medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license .

Risk factors for developing COVID-19: a population-based longitudinal study (COVIDENCE UK)

Hayley Holt[1,2,3]*
Mohammad Talaei[1]*
Matthew Greenig[1]
Dominik Zenner[1]
Jane Symons[4]
Clare Relton[1]
Katherine S Young[5,6]
Molly R Davies[5]
Katherine N Thompson[5]
Jed Ashman[1]
Sultan Saeed Rajpoot[1]
Ahmed Ali Kayyale[1]
Sarah El Rifai[1]
Philippa J. Lloyd[1]
David A. Jolliffe[1,3]
Sarah Finer[1]
Stamatina Ilidriomiti[1]
Alec Miners[7]
Nicholas S. Hopkinson[8]
Bodrul Alam[9]
Paul E Pfeffer[1]
David McCoy[1]
Gwyneth A Davies[10]
Ronan A Lyons[10]
Christopher J Griffiths[1,2]
Frank Kee[11]
Aziz Sheikh[12]
Gerome Breen[5]
Seif O Shaheen[1]*
Adrian R Martineau[1,2,3]*†

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2021.03.27.21254452; this version posted March 30, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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-NC-ND 4.0 International license .

[1]Institute of Population Health Sciences, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[2]Asthma UK Centre for Applied Research, Queen Mary University of London, London, UK

[3]Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK

[4]Jane Symons Media, London

[5]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

[6]NIHR Maudsley Biomedical Research Centre, King's College London, UK

[7]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK

[8]National Heart and Lung Institute, Imperial College London, UK

[9]Edenfield Road Surgery, Rochdale, UK

[10]Population Data Science, Swansea University Medical School, Singleton Park, Swansea, UK

[11]Centre for Public Health Research (NI), Queen's University Belfast, Belfast, UK

[12]Usher Institute, University of Edinburgh, Edinburgh, UK

* Contributed equally

† To whom correspondence should be addressed at the Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark St, London E1 2AT, UK

Tel: +44 207 882 2551 | Fax: +44 207 882 2552 | Email: a.martineau@qmul.ac.uk

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Summary

Background: Risk factors for severe COVID-19 include older age, male sex, obesity, Black or Asian ethnicity and underlying medical conditions. Whether these factors also influence susceptibility to developing COVID-19 is uncertain.

Methods: We undertook a prospective, population-based cohort study (COVIDENCE UK) from 1st May 2020 to 5th February 2021. Baseline information on potential risk factors was captured by an online questionnaire. Monthly follow-up questionnaires captured incident COVID-19. We used logistic regression models to estimate multivariable-adjusted odds ratios (aORs) for associations between potential risk factors and risk of COVID-19.

Findings: We recorded 446 incident cases of COVID-19 in 15,227 participants (2.9%). Increased risk of developing COVID-19 was independently associated with Asian/Asian British *vs*. White ethnicity (aOR 2.31, 95% CI 1.35-3.95), household overcrowding (aOR per additional 0.5 people/bedroom 1.26, 1.11-1.43), any *vs*. no visits to/from other households in previous week (aOR 1.33, 1.07-1.64), number of visits to indoor public places (aOR per extra visit per week 1.05, 1.01-1.09), frontline occupation excluding health/social care *vs.* no frontline occupation (aOR 1.49, 1.12-1.98), and raised body mass index (BMI) (aOR 1.51 [1.20-1.90] for BMI 25.0-30.0 kg/m^2 and 1.38 [1.05-1.82] for BMI >30.0 kg/m^2 *vs.* BMI <25.0 kg/m^2). Atopic disease was independently associated with decreased risk (aOR 0.76, 0.59-0.98). No independent associations were seen for age, sex, other medical conditions, diet, or micronutrient supplement use.

Interpretation: After rigorous adjustment for factors influencing exposure to SARS-CoV-2, Asian/Asian British ethnicity and raised BMI were associated with increased risk of developing COVID-19, while atopic disease was associated with decreased risk.

Funding: Barts Charity, Health Data Research UK

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Introduction

COVID-19 has taken a heavy toll on the health of populations globally.[1-3] Risk factors for severe and fatal disease include male sex, Black or Asian ethnic origin, obesity, deprivation and a range of comorbidities including diabetes mellitus, cardiovascular disease, chronic obstructive pulmonary disease (COPD) and hypertension.[4,5] However, there is growing evidence from population-based studies to suggest that at least some risk factors for developing COVID-19 - irrespective of severity - may be distinct from those which predispose to disease at the most severe end of the spectrum. For example, population-based studies in both the US and the UK have reported that risk of COVID-19 is higher in younger vs. older adults [6], a finding supported by serology studies in the UK and Switzerland reporting higher prevalence of antibodies to SARS-CoV-2 in younger vs older adults.[7,8] Again, in contrast to studies reporting that diabetes, heart disease and hypertension are risk factors for severe disease, the presence of pre-existing health conditions has been reported to associate with decreased, rather than increased, risk of SARS-CoV-2 sero-positivity in a population-based study conducted in the UK.[7]

These apparently paradoxical associations are potentially attributable to changes in behaviour in response to the pandemic, whereby people at greater risk of severe disease because of older age or presence of comorbidities may reduce social contact and visits to indoor public places in order to reduce their exposure to SARS-CoV-2. However, to our knowledge, studies to investigate whether behaviours influencing risk of such exposure might partly explain associations between older age, presence of comorbidities and lower risk of developing COVID-19 are lacking. Such studies could potentially shed light on other controversies relating to risk factors for developing COVID-19, such as the extent to which ethnic differences in disease susceptibility can be explained by behavioural, occupational and socio-economic factors,[9] and whether lifestyle, diet, and use of micronutrient supplements may influence risk of developing COVID-19.[10,11]

In order to address this knowledge gap, we established a new longitudinal study (COVIDENCE UK) at the start of the pandemic, with the specific aim of capturing detailed information on a very wide range of potential risk factors for COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Sufficient incident cases of test-confirmed COVID-19 have now accumulated to allow us to evaluate how a comprehensive panel of demographic, socio-economic, lifestyle, dietary, pharmacological and comorbidity factors relate to the risk of developing COVID-19.

Methods

Study design, setting and participants

COVIDENCE UK (Longitudinal population-based observational study of coronavirus disease in the UK population, www.qmul.ac.uk/covidence) is a prospective cohort study with four main objectives, namely to: i) determine risk factors for incident COVID-19 in the UK population; ii) characterise the natural history of COVID-19 in the UK population; iii) evaluate the impact of COVID-19 on the physical and mental health of the UK population; and iv) provide a resource from which to identify potential participants for future clinical trials of interventions to reduce incidence and/or severity of COVID-19 and other acute respiratory infections. Inclusion criteria were age 16 years or more and residence in the UK at the point of enrolment; there were no exclusion criteria. Participants were invited via a national media campaign to complete an on-line baseline questionnaire to capture information on potential symptoms of COVID-19 experienced since 1st February 2020, results of any COVID-19 tests, and details of a wide range of potential risk factors for COVID-19, as described below. Follow-up questionnaires were administered at monthly intervals to capture incident test-confirmed COVID-19 as well as potential symptoms of COVID-19. The study was launched on 1st May 2020, and this paper reports findings of analysis of data collected up to 5th February 2021.

