medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

TITLE PAGE

Title

Characterising patterns of COVID-19 and long COVID symptoms: Evidence from nine UK longitudinal studies

Authors

Ruth C. E. Bowyer 1, 2
Charlotte Huggins 3
Renin Toms 4, 5
Richard J. Shaw 6
Bo Hou 7
Ellen J. Thompson 1
Alex Kwong 8, 9
Dylan Williams 10, 11
Milla Kibble 1, 4, 8, 12
George B. Ploubidis 14
Nic Timpson 4, 8
Jonathan A. C. Sterne 4, 15, 16
Nish Chaturvedi 10
Claire J. Steves 1, 13
Kate Tilling 8
Richard J. Silverwood 14
on behalf of the CONVALESCENCE Study

Affiliations

1. Department of Twin Research and Genetic Epidemiology, School of Life Course & Population Sciences, King's College London, London, UK
2. AI For Science & Government, Alan Turing Institute, London, UK
3. Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
4. Population Health Sciences, Bristol Medical School, University of Bristol, UK
5. Population Wellbeing, School of Health Sciences, Cardiff Metropolitan University, UK
6. MRC/CSO Social and Public Health Sciences Unit, University of Glasgow, Glasgow, UK
7. Bradford Institute for Health Research, Bradford Teaching Hospitals NHS Foundation Trust, Bradford, UK
8. MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK
9. Division of Psychiatry, University of Edinburgh, Edinburgh, UK
10. MRC Unit for Lifelong Health and Ageing at UCL, University College London, London, UK
11. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
12. Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK
13. Department of Ageing and Health, Guy's and St Thomas' NHS Foundation Trust, London, UK

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

14. Centre for Longitudinal Studies, UCL Social Research Institute, University College London, London, UK
15. NIHR Bristol Biomedical Research Centre, Bristol, UK
16. Health Data Research UK South West, UK

Word count: 4,445

ABSTRACT

Multiple studies across global populations have established the primary symptoms characterising COVID-19 (Coronavirus Disease 2019) and long COVID. However, as symptoms may also occur in the absence of COVID-19, a lack of appropriate controls has often meant that specificity of symptoms to acute COVID-19 or long COVID, and the extent and length of time for which they are elevated after COVID-19, could not be examined. We analysed individual symptom prevalences and characterised patterns of COVID-19 and long COVID symptoms across nine UK longitudinal studies, totalling over 42,000 participants. Conducting latent class analyses separately in three groups ('no COVID-19', 'COVID-19 in last 12 weeks', 'COVID-19 > 12 weeks ago'), the data did not support the presence of more than two distinct symptom patterns, representing high and low symptom burden, in each group. Comparing the high symptom burden classes between the 'COVID-19 in last 12 weeks' and 'no COVID-19' groups we identified symptoms characteristic of acute COVID-19, including loss of taste and smell, fatigue, cough, shortness of breath and muscle pains or aches. Comparing the high symptom burden classes between the 'COVID-19 > 12 weeks ago' and 'no COVID-19' groups we identified symptoms characteristic of long COVID, including fatigue, shortness of breath, muscle pain or aches, difficulty concentrating and chest tightness. The identified symptom patterns among individuals with COVID-19 > 12 weeks ago were strongly associated with self-reported length of time unable to function as normal due to COVID-19 symptoms, suggesting that the symptom pattern identified corresponds to long COVID. Building the evidence base regarding typical long COVID symptoms will improve diagnosis of this condition and the ability to elicit underlying biological mechanisms, leading to better patient access to treatment and services.

Abstract word count: 263

KEY WORDS

Clustering; COVID-19; Long COVID; Longitudinal studies; Symptom patterns

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

INTRODUCTION

Hundreds of millions of people worldwide have now been infected with SARS-CoV-2, the coronavirus responsible for the COVID-19 (coronavirus disease 2019) pandemic (1, 2). SARS-CoV-2 enters the body via the Angiotensin-converting enzyme 2 (ACE-2) receptor; as ACE2 is located on cells across multiple body sites the virus has the capacity to infect and damage cells within multiple organs (3, 4). This is reflected in the variety of symptoms associated with acute COVID-19 (signs and symptoms of COVID-19 for up to 4 weeks), ongoing symptomatic COVID-19 (signs and symptoms of COVID-19 from 4-12 weeks) and in post-COVID-19 syndrome (where signs and symptoms developing during or after COVID-19 infection continue for more than 12 weeks and are not explained by an alternative diagnosis) (5). Both ongoing symptomatic COVID-19 and post-COVID-19 syndrome are regarded under the umbrella of long COVID, as patient advocates prefer the condition to be termed (6). However, for the purpose of understanding difference in symptomology at various stages of illness, we consider only symptoms greater than 12 weeks as long COVID as long-term symptoms are likely to have a more detrimental impact on quality of life.

Whilst multiple studies across global populations have established the primary symptoms characterising long COVID to include fatigue, shortness of breath, cough, cognitive impairment and anosmia, a plethora of persistent symptoms have been reported by patients (7, 8). Better understanding of symptoms which characterise long COVID – or subvariants thereof and thus whether it is meaningful to describe long COVID as one syndrome (1) – may improve diagnostic precision and help elicit underlying mechanisms to better target therapy via patient-centred strategies (3, 9-14). Further, as these symptoms often also occur in the absence of COVID-19, the lack of appropriate controls has often meant that specificity of symptoms to acute COVID-19 or long COVID could not be examined. Inadequate consideration of cohort selection biases, particularly where participants have been recruited via support groups, may undermine generalisability of findings and therefore their utility in guiding clinical practice (15).

We aimed to characterise patterns of symptoms in individuals who had experienced COVID-19, before and after twelve weeks of illness onset, as well as those who had not, across nine UK longitudinal studies, to shed light on specific symptom patterns of COVID-19 and long COVID. We then examined how patterns differed by key factors such as sex, age and (for long COVID) self-reported functional limitation following COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

METHODS

Data

The UK National Core Studies – Longitudinal Health and Wellbeing programme (https://www.ucl.ac.uk/covid-19-longitudinal-health-wellbeing/) combines data from multiple UK population-based longitudinal studies and electronic health records to conduct analyses that allow researchers to investigate pandemic-related changes in population health. As symptom persistence is poorly captured in electronic health records, we performed co-ordinated standardised analyses across multiple longitudinal studies. This approach minimises methodological heterogeneity and maximises comparability, while appropriately accounting for study designs and characteristics of individual datasets. Meta-analyses of key study-specific estimates were performed, maximising statistical power and representativeness.

We analysed data from nine UK longitudinal studies. Four of the studies are birth cohorts, containing participants of a similar age (age-homogeneous): National Child Development Study (NCDS; born 1958) (16, 17), British Cohort Study (BCS70; born 1970) (17, 18), Next Steps (NS; born 1989-90) (17, 19) and Millennium Cohort Study (MCS; born 2000-02) (17, 20). The remaining five studies covered a wider range of ages (age-heterogeneous): Avon Longitudinal Study of Parents and Children (ALSPAC)[1] (21, 22), TwinsUK (23, 24), Born in Bradford (BiB) (25, 26), Understanding Society (USoc) (27, 28) and Generation Scotland (GS) (29, 30). Full details of the studies are provided in Table S1 (Supplementary Material), with ethics and data access statements in Table S2 (Supplementary Material).

