (2.8%, [2.6-2.9])	777 (4.9%, [4.6-5.3])	288 (6.5%, [5.8-7.2])	1979
	Hypertension (high blood pressure)	3004 (9.1%, [8.8-9.4])	2165 (13.7%, [13.2-14.2])	644 (14.5%, [13.5-15.6])	5813
	Kidney disease	146 (0.4%, [0.4-0.5])	106 (0.7%, [0.6-0.8])	47 (1.1%, [0.8-1.4])	299
	Liver disease	158 (0.5%, [0.4-0.6])	159 (1%, [0.9-1.2])	54 (1.2%, [0.9-1.6])	371
	Anemia	87 (0.3%, [0.2-0.3])	84 (0.5%, [0.4-0.7])	32 (0.7%, [0.5-1])	203
	Asthma	381 (1.2%, [1-1.3])	432 (2.7%, [2.5-3])	145 (3.3%, [2.8-3.8])	958
	Other lung condition	3264 (9.9%, [9.6-10.2])	2001 (12.7%, [12.2-13.2])	1113 (25.1%, [23.8-26.4])	6378
	Cancer	347 (1%, [0.9-1.2])	435 (2.8%, [2.5-3])	362 (8.2%, [7.4-9])	1144
	Condition affecting the brain and nerves	284 (0.9%, [0.8-1])	208 (1.3%, [1.2-1.5])	65 (1.5%, [1.2-1.9])	557
	A weakened immune system/reduced ability to deal with infections	220 (0.7%, [0.6-0.8])	211 (1.3%, [1.2-1.5])	65 (1.5%, [1.2-1.9])	496
	Depression	876 (2.6%, [2.5-2.8])	917 (5.8%, [5.5-6.2])	364 (8.2%, [7.4-9])	2157
	Anxiety	2139 (6.5%, [6.2-6.7])	2057 (13%, [12.5-13.6])	695 (15.6%, [14.6-16.7])	4891
	Psychiatric disorder	3504 (10.6%, [10.3-10.9])	2940 (18.6%, [18-19.2])	998 (22.5%, [21.3-23.7])	7442
	None of these	189 (0.6%, [0.5-0.7])	188 (1.2%, [1-1.4])	76 (1.7%, [1.4-2.1])	453