Sponsorship, registration, ethics and reporting

The study was sponsored by Queen Mary University of London and approved by Leicester South Research Ethics Committee (ref 20/EM/0117). It is registered with ClinicalTrials.gov (NCT04330599).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Outcomes

The primary outcome was incidence of test-confirmed COVID-19, as defined by a self-reported positive result from PCR or lateral flow testing of eluate from a nose or throat swab for SARS-CoV-2. Those who were not tested were assumed to be test negative. The secondary outcome was incidence of symptom-defined probable COVID-19, with casehood defined using the algorithm described by Menni and colleagues,[12] based on age, sex and self-reported loss of smell/taste, significant/severe persistent cough, severe fatigue and skipped meals (see supplementary Appendix for further details). This outcome was included in order to address potential under-ascertainment of COVID-19 arising from use of test-confirmed COVID-19 as an outcome measure, which would not have captured episodes where testing was not done, potentially introducing collider bias.[13] In order to minimise the potential for reverse causality to explain associations observed, outcomes occurring within 30 days of enrolment were censored. At enrolment, participants were asked to provide details of potential symptoms of COVID-19 experienced since 1st February 2020, and results of any PCR or lateral flow tests for SARS-CoV-2 performed on eluates from nose / throat swabs to date.

Independent variables

At enrolment, participants were asked to complete an on-line questionnaire capturing information about their socio-demographic characteristics, type of occupation, lifestyle, weight, height, longstanding medical conditions, medication use, vaccination status, diet and supplemental micronutrient intake (for baseline questionnaire, see Table S1, Supplementary Appendix). Monthly on-line follow-up questionnaires (Table S2, Supplementary Appendix) captured incident test-confirmed COVID-19 and potential symptoms of COVID-19.

Sample size

The sample size required to detect an odds ratio of at least 1.08 (effect size) for a binary exposure variable with maximum variability (probability = 0.50 changing to 0.52) and correlated with other variables in the model (R^2 = 0.5), with a power of

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

90% using a one-sided test with 5% significance level was estimated as 10,721, using the 'powerlog' program in Stata 14.2 (College Station, TX). Assuming 10% censoring at baseline (prevalent COVID-19 and missing) and 20% loss to follow-up, we aimed to recruit a minimum of 14,890 participants. No upper limit for sample size was specified.

Statistical methods

Statistical analyses were performed using Stata 14.2. Putative risk factors for COVID-19 were selected *a priori* and classified into the following groups: socio-demographic, occupational and lifestyle factors; longstanding medical conditions, medication use and vaccination status; and diet and supplemental micronutrient intake. To produce patient-level covariates for each class of medications investigated, participant answers were mapped to drug classes listed on the British National Formulary (BNF) or the DrugBank and Electronic Medicines Compendium databases if not explicitly listed on the BNF; further details of the computational methods used to achieve this are presented in supplementary Appendix. Index of Multiple Deprivation (IMD) 2019 scores were assigned according to participants' postcodes, and categorised into quartiles.

Participants who reported definite COVID-19 prior to enrolment, or who were classified as having had symptom-defined probable COVID-19 prior to enrolment on the basis of self-reported symptoms, were excluded from prospective analyses. Logistic regression models were used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for each individual factor, first in a crude model, then in 'minimally adjusted' models. For the primary outcome of test-confirmed COVID-19, the minimally adjusted model included age (six categories), sex (male/female), duration of participation and frequency of COVID-19 testing. For the secondary outcome of symptom-defined probable COVID-19, the minimally adjusted model included age (six categories), sex and duration of participation only. Factors associating independently with each outcome at the 10% significance level were evaluated for collinearity using Cramer's V statistic[14] for all pairwise combinations of the covariates and clustering with average-linkage hierarchical clustering using 1 - V as a dissimilarity metric. The resultant heat-maps were reviewed, and where a

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

cluster of highly correlated covariates was identified, one variable within each cluster that was deemed to relate most closely and plausibly to COVID-19 risk was selected for inclusion in the multivariable model. Covariate associations with age were not deemed to warrant exclusion from the multivariable model, since all covariates investigated in the collinearity analysis exhibited a significant relationship with COVID-19 in the minimally-adjusted model that controlled for age. Variables that did not exhibit clustering in the heat-maps were all included in the final model, along with age, sex, IMD score quartile, duration of participation and frequency of COVID-19 testing. Correction for multiple comparisons was not applied, on the grounds that we were testing *a priori* hypotheses for all risk factors investigated.[15] For the primary outcome, two sensitivity analyses were performed: one excluded participants who received one or more doses of COVID-19 vaccine, and the other excluded those who were randomised to receive vitamin D supplementation as part of a nested clinical trial that was initiated during follow-up.

Role of the funding source

Barts Charity and Health Data Research UK had no role in study design, data analysis, data interpretation, or writing of the report. MT, HH and MG had access to raw data. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication.

Results

Of the 17,558 participants who completed the COVIDENCE UK baseline questionnaire on or before 2nd November 2020, we excluded those who were identified as already having had test-confirmed and/or symptom-defined probable COVID-19 (n=1,613) and those who did not complete at least one monthly follow-up questionnaire (n=854), leaving 15,227 participants eligible for inclusion in this prospective analysis (Figure S1, Supplementary Appendix). Selected baseline characteristics are presented in Table 1. Mean age was 59.4 years (range 16.0 to 94.4 years), 69.8% were female, and 94.9% identified their ethnic origin as White.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

The geographical distribution of COVIDENCE UK participants aligned closely with that of incident COVID-19 in the UK (Figure S2, Supplementary Appendix).