Information relating to COVID-19 and symptoms was obtained from questionnaires completed by study participants between July 2020 and September 2021 (periods differed by study).

Variables

Here we provide an overview of the variables used in the analysis. Further details of how information was captured and variables derived in each study are provided in Methods S1 (Supplementary Material).

Symptoms: In each study, respondents reported the presence of different individual symptoms, such as fever, cough and sore throat, regardless of whether they attributed these symptoms to any specific cause. We considered a “core” set of symptoms which were almost all available in all studies to aid between-study comparability and a “maximal” set of symptoms to allow broader consideration of symptoms. The symptoms included in each set are shown in Table 1. The period over which the presence of symptoms was reported also differed by study, between two weeks and two months. In two studies (TwinsUK, USoc), symptoms were observed at multiple timepoints for each individual with the presence or absence of each symptom derived for each symptom timepoint.

[1] We included both the ALSPAC cohort (ALSPAC-G1; born 1991-92) and their parents (ALSPAC-G0) as a pooled sample.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Functional limitation following COVID-19: This was asked about in NCDS, BCS70, NS, MCS and TwinsUK only using the question *"For how long were you unable to function as normal due to Coronavirus symptoms?"* (or a subtle variation thereof).

COVID-19: Prior or current COVID-19 was self-reported in all studies. Among individuals reporting prior or current COVID-19, time since COVID-19 onset at the point of symptom reporting was derived using the date of the symptom timepoint and the reported date of COVID-19 onset (complete date or month and year only depending on study). For cohorts unable to derive time since COVID-19 onset (USoc, GS), self-reported symptom length was instead used. We derived a COVID-19 status indicator (time-varying for studies with multiple symptom timepoints (TwinsUK, USoc)) using information on prior COVID-19, time since COVID-19, and functional limitation at 12 weeks post-COVID-19, with categories:

1. No COVID-19
2. COVID-19 in last 12 weeks
3. COVID-19 > 12 weeks ago + no functional limitation at 12 weeks
4. COVID-19 > 12 weeks ago + functional limitation at 12 weeks

In studies where data on functional limitation were not collected (ALSPAC, BiB, USoc, GS), categories 3 and 4 could not be differentiated so were pooled. In studies which only collected symptom data for participants who reported prior COVID-19 (USoc, GS), category 1 was not present and only individuals with prior COVID-19 were analysed. We emphasise that these categories capture only time since reported COVID-19 at the point of symptom reporting and self-reported functional limitation at 12 weeks; they do not necessarily suggest that reported symptoms were due to COVID-19, which is why we make use of the symptoms reported in the 'No COVID-19' group.

Sex: Sex (male/female) was obtained from responses to the same or earlier questionnaires.

Age: Age at each symptom timepoint was derived from the date of the symptom timepoint and the date of birth reported at the same or earlier questionnaires (age-heterogeneous studies) or the known common date of birth (age-homogeneous studies).

Statistical analysis

Individual symptom analyses: For each available symptom within each study, the number and percentage of participants reporting the symptom within each COVID-19 group were tabulated. Prevalences of symptoms within the core symptom set were subsequently combined across studies using random-effects meta-analysis (restricted to studies with functional limitation information when considering categories where this was required). Logistic regression (for studies with a single symptom timepoint), logistic generalised estimation equations (GEE) with clustering by participant identifier and an unstructured correlation matrix (USoc) or fixed effects logistic regression (TwinsUK; due to non-convergence of GEE approach) were used to estimate odds ratios (ORs) comparing symptom presence in the COVID-19 groups. The 'no COVID-19' category was considered as the reference group, except in studies where symptom data were only available for participants who had reported prior COVID-19 (GS, USoc). In such cases the 'COVID in last 12 weeks' category was considered the baseline group. Models were adjusted for sex (male/female), age (age-heterogeneous studies only; continuous) and calendar time (for most studies, month). Survey design

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

weights (where necessary) and non-response weights (where available) were used. To be included in these analyses, study participants needed to have observed data on a given symptom (plus calendar time and age, though these were fully observed). ORs for symptoms within the core symptom set were subsequently combined across studies using random-effects meta-analysis (restricted to studies with a 'no COVID-19' category to allow a consistent reference group and to studies with functional limitation information when considering categories where this was required). These analyses were intended to be descriptive, providing an exploration of the symptom data prior to undertaking the clustering analyses.

Symptom clustering analyses: We conducted, within each study, latent class analyses (LCAs) of reported symptoms separately within each category of the previously derived COVID-19 status indicator. This was undertaken separately in primary (core symptom set) and secondary (maximal symptom set) analyses following an identical procedure. Due to small numbers in the 'COVID-19 > 12 weeks ago + functional limitation at 12 weeks' group, this category was combined with the 'COVID-19 > 12 weeks ago + no functional limitation at 12 weeks' category to form a 'COVID-19 > 12 weeks ago' category. All individuals within each study with observed data on symptoms from at least one wave of data collection were included in the LCA.

We fitted LCAs of symptoms with increasing numbers of classes, from 1 to 5, unless non-convergence occurred first. Where available, calendar time (in months for most studies) of symptom observation or wave of data collection was allowed to affect latent class membership. Study design weights (if applicable) and non-response weights (if available) were utilised. Sufficient different starting values were used to ensure that the obtained maximum likelihood solution was replicated. Full information maximum likelihood was used to handle a small amount of missing symptom data in some studies. For each obtained LCA solution, we noted model fit statistics (Akaike information criterion (AIC), Bayesian information criterion (BIC), adjusted-BIC), entropy (a summary measure of the certainty with which individuals can be allocated to classes) and the percentage of individuals in the smallest class. The optimal number of latent classes in each LCA was determined through consideration of the model fit statistics. The information criteria were plotted against the number of classes and the optimal number chosen through identification of a point of inflection in the BIC curve (31), with the additional criterion that the smallest class must be >5% of the total sample. The LCA outputs of primary interest were the number of classes supported by the data in each COVID-19 group and the probability of each symptom within each latent class (i.e. how the symptom pattern could be characterised). Formal quantitative cross-study synthesis (e.g. meta-analysis) of the symptom patterns was not undertaken, with a more qualitative approach utilised.

Associations with symptom patterns: For the core symptom set findings, subsequent analyses used logistic regression to examine how patterns differed by sex, age group (age-heterogeneous studies only) and self-reported functional limitation post-COVID onset (individuals with COVID-19 onset > 12 weeks ago only). Given the high entropy values observed, participants were allocated to their most likely latent class (symptom pattern) according to their posterior probabilities of class membership. Inclusion in each of these analyses had the additional requirement of complete data on the relevant variable. ORs were subsequently combined across studies using random-effects meta-analysis.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

RESULTS

Across the nine studies we analysed a total of over 42,000 individuals, considering between 16 and 28 symptoms reported across 15 months of the pandemic (July 2020 to September 2021).

