



OPEN

## Distinct clinical symptom patterns in patients hospitalised with COVID-19 in an analysis of 59,011 patients in the ISARIC-4C study

Jonathan E. Millar<sup>1</sup>, Lucile Neyton<sup>1</sup>, Sohan Seth<sup>2</sup>, Jake Dunning<sup>3,4</sup>, Laura Merson<sup>5,6</sup>, Srinivas Murthy<sup>7</sup>, Clark D. Russell<sup>8</sup>, Sean Keating<sup>9</sup>, Maaïke Swets<sup>1,10</sup>, Carole H. Sudre<sup>11</sup>, Timothy D. Spector<sup>12</sup>, Sebastien Ourselin<sup>11</sup>, Claire J. Steves<sup>12</sup>, Jonathan Wolf<sup>13</sup>, Annemarie B. Docherty<sup>14</sup>, Ewen M. Harrison<sup>14</sup>, Peter J. M. Openshaw<sup>4</sup>, Malcolm G. Semple<sup>15</sup>✉, J. Kenneth Baillie<sup>1</sup>✉ & ISARIC-4C \*

COVID-19 is clinically characterised by fever, cough, and dyspnoea. Symptoms affecting other organ systems have been reported. However, it is the clinical associations of different patterns of symptoms which influence diagnostic and therapeutic decision-making. In this study, we applied clustering techniques to a large prospective cohort of hospitalised patients with COVID-19 to identify clinically meaningful sub-phenotypes. We obtained structured clinical data on 59,011 patients in the UK (the ISARIC Coronavirus Clinical Characterisation Consortium, 4C) and used a principled, unsupervised clustering approach to partition the first 25,477 cases according to symptoms reported at recruitment. We validated our findings in a second group of 33,534 cases recruited to ISARIC-4C, and in 4,445 cases recruited to a separate study of community cases. Unsupervised clustering identified distinct sub-phenotypes. First, a core symptom set of fever, cough, and dyspnoea, which co-occurred with additional symptoms in three further patterns: fatigue and confusion, diarrhoea and vomiting, or productive cough. Presentations with a single reported symptom of dyspnoea or confusion were also identified, alongside a sub-phenotype of patients reporting few or no symptoms. Patients presenting with gastrointestinal symptoms were more commonly female, had a longer duration of symptoms before presentation, and had lower 30-day mortality. Patients presenting with confusion, with or without core symptoms, were older and had a higher unadjusted mortality. Symptom sub-phenotypes were highly consistent in replication analysis within the ISARIC-4C study. Similar patterns were externally verified in patients from a study of self-reported symptoms of mild disease. The large scale of the ISARIC-4C study enabled robust, granular discovery and replication. Clinical interpretation is

<sup>1</sup>Division of Functional Genetics and Development, Roslin Institute, University of Edinburgh, Easter Bush Campus, Midlothian, Edinburgh EH25 9RG, UK. <sup>2</sup>Institute for Adaptive and Neural Computation, School of Informatics, University of Edinburgh, Edinburgh, UK. <sup>3</sup>National Infection Service, Public Health England, London, UK. <sup>4</sup>National Heart and Lung Institute, Imperial College London, London, UK. <sup>5</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, ISARIC Global Support Centre, University of Oxford, Oxford, UK. <sup>6</sup>Infectious Diseases Data Observatory, Centre for Tropical Medicine and Global Health, University of Oxford, Oxford, UK. <sup>7</sup>BC Children's Hospital, University of British Columbia, Vancouver, Canada. <sup>8</sup>Centre for Inflammation Research, The Queen's Medical Research Institute, University of Edinburgh, Edinburgh, UK. <sup>9</sup>Intensive Care Unit, Royal Infirmary of Edinburgh, Edinburgh, UK. <sup>10</sup>Department of Infectious Diseases, Leiden University Medical Center, Leiden, The Netherlands. <sup>11</sup>School of Biomedical and Imaging Sciences, King's College London, London, UK. <sup>12</sup>Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. <sup>13</sup>ZOE Global Ltd, London, UK. <sup>14</sup>Centre for Medical Informatics, Usher Institute, University of Edinburgh, Edinburgh, UK. <sup>15</sup>NIHR Health Protection Research Unit in Emerging and Zoonotic Infections, Institute of Infection, Veterinary and Ecological Sciences, Faculty of Health and Life Sciences, University of Liverpool, Liverpool, UK. \*These authors contributed equally: Jonathan E. Millar and Lucile Neyton. ✉A list of authors and their affiliations appears at the end of the paper. ✉email: M.G.Semple@liverpool.ac.uk; j.k.baillie@ed.ac.uk

necessary to determine which of these observations have practical utility. We propose that four sub-phenotypes are usefully distinct from the core symptom group: gastro-intestinal disease, productive cough, confusion, and pauci-symptomatic presentations. Importantly, each is associated with an in-hospital mortality which differs from that of patients with core symptoms.

Coronavirus-19 disease (COVID-19) is characterised by a triad of symptoms: cough, fever and dyspnoea. However, it is clear that COVID-19 is not a homogeneous clinical entity. Important biological differences exist between sub-phenotypes, as is seen in other forms of critical illness including sepsis<sup>1</sup>, pancreatitis<sup>2</sup>, and dengue<sup>3</sup>. Remarkably, this is already evident for COVID-19: highly significant sub-group effects have been seen in trials demonstrating the efficacy of dexamethasone<sup>4,5</sup> and tocilizumab<sup>6</sup>, and progress has been made to uncover their underlying biology<sup>7</sup>. Similarly, genome-wide association studies have identified a genetic basis for differences in the host response to SARS-CoV-2 infection<sup>8</sup>.

The recognition of clinical similarities between patients is a fundamental unit of medical progress. Grouping patients enables us to select appropriate diagnostic tests, predict response to therapy, and to prognosticate. Symptoms of disease are the most basic characteristics available to clinicians for stratifying patients with any illness. Simple machine learning methods can reveal patterns in clinical symptoms<sup>9</sup> with diagnostic and therapeutic relevance<sup>10</sup>.

The International Severe Acute Respiratory Infection Consortium Clinical Characterisation Study (ISARIC-4C) is an ongoing study of patients admitted to more than 260 acute hospitals in the United Kingdom. We employed unsupervised machine learning techniques in a large, prospective cohort of patients with SARS-CoV-2, admitted during the ‘first wave’, to better characterise the symptoms of COVID-19 and to identify sub-phenotypes based on symptoms.

## Results

For the initial cohort, we studied the first 33,468 patients enrolled to the ISARIC-4C study, of which 7,991 had fully missing symptom data and were excluded from our analysis (Figure S1). The baseline characteristics of included patients ( $n = 25,477$ ) are detailed in Table S1. Overall, the median age was 73 (59–83) years and the majority of patients were male ( $n = 15,046$  (59%)). On average, individuals presented to hospital 4 (1–8) days after the onset of symptoms.