A total of 446 participants experienced at least one episode of PCR- or lateral flow test-confirmed COVID-19 during 2,613,921 person-days of follow-up, of whom 32 were hospitalised. We calculated crude and minimally-adjusted ORs for associations between risk of test-confirmed COVID-19 and socio-demographic, occupational and lifestyle factors (Table 2); longstanding conditions, medication use and vaccination status (Table 3); and diet and supplemental micronutrient intake (Table 4). After adjustment for age, sex, duration of participation and testing frequency ('minimal adjustment'), the following factors were found to associate with increased risk of COVID-19 with P <0.10: Asian / Asian British vs. White ethnic origin, housing type (paying mortgage and 'other' vs. owning own home), frontline vs. non-frontline worker status, household overcrowding (>0.5 vs. ≤0.5 people per bedroom), any visit vs. no visits to/from other households in the previous week, presence vs. absence of schoolchildren and working-age adults in the household, living with others vs. living alone, having vs. not having a dog at home, any vs. no travel to work or place of study in the week preceding questionnaire completion, number of visits to shops or other indoor public places per week (quartiles 2,3,4 vs quartile 1), alcohol consumption (15-21 vs 0 U/week); sleep duration (7 or ≥9 vs 8 hours per night); raised body mass index (BMI) (>25.0 vs ≤25.0 kg/m^2) and history vs. no history of BCG vaccination. The following factors associated with decreased risk of COVID-19 with P<0.10 after minimal adjustment: age ≥60 vs. 16-29.99 years, education to college or postgraduate level vs primary / secondary level, shielding vs. non-shielding, low-impact physical activity (≥2 vs 0 hours/week), presence vs. absence of asthma diagnosis, presence vs. absence of atopic disease (defined by atopic eczema/dermatitis and/or hayfever/allergic rhinitis), use vs. no use of systemic immunosuppressants, inhaled corticosteroids and bronchodilators (defined as β-2-adrenoreceptor agonists or anticholinergics), higher intake of fruit and vegetables (top vs bottom quartiles) and use vs. no use of vitamin D supplements.

All factors associating with test-confirmed COVID-19 with P <0.10 in the minimally-adjusted model were then assessed for collinearity: the resultant heat-map (Figure S3, Supplementary Appendix) revealed a high degree of collinearity between the

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

number of working-age adults in the household, multi-generational households and number of people per bedroom. Since household over-crowding (as indicated by number of people per bedroom) was deemed to relate most closely to SARS-CoV-2 exposure risk, the other two independent variables were excluded from the multivariable model.

Table 5 presents fully-adjusted ORs for associations between potential risk factors for test-confirmed COVID-19. The final multivariable model adjusted mutually for age, sex, duration of participation, test frequency, ethnicity, highest educational level attained, IMD rank, household income, housing type, number of people per bedroom, presence of schoolchildren at home, presence of a dog in the household, shielding, visits to/from other households, visits to shops and other indoor places, travel to work or study, frontline worker status, low-impact physical activity, alcohol intake, BMI, history of asthma, history of atopic disease, use of systemic immunosuppressants, use of inhaled corticosteroids, use of bronchodilators, BCG vaccination status, intake of fruit, vegetables and salads, and intake of supplemental vitamin D. Increased risk of developing COVID-19 was independently associated with Asian/Asian British *vs*. White ethnicity (aOR 2.31, 95% CI 1.35-3.95), household overcrowding (aOR per additional 0.5 people/bedroom 1.26, 1.11-1.43), any *vs*. no visits to/from other households in previous week (aOR 1.33, 1.07-1.64), number of visits to indoor public places (aOR per extra visit per week 1.05, 1.01-1.09), frontline occupation outside health/social care *vs.* no frontline occupation (aOR 1.49, 1.12-1.98), and raised BMI (aOR 1.51 [1.20-1.90] for BMI 25.0-30.0 kg/m^2 and 1.38 [1.05-1.82] for BMI >30.0 kg/m^2 *vs.* BMI <25.0 kg/m^2). Lower risk of test-confirmed COVID-19 was independently associated with history of atopic disease (aOR 0.76, 0.59-0.98) and taking systemic immunosuppressants (aOR 0.43, 0.19-0.95).

We performed two sensitivity analyses to explore the robustness of the multivariable results. The first (Table S3, Supplementary Appendix) excluded 3202 participants who received one or more doses of COVID-19 vaccine before the date of the data download (5th February 2021). The second (Table S4, Supplementary Appendix) excluded 3813 participants who commenced vitamin D supplements after enrolment in the cohort due to participation in a clinical trial. Both analyses yielded similar findings to those presented above.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

We also performed an exploratory analysis to determine whether COVID-19 risk differed for participants with atopic vs. non-atopic asthma endotypes, as defined by the presence or absence of 'atopic disease' (atopic eczema/dermatitis and/or hayfever/allergic rhinitis). This showed a reduced risk of COVID-19 for participants with atopic asthma (aOR 0.64, 0.42-0.96), but not for those with non-atopic asthma (aOR 0.91, 0.62-1.33), as compared to participants without atopic disease or asthma. These effect estimates did not materially change after further adjustment for inhaled corticosteroids (Table S5, Supplementary Appendix).

We then proceeded to investigate determinants of symptom-defined probable COVID-19, with casehood ascribed using an algorithm published by Menni et al.[12] In the subset of COVIDENCE UK participants who had one or more tests for COVID-19, this case definition had sensitivity and specificity for test-confirmed COVID-19 of 0.68 and 0.78, respectively, with an area under the receiver operating characteristic curve of 0.79 (95% CI, 0.76 to 0.82). Potential risk factors associating with probable symptom-defined COVID-19 with P<0.10 in a minimally-adjusted model (i.e. adjusting for age, sex and duration of participation) were included in the multivariable model presented in Table S6, Supplementary Appendix. Increased risk of probable symptom-defined COVID-19 was independently associated with Asian / Asian British vs. White ethnicity, housing type (having a mortgage vs. home-ownership), household overcrowding (>0.5 vs. ≤0.5 people per bedroom), frontline occupation outside of health or social care, use of cod liver oil supplements, poorer self-assessed general health, and use of selective serotonin reuptake inhibitors, while lower risk of probable symptom-defined COVID-19 was independently associated with greater age (age ≥50 vs. 16-29.99 years).

Discussion

In this large, prospective population-based study evaluating a diverse array of potential risk factors for developing COVID-19, we found that Asian/Asian British ethnicity, household overcrowding, indoor social mixing, employment as a frontline worker outside of health and social care, and being overweight or obese were all independently associated with an increased risk of test-confirmed COVID-19.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Associations with household overcrowding and visits to indoor public places showed dose-response relationships, supporting a causal interpretation. History of atopic disease and use of systemic immunosuppressant medication were independently associated with decreased risk of test-positive disease. No statistically significant independent associations with disease risk were seen for other factors investigated, including age, sex, diet, supplemental micronutrient intake, and other longstanding conditions and medications.

This study sheds new light on the degree of overlap between risk factors for developing COVID-19 (irrespective of severity) *vs.* risk factors for developing severe or fatal disease specifically. Our finding that people of Asian/Asian British ethnic origin are at increased risk of developing COVID-19 is consistent with reports of increased susceptibility and disease severity in this group.[4,7,16] One limitation of previous studies investigating ethnic variation in COVID-19 risk is that they did not adjust for behaviours influencing SARS-CoV-2 exposure, such as visits to other households and indoor public places. In our study, increased risk of developing COVID-19 in people of Asian/Asian British ethnic origin was not explained by such behaviours, nor by social deprivation, domestic overcrowding, occupation, BMI, or comorbidities. There is therefore an urgent need to explain ethnic disparities in risk of developing COVID-19 so that preventive strategies can be deployed. The association between raised BMI and increased susceptibility to COVID-19 that we found is consistent with studies identifying obesity as a risk factor for both susceptibility to, and severe outcomes of, COVID-19.[4,17,18] It would appear that immune dysregulation associated with obesity may increase susceptibility to infection as well as disease severity.