Individual symptom analyses

For each study, the number and percentage of participants reporting each symptom within each COVID-19 group are reported in Table S4 (Supplementary Material); meta-analysed prevalences are presented in Fig. 1 with the underlying data in Table S5 (Supplementary Material). An important observation is that individuals within the 'no COVID-19' group reported moderate levels of many symptoms, including some of those commonly associated with COVID-19. For example, across all studies headaches were reported by between 20.7% (95% confidence interval (CI) 15.9%, 25.9%) of those with no reported COVID-19 in each study, fatigue by 20.0% (17.7%, 22.5%) and muscle or body aches or pains by 12.5% (9.4%, 16.0%).

Estimated ORs comparing symptom presence in the COVID-19 groups within each study are also reported in Table S4 (Supplementary Material); meta-analysed ORs are presented in Fig. 2 with the underlying data in Table S6 (Supplementary Material). A small proportion of ORs were not estimable due to low symptom prevalence in one or more COVID-19 groups. Whilst there was considerable heterogeneity between studies, many symptoms had a higher prevalence in the 'COVID-19 in last 12 weeks' group than in the 'no COVID-19' group, with this being most marked for loss of smell (meta-analysed OR 28.6; 95% CI 16.6, 49.2), loss of taste (20.5; 15.4, 27.4), fever (5.5; 4.3, 7.1), cough (3.6; 2.0, 6.3) and shortness of breath (3.1; 2.6, 3.8). The relative prevalence of all symptoms was lower in the 'COVID-19 > 12 weeks ago + no functional limitation at 12 weeks' group than in the 'COVID-19 in last 12 weeks', though remained particularly elevated (relative to the 'no COVID-19' group) for loss of smell (6.8; 4.4, 10.5) and loss of taste (4.2; 3.1, 5.8), with other ORs no greater than 2. In the 'COVID-19 > 12 weeks ago + functional limitation at 12 weeks' group there were again many symptoms with raised prevalence relative to the 'no COVID-19' group including, in addition to loss of taste (33.1; 9.8, 111.5) and loss of smell (26.3; 7.7, 89.2), fatigue (13.7; 6.9, 27.3), shortness of breath (11.9; 5.3, 26.6), muscle pain or aches (9.7; 6.0, 15.8), chest tightness (7.3; 4.3, 12.3), memory loss (6.3; 3.1, 13.0) and difficulty concentrating (6.2; 1.4, 28.0).

Symptom clustering analyses

For the primary analysis using the core symptom set, the data did not support more than two latent classes (symptom patterns) in each COVID-19 group within each study (LCA model fit statistics in Table S6 (Supplementary Material))[2]. The probability of each symptom within each symptom pattern are shown in Fig. S1 (Supplementary Material) for each study. In each instance, we identified one pattern ('symptom pattern 1' in the figures) that was characterised by a generally low prevalence of symptoms. A second pattern ('symptom pattern 2') was characterised by a higher prevalence of many symptoms, though precisely which symptoms had particularly high prevalence differed by COVID-19 group. The general similarity of symptom pattern 1 across the COVID-19 groups in each

[2] Analyses were not possible for either group with COVID-19 in BiB due to insufficient sample size.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

study suggests that this pattern identifies subgroups of similar individuals who, although they may have non-negligible probability of common symptoms such as a runny nose or a headache, were essentially well. The higher symptom burden within symptom pattern 2 therefore identifies individuals who are unwell, either due to non-COVID-19-related reasons (in the case of the 'no COVID-19' group) or due to a combination of COVID-19- and non-COVID-19-related reasons (as in the 'COVID-19 in last 12 weeks' and 'COVID-19 > 12 weeks ago' groups). Through a comparison of symptom pattern 2 between the two groups with COVID-19 and the no COVID-19 group we can explore which symptoms have the greatest excess probability relative to the COVID-19-free population, allowing us to identify symptoms that are typical of more acute COVID-19 and of long COVID.

Such comparisons can be more easily made using plots of the absolute probability differences, presented and in Fig. S2 (Supplementary Material) for each study[3]. To further aid cross-study interpretation of these findings, we have plotted the symptom probability differences for all available studies together on a single heatmap (Fig. 3). Although there was variability between studies, some common features were observed. The symptoms most consistently observed to have excess probability among individuals with COVID-19 in the last 12 weeks were loss of taste, loss of smell, fatigue, cough (particularly dry cough), shortness of breath, muscle pains or aches, fever, headaches and difficulty concentrating. The 95% CIs for these excess probabilities almost always excluded the null within each study, providing compelling evidence that these symptoms can be considered characteristic of more acute COVID-19. The symptoms most consistently observed to have excess probability among individuals with COVID-19 > 12 weeks ago were fatigue, shortness of breath, muscle pain or aches, difficulty concentrating, chest tightness, loss of smell, memory loss and loss of taste. The 95% CIs for these excess probabilities did not always exclude the null within each study, but the consistency of the findings across the cohorts provides strong evidence that these symptoms can be considered characteristic of long COVID.

In the secondary analysis using the maximal symptom set, the data again did not support more than two latent classes (symptom patterns) in each COVID-19 group within each study (Table S8 (Supplementary Material))[4]. The probability of each symptom within each symptom pattern are shown in Fig. S3 (Supplementary Material) for each study, with the absolute probability differences plotted in Fig. S4 (Supplementary Material). Additional symptoms observed to have excess probability among individuals with COVID-19 in the last 12 weeks were chills and heaviness in arms or legs. Among individuals with COVID-19 > 12 weeks ago, only heaviness in arms or legs was additionally identified.

Associations with symptom patterns

Estimated associations with symptom patterns are shown in Table S8 (Supplementary Material) for each study; meta-analysed ORs are presented in Fig. 4 with the underlying data in Table S10 (Supplementary Material). The majority of individuals who had COVID-19 were unable to function as normal for less than two weeks (between 58.9% and 88.0% across the studies), with relatively few unable to function as normal for 12 weeks or more (1.5% to 7.4%). Across almost all studies there

[3] With the exception of GS and USoc which lack a no COVID-19 group to act as a comparator.
[4] Analyses were again not possible for either group with COVID-19 in BiB due to insufficient sample size.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

was consistent evidence that symptom pattern 2 (corresponding to a higher symptom burden) was more common among females in each of the COVID-19 groups (e.g. meta-analysed OR 1.6; 95% CI 1.4, 1.9 in the COVID-19 > 12 weeks ago group). In the age-heterogeneous studies there was evidence that symptom pattern 2 was more common at younger ages in the no COVID-19 group and, to a lesser extent, in the COVID-19 > 12 weeks ago group. Findings relating to functional limitation following COVID-19 were clear and consistent: the prevalence of symptom pattern 2 was greater for individuals who were unable to function as normal for longer, being particularly high in those who were unable to function for 12 weeks or more (meta-analysed OR 8.7; 95% CI 5.4, 14.2 relative to always being able to function as normal).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

DISCUSSION

We have characterised patterns of COVID-19 and long COVID symptoms across nine UK longitudinal studies and examined how patterns differed by key factors such as sex, age and (for long COVID) self-reported functional limitation following COVID-19.