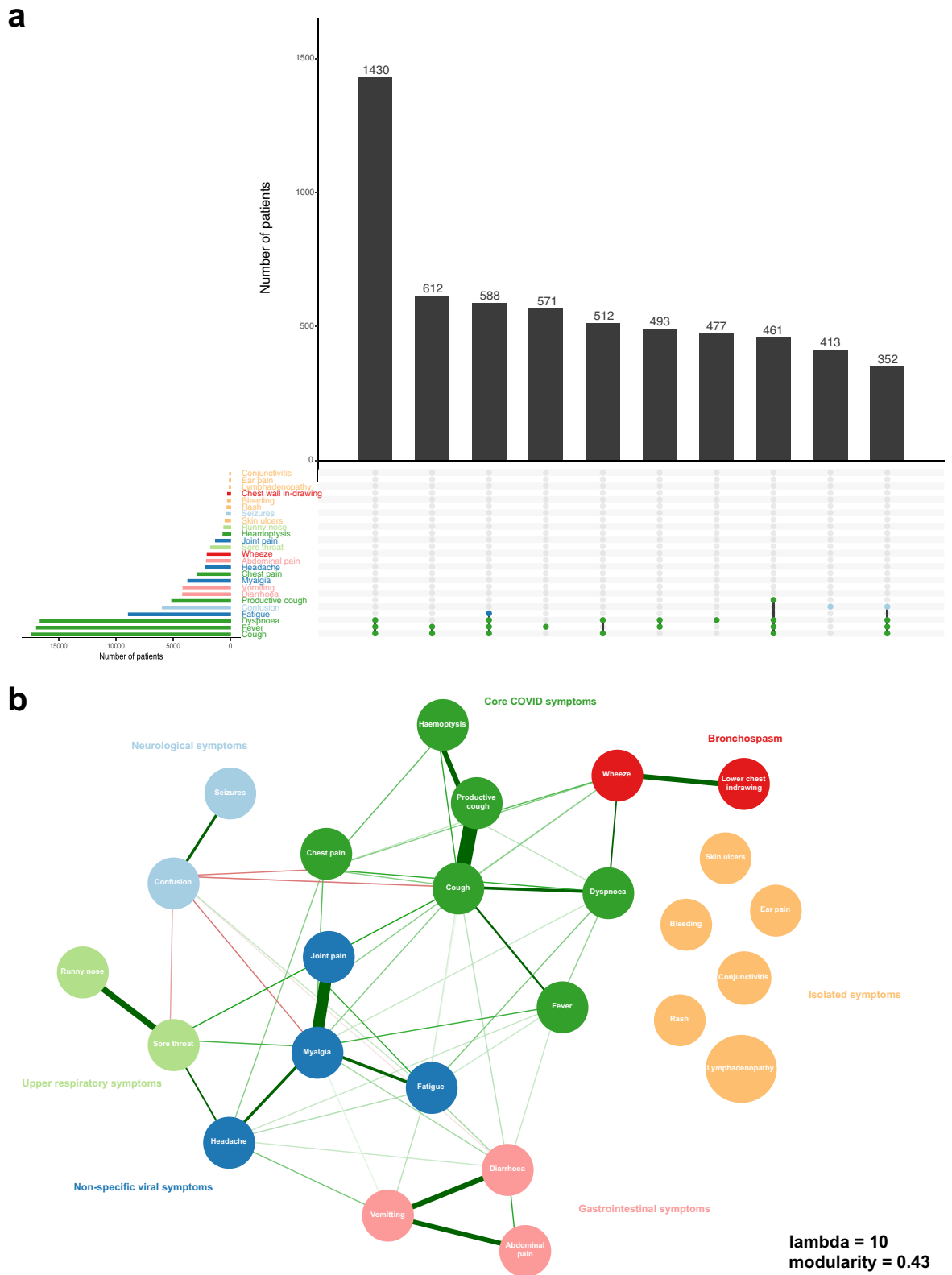
**Symptom prevalence and the relationship between symptoms.** Cough was the most prevalent symptom (68.0%, 95% CI 67.5–68.6%), followed by fever (66.4%, 95% CI 65.8–67%), and dyspnoea (65.2%, 95% CI 64.6–65.8%) (Fig. 1a and Table S2). Furthermore, these were the only symptoms to be reported by more than half of participants. The prevalence of individual symptoms varied with age (Figure S2). Fever was less marked at the extremes of age, an observation which was also evident for dyspnoea, and, with the exception of those aged > 90 years, for cough. Similarly, rash and runny nose were limited mostly to those aged < 20 years, especially to those aged under 10 years. In sum, there were 4,335 unique symptom combinations in the cohort, the most frequent being fever, cough, and dyspnoea ( $n = 1,430$  (5.6%)) (Fig. 1a).

To explore the relationships between symptoms, we fit an Ising model, employing L1-regularised logistic regression. The majority of symptoms exhibited some degree of conditional dependence with at least one other, however, there were several that occurred independently: skin ulcers, rash, bleeding, lymphadenopathy, ear pain, and conjunctivitis (Fig. 1b). All of which had a low prevalence (< 5%). Uniquely, confusion was negatively associated with cough, myalgia, sore throat, and diarrhoea. Groupings of symptoms with interconnected, positive conditional dependencies were appreciable from inspection of the network graph. To formalise communities, we used a short random walk algorithm. Excluding the 6 orphan symptoms, 6 distinct communities were identified. These include: core COVID-19 (fever, cough, and dyspnoea), upper respiratory, bronchospasm, gastrointestinal (GI), neurological, and non-specific viral symptom sets (Fig. 1b).

**Symptom sub-phenotype derivation.** To identify symptom sub-phenotypes within the study cohort, we performed unsupervised partitional clustering. An a priori assessment suggested that 7 clusters was the optimal solution by majority assessment. This combined the inflection points in the decline in total sum of squares and the rise in gap statistic, with the nearest peak in average silhouette width (Figure S3).

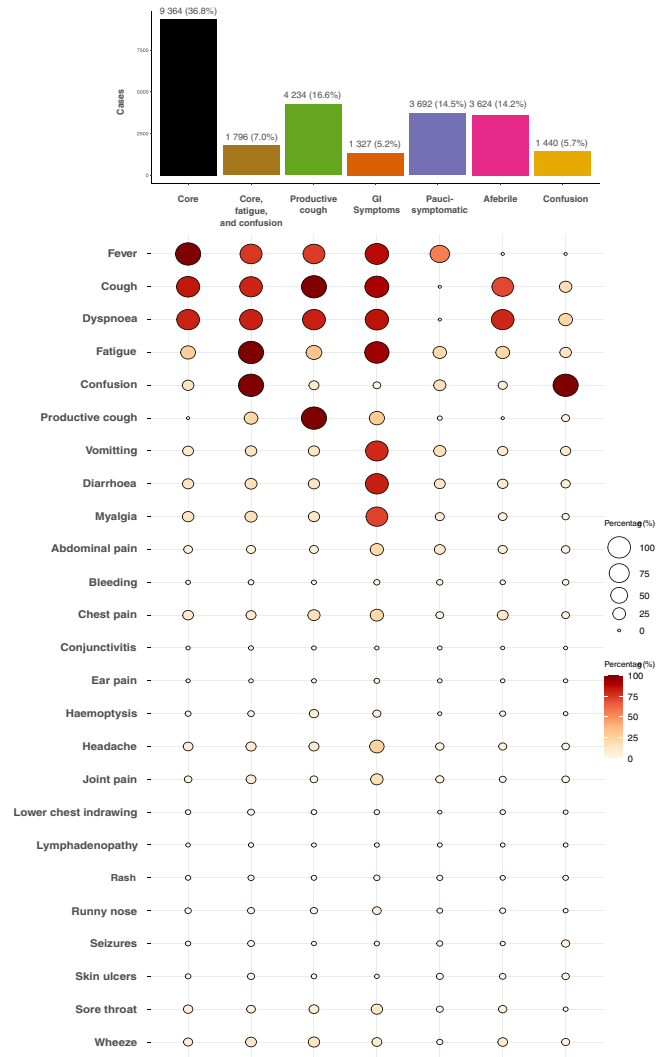
The patterns of symptoms within the seven clusters are shown in Fig. 2a. Based on the central (medoid) case, we characterised the sub-phenotypes as: core COVID-19 symptoms (fever, cough, and dyspnoea); core symptoms plus fatigue and confusion; productive cough; gastrointestinal (GI) symptoms; pauci-symptomatic (no single symptom being present in > 50% of the cluster membership); afebrile; and confusion. The core symptoms sub-phenotype accounted for the largest number of patients ( $n = 9,364$  (36.8%)) and the GI symptoms sub-phenotype the fewest ( $n = 1,327$  (5.2%)). Measures of internal validity and stability are presented in Figure S4a.

**Sub-phenotype sensitivity analyses.** To examine the implications of our handling of missing data, we performed a sensitivity analysis by clustering only cases with fully complete symptom data ( $n = 12,712$  (49.9%)). This analysis retained the cluster structure, except for the afebrile sub-phenotype, in which the medoid case exhibited dyspnoea alone (Figure S4b). The simple agreement between iterations was 89.2%, with a Cohen’s kappa of 0.86 ( $p < 0.001$ ). Patients with fully missing data ( $n = 7,991$ ) were of a similar age (74 (59–84) years) and sex (57.6% male). We conducted three additional analyses to explore the contribution of; age, heterogeneity of symptom recording, and the time from symptom onset to study enrolment. For age, we replicated our clustering pipeline

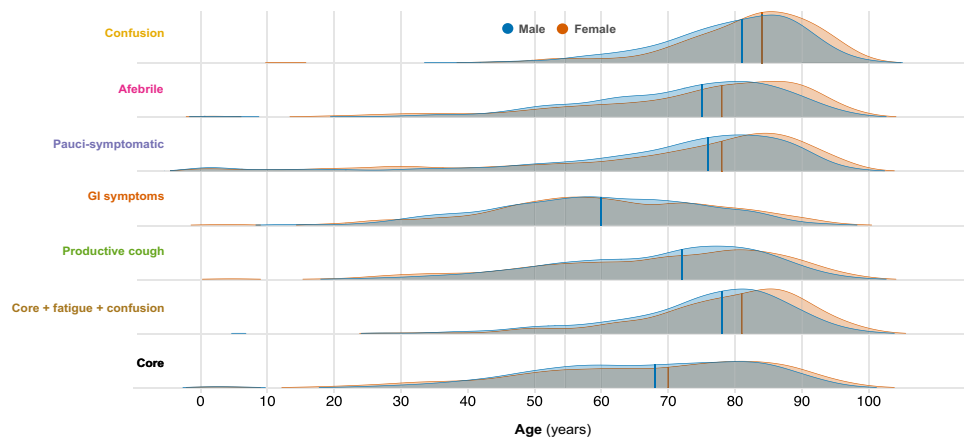


**Figure 1.** Symptom prevalence and relationships. (a) Symptom prevalence. Upset plot, intersections describe the ‘top 10’ symptom combinations within the cohort. The upper graph charts the total number of patients exhibiting these symptom sets. The lower graph charts the total number of patients with each symptom in the cohort. (b) Symptom network graph, derived using eLasso. Lines between symptom nodes illustrate conditional dependencies. A thicker line width and darker hue represents a stronger positive conditional dependence. Red lines represent a negative conditional dependence.