By contrast, a number of established risk factors for severe and fatal disease, including older age, male sex and underlying conditions such as diabetes, heart disease, COPD and hypertension, were not associated with risk of developing COVID-19 in our study. In keeping with reports from the UK[7] and elsewhere,[8] we found younger age to be associated with increased risk of developing COVID-19 in crude and minimally-adjusted models. However, this association did not persist after adjustment for multiple potential confounders, including behaviours related to social mixing, suggesting that lower incidence of COVID-19 in older adults in our study may be explained by reductions in social contact. We did not see a difference in disease

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

risk for people with diabetes, heart disease or hypertension. Whilst this contrasts with a study reporting lower prevalence of SARS-CoV-2 seropositivity among people with these underlying conditions, that study did not adjust, as we did, for behaviours influencing exposure to infection.[7] The only long-standing conditions associated with disease risk in our study were atopic diseases, which were associated with reduced risk of disease, particularly among those who also had asthma. This may reflect decreased expression of *ACE2*, the SARS-CoV-2 receptor, which has been reported in people with both high levels of allergic sensitisation and asthma.[19]

Our study has several strengths. COVIDENCE UK was set up with the specific purpose of investigating incident COVID-19, and consequently our questionnaires were specifically designed to capture contemporaneous and granular detail on potential risk factors, including behaviours influencing risk of exposure to SARS-CoV-2. Our finding that visits to other households and indoor public places were associated with increased risk of disease supports the case for restricting such activities as a public health strategy to control disease. Our low rates of loss to follow-up reflect the very high degree of participant engagement with the COVIDENCE UK study. Our ability to identify episodes of milder disease affords potential insights into susceptibility factors as well as severity factors, and sets our study apart from long-established cohort studies in which assessment of risk factors may be temporally remote, and capture of outcomes is limited to events that are fatal or that precipitate hospitalisation. Our prospective design, coupled with censoring events occurring within 30 days of enrolment minimises the potential for reverse causation to explain associations observed.

Our study also has limitations. Use of test-confirmed COVID-19 as our primary outcome may have resulted in under-ascertainment of disease, particularly early in the pandemic (when testing capacity was particularly limited) and among people with less access to testing services; this might introduce collider bias.[13] We addressed this limitation by including a secondary outcome of symptom-defined probable COVID-19, which did not rely on access to testing. A second issue relates to under-representation of ethnic minorities, particularly people of Black, African and Caribbean ethnic origin; a lack of statistical power may explain why we did not confirm an increased risk of disease in these groups. Third, as with any observational study, we cannot exclude the possibility that the associations we report

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

may be explained by residual and/or unmeasured confounding. We have sought to minimise this by comprehensive capture of, and rigorous adjustment for, multiple potential confounders of relationships between putative risk factors and our primary outcome.

In conclusion, this population-based longitudinal study conducted in UK adults found that increased risk of developing COVID-19 associated independently with Asian/Asian British ethnicity, household overcrowding, visits to other households and other indoor public places, frontline occupation outside of health or social care, and increased BMI, after rigorous adjustment for multiple confounders. Atopic diseases, and especially atopic asthma, were associated with decreased risk. In contrast to studies investigating risk factors for severe disease, older age, male sex and other comorbidities were not associated with increased risk of developing COVID-19.

Acknowledgements

This study was supported by a grant from Barts Charity to ARM and CJG (ref MGU0466). The work was carried out with the support of BREATHE - The Health Data Research Hub for Respiratory Health [MC_PC_19004] in partnership with SAIL Databank. BREATHE is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK. MT is supported by a grant from the Rosetrees Trust and The Bloom Foundation (ref: M771). The views expressed are those of the authors and not necessarily those of Barts Charity, BREATHE or Health Data Research UK. We thank all the people who participated in the COVIDENCE UK study, and the following individuals and organisations who supported study recruitment: Asthma UK, the British Heart Foundation, the British Lung Foundation, the British Obesity Society, Cancer Research UK, Diabetes UK, Future Publishing Ltd, Kidney Care UK, Kidney Wales, Mumsnet, the National Kidney Federation, the National Rheumatoid Arthritis Society, the North West London Health Research Register (DISCOVER), Primary Immunodeficiency UK, the Race Equality Foundation, SWM Health Ltd, the Terence Higgins Trust and Vasculitis UK.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Author Contributions

ARM wrote the study protocol, with input from HH, MT, CR, GB and SOS. HH, MT, JS, CR, KY, MD, KT, SF, SI, AM, PP, DM, GD, RL, CJG, FK, AS, GB, SAS and ARM contributed to questionnaire development and design. HH co-ordinated and managed the study, with input from ARM, MT, JS and SOS. HH, JS, ARM, SOS, NSH and BA supported recruitment. HH, MT, MG, MD, KT, SSR, AAK, SER, PJL and DAJ contributed to data management and coding medication data. Statistical analyses were done by MT, with input from SOS, ARM, MG and HH. ARM wrote the first draft of the report. All authors revised it critically for important intellectual content, gave final approval of the version to be published, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work were appropriately investigated and resolved.

Competing Interests

All authors have completed the ICMJE uniform disclosure form. No author has had any financial relationship with any organisations that might have an interest in the submitted work in the previous three years. No author has had any other relationship, or undertaken any activity, that could appear to have influenced the submitted work.

Transparency Declaration

ARM is the manuscript's guarantor and he affirms that this is an honest, accurate, and transparent account of the study.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Table 1. Selected cohort characteristics (n=15,227)

Variable name	Categories	Number (%)
Age, years	16-29.99	576 (3.8)
	30-39.99	988 (6.5)
	40-49.99	1787 (11.7)
	50-59.99	3361 (22.1)
	60-69.99	5105 (33.5)
	≥70.00	3410 (22.4)
Sex	Female	10630 (69.8)
	Male	4597 (30.2)
Ethnicity	Asian/Asian British	281 (1.8)
	Black/African/Caribbean/Black British	96 (0.6)
	Mixed/Multiple/other ethnic groups	401 (2.6)
	White	14449 (94.9)
Country of residence	England	13463 (88.4)
	Northern Ireland	294 (1.9)
	Scotland	915 (6.0)
	Wales	555 (3.6)
Household income sufficient for basic needs	Yes	14209 (93.3)
	Mostly / Sometimes / No	1016 (6.7)
Housing	Owns own home	9326 (61.3)
	Mortgage	3743 (24.6)
	Privately renting	1069 (7.0)
	Renting from council	477 (3.1)
	Others	610 (4.0)
Number of people per bedroom	≤0.50	5836 (38.6)
	>0.50-0.99	4179 (27.6)
	1.00-1.99	4766 (31.5)
	≥2.00	333 (2.2)
Highest educational level attained	Primary/Secondary	1649 (10.8)
	Higher/Further (A levels)	2233 (14.7)
	College	6720 (44.2)
	Post-graduate	4610 (30.3)
Occupational status	Employed	5349 (35.1)
	Self-employed	1369 (9.0)
	Retired	6951 (45.6)
	Furloughed	365 (2.4)
	Unemployed	267 (1.8)
	Student	298 (2.0)
	Other	628 (4.1)
Frontline worker	No	12474 (82.0)
	Health or social care worker	1167 (7.7)
	Other frontline worker	1572 (10.3)
Body mass index, kg/m^2	<25	7431 (48.9)
	25-30	4829 (31.8)
	>30	2929 (19.3)
Self-reported general health	Excellent	3071 (20.2)
	Very good	6024 (39.6)
	Good	4048 (26.6)
	Fair	1633 (10.7)
	Poor	449 (2.9)
Tobacco smoking history	Never-smoker	8529 (56.0)
	Ex-smoker	5862 (38.5)
	Current smoker	836 (5.5)
Alcohol consumption in week prior to questionnaire completion, units	None	4271 (28.1)
	1-7	5406 (35.5)
	8-14	2961 (19.4)
	15-21	1417 (9.3)
	22-28	657 (4.3)
	>28	513 (3.4)