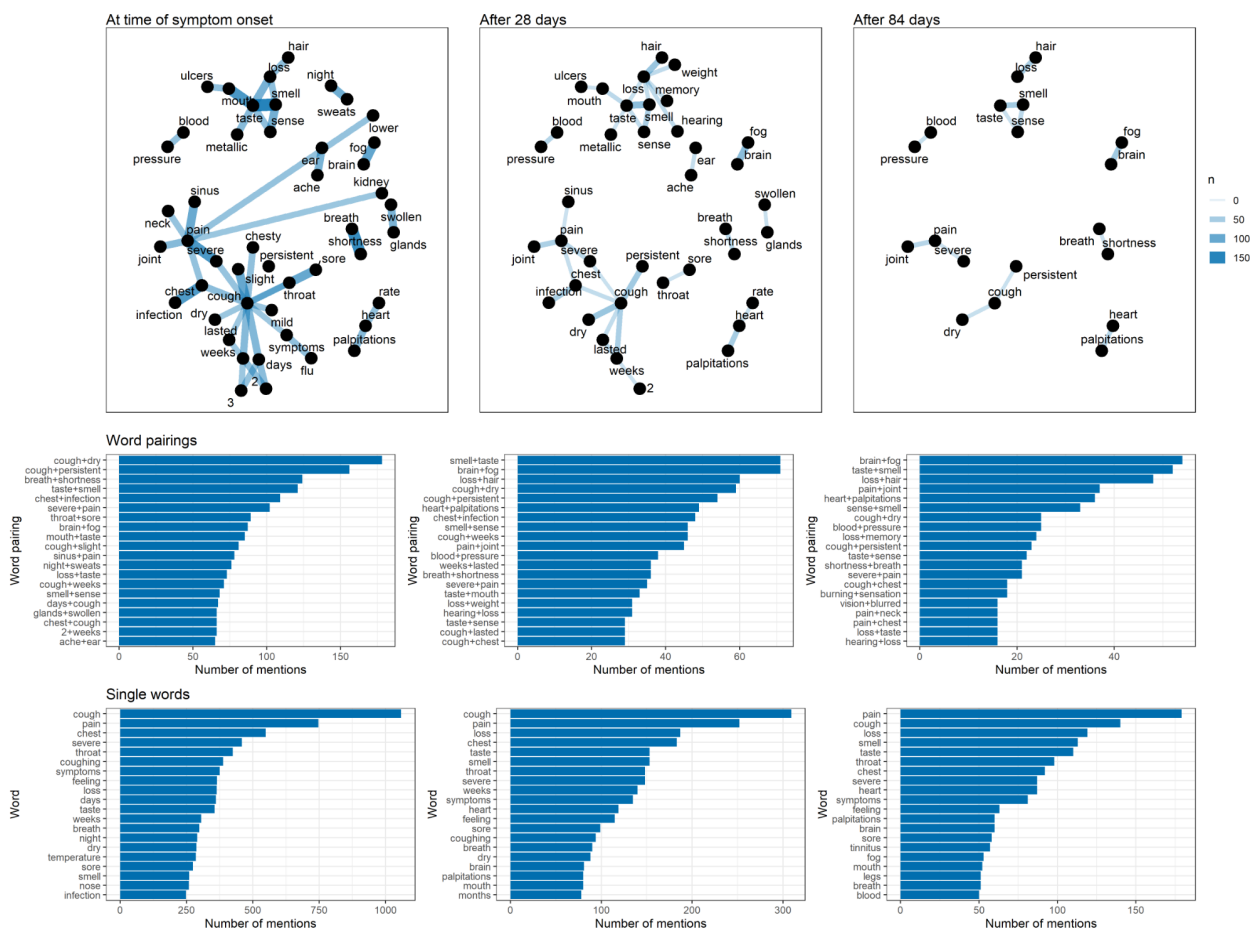
**Figure S1** Study population flow chart