**a**



**b**



**Figure 2.** Symptom clusters. **(a)** Cluster identities, proportions, and patterns. Data are presented as count (percentage). **(b)** Distribution of age by symptom cluster. Density plots, solid lines represent the median age.

in the subset of the primary cohort aged  $\geq 70$  years (Table S3). In this instance, in the absence of confusion as a discriminatory symptom, we remained able to retrieve the core, productive cough, GI symptoms, pauci-symptomatic, and afebrile sub-phenotypes. For heterogeneity of symptom recording, we grouped the clustered primary cohort by recruiting site (sites recruiting  $\geq 50$  patients, 122 sites,  $n = 24,279$ ), and analysed the proportion of patients reporting each of the 9 most discriminatory symptoms by site (Figure S5). All, except for myalgia, had a co-efficient of variation  $< 0.35$ . Finally, for time from symptom onset to study enrolment, we divided the clustered primary cohort into its deciles. We then visualised the proportion of cases assigned to each sub-phenotype per decile (Figure S6). All sub-phenotypes were present in each decile.

**Sub-phenotype replication.** We replicated our clustering in two independent cohorts. First, we repeated our clustering approach on a cohort of patients enrolled to the ISARIC CCP-UK after those in the primary analysis and up until 7<sup>th</sup> July, 2020 ( $n = 35,446$ ). Of these, 1,912 (5.4%) had fully missing symptom data and were excluded from analysis. Clustering returned identical sub-phenotypes, with the exception of the afebrile cluster, in which cough was no longer implicated (Figure S7). This cluster was also reduced in relative size (14.2% to 6.7%). Second, our clustering approach was replicated in an outpatient cohort (COVID Symptom Study,  $n = 4,445$ ). This study records several overlapping or closely associated symptoms. Despite differences in study design and population, similar sub-phenotypes are discernible, including GI (cluster 1), pauci-symptomatic (cluster 2), and confusion (cluster 5) (Figure S8).

**Association of sub-phenotypes with patient characteristics.** Compared to the primary cohort average (73 years (59–83)), those in the GI symptoms sub-phenotype were younger (60 years (49–72)), while those in the core symptoms, fatigue, and confusion (79 years (70–85)) or confusion sub-phenotypes (82 years (75–88)), were older (Fig. 2b). The GI symptoms sub-phenotype also had the highest proportion of female patients (628 (47%)). These differences were accompanied by variations in the median time from symptom onset to hospital admission between sub-phenotypes ( $p < 0.001$ ) (Fig. 3a), and similarly, in the burden of co-morbidity (Fig. 3b).

To quantify these differences, we used a multinomial logistic regression model. Taking the largest sub-phenotype (core symptoms) as our reference, we included major demographic variables and co-morbidities in the analysis (Fig. 3c and Table S3). By comparison, those in the GI symptoms (RRR 0.66, 95% CI 0.60 to 0.73,  $p < 0.001$ ), afebrile (RRR 0.84, 95% CI 0.76 to 0.92,  $p < 0.001$ ), confusion (RRR 0.79, 95% CI 0.70 to 0.90,  $p < 0.001$ ), and pauci-symptomatic (RRR 0.78, 95% CI 0.72 to 0.86,  $p < 0.001$ ) sub-phenotypes, were more likely to be female (Fig. 3c). Patients assigned to the productive cough sub-phenotype were more likely to suffer from chronic pulmonary diseases (RRR 2.07, 95% CI 1.84 to 2.33,  $p < 0.001$ ) and asthma (RRR 1.56, 95% CI 1.37 to 1.77,  $p < 0.001$ ) (Fig. 3c).

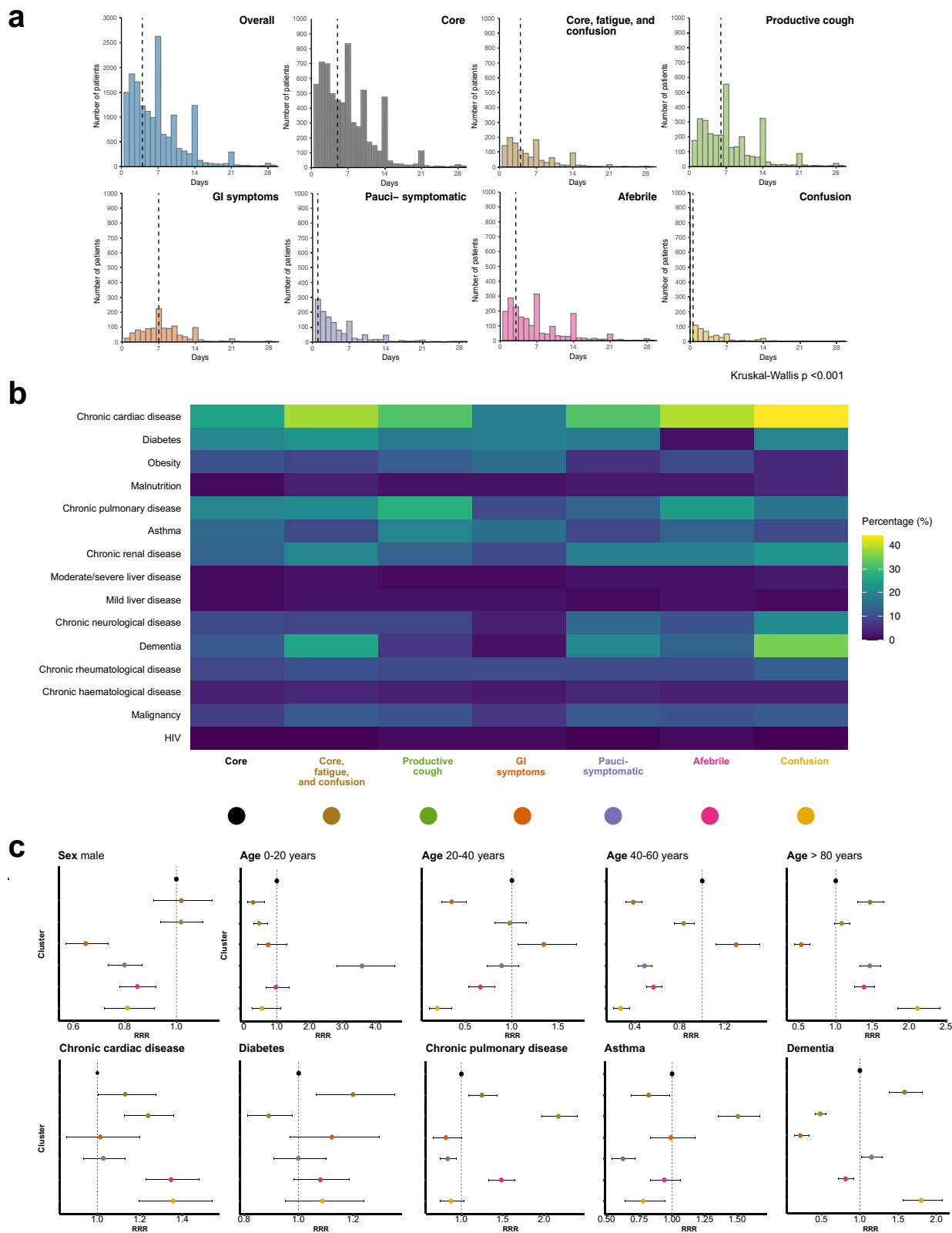
**Association between sub-phenotype and outcome.** Overall, the unadjusted in-hospital mortality was 35%, with 24% having an incomplete hospital episode at the end of follow-up (a further 474 patients were excluded due to conflicting outcome data leaving 18,884 available for analysis). Outcomes, stratified by sub-phenotype, are detailed in Table S1. To assess differences in mortality between sub-phenotypes, we first compared Kaplan–Meier curves (log-rank test,  $p < 0.001$ ) (Fig. 4a). For the largest sub-phenotype, core symptoms, unadjusted in-hospital mortality was 33%. The lowest mortality was found in the GI symptoms sub-phenotype (18%) and the highest in the core, fatigue, and confusion sub-phenotype (53%). Subsequently, we used a Cox proportional hazards model to account for the influence of age and sex (Fig. 3b). Those in the core, fatigue, and confusion sub-phenotype remained at the highest risk of death when compared to those with core symptoms, (HR 1.26, 95% CI 1.15–1.37,  $p < 0.001$ ). However, membership of the confusion cluster was no longer associated with an increased risk of death (HR 0.92, 95% CI 0.83–1.02,  $p = 0.096$ ). Those in the productive cough, GI, and pauci-symptomatic sub-phenotypes continued to attract a lower risk of death (Fig. 4b). Given that there was evidence of variation in risk over time for some sub-phenotypes, we performed Restricted Mean Survival Time (RMST) analyses (Table S4). By this method, males in the core, fatigue, and confusion sub-phenotype had the poorest survival compared to core symptoms, mean survival difference at 30-days -1.9 days (95% CI -2.8–1.1).