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Table 2. Socio-demographic, occupational and lifestyle factors and risk of test-confirmed COVID-19: crude and minimally-adjusted odds ratios

		Number (%) with incident test-confirmed COVID-19	Total number	Crude odds ratio (95% CI)	Minimally adjusted odd ratio (95% CI)[1]
Age, years	16-29.99	26 (4.5)	15227	1.00	1.00
	30-39.99	47 (4.8)		1.06 (0.65-1.73)	1.08 (0.65-1.78)
	40-49.99	94 (5.3)		1.17 (0.75-1.83)	1.24 (0.79-1.95)
	50-59.99	135 (4.0)		0.89 (0.58-1.36)	0.96 (0.62-1.49)
	60-69.99	97 (1.9)		0.41 (0.26-0.64)	0.49 (0.31-0.77)
	≥70.00	47 (1.4)		0.30 (0.18-0.48)	0.39 (0.24-0.64)
Sex	Female	329 (3.1)	15227	1.00	1.00
	Male	117 (2.5)		0.82 (0.66-1.01)	1.08 (0.86-1.34)
Ethnicity	Asian/Asian British	18 (6.4)	15227	2.39 (1.46-3.88)	1.99 (1.20-3.30)
	Black/African/Caribbean/Black British	6 (6.3)		2.32 (1.01-5.34)	1.98 (0.84-4.65)
	Mixed/Multiple/other ethnic groups	19 (4.7)		1.73 (1.08-2.78)	1.40 (0.86-2.28)
	White	403 (2.8)		1.00	1.00
Highest educational level attained	Primary/Secondary	51 (3.1)	15212	1.00	1.00
	Higher/further (A levels)	69 (3.1)		1.00 (0.69-1.44)	0.86 (0.59-1.25)
	College	197 (2.9)		0.95 (0.69-1.29)	0.75 (0.54-1.03)
	Post-graduate	129 (2.8)		0.90 (0.65-1.25)	0.62 (0.44-0.87)
Index of multiple deprivation (IMD) rank, quartiles	Q1 (most deprived)	131 (3.6)	15152	1.30 (1.01-1.69)	1.22 (0.93-1.60)
	Q2	108 (2.8)		1.02 (0.78-1.34)	0.97 (0.73-1.28)
	Q3	97 (2.5)		0.90 (0.68-1.19)	0.89 (0.67-1.19)
	Q4 (least deprived)	107 (2.8)		1.00	1.00
Household income sufficient for basic needs	Yes	416 (2.9)	15225	1.00	1.00
	Mostly / Sometimes / No	30 (3.0)		1.01 (0.69-1.47)	0.85 (0.58-1.25)
Claiming universal credit	No	425 (2.9)	15178	1.00	1.00
	Yes	19 (4.7)		1.65 (1.03-2.64)	1.25 (0.77-2.03)
Housing	Owns own home	189 (2.0)	15225	1.00	1.00
	Mortgage	173 (4.6)		2.34 (1.90-2.89)	1.34 (1.04-1.74)
	Privately Renting	40 (3.7)		1.88 (1.33-2.66)	1.17 (0.79-1.73)
	Renting from council	16 (3.4)		1.68 (1.00-2.82)	1.20 (0.70-2.07)
	Others	28 (4.6)		2.33 (1.55-3.49)	1.54 (0.95-2.49)
Frontline worker	No	276 (2.2)	15213	1.00	1.00
	Health or social care worker	90 (7.7)		3.69 (2.89-4.72)	1.68 (1.28-2.21)
	Other frontline worker	80 (5.1)		2.37 (1.84-3.06)	1.85 (1.42-2.42)
Shielding	No	423 (3.0)	15171	1.00	1.00
	Yes	23 (2.0)		0.64 (0.42-0.98)	0.63 (0.41-0.97)
Number of people per bedroom	≥0.50	98 (1.7)	15114	1.00	1.00
	>0.50-0.99	110 (2.6)		1.58 (1.20-2.08)	1.34 (1.01-1.78)
	1-1.99	213 (4.5)		2.74 (2.15-3.49)	1.91 (1.46-2.50)
	≥2.00	19 (5.7)		3.54 (2.14-5.87)	2.41 (1.40-4.13)
Any visits to/from other households in week prior to questionnaire completion	No	202 (2.7)	15217	1.00	1.00
	Yes	244 (3.2)		1.21 (1.00-1.46)	1.46 (1.19-1.80)
Pre-school children (0-4 years old) at home with participant	No	416 (2.8)	15191	1.00	1.00
	Yes	30 (5.9)		2.15 (1.47-3.15)	1.19 (0.77-1.84)
Schoolchildren (5-15 years old) at home with participant	No	351 (2.6)	15182	1.00	1.00
	Yes	95 (5.5)		2.18 (1.73-2.75)	1.42 (1.08-1.87)
Working-age adult (16-64 years old) at home with participant	No	117 (1.6)	15183	1.00	1.00
	Yes	329 (4.2)		2.74 (2.21-3.39)	1.86 (1.45-2.41)
Number of different generations in household	Living alone	46 (1.6)	15227	1.00	1.00
	Single generation	230 (2.8)		1.78 (1.29-2.45)	1.66 (1.19-2.30)
	Two generations	164 (4.1)		2.62 (1.88-3.64)	2.08 (1.48-2.94)
	Three generations	6 (5.0)		3.27 (1.37-7.83)	2.54 (1.04-6.23)
Cat at home	No	344 (2.9)	15215	1.00	1.00
	Yes	102 (3.1)		1.06 (0.85-1.33)	0.88 (0.70-1.12)
Dog at home	No	317 (2.7)	15215	1.00	1.00
	Yes	129 (3.6)		1.33 (1.08-1.64)	1.21 (0.97-1.50)
Number of public transport journeys in week prior to questionnaire completion	0	393 (2.9)	15175	1.00	1.00
	1-5	35 (3.0)		1.02 (0.72-1.45)	1.12 (0.78-1.61)
	≥6	17 (3.7)		1.29 (0.79-2.12)	1.37 (0.82-2.28)
Travel to work/study in week prior to questionnaire completion	No	37 (2.5)	14483	1.00	1.00
	Yes	318 (3.8)		1.54 (1.09-2.18)	1.71 (1.20-2.44)
	Not currently working or studying	83 (1.8)		0.72 (0.49-1.06)	1.24 (0.82-1.88)
Number of visits to shops and other indoor public places in week prior to	Q1	33 (1.5)	15207	1.00	1.00
	Q2	151 (3.0)		2.05 (1.40-2.99)	2.23 (1.51-3.29)