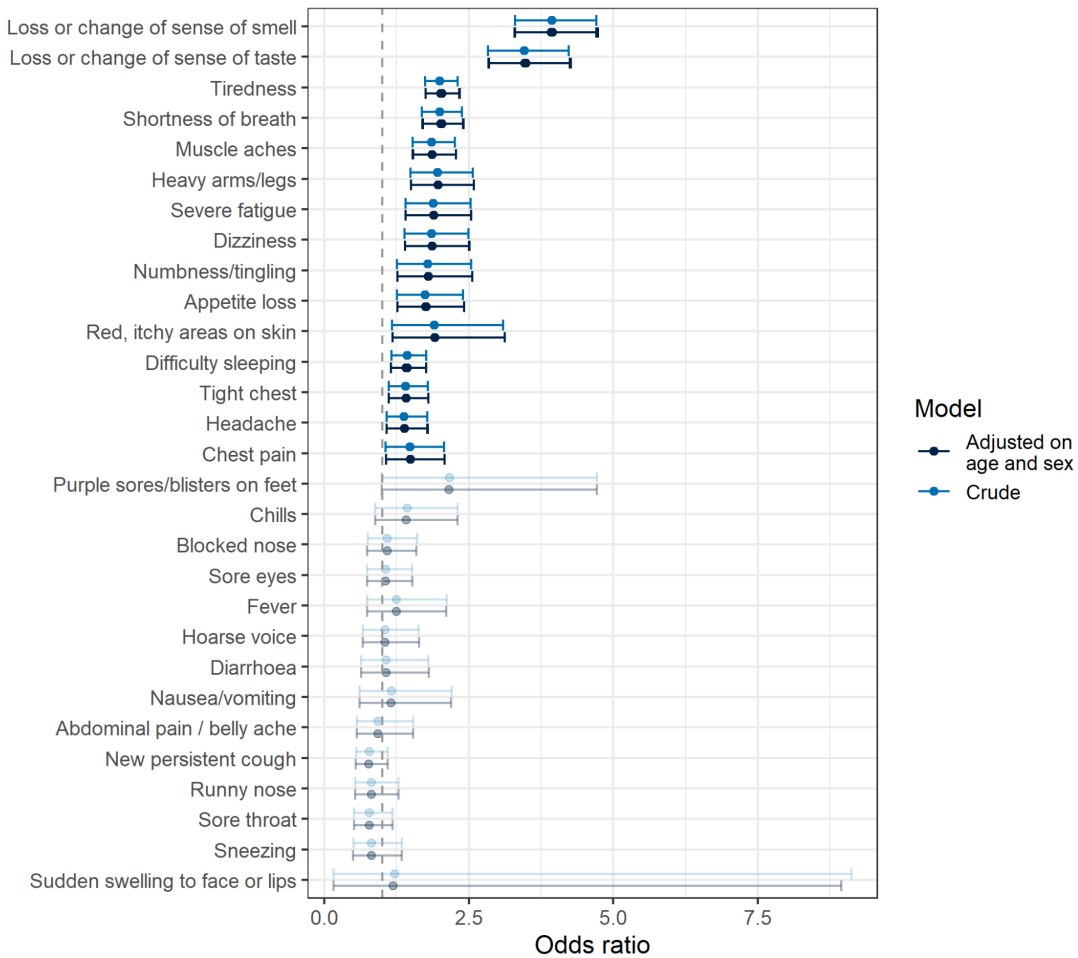
**Figure S2** Plots showing (top row) modelled probability of persistent symptoms at 12 weeks as a function of age and sex, using generalised additive models with splines on age and interactions between age and sex (top row), and (bottom row) weighted prevalence of persistent symptoms at 12 weeks. From left to right, plots show (i) probability of experiencing COVID-19 symptoms in the population, (ii) probability of experiencing persistent symptoms at 12 weeks in the population, and (iii) probability of experiencing persistent symptoms at 12 weeks, conditional on symptomatic infection. Conditional on symptomatic infection (likelihood of which decreases after age 50, see central panels), risk of persistent symptoms increases linearly with age (right panels).



**Figure S3** Analysis of the free text responses to COVID-19 symptom questions in REACT-2. A total of 8,374 respondents left free-text responses, of whom 1,860 reported the described symptoms persisting for 12 weeks or more. From left to right, the panels show (i) analysis of symptoms occurring at time of symptom onset; (ii) analysis of symptoms occurring at 4 weeks after symptom onset, and (iii) analysis of symptoms occurring at 12 weeks after symptom onset. The top two rows of panels visualise within-response word co-occurrence, as networks (top panels) and simple bar plots of the top 20 most frequently co-occurring pairs of words (middle panels). The bottom row shows the most commonly occurring single words.