## Discussion

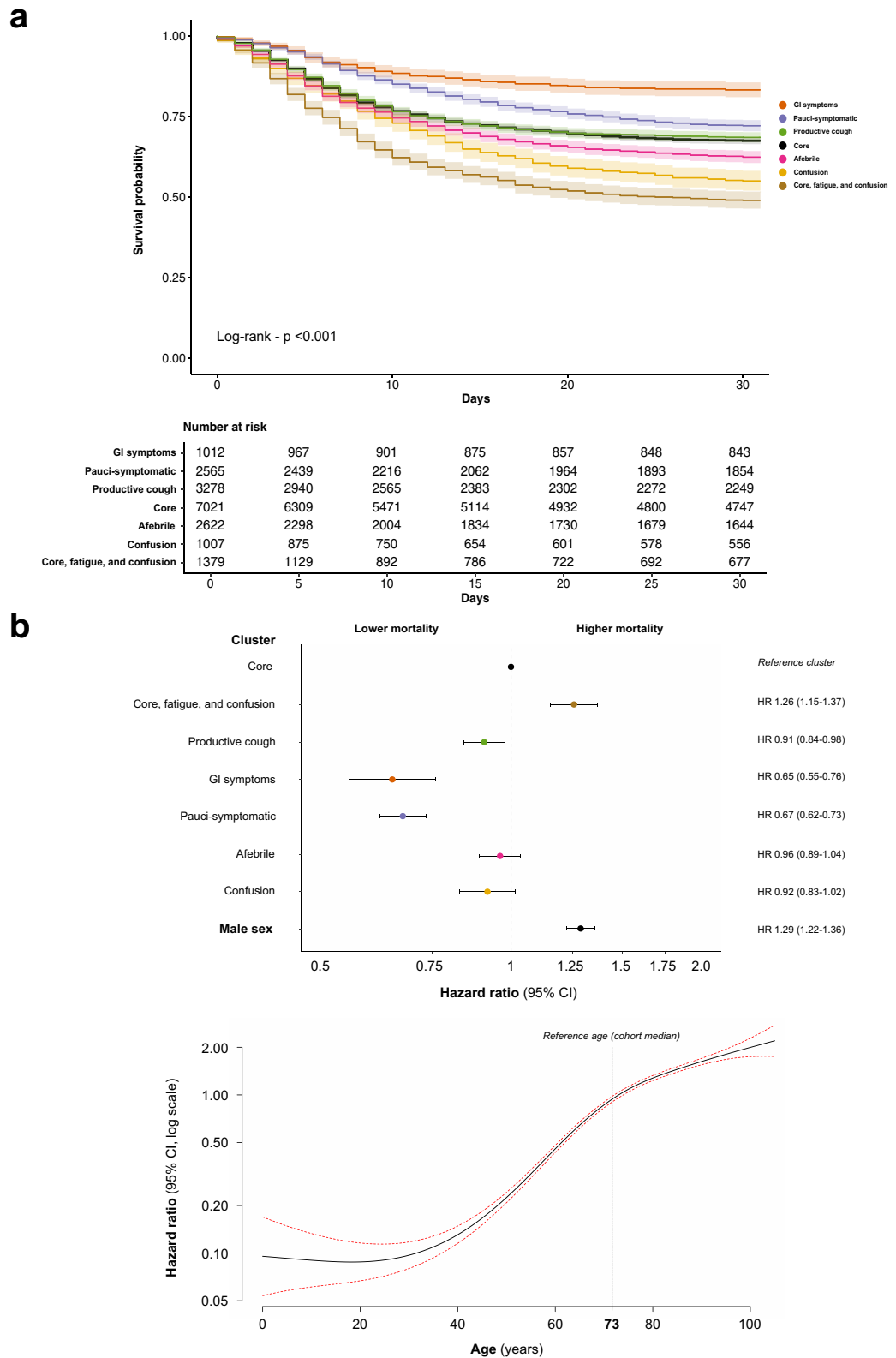
This study identifies distinct symptom sub-phenotypes in a large cohort of hospitalised patients with COVID-19. These sub-phenotypes are internally robust and reproducible. This report also provides one of the largest datasets of symptom prevalence in hospitalised patients with COVID-19 to date. Knowledge of distinct symptom sub-phenotypes has potential importance for our understanding of COVID-19. Two groups in our analysis have distinct clinical trajectories: those presenting with GI symptoms, and those presenting with confusion. Those in the GI cluster tended to be younger, more likely female, presented to hospital later, and had a higher probability of survival. Conversely, those with confusion (with or without fever, cough, and dyspnoea), were older, presented earlier, and had poorer outcomes. These data may be important for refining risk-prediction at the time of hospital admission. Similarly, the identification of a sub-phenotype in which patients had few symptoms other than confusion has implications for defining cases and for targeting testing, particularly in elderly patients. Importantly, these clusters have divergent outcomes from those with core COVID-19 symptoms.

The pooled symptom prevalences reported in large meta-analyses of independent studies<sup>11,12</sup> are broadly consistent with ours, considering the higher severity of illness and more advanced age in our cohort. However, the prevalence of GI symptoms and of confusion was higher in our study. In an enlarged international cohort study, which includes the patients used in this analysis, estimates of symptom prevalence were largely similar<sup>13</sup>.

Several of our sub-phenotypes are consistent with patterns highlighted in isolation by observational studies. Confusion, as the predominant symptom, has been reported in older adults presenting to hospital with



**Figure 3.** Patient characteristics and symptom clusters. **(a)** Time from symptom onset to hospital admission. Data are presented as counts in singleday bins. Vertical dashed lines represent the median time (days). **(b)** Co-morbidities by symptom cluster. Percentage of individuals with co-morbidity at time of admission. **(c)** Association between patient characteristics and symptom cluster membership. Multinomial regression, presented as relative risk ratio (95% confidence interval). Core symptoms chosen as reference cluster. The age group 60-80 years serves as the reference group for age.



**Figure 4.** Symptom clusters and survival. **(a)** Unadjusted 30-day in-hospital mortality by cluster. Kaplan-Meier curves, those discharged before 30 days were assumed to have survived until the end of follow-up. **(b)** The risk of 30 day in-hospital mortality, adjusted for age and sex. Upper panel, Forest plot, showing results of a Cox proportional hazards model. Lower panel, Hazard ratio associated with varying age, fitted with psplines. Data are presented as hazard ratio (HR) and 95% confidence intervals. Red dotted line - 95% CI Core symptoms serves as the reference symptom cluster. The cohort median age serves as the reference age.

COVID-19<sup>14–16</sup>. Similar to our findings, this pattern has been associated with a higher risk of in-hospital death. Likewise, the GI sub-phenotype identified by our study is reflective of an enteric form of COVID-19 described in various cohorts<sup>17</sup>. Gastrointestinal infection occurred in SARS<sup>18</sup>, MERS<sup>19</sup>, and was reported in early descriptions of COVID-19<sup>20</sup>. Single-cell transcriptomic analysis from ileum and colon has demonstrated the ability of SARS-CoV-2 to infect enterocytes<sup>21</sup>. Consistent with our findings, observational studies of patients with predominant GI symptoms have associated this presentation with a milder course of illness and better survival<sup>22</sup>.

These data may offer a method for predictive enrichment in clinical trials<sup>23</sup>. Predictive enrichment based on sub-phenotypes is expected to perform best where causal relationships exist between fundamental biological or genomic features of disease and the clinical manifestations of severe illness<sup>24</sup>. Similar relationships have been described in conditions as diverse as schizophrenia<sup>25</sup> and asthma<sup>26</sup>. Sub-phenotypes based on symptoms, may when combined with other clinical characteristics, offer a means of predictive enrichment in the absence of biochemically distinguishable sub-groups or in resource poor settings or where a highly pragmatic basis for study enrolment is employed.