In analyses of individual symptoms, we found replication of known symptoms of COVID-19, in that fever, cough and loss of smell and taste all had highest prevalence in the group with COVID within the past 12 weeks. This suggests that despite using self-reported COVID, and asking people to recall symptoms over varying periods, our results have face validity. The prevalence of some symptoms varied across studies; this could be due to seasonality, the different variants during different stages of the pandemic or just between-study differences in age, geography or other factors. The prevalence of runny nose and sneezing did not seem to differ between those with and without COVID-19, or those with COVID-19 within the past 12 weeks or longer ago, suggesting that these symptoms tend not to be COVID-19-specific.

In the symptom clustering analyses, the data did not support more than two symptom patterns among any of the COVID-19 groups in any of the studies, though relatively small sample sizes in these groups may have affected our ability to identify further symptom patterns of low prevalence. Other studies using differing clustering methods and study designs have found greater than two symptom patterns annotated as distinct symptoms sets when studying acute COVID-19 (32, 33) and long COVID (34). However, some studies have similarly found two symptom patterns to best fit the data: in the Norwegian Mother, Father and Child Cohort Study, Caspersen et al 2022 (1) analysed 73,727 adults followed throughout the pandemic and observed distinct patterns of post-acute symptoms characterised as 'neurocognitive' and 'cardiorespiratory'. The symptoms were captured at 12 months post-infection, so it could be that symptoms become more disaggregated at a longer time interval since initial infection. Reflecting our results, Peluso et al (35) observed patients to group into two clusters, one reflecting high symptom prevalence and the other representing low, although in their approach they first aggregated reported symptoms to seven domains. Considered with our results, this does not support the idea that long COVID may be multiple syndromes discernible by their difference in symptom pattern.

Symptom pattern 2 (characterised by higher symptom burden) was generally more common among individuals with COVID-19 in the last 12 weeks or COVID-19 > 12 weeks ago than among those with no COVID-19, suggesting that, whilst there is a significant symptom burden among those who have never had COVID-19, this is greater in those who have had COVID-19. The presence of a substantial group of individuals without COVID-19 reporting a relatively high symptom burden emphasises the importance of a control group in analyses of COVID-19 symptoms.

Although symptom pattern 2 corresponded to a generally higher symptom burden, the precise symptom profile differed between COVID-19 groups. Although there was some between-study variability, the symptoms identified through comparison of the 'COVID-19 in last 12 weeks' and 'no COVID-19' groups as being characteristics of acute COVID-19 and ongoing symptomatic COVID-19 were loss of taste, loss of smell, fatigue, cough (particularly dry cough), shortness of breath, muscle pains or aches, fever, headaches and difficulty concentrating. Several meta-analyses have reported similarly, with fatigue, cough and alterations to taste and smell being characteristic of acute COVID-19 (36-38).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Symptoms identified through comparison of the 'COVID-19 > 12 weeks ago' and 'no COVID-19' groups as being characteristic of long COVID were fatigue, shortness of breath, muscle pain or aches, difficulty concentrating, chest tightness, loss of smell, memory loss and loss of taste. These symptoms are comparable to those identified in the existing literature. A clinical review by Crook and colleagues similarly found shortness of breath, impaired cognition, chest pain and in particular fatigue to characterise long COVID (3). Similarly, a meta-analysis by Martimbianco et al. (14) observed chest pain, fatigue, shortness of breath and cough as the symptoms characterising long COVID. However, these studies were not able to incorporate a 'no COVID-19' group to account for baseline population symptoms which may account for the differences observed.

Symptom pattern 2 (characterised by higher symptom burden) was found to be more common in females in all the COVID-19 groups. While this could be interpreted as females having a higher (true) underlying symptom burden than males in each of these groups, an alternative explanation could be differential reporting of (ostensibly similar) symptoms between males and females. Unlike in studies of health care use, differential health-seeking behaviour is unlikely to provide an explanation as symptom information was requested of all participants in each study. Decrease in symptom burden has been previously observed in older age groups compared with younger (e.g. (32)), reflecting our results here. The identified symptom patterns among individuals with COVID-19 > 12 weeks ago were found to be strongly associated with length of time unable to function as normal due to COVID-19 symptoms. This shows that the symptom pattern identified by the LCA relates closely to long COVID.

There are many strengths to our work. Working across multiple studies with different geographic and demographic characteristics allowed us to compare findings and draw more robust conclusions. Co-ordinated standardised analyses minimised methodological heterogeneity and maximised comparability, while appropriately accounting for study designs and characteristics of individual datasets. Meta-analyses of key study-specific estimates maximised statistical power and representativeness. Focussing on functional limitation due to COVID-19 symptoms in order to identify long COVID led to small numbers which caused analytical problems, but we successfully overcame this through a novel application of LCA which allowed us to identify symptoms characteristic of long COVID (as well as acute COVID-19). This was only possible due to the inclusion of a control group of COVID-19-free individuals, which has been a limitation of previous research (15).

There are also some limitations. Although working across multiple studies has its benefits, between-study variability in structure and data availability (particularly which symptoms were reported, how and when) added considerable complexity to the analysis. While each study had a reasonable total analytic sample size, as analyses were conducted separately in COVID-19 groups, small sample sizes, particularly in the groups with COVID-19, may have affected analyses, in particular our ability to identify symptom patterns of low prevalence. This may have been exacerbated by the relatively limited number of symptoms enquired about in some studies, reflected in the core symptom set considered, though analyses using additional symptoms were possible in a smaller number of studies. Further partitioning of the 'COVID-19 in last 12 weeks' group into those who had had COVID-19 in the last 4 weeks and those who had had COVID-19 4 to 11 weeks ago would have allowed separate analyses relating to acute COVID-19 and ongoing symptomatic COVID-19, but this was not possible due to low numbers of relevant participants. Because we relied on self-reported COVID-19 status, individuals who had asymptomatic COVID-19 or who had COVID-19 which was misattributed

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

to another cause may have been incorrectly classified as never having had COVID-19. If these misclassified true COVID-19 cases had COVID-19-related symptoms when we observed them, this could attenuate the differences in symptoms between the COVID-19 and no COVID-19 groups, making our findings conservative. We carried out complete-case analyses and were only able to apply non-response weights in studies where these were available. In the remaining studies (and potentially to some extent even in studies with non-response weights due to residual bias), if individuals with more debilitating symptoms were more/less likely to respond to a questionnaire, we would over/under-estimate the prevalence of symptoms. However, unless this happened differentially with respect to COVID-19 status, this would not bias our estimates of differences between COVID-19 groups. Vaccination status was not considered – and given the timing of symptom data collection relative to the vaccination programme rollout would not have been relevant for many studies – but could be of interest in future research. Finally, because data were collected prior to the emergence and dominance of the Omicron variant in the UK, findings may not be generalisable to the current UK circumstances (39).