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

		Number (%) with incident test-confirmed COVID-19	Total number	Crude odds ratio (95% CI)	Minimally adjusted odd ratio (95% CI)[1]
questionnaire completion, quartiles	Q3	118 (3.3)		2.25 (1.52-3.32)	2.77 (1.85-4.14)
	Q4	144 (3.2)		2.16 (1.48-3.17)	3.38 (2.25-5.07)
Tobacco smoking status	Never-smoker	247 (2.9)	15227	1.00	1.00
	Ex-smoker	165 (2.8)		0.97 (0.80-1.19)	1.09 (0.89-1.35)
	Current smoker	34 (4.1)		1.42 (0.99-2.05)	1.24 (0.85-1.82)
Regular exposure to environmental tobacco smoke at home or in car	No	433 (2.9)	15226	1.00	1.00
	Yes	13 (3.9)		1.35 (0.77-2.37)	1.17 (0.66-2.09)
Vaping status	Never-vaper	410 (2.9)	15185	1.00	1.00
	Ex-vaper	16 (3.2)		1.12 (0.67-1.86)	0.89 (0.53-1.50)
	Current vaper	19 (4.0)		1.40 (0.88-2.24)	1.04 (0.64-1.69)
Alcohol consumption in week prior to questionnaire completion, units	None	119 (2.8)	15225	1.00	1.00
	1-7	162 (3.0)		1.08 (0.85-1.37)	1.08 (0.85-1.38)
	8-14	85 (2.9)		1.03 (0.78-1.37)	1.07 (0.80-1.43)
	15-21	51 (3.6)		1.30 (0.93-1.82)	1.40 (0.99-1.98)
	22-28	17 (2.6)		0.93 (0.55-1.55)	1.10 (0.65-1.87)
	>28	12 (2.3)		0.84 (0.46-1.52)	0.95 (0.52-1.76)
Hours of vigorous physical exercise in week prior to questionnaire completion[2]	0	171 (2.9)	15176	1.00	1.00
	1-3	163 (2.9)		0.98 (0.79-1.22)	0.95 (0.75-1.18)
	≥4	109 (2.9)		0.99 (0.78-1.27)	1.03 (0.80-1.32)
Hours of light physical exercise in week prior to questionnaire completion[3]	0-4	189 (3.6)	15192	1.00	1.00
	5-9	140 (2.8)		0.76 (0.61-0.95)	0.87 (0.69-1.10)
	≥10	117 (2.4)		0.64 (0.51-0.81)	0.89 (0.70-1.14)
Hours of lower-impact physical activity in week prior to questionnaire completion[4]	0	274 (3.1)	15176	1.00	1.00
	1	88 (3.1)		0.97 (0.76-1.24)	0.90 (0.70-1.16)
	≥2	83 (2.3)		0.73 (0.57-0.94)	0.76 (0.59-0.98)
Estimated average hours of sleep per night in month prior to questionnaire completion[5]	≤6	41 (3.1)	15223	1.22 (0.86-1.73)	1.15 (0.80-1.64)
	7	113 (3.1)		1.20 (0.94-1.53)	1.24 (0.97-1.59)
	8	159 (2.6)		1.00	1.00
	≥9	133 (3.3)		1.28 (1.02-1.62)	1.26 (0.99-1.60)

1, adjusted for age, sex, duration of participation and test frequency

2, defined as exercise of sufficient intensity to make the participant breathless or to raise their heart rate significantly, such as heavy physical work, strenuous gardening (e.g. vigorous digging, landscaping) swimming, jogging, aerobics, football, tennis, cycling, gym workout, reported to the nearest hour

3, defined as exercise that did not make the participant breathless, such as light gardening, walking, including walking for pleasure or exercise, walking to the shops, walking to work, reported to the nearest hour

4, defined as exercise to improve flexibility or core strength such as yoga, tai chi or pilates, reported to the nearest hour

5, reported to the nearest hour

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Table 3. Underlying conditions, medication use, vaccination status and risk of test-confirmed COVID-19: crude and minimally-adjusted odds ratios