This study has some limitations. First, symptoms were sought only at hospital admission, potentially disorienting patients to recall bias given the time between symptom onset and presentation. Similarly, patients may have elected to describe symptoms at that time or the gamut of symptoms since the onset of illness. Additionally, a small number of symptoms now known to be associated with COVID-19, namely anosmia and ageusia<sup>27</sup>, were not recorded. Therefore, these data may not be generalisable to individuals with milder disease who do not require admission to hospital. Second, the patients analysed in this study were admitted in the ‘first wave’ of COVID-19, prior to the emergence of alternative variants of SARS-CoV-2, thus limiting its generalisability. However, large-scale analysis failed to find a difference in reported symptoms after the emergence of the B.1.1.7 variant in the UK<sup>28</sup>. Third, our study had missing data. In handling missing symptom records, we sought to retain as much data as possible given its informative missingness. This approach was robust to a sensitivity analysis. For survival data this was more challenging, and several methods of analysis were employed to reduce the risk of bias. Fourth, a limitation of partitioning clustering is the need to pre-specify the number of clusters. In a large, heterogeneous population, this requires investigators to strike a balance between parsimony and granularity. Each patient is unique, and there were 4,335 symptom combinations in our cohort; hence, the most granular clustering would reveal 4,335 distinct patterns of disease. Our purpose in grouping these patients is to reveal clinical patterns that will have practical utility. As we have argued previously, the question should not be “How many clusters exist?”, but rather, “Which clusters are potentially useful?”<sup>24</sup>. The results of our clustering analysis may also have been confounded by structural issues in the data, heterogeneity of symptom recording and differences in time from symptom onset to study enrolment being two examples. Where these were recognisable, we performed sensitivity analyses. Finally, statistical modelling in this study corrected for a limited number of co-variables. With respect to differences in patient characteristics and symptom cluster, it may be that unmodelled demographic or clinical variables account partially or wholly for the variations which we observed. Likewise, for survival analysis, we adjusted only for age and sex, both of which we know to have been largely complete and not subject to significant confounding. Additionally, for some clusters there was evidence of variation in risk of death with time. This violates the assumption of proportional hazards and may limit the interpretation of the log-rank test and hazard ratios. We attempted to defend against this violation by performing RMST analyses, which confirmed the directions of effect.

In summary, our study of 59,001 hospitalised patients with COVID-19 identified distinct symptom sub-phenotypes whose character and outcome differed significantly from those with the core symptoms of COVID-19. These observations improve our understanding of COVID-19 and have implications for clinical diagnosis, risk prediction, and future mechanistic and clinical studies.

## Methods

**Study design, setting, and population.** The ISARIC Coronavirus Clinical Characterisation Consortium (4C) study is an ongoing prospective cohort study, involving 260 acute hospital sites in England, Scotland, and Wales. The study builds on an international consensus protocol for investigation of new infectious diseases, the International Severe Acute Respiratory Infection Consortium/World Health Organisation Clinical Characterisation Protocol (ISARIC/WHO CCP)<sup>29</sup>, designed to enable internationally harmonised clinical research during outbreaks<sup>30</sup>. The protocol, revision history, case report form, information leaflets, consent forms, and details of the Independent Data and Material Access Committee are available at <https://isaric4c.net>. The UK study was approved by the South Central—Oxford C Research Ethics Committee (13/SC/0149) and by the Scotland A Research Ethics Committee (20/SS/0028). The study was conducted in accordance with the Declaration of Helsinki 1964 and all its subsequent amendments. Informed consent to participate was obtained from all patients or an appropriate consultee. This study is reported in compliance with the TRIPOD guidelines<sup>31</sup>.

Patients included in the primary cohort were admitted to hospital between 6<sup>th</sup> February and 8<sup>th</sup> May, 2020. Inclusion criteria were all patients admitted to a participating hospital with laboratory proven or clinically highly suspected SARS-CoV-2 infection. Reverse transcription-PCR was the sole method of testing available during the study period. In the original CCP, the inclusion of patients with clinically suspected infection reflects the design of this study as a preparedness protocol where laboratory tests may not be available, but in the context of this outbreak in the UK, site training emphasised the importance of enrolling only laboratory-confirmed cases. Patients who were admitted to hospital for an unrelated condition but who subsequently tested positive for SARS-CoV-2 were also included.

**Data collection.** Data were collected on a case report form, developed by ISARIC and WHO in advance of this outbreak. From admission, data were uploaded to an electronic database (REDCap, Vanderbilt University,



US; hosted by University of Oxford, UK). We recorded demographic details as well as patient co-morbidities, in-hospital clinical course, treatments, and outcomes. The presence or absence of a pre-defined list of 25 symptoms was assessed at study enrolment. This was performed by research nurses and assistants, either by direct questioning or where this was not possible by review of electronic health records. Symptoms were: fever, cough, productive cough, haemoptysis, dyspnoea, wheeze, chest wall in-drawing, chest pain, fatigue, myalgia, joint pain, vomiting, diarrhoea, abdominal pain, headache, confusion, seizures, lymphadenopathy, ear pain, sore throat, runny nose, conjunctivitis, bleeding, rash, and skin ulceration. Similarly, the presence of pre-defined co-morbidities was recorded. These were: asthma, diabetes (type 1 and type 2), chronic cardiac disease, chronic haematological disease, chronic kidney disease, chronic neurological disease, chronic pulmonary disease (excluding asthma), dementia, HIV, malignancy, malnutrition, mild liver disease, moderate or severe liver disease, obesity, chronic rheumatological disease, and smoking history. Outcome data were collected for admission to critical care (Intensive Care Unit or High Dependency Unit), the use of invasive mechanical ventilation (IMV), and in-hospital mortality.

**Statistical analysis.** Continuous data are summarised as median (inter-quartile range). Categorical data are summarized as frequency (percentage). Prevalence is reported as percentage (95% confidence interval). Confidence intervals were calculated for a binomial proportion using the Clopper-Pearson exact method. We analysed data using R (R Core Team, Version 4.0.0, Vienna, Austria). P values < 0.05 were deemed significant.

*Missing data*—Given the extraordinary circumstances in which this study was conducted there was a large amount of missing data. No attempt at multiple imputation was made. In respect of symptom data, in many cases the presence of a positive symptom(s) was recorded, with the remainder missing. Exploration of the structure of missing symptom data (Supplementary Figure S1) suggests that this did not occur at random. In such cases, missing symptoms were recoded as being absent i.e., recoded as '0'. Patients with fully missing data were excluded from the analysis.

*Symptom network analysis*—To explore the relationship between the 25 recorded symptoms we fit an Ising model, using L1-regularised logistic regression with model selection by Extended Bayesian Information criteria (EBIC)<sup>32</sup>, using the R package *IsingFit* (version 0.3.1). A  $\lambda$  value of 10 was chosen to minimise spurious conditional dependencies. The  $\lambda$  value acts as a tuning parameter, when multiplied by the sum of squares of the coefficients in the model, to produce a 'shrinkage' penalty. The partition of symptoms into communities was formalised using a short random walks method, with the R package *igraph* (version 1.2.6)<sup>33</sup>.

*Unsupervised partitional clustering*—Symptom data for each patient, encoded as binary responses, were used to derive a Jaccard distance matrix. This was then supplied as the input to a k-medoids clustering algorithm. This is an unsupervised partitional algorithm that seeks to divide the sample into  $k$  clusters, where the arbitrary distance between any individual case and the case chosen as the centre of a cluster (medoid) is minimised. In this study we used a variation of this algorithm, Clustering for Large Applications (CLARA), with the R package *cluster* (version 2.1.0)<sup>34</sup>. CLARA, for the optimisation of computational runtime, performs iterations of k-medoids clustering on subsets of the data and selects the best performing result. We clustered 100 random sub-samples each consisting of approximately 10% of the analysed population ( $n = 2,500$ ). Each sub-sample was used to select  $k$  medoid cases, after which every case in the dataset was assigned to the nearest medoid. The iteration in which the mean of the dissimilarities between cases and the nearest medoid was lowest was selected. Random sampling was performed deterministically to ensure consistency between our primary analysis and validation steps. Clustering was performed agnostic of patient demographics or outcome.