Conclusions

Across nine UK longitudinal studies we identified patterns of symptoms in individuals with and without COVID-19 which allowed us to discern symptoms characteristic of acute COVID-19 and long COVID. The symptoms we identified largely replicated those previously identified in the literature. Building the evidence base regarding typical long COVID symptoms will improve diagnosis of this condition and the ability to elicit underlying biological mechanisms, leading to better patient access to treatment and services.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

STATEMENTS & DECLARATIONS

FUNDING

This work was supported by the National Core Studies, an initiative funded by UKRI, NIHR and the Health and Safety Executive. The COVID-19 Longitudinal Health and Wellbeing National Core Study was funded by the Medical Research Council (MC_PC_20030 and MC_PC_20059). The research was also underpinned by additional NIHR funding (COV-LT-0009).

Data gathered from questionnaire(s) was provided by Wellcome Longitudinal Population Study (LPS) COVID-19 Steering Group and Secretariat (221574/Z/20/Z).

The 1946 National Child Development Study, the 1970 British Cohort Study, Next Steps and the Millennium Cohort Study are supported by the Centre for Longitudinal Studies Resource Centre 2015-20 grant (ES/M001660/1) and a host of other co-funders. The COVID-19 data collections in these five cohorts were funded by the UKRI grant Understanding the economic, social and health impacts of COVID-19 using lifetime data: evidence from 5 nationally representative UK cohorts (ES/V012789/1).

Born in Bradford (BiB) receives core infrastructure funding from the Wellcome Trust (WT101597MA), and a joint grant from the UK Medical Research Council (MRC) and UK Economic and Social Science Research Council (ESRC) (MR/N024397/1),the British Heart Foundation (BHF) (CS/16/4/32482), and The Health Foundation COVID-19 award (2301201). The National Institute for Health Research Yorkshire and Humber Applied Research Collaboration (ARC) (NIHR200166), and Clinical Research Network both provide support for BiB research. Born in Bradford is only possible because of the enthusiasm and commitment of the children and parents in BiB. We are grateful to all the participants, health professionals, schools and researchers who have made Born in Bradford happen.

The UK Medical Research Council and Wellcome (Grant Ref: 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf). We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. Please note that the study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool" and reference the following webpage: http://www.bristol.ac.uk/alspac/researchers/our-data/. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. Part of this data was collected using REDCap, see the REDCap website for details https://projectredcap.org/resources/citations/

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Wellcome Trust Clinical Research Facility, Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award “STratifying

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). Generation Scotland is funded by the Wellcome Trust (216767/Z/19/Z) and (221574/Z/20/Z).

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The Understanding Society COVID-19 study is funded by the Economic and Social Research Council (ES/K005146/1) and the Health Foundation (2076161). The research data are distributed by the UK Data Service.

TwinsUK receives funding from the Wellcome Trust (WT212904/Z/18/Z), the National Institute for Health Research (NIHR) Biomedical Research Centre based at Guy's and St Thomas' NHS Foundation Trust and King's College London. The TwinsUK COVID-19 personal experience study was funded by the King's Together Rapid COVID-19 Call award, under the projects original title 'Keeping together through coronavirus: The physical and mental health implications of self-isolation due to the Covid-19'. TwinsUK is also supported by the Chronic Disease Research Foundation and Zoe Global Ltd. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

NJT is a Wellcome Trust Investigator (202802/Z/16/Z), is the PI of the Avon Longitudinal Study of Parents and Children (MRC & WT 217065/Z/19/Z), is supported by the University of Bristol NIHR Biomedical Research Centre, the MRC Integrative Epidemiology Unit (MC_UU_00011/1) and works within the CRUK Integrative Cancer Epidemiology Programme (C18281/A29019). ASFK acknowledges funding from the ESRC (ES/V011650/1). MK is supported by the Medical Research Council (MR/W021315/1).

GBP acknowledges funding from the Economic and Social Research Council (ES/V012789/1). KT works in a Unit that is supported by the University of Bristol and UK Medical Research Council (MC_UU_00011/3). NC is supported by funding from the UK Medical Research Council (MC_UU_00019/2).

COMPETING INTERESTS

NC serves and gets paid for sitting on data safety and monitoring committees for drug trials sponsored by AstraZeneca. CJS acts as a consultant for ZOE Ltd for design of ZOE health studies. All other authors report no competing interests.

AUTHOR CONTRIBUTIONS

All authors contributed to the study conception and design. Material preparation and data collection were performed within-cohort as listed above. Data analysis was performed by RB, RT, CH, RJS, BH, AK, EJT, RS and KT. The first draft of the manuscript was written by RS with additional writing from KT and RB, and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY

Study specific data availability details are indicated in **Table S2** (Supplementary Material).

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

ETHICS APPROVAL

Study specific details are indicated in **Table S2** (Supplementary Material).

CONSENT TO PARTICIPATE

Informed consent was obtained from all individual participants included in the study.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

REFERENCES

1. Caspersen IH, Magnus P, Trogstad L. Excess risk and clusters of symptoms after COVID-19 in a large Norwegian cohort. European Journal of Epidemiology. 2022.
2. World Health Organization. WHO Coronavirus (COVID-19) Dashboard 2022 [Available from: https://covid19.who.int/].
3. Crook H, Raza S, Nowell J, Young M, Edison P. Long covid—mechanisms, risk factors, and management. BMJ. 2021;374:n1648.
4. Iwasaki M, Saito J, Zhao H, Sakamoto A, Hirota K, Ma D. Inflammation Triggered by SARS-CoV-2 and ACE2 Augment Drives Multiple Organ Failure of Severe COVID-19: Molecular Mechanisms and Implications. Inflammation. 2021;44(1):13-34.
5. National Institute for Health and Care Excellence (NICE), Scottish Intercollegiate Guidelines Network (SIGN), Royal College of General Practitioners (RCGP). COVID-19 rapid guideline: managing the long-term effects of COVID-19. (https://www.nice.org.uk/guidance/ng188/resources/covid19-rapid-guideline-managing-the-longterm-effects-of-covid19-pdf-51035515742). 2022.
6. Callard F, Perego E. How and why patients made Long Covid. Social Science & Medicine. 2021;268:113426.
7. Iqbal A, Iqbal K, Arshad Ali S, Azim D, Farid E, Baig MD, et al. The COVID-19 Sequelae: A Cross-Sectional Evaluation of Post-recovery Symptoms and the Need for Rehabilitation of COVID-19 Survivors. Cureus. 2021;13(2):e13080-e.
8. Mandal S, Barnett J, Brill SE, Brown JS, Denneny EK, Hare SS, et al. 'Long-COVID': a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. Thorax. 2021;76(4):396.
9. Whitaker M, Elliott J, Chadeau-Hyam M, Riley S, Darzi A, Cooke G, et al. Persistent symptoms following SARS-CoV-2 infection in a random community sample of 508,707 people. medRxiv. 2021:2021.06.28.21259452.
10. Matta J, Wiernik E, Robineau O, Carrat F, Touvier M, Severi G, et al. Association of Self-reported COVID-19 Infection and SARS-CoV-2 Serology Test Results With Persistent Physical Symptoms Among French Adults During the COVID-19 Pandemic. JAMA Internal Medicine. 2022;182(1):19-25.
11. Moreno-Pérez O, Merino E, Leon-Ramirez J-M, Andres M, Ramos JM, Arenas-Jiménez J, et al. Post-acute COVID-19 syndrome. Incidence and risk factors: A Mediterranean cohort study. J Infect. 2021;82(3):378-83.
12. Osikomaiya B, Erinoso O, Wright KO, Odusola AO, Thomas B, Adeyemi O, et al. 'Long COVID': persistent COVID-19 symptoms in survivors managed in Lagos State, Nigeria. BMC Infectious Diseases. 2021;21(1):304.
13. Augustin M, Schommers P, Stecher M, Dewald F, Gieselmann L, Gruell H, et al. Post-COVID syndrome in non-hospitalised patients with COVID-19: a longitudinal prospective cohort study. The Lancet Regional Health - Europe. 2021;6:100122.
14. Cabrera Martimbianco AL, Pacheco RL, Bagattini ÂM, Riera R. Frequency, signs and symptoms, and criteria adopted for long COVID-19: A systematic review. International Journal of Clinical Practice. 2021;75(10):e14357.
15. Amin-Chowdhury Z, Ladhani SN. Causation or confounding: why controls are critical for characterizing long COVID. Nature Medicine. 2021;27(7):1129-30.
16. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int J Epidemiol. 2006;35(1):34-41.
17. Brown M, Goodman A, Peters A, Ploubidis GB, Sanchez A, Silverwood R, et al. COVID-19 Survey in Five National Longitudinal Studies: Waves 1, 2 and 3 User Guide (Version 3). London: UCL Centre for Longitudinal Studies and MRC Unit for Lifelong Health and Ageing; 2021.
18. Elliott J, Shepherd P. Cohort Profile: 1970 British Birth Cohort (BCS70). International Journal of Epidemiology. 2006;35(4):836-43.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