		Number (%) with incident test-confirmed COVID-19	Total number	Crude odds ratio (95% CI)	Minimally adjusted odds ratio (95% CI)[1]
Underlying conditions					
Self-rated general health	Excellent	89 (2.9)	15225	1.00	1.00
	Very good	174 (2.9)		1.00 (0.77-1.29)	1.06 (0.81-1.38)
	Good	120 (3.0)		1.02 (0.77-1.35)	1.03 (0.77-1.37)
	Fair	52 (3.2)		1.10 (0.78-1.56)	0.99 (0.70-1.42)
	Poor	11 (2.4)		0.84 (0.45-1.59)	0.75 (0.40-1.43)
Self-rated anxiety or depression	No	316 (2.8)	15216	1.00	1.00
	Yes	130 (3.2)		1.13 (0.92-1.39)	0.91 (0.73-1.12)
Body mass index, kg/m^2	<25	177 (2.4)	15189	1.00	1.00
	25-30	168 (3.5)		1.48 (1.19-1.83)	1.58 (1.27-1.97)
	>30	100 (3.4)		1.45 (1.13-1.86)	1.37 (1.06-1.77)
Asthma	No	384 (3.0)	15227	1.00	1.00
	Yes	62 (2.5)		0.82 (0.62-1.07)	0.67 (0.51-0.88)
Atopic disease[2]	No	349 (3.1)	15227	1.00	1.00
	Yes	97 (2.5)		0.80 (0.64-1.01)	0.65 (0.51-0.82)
COPD	No	438 (2.9)	15227	1.00	1.00
	Yes	8 (2.5)		0.84 (0.41-1.70)	1.14 (0.55-2.34)
Diabetic status	No diabetes	415 (3.0)	15206	1.00	1.00
	Pre-diabetes	14 (3.1)		1.06 (0.62-1.82)	1.38 (0.79-2.41)
	Type 1 diabetes	3 (2.8)		0.96 (0.30-3.02)	0.81 (0.25-2.62)
	Type 2 diabetes	13 (2.1)		0.70 (0.40-1.22)	0.92 (0.52-1.62)
Heart disease[3]	No	432 (3.0)	15227	1.00	1.00
	Yes	14 (2.4)		0.80 (0.47-1.38)	1.28 (0.73-2.25)
Arterial disease[4]	No	431 (3.0)	15227	1.00	1.00
	Yes	15 (1.9)		0.63 (0.38-1.07)	0.91 (0.53-1.57)
Hypertension	No	378 (3.2)	15227	1.00	1.00
	Yes	68 (2.1)		0.64 (0.50-0.84)	0.88 (0.67-1.17)
Kidney disease	No	441 (3.0)	15227	1.00	1.00
	Yes	5 (1.6)		0.54 (0.22-1.31)	0.62 (0.25-1.52)
Major neurological conditions[5]	No	440 (3.0)	15227	1.00	1.00
	Yes	6 (1.5)		0.50 (0.22-1.12)	0.62 (0.27-1.41)
Cancer	Never	410 (3.0)	15227	1.00	1.00
	Past (cured or in remission)	33 (2.6)		0.87 (0.61-1.25)	1.12 (0.77-1.63)
	Present (active treatment in progress)	3 (2.2)		0.72 (0.23-2.27)	0.89 (0.28-2.87)
Immunodeficiency[6]	No	446 (2.9)	15140	1.00	1.00
	Yes	0 (0.0)		NA	NA
Medications					
Statins	No	398 (3.2)	15227	1.00	1.00
	Yes	48 (1.8)		0.57 (0.42-0.77)	0.93 (0.67-1.29)
ACE inhibitors	No	415 (3.0)	15227	1.00	1.00
	Yes	31 (2.1)		0.68 (0.47-0.99)	0.90 (0.62-1.32)
Proton pump inhibitors	No	395 (3.0)	15227	1.00	1.00
	Yes	51 (2.4)		0.80 (0.60-1.08)	0.88 (0.65-1.20)
Inhaled corticosteroids	No	425 (3.0)	15227	1.00	1.00
	Yes	21 (2.3)		0.75 (0.48-1.17)	0.64 (0.41-1.02)
Systemic Immunosuppressants	No	437 (3.0)	15227	1.00	1.00
	Yes	9 (1.3)		0.43 (0.22-0.84)	0.39 (0.20-0.77)
Selective serotonin re-uptake inhibitors (SSRIs)	No	401 (2.8)	15227	1.00	1.00
	Yes	45 (4.2)		1.50 (1.09-2.05)	1.19 (0.86-1.64)
Non-SSRI antidepressants	No	424 (2.9)	15227	1.00	1.00
	Yes	22 (3.4)		1.19 (0.77-1.84)	1.17 (0.75-1.83)
Angiotensin receptor blockers	No	427 (3.0)	15227	1.00	1.00
	Yes	19 (2.2)		0.74 (0.46-1.17)	1.01 (0.62-1.63)
Vitamin K antagonists	No	444 (2.9)	15227	1.00	1.00
	Yes	2 (1.8)		0.61 (0.15-2.46)	0.73 (0.18-3.06)
Beta-blockers	No	424 (3.0)	15227	1.00	1.00
	Yes	22 (2.0)		0.67 (0.43-1.03)	0.91 (0.59-1.43)
Thiazides	No	437 (3.0)	15227	1.00	1.00
	Yes	9 (1.8)		0.61 (0.31-1.19)	0.81 (0.41-1.60)
H2-receptor antagonists	No	442 (2.9)	15227	1.00	1.00
	Yes	4 (4.6)		1.60 (0.59-4.39)	1.36 (0.48-3.83)
Calcium channel blockers	No	417 (3.0)	15227	1.00	1.00
	Yes	29 (2.0)		0.64 (0.44-0.93)	0.91 (0.61-1.36)

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

		Number (%) with incident test-confirmed COVID-19	Total number	Crude odds ratio (95% CI)	Minimally adjusted odds ratio (95% CI)[1]
Inhaled bronchodilators[7]	No	14 (2.0)	15227	1.00	1.00
	Yes	411 (3.0)		0.84 (0.59-1.20)	0.54 (0.31-0.94)
Non-steroidal anti-inflammatory drugs	No	421 (3.0)	15227	1.00	1.00
	Yes	25 (2.0)		0.66 (0.44-0.99)	0.84 (0.55-1.28)
Sodium-glucose co-transporter-2 (SGLT2) inhibitors	No	444 (2.9)	15227	1.00	1.00
	Yes	2 (2.5)		0.86 (0.21-3.51)	0.95 (0.23-4.00)
Anti-platelet drugs	No	422 (3.0)	15227	1.00	1.00
	Yes	24 (2.5)		0.83 (0.55-1.26)	1.29 (0.83-1.99)
Sex hormone therapy	No	401 (2.8)	15227	1.00	1.00
	Yes	45 (4.0)		1.44 (1.05-1.97)	1.19 (0.85-1.65)
Paracetamol	No	430 (2.9)	15227	1.00	1.00
	Yes	16 (2.5)		0.84 (0.51-1.39)	0.89 (0.53-1.50)
Metformin	No	434 (2.9)	15227	1.00	1.00
	Yes	12 (2.8)		0.94 (0.52-1.67)	1.28 (0.71-2.31)
Bisphosphonates	No	441 (2.9)	15227	1.00	1.00
	Yes	5 (1.9)		0.65 (0.27-1.58)	0.76 (0.31-1.90)
Vaccinations					
BCG-vaccinated	No	40 (2.0)	15197	1.00	1.00
	Yes	366 (3.1)		1.55 (1.12-2.16)	1.36 (0.97-1.90)
	Unsure	38 (2.5)		1.23 (0.79-1.93)	1.26 (0.80-2.00)
MMR-vaccinated	No	131 (2.6)	11451	1.00	1.00
	Yes	226 (3.5)		1.38 (1.11-1.71)	1.04 (0.81-1.32)

BCG, *M. bovis* Bacille Calmette Guérin; MMR, measles mumps and rubella

1, adjusted for age, sex, duration of participation and test frequency

2, defined by atopic eczema/dermatitis and/or hayfever/allergic rhinitis

3, defined as coronary artery disease or heart failure

4, defined as ischaemic heart disease, peripheral vascular disease or cerebrovascular disease

5, defined as stroke, transient ischaemic attack, dementia, Parkinson's disease, multiple sclerosis or motor neurone disease

6, defined as HIV, primary immune deficiency or other immunodeficiency

7, defined as β-2-adrenoceptor agonists or anticholinergics

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Table 4. Diet, supplemental micronutrient intake and risk of test-confirmed COVID-19: crude and minimally-adjusted odds ratios