The optimal number of clusters to specify to the algorithm was derived from a 'majority' assessment of three measures; total within sum of squares, average silhouette width, and gap statistic<sup>35</sup>, in which a parsimonious solution was preferred. To assess the stability of clusters, we employed a non-parametric bootstrap-based strategy using the R package *fpc* (version 2.2.8)<sup>36</sup>. This generated 1000 new datasets by randomly drawing samples from the initial dataset with replacement and applying the same clustering technique to each. Clustering results were then compared for each cluster identified in the primary analysis and the most similar cluster identified for each random re-sampling. A mean value for the Jaccard coefficient, for the sum of the comparisons, was generated for each cluster present in the primary analysis. As a sensitivity analysis for our treatment of missing data, clustering was repeated on patients with only fully complete symptom data. Cluster allocations for individuals partitioned by both iterations were then compared using Cohen's kappa and simple percentage agreement with the R package *irr* (version 0.84.1). We also performed sensitivity analyses for the effect of age, recruitment site, and time from symptom onset to study enrolment.

*Replication.* We replicated our clustering in two independent datasets, one internal and one external. Internally, we used symptom data for patients enrolled to ISARIC CCP-UK after those included in the primary cohort and until 7<sup>th</sup> July, 2020 (secondary cohort). These data were processed and analysed as for the primary cohort. Missing data were treated in the same fashion. Externally, our clustering strategy was replicated independently in a sub-sample of users from the COVID Symptom Study app (developed by Zoe Global Ltd. with input from scientists and clinicians from King's College London and Massachusetts General Hospital). Individuals with confirmed SARS-CoV-2 laboratory results, registering healthy on the app, with symptom duration of more than 7 days were included, considering the presentation at symptom peak to build the clustering<sup>37</sup>.

*Multinomial regression modelling.* To quantify the relationship between demographic factors, co-morbidities, and cluster membership we built a multinomial logistic regression model with the R package *nnet* (version 7.3.14). For binary variables, a missing value was assumed to correspond to the absence of a given co-morbidity. Individuals with missing values for age and sex were excluded from this analysis. The dependent variable was

symptom cluster. The independent variables were; sex, age (categorical), chronic cardiac disease, chronic pulmonary disease, asthma, chronic kidney disease, chronic neurological disease, malignancy, chronic haematological disease, HIV, chronic rheumatological disease, dementia, malnutrition, chronic liver disease, and diabetes. Results are summarised as relative risk ratio (RRR) and 95% CI. The average time from self-reported symptom onset to hospital admission in each cluster were compared using the Kruskal–Wallis test.

**Survival analysis.** To examine survival differences between symptom clusters we employed several methods. Kaplan–Meier curves for 30-day in-hospital mortality were calculated and compared using a log-rank test, with the R package *survival* (version 3.2.7). Individuals reported as being dead but with no outcome date were excluded. All individuals not reported as dead were presumed alive. Discharged individuals were retained within the at-risk set until the end of follow-up; thus, discharge was not a competing risk for death. Survival time was defined as the time in days between hospital admission and the reported outcome date. We then fitted a Cox proportional hazards model to the data with the a priori inclusion of age and sex as co-variables, using the R package *survival*. Given the non-linear effect of age on the risk of death we fitted age with a penalised smoothing spline. Results are reported as hazard ratio (HR) and 95% CI. In anticipation that the risk of death per cluster varied over time, we also calculated restricted mean survival times at 30 days, insuring the analysis against violations of proportional hazards assumptions. This was performed with the R package *survRM2* (version 1.0.3). Results are reported as mean survival difference (days) and 95% CI.

Received: 9 April 2021; Accepted: 14 February 2022

Published online: 27 April 2022

## References

- Seymour, C. W. *et al.* Derivation, validation, and potential treatment implications of novel clinical phenotypes for sepsis. *JAMA* **321**, 2003–2017 (2019).
- Neyton, L.P.A., *et al.* Molecular Patterns in Acute Pancreatitis Reflect Generalizable Endotypes of the Host Response to Systemic Injury in Humans. *Ann. Surg.* Online Ahead of Print (2020).
- Macedo Hair, G., Fonseca Nobre, F. & Brasil, P. Characterization of clinical patterns of dengue patients using an unsupervised machine learning approach. *BMC Infect Dis.* **19**, 649 (2019).
- Horby, P. *et al.* Dexamethasone in hospitalized patients with Covid-19. *N. Engl. J. Med.* **384**, 693–704 (2021).
- Angus, D. C. *et al.* Effect of hydrocortisone on mortality and organ support in patients with severe COVID-19: The REMAP-CAP COVID-19 corticosteroid domain randomized clinical trial. *JAMA* **324**, 1317–1329 (2020).
- Gordon, A. C. *et al.* Interleukin-6 receptor antagonists in critically ill patients with Covid-19. *N. Engl. J. Med.* **384**, 1491–1502 (2021).
- Sinha, S., *et al.* Dexamethasone modulates immature neutrophils and interferon programming in severe COVID-19. *Nat Med* Online Ahead of Print (2021).
- Pairo-Castineira, E. *et al.* Genetic mechanisms of critical illness in COVID-19. *Nature* **591**, 92–98 (2021).
- Hall, D.P., *et al.* Network analysis reveals distinct clinical syndromes underlying acute mountain sickness. *PLoS One* **9**, e81229 (2014).
- Roach, R. C. *et al.* The 2018 Lake Louise Acute Mountain Sickness Score. *High Alt. Med. Biol.* **19**, 4–6 (2018).
- Grant, M. C. *et al.* The prevalence of symptoms in 24,410 adults infected by the novel coronavirus (SARS-CoV-2; COVID-19): A systematic review and meta-analysis of 148 studies from 9 countries. *PLoS ONE* **15**, e0234765 (2020).
- Booth, A. *et al.* Population risk factors for severe disease and mortality in COVID-19: A global systematic review and meta-analysis. *PLoS ONE* **16**, e0247461 (2021).
- COVID-19 symptoms at hospital admission vary with age and sex: results from the ISARIC prospective multinational observational study. *Infection* **49**, 889–905 (2021)
- Kennedy, M. *et al.* Delirium in older patients with COVID-19 presenting to the emergency department. *JAMA Netw. Open* **3**, e2029540 (2020).
- Zazzara, M. B. *et al.* Probable delirium is a presenting symptom of COVID-19 in frail, older adults: A cohort study of 322 hospitalised and 535 community-based older adults. *Age Ageing* **50**, 40–48 (2021).
- Annweiler, C. *et al.* National french survey of coronavirus disease (COVID-19) symptoms in people aged 70 and over. *Clin. Infect Dis.* **72**, 490–494 (2021).
- Villapol, S. Gastrointestinal symptoms associated with COVID-19: Impact on the gut microbiome. *Transl Res* **226**, 57–69 (2020).
- Ding, Y. *et al.* Organ distribution of severe acute respiratory syndrome (SARS) associated coronavirus (SARS-CoV) in SARS patients: Implications for pathogenesis and virus transmission pathways. *J. Pathol.* **203**, 622–630 (2004).
- Zhou, J. *et al.* Human intestinal tract serves as an alternative infection route for Middle East respiratory syndrome coronavirus. *Sci. Adv.* **3**, 966 (2017).
- Gao, Q. Y., Chen, Y. X. & Fang, J. Y. 2019 Novel coronavirus infection and gastrointestinal tract. *J. Dig. Dis.* **21**, 125–126 (2020).
- Qi, F., Qian, S., Zhang, S. & Zhang, Z. Single cell RNA sequencing of 13 human tissues identify cell types and receptors of human coronaviruses. *Biochem. Biophys. Res. Commun.* **526**, 135–140 (2020).
- Han, C. *et al.* Digestive Symptoms in COVID-19 Patients With Mild Disease Severity: Clinical Presentation, Stool Viral RNA Testing, and Outcomes. *Am J Gastroenterol* **115**, 916–923 (2020).
- Prescott, H. C., Calfee, C. S., Thompson, B. T., Angus, D. C. & Liu, V. X. Toward smarter lumping and smarter splitting: Rethinking strategies for sepsis and acute respiratory distress syndrome clinical trial design. *Am. J. Respir. Crit. Care Med.* **194**, 147–155 (2016).
- Russell, C. D. & Baillie, J. K. Treatable traits and therapeutic targets: Goals for systems biology in infectious disease. *Curr. Opin. Syst. Biol.* **2**, 140–146 (2017).
- DeRosse, P. *et al.* The genetics of symptom-based phenotypes: Toward a molecular classification of schizophrenia. *Schizophr. Bull.* **34**, 1047–1053 (2008).
- Wenzel, S. E. Asthma phenotypes: the evolution from clinical to molecular approaches. *Nat. Med.* **18**, 716–725 (2012).
- Carrillo-Larco, R. M. & Altez-Fernandez, C. Anosmia and dysgeusia in COVID-19: A systematic review. *Wellcome Open Res.* **5**, 94 (2020).
- Graham, M.S., *et al.* Changes in symptomatology, reinfection, and transmissibility associated with the SARS-CoV-2 variant B.1.1.7: An ecological study. *Lancet Public Health* **6**, e335–e345 (2021).