19. Calderwood L, Sanchez C. Next Steps (formerly known as the Longitudinal Study of Young People in England). Open Health Data. 2016;4(1):e2.
20. Joshi H, Fitzsimons E. The Millennium Cohort Study: the making of a multi-purpose resource for social science and policy. Longitudinal and Life Course Studies. 2016;7(4):409-30.
21. Boyd A, Golding J, Macleod J, Lawlor DA, Fraser A, Henderson J, et al. Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. International Journal of Epidemiology. 2013;42(1):111-27.
22. Fraser A, Macdonald-Wallis C, Tilling K, Boyd A, Golding J, Davey Smith G, et al. Cohort Profile: The Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. International Journal of Epidemiology. 2013;42(1):97-110.
23. Verdi S, Abbasian G, Bowyer RCE, Lachance G, Yarand D, Christofidou P, et al. TwinsUK: The UK Adult Twin Registry Update. Twin Research and Human Genetics. 2019;22(6):523-9.
24. Suthahar A, Sharma P, Hart D, García M, Horsfall R, Bowyer R, et al. TwinsUK COVID-19 personal experience questionnaire (CoPE): wave 1 data capture April-May 2020 [version 1; peer review: awaiting peer review]. Wellcome Open Research. 2021;6(123).
25. Wright J, Small N, Raynor P, Tuffnell D, Bhopal R, Cameron N, et al. Cohort Profile: The Born in Bradford multi-ethnic family cohort study. International Journal of Epidemiology. 2013;42(4):978-91.
26. Dickerson J, Bird PK, McEachan RRC, Pickett KE, Waiblinger D, Uphoff E, et al. Born in Bradford's Better Start: an experimental birth cohort study to evaluate the impact of early life interventions. BMC Public Health. 2016;16(1):711.
27. Buck N, McFall S. Understanding Society: design overview. Longitudinal and Life Course Studies; Vol 3, No 1 (2012): Longitudinal and Life Course StudiesDO - 1014301/llcsv3i1159. 2011.
28. University of Essex, Institute for Social and Economic Research. Understanding Society: COVID-19 Study, 2020-2021. [data collection]. 11th Edition. UK Data Service. SN: 8644. 2021.
29. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. International Journal of Epidemiology. 2013;42(3):689-700.
30. Fawns-Ritchie C, Altschul D, Campbell A, Huggins C, Nangle C, Dawson R, et al. CovidLife: a resource to understand mental health, well-being and behaviour during the COVID-19 pandemic in the UK [version 1; peer review: 1 approved]. Wellcome Open Research. 2021;6(176).
31. Nylund KL, Asparouhov T, Muthén BO. Deciding on the Number of Classes in Latent Class Analysis and Growth Mixture Modeling: A Monte Carlo Simulation Study. Structural Equation Modeling: A Multidisciplinary Journal. 2007;14(4):535-69.
32. Cheng X, Wan H, Yuan H, Zhou L, Xiao C, Mao S, et al. Symptom Clustering Patterns and Population Characteristics of COVID-19 Based on Text Clustering Method. Front Public Health. 2022;10:795734-.
33. Sudre Carole H, Lee Karla A, Ni Lochlainn M, Varsavsky T, Murray B, Graham Mark S, et al. Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app. Science Advances.7(12):eabd4177.
34. Chopra N, Chowdhury M, Singh AK, Ma K, Kumar A, Ranjan P, et al. Clinical predictors of long COVID-19 and phenotypes of mild COVID-19 at a tertiary care centre in India. Drug Discoveries & Therapeutics. 2021;15(3):156-61.
35. Peluso MJ, Kelly JD, Lu S, Goldberg SA, Davidson MC, Mathur S, et al. Persistence, Magnitude, and Patterns of Postacute Symptoms and Quality of Life Following Onset of SARS-CoV-2 Infection: Cohort Description and Approaches for Measurement. Open Forum Infectious Diseases. 2022;9(2):ofab640.
36. Iqbal FM, Lam K, Sounderajah V, Clarke JM, Ashrafian H, Darzi A. Characteristics and predictors of acute and chronic post-COVID syndrome: A systematic review and meta-analysis. EClinicalMedicine. 2021;36:100899.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

37. Zhu J, Ji P, Pang J, Zhong Z, Li H, He C, et al. Clinical characteristics of 3062 COVID-19 patients: A meta-analysis. Journal of Medical Virology. 2020;92(10):1902-14.
38. Olumade TJ, Uzairue LI. Clinical characteristics of 4499 COVID-19 patients in Africa: A meta-analysis. Journal of Medical Virology. 2021;93(5):3055-61.
39. Menni C, Valdes AM, Polidori L, Antonelli M, Penamakuri S, Nogal A, et al. Symptom prevalence, duration, and risk of hospital admission in individuals infected with SARS-CoV-2 during periods of omicron and delta variant dominance: a prospective observational study from the ZOE COVID Study. The Lancet.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

TABLES AND FIGURES

Table 1. Core and maximal symptom sets.