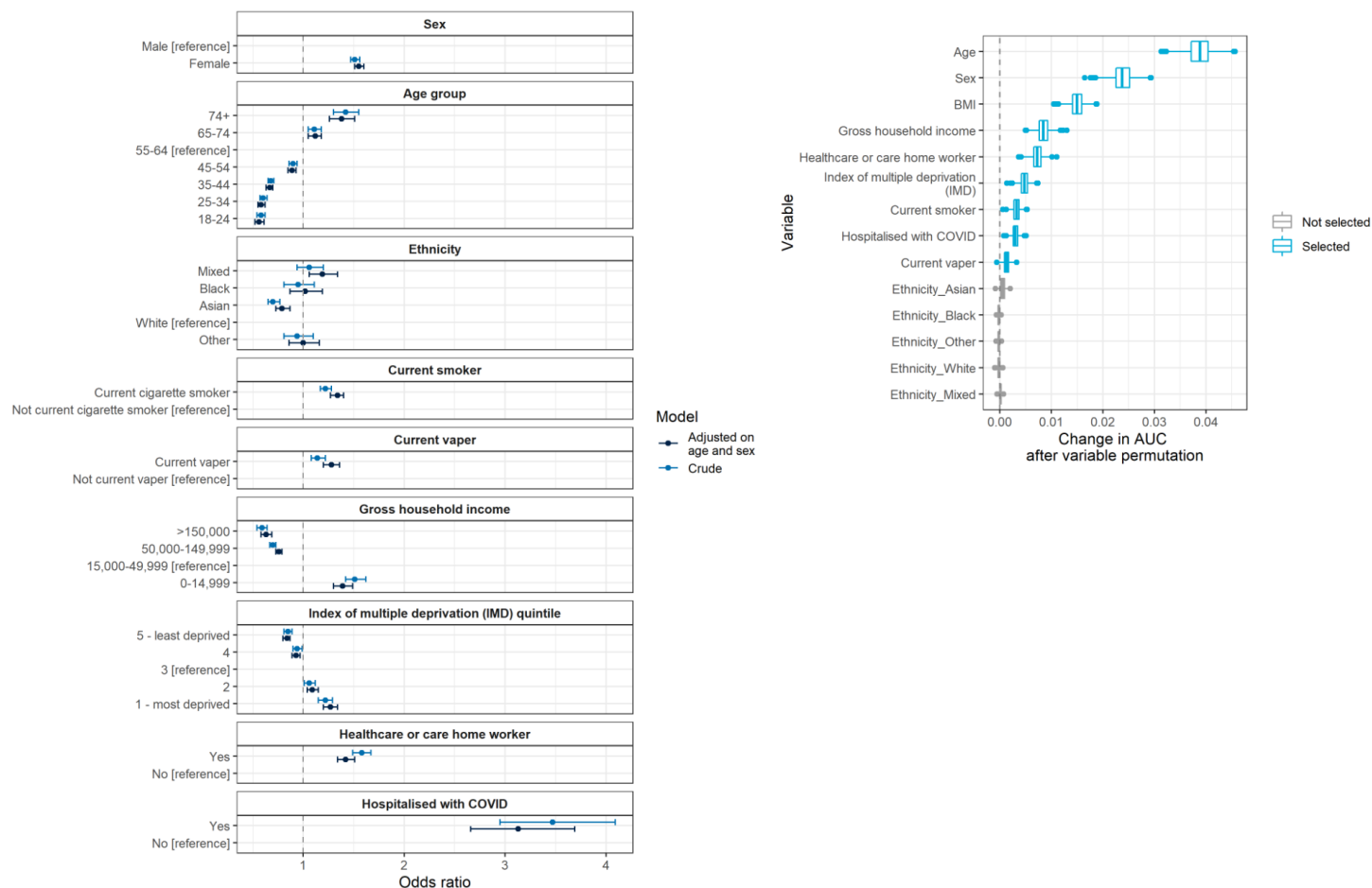


**Figure S4** Results of logistic model with self-reported PCR positivity as the binary outcome variable, and the 29 surveyed symptoms as binary independent variables, among a 30% variable selection holdout data set. The 15 variables that were identified as positively predictive of PCR positivity (shown in bold colours) were taken forward as a sensitivity analysis of prevalence of persistent symptoms beyond 12 weeks, and as outcome variables in logistic regression of risk factors for persistence, among this reduced symptom set.





**Figure S5** Plots showing the results of predictive modelling, with one or more of a subset of 15 symptoms at 12 weeks (y/n) as the binary outcome variable. Left panel shows odds ratios from logistic regression models, both crude and adjusted on sex and age. Right panel shows the mean contribution in area under the curve (AUC) that each variable makes to a multivariable boosted tree model, derived using variable permutation. Age, sex, BMI, smoking status and prior hospitalisation with COVID-19 are the strongest predictors of persistent symptoms in multivariable modelling, while Asian ethnicity is associated with a lower risk of persistent symptoms at 12 weeks; Asian ethnicity is selected as a predictor in multivariable boosted tree modelling.



**Figure S6** Pairwise co-occurrence heatmaps of symptoms at time of symptom onset (left) and 12 weeks after symptom onset (right), among 76,155 symptomatic participants. Deeper blue colour indicates that more people had both of the paired symptoms at symptom onset or 12 weeks post symptom onset; bar plots indicate marginal counts of participants with each symptom at symptom onset and at 12 weeks post symptom onset.