29. Dunning, J. W. *et al.* Open source clinical science for emerging infections. *Lancet Infect. Dis.* **14**, 8–9 (2014).
30. Global outbreak research: Harmony not hegemony. *Lancet Infect. Dis.* **20**, 770–772 (2020).
31. von Elm, E. *et al.* The strengthening of reporting of observational studies in epidemiology (STROBE) statement: Guidelines for reporting observational studies. *Ann. Intern. Med.* **147**, 573–577 (2007).
32. van Borkulo, C. D. *et al.* A new method for constructing networks from binary data. *Sci. Rep.* **4**, 5918 (2014).
33. Pons, P. & Latapy, M. Computing communities in large networks using random walks, in *International symposium on computer and information sciences*, 284–293 (Springer, 2005).
34. Kaufman, L. & Rousseeuw, P. J. *Finding groups in data: An introduction to cluster analysis* (Wiley, 2009).
35. Tibshirani, R., Walther, G. & Hastie, T. Estimating the number of clusters in a data set via the gap statistic. *J. R. Stat. Soc. B Stat. Methodol.* **63**, 411–423 (2001).
36. Hennig, C. Cluster-wise assessment of cluster stability. *Comp. Stat. Data Anal.* **52**, 258–271 (2007).
37. Sudre, C.H., *et al.* Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app. *Sci Adv* **7** (2021).

### Author contributions

J.E.M. and L.N. performed the analyses and constructed the draft manuscript. SS assisted with the design and performance of the study analyses. J.D., L.M., S.M., C.D.R., S.K., M.S. critically revised the draft manuscript. C.H.S., T.D.S., S.O., C.J.S., J.W. performed the validation analysis on the COVID Symptom Study dataset and assisted with the preparation of the draft manuscript. A.B.D., E.M.H. assisted with data curation and with the design and conduct of the study analyses. P.J.M.O., M.G.S., J.K.B. conceived the study and undertook critical revision of the draft manuscript. All authors reviewed and approve the final manuscript.

### Competing interests

All authors have completed the ICMJE uniform disclosure form publicly available at [www.icmje.org/coi\\_disclosure.pdf](http://www.icmje.org/coi_disclosure.pdf) and declare: during the conduct of the study; JW reports he is an employee of ZOE Global Ltd; TDS reports he has acted as a consultant for ZOE Global Ltd; PJMO reports personal fees from Consultancy, grants from MRC, grants from EU Grant, grants from NIHR Biomedical Research Centre, grants from MRC/GSK, grants from Wellcome Trust, grants from NIHR (HPRU), grants from NIHR Senior Investigator, personal fees from European Respiratory Society, grants from MRC Global Challenge Research Fund, outside the submitted work; and The role of President of the British Society for Immunology was an unpaid appointment but my travel and accommodation at some meetings is provided by the Society; AMD reports grants from Department of Health and Social Care, during the conduct of the study; grants from Wellcome Trust, outside the submitted work; JKB reports grants from DHSC National Institute of Health Research UK, grants from Medical Research Council UK, grants from Wellcome Trust, grants from Fiona Elizabeth Agnew Trust, grants from Intensive Care Society, grants from Chief Scientist Office, during the conduct of the study; MGS reports grants from DHSC National Institute of Health Research UK, grants from Medical Research Council UK, grants from Health Protection Research Unit in Emerging & Zoonotic Infections, University of Liverpool, during the conduct of the study; other from Integrum Scientific LLC, Greensboro, NC, USA, outside the submitted work; the remaining authors declare no competing interests; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-08032-3>.

**Correspondence** and requests for materials should be addressed to M.G.S. or J.K.B.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

### ISARIC-4C

### Consortium Lead Investigator

J. Kenneth Baillie<sup>1</sup>

**Chief Investigator****Malcolm G. Semple<sup>15</sup>****Co-Lead Investigator****Peter J. M. Openshaw<sup>4</sup>****ISARIC Clinical Coordinator****Gail Carson****Co-Investigators**

**Beatrice Alex<sup>16</sup>, Benjamin Bach<sup>16</sup>, Wendy S. Barclay<sup>17</sup>, Debby Bogaert<sup>8</sup>, Meera Chand<sup>18</sup>, Graham S. Cooke<sup>19</sup>, Annemarie B. Docherty<sup>14</sup>, Jake Dunning<sup>3</sup>, Anna da Silva Filipe<sup>20</sup>, Tom Fletcher<sup>21</sup>, Christopher A. Green<sup>22</sup>, Ewen M. Harrison<sup>14</sup>, Julian A. Hiscox<sup>23</sup>, Antonia YW Ho<sup>24</sup>, Peter W. Horby<sup>25</sup>, Samreen Ijaz<sup>26</sup>, Saye Khoo<sup>27</sup>, Paul Klenerman<sup>28</sup>, Andrew Law<sup>1</sup>, Wei Shen Lim<sup>29</sup>, Alexander J. Mentzer<sup>30</sup>, Laura Merson<sup>5</sup>, Alison M. Meynert<sup>31</sup>, Mahdad Noursadeghi<sup>32</sup>, Shona C. Moore<sup>33</sup>, Massimo Palmarini<sup>24</sup>, William A. Paxton<sup>23</sup>, Georgios Pollakis<sup>23</sup>, Nicholas Price<sup>34</sup>, Andrew Rambaut<sup>35</sup>, David L. Robertson<sup>24</sup>, Clark D. Russell<sup>8</sup>, Vanessa Sancho-Shimizu<sup>36</sup>, Janet T. Scott<sup>24</sup>, Louise Sigfrid<sup>5</sup>, Tom Solomon<sup>15</sup>, Shiranee Srischandran<sup>19</sup>, David Stuart<sup>37</sup>, Charlotte Summers<sup>38</sup>, Richard S. Tedder<sup>39</sup>, Emma C. Thomson<sup>24</sup>, Ryan S. Thwaites<sup>4</sup>, Lance C. W. Turtle<sup>15</sup> & Maria Zambon<sup>40</sup>**

<sup>16</sup>School of Informatics, University of Edinburgh, Edinburgh, UK. <sup>17</sup>Section of Molecular Virology, Imperial College London, London, UK. <sup>18</sup>Antimicrobial Resistance and Hospital Acquired Infection Department, Public Health England, London, UK. <sup>19</sup>Department of Infectious Disease, Imperial College London, London, UK. <sup>20</sup>Centre for Virus Research, Sir Michael Stoker Building, University of Glasgow, Glasgow, UK. <sup>21</sup>Liverpool School of Tropical Medicine, Liverpool, UK. <sup>22</sup>Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. <sup>23</sup>Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, Liverpool, UK. <sup>24</sup>MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, UK. <sup>25</sup>Centre for Tropical Medicine and Global Health, Nuffield Department of Medicine, University of Oxford, Oxford, UK. <sup>26</sup>Virology Reference Department, National Infection Service, Public Health England, London, UK. <sup>27</sup>Department of Pharmacology, University of Liverpool, Liverpool, UK. <sup>28</sup>Nuffield Department of Medicine, Peter Medawar Building for Pathogen Research, University of Oxford, Oxford, UK. <sup>29</sup>Nottingham University Hospitals NHS Trust, Nottingham, UK. <sup>30</sup>Nuffield Department of Medicine, John Radcliffe Hospital, Oxford, UK. <sup>31</sup>MRC Human Genetics Unit, MRC, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. <sup>32</sup>Division of Infection and Immunity, University College London, London, UK. <sup>33</sup>Institute of Infection and Global Health, University of Liverpool, Liverpool, UK. <sup>34</sup>Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, School of Immunology and Microbial Sciences, King's College London, London, UK. <sup>35</sup>Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK. <sup>36</sup>Department of Pediatrics and Virology, St Mary's Medical School Bldg, Imperial College London, London, UK. <sup>37</sup>Division of Structural Biology, The Wellcome Centre for Human Genetics, University of Oxford, Oxford, UK. <sup>38</sup>Department of Medicine, University of Cambridge, Cambridge, Cambridgeshire, UK. <sup>39</sup>Blood Borne Virus Unit, Virus Reference Department, National Infection Service, Public Health England, London, UK. <sup>40</sup>Public Health England, London, UK.