Symptom	Core set	Maximal set
Fever	x	x
Cough	x	x
Sore throat	x	x
Chest tightness	x	x
Shortness of breath	x	x
Runny and/or blocked stuffy nose	x	x
Muscle or body aches	x	x
Fatigue	x	x
Diarrhoea	x	x
Loss of smell and/or taste	x	x
Nausea and/or vomiting	x	x
Raised, red, itchy areas on the skin	x	x
Sneezing	x	x
Headaches	x	x
Difficulty concentrating	x	x
Memory loss	x	x
Severe fatigue		x
Decrease in appetite		x
Abdominal pain/tummy ache		x
Sore eyes		x
Hoarse voice		x
Dizziness		x
Chest pain		x
Chills		x
Difficulty sleeping		x
Numbness or tingling somewhere in the body		x
Feeling of heaviness in arms or legs		x
Sudden swelling of the face or lips		x

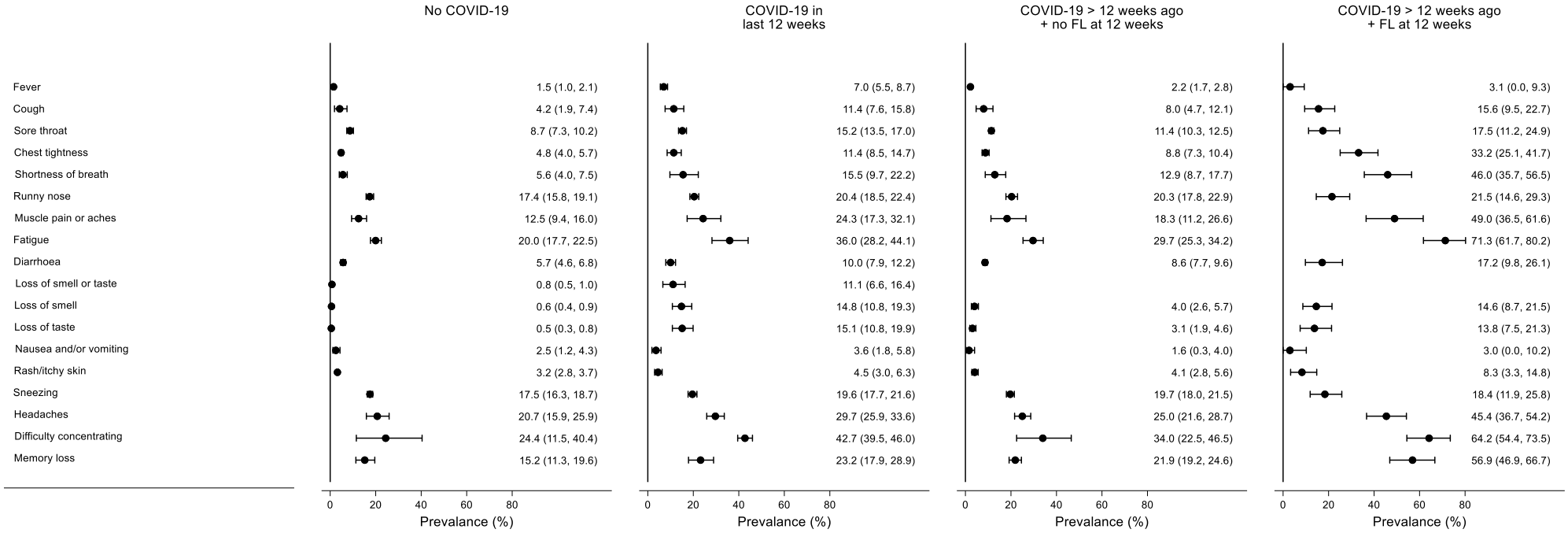

Figure 1. Meta-analysis of prevalence of each symptom for each COVID-19 group. Bars represent 95% confidence intervals. Estimates are from random effects meta-analyses of study-specific estimates. FL: function limitation.

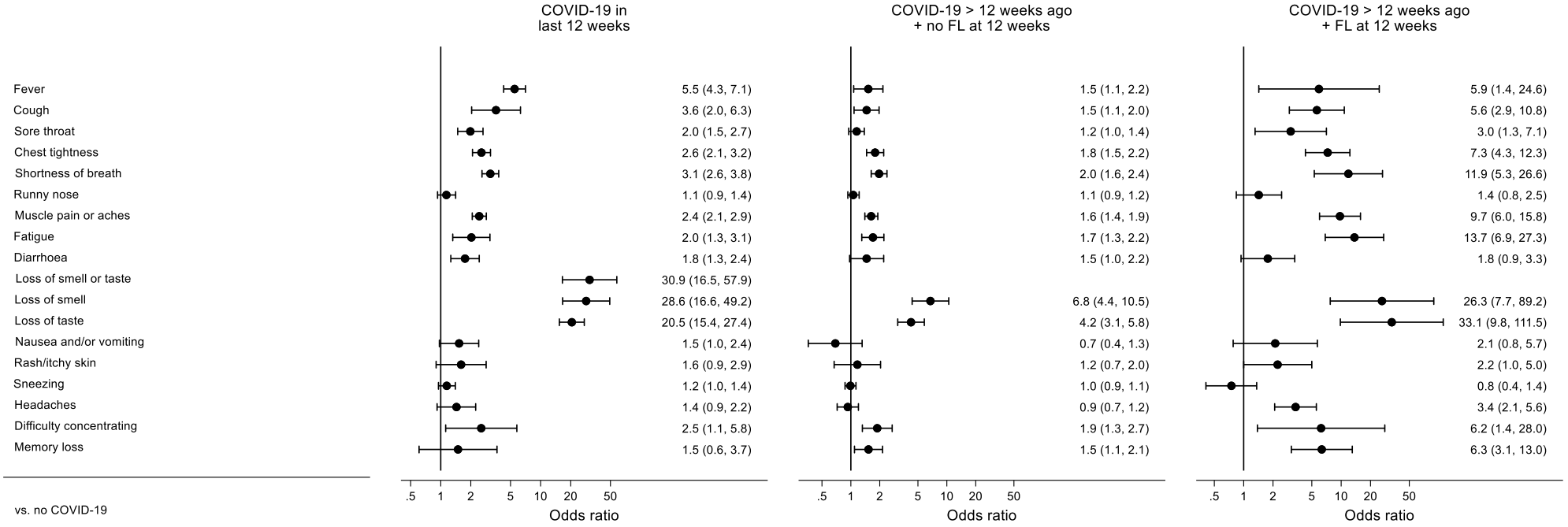

Figure 2. Meta-analysis of odds ratio of each symptom for each COVID-19 group relative to the no COVID-19 group. Bars represent 95% confidence intervals. Estimates are from random effects meta-analyses of study-specific estimates. FL: function limitation.

medRxiv preprint doi: https://doi.org/10.1101/2022.06.20.22275994; this version posted June 20, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

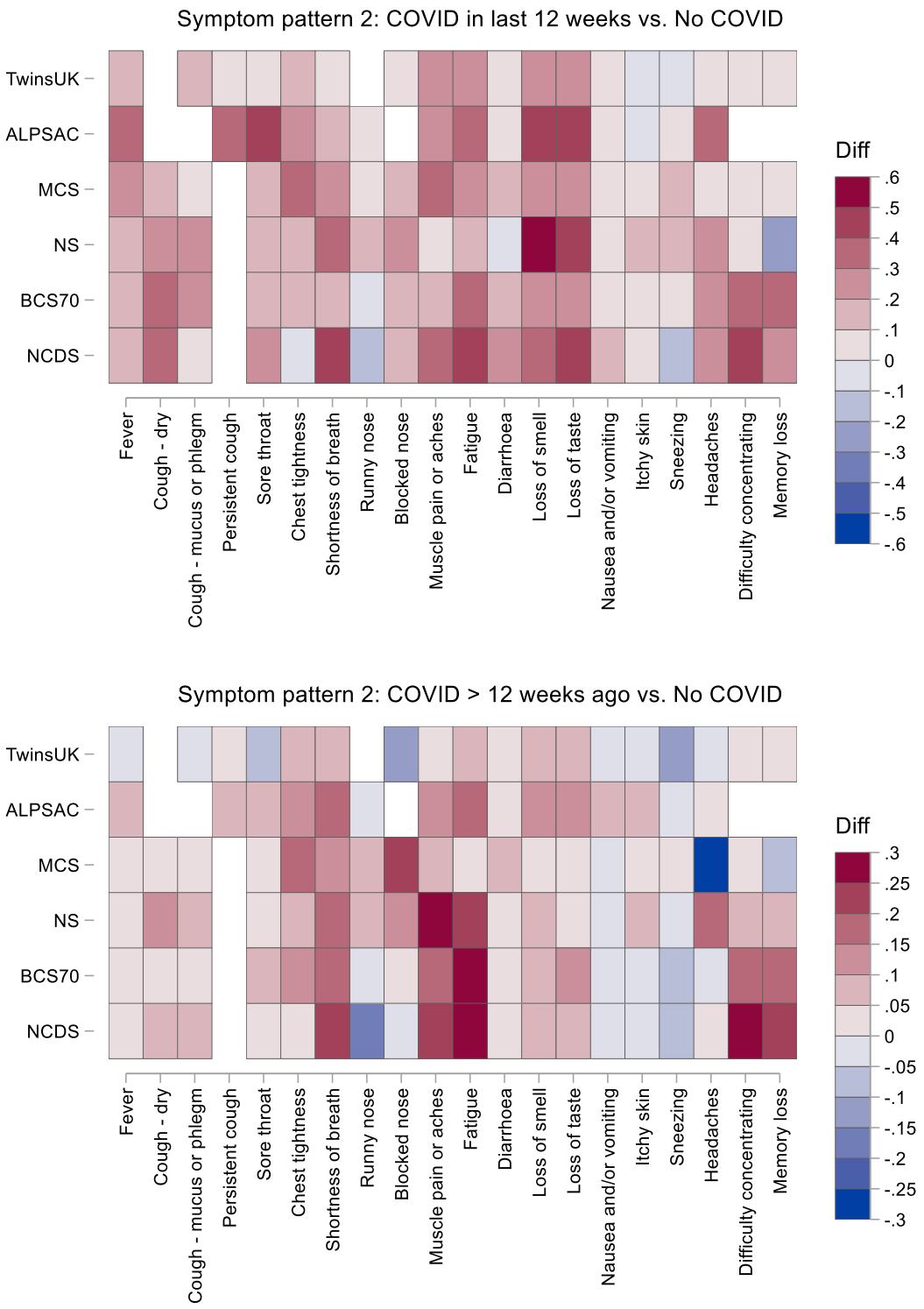

Figure 3. Heatmap of symptom probability differences comparing the two COVID-19 groups with the no COVID-19 group across all studies. NCDS: 1958 National Child Development Study; BCS1970: 1970 British Cohort Study; NS: Next Steps; MCS: Millennium Cohort Study; ALSPAC: Avon Longitudinal Study of Parents and Children. Notes: White cells indicate that the symptom was not asked about in in that study; ALSPAC results for “loss of smell or taste” are duplicated in the “loss of smell” and “loss of taste” cells; TwinsUK results for “confusion” are duplicated in the “difficulty concentrating” and “memory loss” cells.

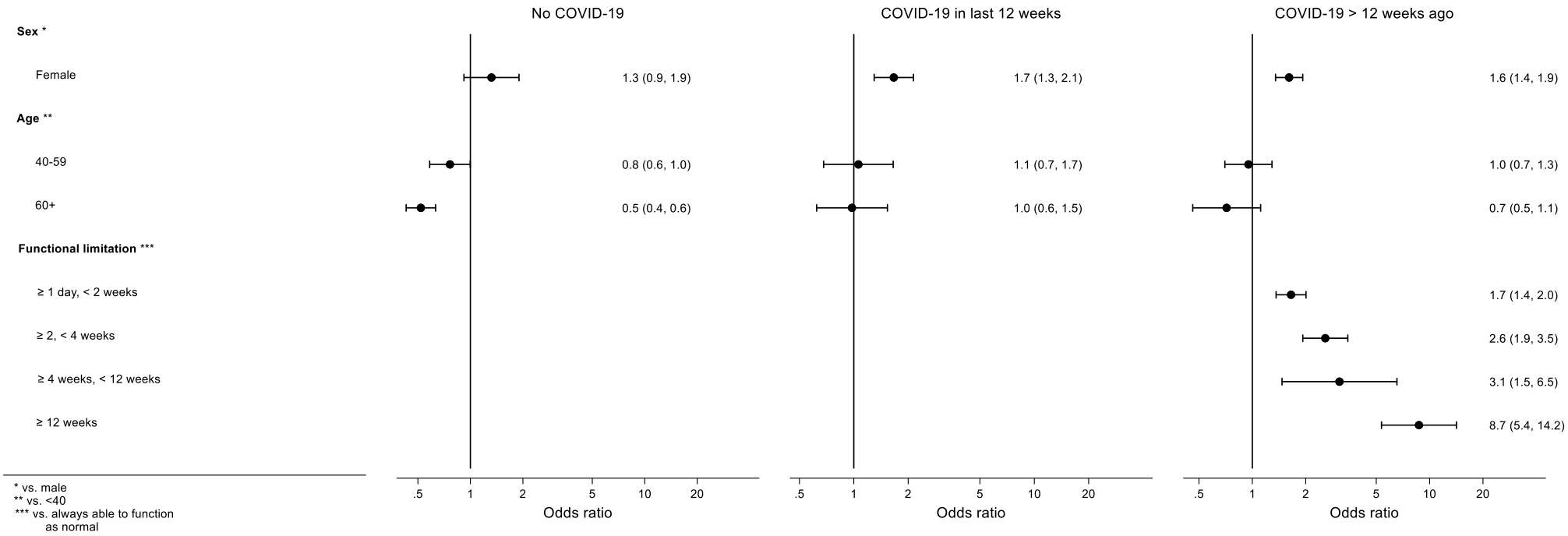

Figure 4. Meta-analysis of associations with symptom pattern 2 (vs. symptom pattern 1) across studies. Bars represent 95% confidence intervals. Estimates are from random effects meta-analyses of study-specific estimates.