**Project Managers****Hayley Hardwick<sup>41</sup>, Chloe Donohue<sup>41</sup>, Jane Ewins<sup>41</sup>, Wilna Oosthuizen<sup>41</sup> & Fiona Griffiths<sup>41</sup>****Data Analysts****Lisa Norman<sup>41</sup>, Riinu Pius<sup>41</sup>, Tom M. Drake<sup>41</sup>, Cameron J. Fairfield<sup>41</sup>, Stephen Knight<sup>41</sup>, Kenneth A. Mclean<sup>41</sup>, Derek Murphy<sup>41</sup> & Catherine A. Shaw<sup>41</sup>****Data and Information System Manager****Jo Dalton<sup>41</sup>, Michelle Girvan<sup>41</sup>, Egle Saviciute<sup>41</sup>, Stephanie Roberts<sup>41</sup>, Janet Harrison<sup>41</sup>, Laura Marsh<sup>41</sup> & Marie Connor<sup>41</sup>**

## Data integration and presentation

Gary Leeming<sup>41</sup> & Ross Hendry<sup>41</sup>

## Material Management

William Greenhalf<sup>41</sup>, Victoria Shaw<sup>41</sup> & Sarah McDonald<sup>41</sup>

## Local Principal Investigators

Kayode Adeniji<sup>41</sup>, Daniel Agranoff<sup>41</sup>, Ken Agwuh<sup>41</sup>, Dhiraj Ail<sup>41</sup>, Ana Alegria<sup>41</sup>, Brian Angus<sup>41</sup>, Abdul Ashish<sup>41</sup>, Dougal Atkinson<sup>41</sup>, Shahedal Bari<sup>41</sup>, Gavin Barlow<sup>41</sup>, Stella Barnass<sup>41</sup>, Nicholas Barrett<sup>41</sup>, Christopher Bassford<sup>41</sup>, David Baxter<sup>41</sup>, Michael Beadsworth<sup>41</sup>, Jolanta Bernatoniene<sup>41</sup>, John Berridge<sup>41</sup>, Nicola Best<sup>41</sup>, Pieter Bothma<sup>41</sup>, David Brealey<sup>41</sup>, Robin Brittain-Long<sup>41</sup>, Naomi Bulteel<sup>41</sup>, Tom Burden<sup>41</sup>, Andrew Burtenshaw<sup>41</sup>, Vikki Caruth<sup>41</sup>, David Chadwick<sup>41</sup>, Duncan Chambler<sup>41</sup>, Nigel Chee<sup>41</sup>, Jenny Child<sup>41</sup>, Srikanth Chukkambotla<sup>41</sup>, Tom Clark<sup>41</sup>, Paul Collini<sup>41</sup>, Graham Cooke<sup>41</sup>, Catherine Cosgrove<sup>41</sup>, Jason Cupitt<sup>41</sup>, Maria-Teresa Cutino-Moguel<sup>41</sup>, Paul Dark<sup>41</sup>, Chris Dawson<sup>41</sup>, Samir Dervisevic<sup>41</sup>, Phil Donnison<sup>41</sup>, Sam Douthwaite<sup>41</sup>, Ingrid DuRand<sup>41</sup>, Ahilanadan Dushianthan<sup>41</sup>, Tristan Dyer<sup>41</sup>, Cariad Evans<sup>41</sup>, Chi Eziefula<sup>41</sup>, Chrisopher Fegan<sup>41</sup>, Adam Finn<sup>41</sup>, Duncan Fullerton<sup>41</sup>, Sanjeev Garg<sup>41</sup>, Atul Garg<sup>41</sup>, Jo Godden<sup>41</sup>, Arthur Goldsmith<sup>41</sup>, Clive Graham<sup>41</sup>, Elaine Hardy<sup>41</sup>, Stuart Hartshorn<sup>41</sup>, Daniel Harvey, Peter Havalda<sup>41</sup>, Daniel B. Hawcutt<sup>41</sup>, Maria Hobrok<sup>41</sup>, Luke Hodgson<sup>41</sup>, Anita Holme<sup>41</sup>, Anil Hormis<sup>41</sup>, Michael Jacobs<sup>41</sup>, Susan Jain<sup>41</sup>, Paul Jennings<sup>41</sup>, Agilan Kaliappan<sup>41</sup>, Vidya Kasipandian<sup>41</sup>, Stephen Kegg<sup>41</sup>, Michael Kelsey, Jason Kendall<sup>41</sup>, Caroline Kerrison<sup>41</sup>, Ian Kerslake<sup>41</sup>, Oliver Koch<sup>41</sup>, Gouri Koduri<sup>41</sup>, George Koshy, Shondipon Laha<sup>41</sup>, Susan Larkin<sup>41</sup>, Tamas Leiner<sup>41</sup>, Patrick Lillie<sup>41</sup>, James Limb<sup>41</sup>, Vanessa Linnett<sup>41</sup>, Jeff Little<sup>41</sup>, Michael MacMahon<sup>41</sup>, Emily MacNaughton<sup>41</sup>, Ravish Mankregod<sup>41</sup>, Huw Masson<sup>41</sup>, Elijah Matovu<sup>41</sup>, Katherine McCullough<sup>41</sup>, Ruth McEwen<sup>41</sup>, Manjula Meda<sup>41</sup>, Gary Mills<sup>41</sup>, Jane Minton<sup>41</sup>, Mariyam Mirfenderesky<sup>41</sup>, Kavya Mohandas<sup>41</sup>, Quen Mok<sup>41</sup>, James Moon<sup>41</sup>, Elinoor Moore<sup>41</sup>, Patrick Morgan<sup>41</sup>, Craig Morris<sup>41</sup>, Katherine Mortimore<sup>41</sup>, Samuel Moses<sup>41</sup>, Mbiye Mpenge<sup>41</sup>, Rohinton Mulla<sup>41</sup>, Michael Murphy<sup>41</sup>, Megan Nagel<sup>41</sup>, Thapas Nagarajan<sup>41</sup>, Mark Nelson<sup>41</sup>, Igor Otahal<sup>41</sup>, Mark Pais<sup>41</sup>, Selva Panchatsharam<sup>41</sup>, Hassan Paraiso<sup>41</sup>, Brij Patel<sup>41</sup>, Justin Pepperell<sup>41</sup>, Mark Peters<sup>41</sup>, Mandeep Phull<sup>41</sup>, Stefania Pintus<sup>41</sup>, Jagtur Singh Pooni<sup>41</sup>, Frank Post<sup>41</sup>, David Price<sup>41</sup>, Rachel Prout<sup>41</sup>, Nikolas Rae<sup>41</sup>, Henrik Reschreiter<sup>41</sup>, Tim Reynolds<sup>41</sup>, Neil Richardson<sup>41</sup>, Mark Roberts<sup>41</sup>, Devender Roberts<sup>41</sup>, Alistair Rose<sup>41</sup>, Guy Rousseau<sup>41</sup>, Brendan Ryan<sup>41</sup>, Taranprit Saluja<sup>41</sup>, Aarti Shah<sup>41</sup>, Prad Shanmuga<sup>41</sup>, Anil Sharma<sup>41</sup>, Anna Shawcross<sup>41</sup>, Jeremy Sizer<sup>41</sup>, Richard Smith<sup>41</sup>, Catherine Snelson<sup>41</sup>, Nick Spittle<sup>41</sup>, Nikki Staines<sup>41</sup>, Tom Stambach<sup>41</sup>, Richard Stewart<sup>41</sup>, Pradeep Subudhi<sup>41</sup>, Tamas Szakmany<sup>41</sup>, Kate Tatham<sup>41</sup>, Jo Thomas<sup>41</sup>, Chris Thompson<sup>41</sup>, Robert Thompson<sup>41</sup>, Ascanio Tridente<sup>41</sup>, Darell Tupper-Carey<sup>41</sup>, Mary Twagira<sup>41</sup>, Andrew Ustianowski<sup>41</sup>, Nick Vallotton<sup>41</sup>, Lisa Vincent-Smith<sup>41</sup>, Shico Visuvanathan<sup>41</sup>, Alan Vuylsteke<sup>41</sup>, Sam Waddy<sup>41</sup>, Rachel Wake<sup>41</sup>, Andrew Walden<sup>41</sup>, Ingeborg Welters<sup>41</sup>, Tony Whitehouse<sup>41</sup>, Paul Whittaker<sup>41</sup>, Ashley Whittington<sup>41</sup>, Meme Wijesinghe<sup>41</sup>, Martin Williams<sup>41</sup>, Lawrence Wilson<sup>41</sup>, Sarah Wilson<sup>41</sup>, Stephen Winchester<sup>41</sup>, Martin Wiselka<sup>41</sup>, Adam Wolverson<sup>41</sup>, Daniel G. Wooton<sup>41</sup>, Andrew Workman<sup>41</sup>, Bryan Yates<sup>41</sup> & Peter Young<sup>41</sup>

<sup>41</sup>ISARIC 4C Management Team, C/O Roslin Institute, University of Edinburgh, Edinburgh, UK.