		Number (%) with incident test-confirmed COVID-19	Total number	Crude odds ratio (95% CI)	Minimally adjusted odds ratio (95% CI)[1]
Diet					
Dietary restrictions	None	419 (2.9)	15227	1.00	1.00
	Vegetarian[2]	22 (3.2)		1.11 (0.72-1.72)	0.89 (0.57-1.40)
	Vegan[3]	5 (2.2)		0.75 (0.31-1.83)	0.60 (0.24-1.49)
Number of portions of fruit, vegetables, and salad intake per day in week prior to questionnaire completion, quartiles	Q1	78 (3.5)	15188	1.00	1.00
	Q2	158 (3.1)		0.88 (0.67-1.17)	0.90 (0.68-1.20)
	Q3	87 (3.0)		0.85 (0.62-1.16)	0.89 (0.64-1.22)
	Q4	120 (2.4)		0.66 (0.49-0.88)	0.73 (0.54-0.99)
Number of portions of dairy products per day in week prior to questionnaire completion	0-1	114 (2.8)	15183	1.00	1.00
	2	142 (3.2)		1.15 (0.90-1.48)	1.18 (0.91-1.53)
	3-5	99 (2.7)		0.96 (0.73-1.27)	1.06 (0.80-1.41)
	≥6	91 (2.9)		1.03 (0.78-1.37)	1.24 (0.93-1.65)
Total number of portions of fish (oily or white) intake in the week prior to questionnaire completion	Q1	94 (3.8)	15213	1.00	1.00
	Q2	79 (3.2)		0.85 (0.63-1.15)	0.99 (0.73-1.36)
	Q3	111 (2.7)		0.71 (0.53-0.93)	0.88 (0.66-1.18)
	Q4	162 (2.6)		0.70 (0.54-0.90)	0.95 (0.72-1.25)
Portions of oily fish intake in the week prior to questionnaire completion	0	210 (3.7)	15220	1.00	1.00
	1	133 (2.7)		0.72 (0.58-0.90)	0.86 (0.69-1.08)
	2	71 (2.4)		0.64 (0.49-0.85)	0.80 (0.60-1.06)
	≥3	32 (2.1)		0.57 (0.39-0.83)	0.76 (0.52-1.13)
Portions of white fish intake in the week prior to questionnaire completion	0	131 (3.3)	15216	1.00	1.00
	1	170 (2.8)		0.83 (0.66-1.04)	0.99 (0.78-1.26)
	2	92 (2.6)		0.76 (0.58-1.00)	0.95 (0.72-1.26)
	≥3	53 (3.3)		0.98 (0.71-1.35)	1.16 (0.83-1.61)
Number of cups of non-alcoholic fluids per day in the week prior to questionnaire completion, quartiles	Q1	104 (3.0)	15177	1.00	1.00
	Q2	75 (2.9)		0.98 (0.73-1.32)	1.03 (0.76-1.41)
	Q3	116 (2.5)		0.84 (0.64-1.10)	0.81 (0.62-1.07)
	Q4	149 (3.4)		1.14 (0.88-1.47)	0.99 (0.76-1.29)
Micronutrient supplement use					
Multivitamin supplement	No	331 (2.7)	15227	1.00	1.00
	Yes	115 (3.6)		1.33 (1.07-1.65)	1.12 (0.89-1.40)
Vitamin A supplement	No	445 (2.9)	15227	1.00	1.00
	Yes	1 (1.1)		0.38 (0.05-2.73)	0.43 (0.06-3.15)
Vitamin C supplement	No	397 (2.9)	15227	1.00	1.00
	Yes	49 (3.1)		1.09 (0.80-1.47)	1.03 (0.76-1.41)
Vitamin D supplement	No	305 (3.2)	15227	1.00	1.00
	Yes	141 (2.5)		0.80 (0.65-0.98)	0.80 (0.65-0.99)
Zinc supplement	No	425 (2.9)	15227	1.00	1.00
	Yes	21 (2.8)		0.97 (0.62-1.51)	0.93 (0.58-1.46)
Iron supplement	No	423 (2.9)	15227	1.00	1.00
	Yes	23 (4.6)		1.63 (1.06-2.51)	1.27 (0.82-1.98)
Probiotic supplement	No	426 (3.0)	15227	1.00	1.00
	Yes	20 (2.2)		0.72 (0.46-1.13)	0.69 (0.44-1.10)
Fish oil, krill oil or other omega-3 supplements	No	400 (3.0)	15227	1.00	1.00
	Yes	46 (2.7)		0.91 (0.67-1.25)	0.98 (0.71-1.35)
Cod liver oil supplement	No	409 (2.9)	15227	1.00	1.00
	Yes	37 (3.0)		1.03 (0.73-1.45)	1.29 (0.91-1.84)
Garlic or garlic powder (allicin) supplement	No	439 (2.9)	15227	1.00	1.00
	Yes	7 (2.3)		0.77 (0.36-1.63)	0.93 (0.43-2.02)
Selenium supplement	No	445 (3.0)	15227	1.00	1.00
	Yes	1 (0.6)		0.21 (0.03-1.47)	0.20 (0.03-1.46)

1, adjusted for age, sex, duration of participation and test frequency

2, defined as excluding meat and fish, but not eggs or cow's milk, from the diet

3, defined as excluding meat, fish, eggs and cow's milk from the diet

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Table 5. Independent risk factors for test-confirmed COVID-19: final multivariable model (n=14,027)

Characteristic	Categories	Fully-adjusted odds ratio (95% CI)[1]	P for trend
Age, years	16-29.99	1.00	0.13
	30-39.99	1.21 (0.69-2.15)	
	40-49.99	1.24 (0.71-2.17)	
	50-59.99	1.17 (0.68-2.02)	
	60-69.99	0.86 (0.48-1.54)	
	≥70.00	0.79 (0.41-1.50)	
Sex	Female	1.00	--
	Male	0.93 (0.73-1.19)	
Ethnicity	Asian/Asian British	2.31 (1.35-3.95)	--
	Black/African/Caribbean/Black British	1.83 (0.70-4.77)	
	Mixed/Multiple/other ethnic groups	1.54 (0.93-2.55)	
	White	1.00	
Highest educational level attained	Primary/Secondary	1.00	0.09
	Higher/Further (A levels)	0.92 (0.62-1.37)	
	College	0.84 (0.59-1.18)	
	Post-graduate	0.75 (0.52-1.09)	
Index of multiple deprivation (IMD) rank, quartiles	Q1 (most deprived)	1.15 (0.86-1.52)	0.33
	Q2	0.90 (0.67-1.20)	
	Q3	0.88 (0.65-1.18)	
	Q4 (least deprived)	1.00	
Housing	Owns own home	1.00	--
	Mortgage	1.20 (0.92-1.58)	
	Privately renting	0.99 (0.65-1.51)	
	Renting from council	1.01 (0.56-1.82)	
	Others	1.38 (0.84-2.28)	
Number of people per bedroom[2]	≤0.50	1.00	<0.001
	>0.50-0.99	1.22 (0.91-1.64)	
	1.00-1.99	1.64 (1.23-2.19)	
	≥2.00	2.03 (1.15-3.58)	
Schoolchildren (aged 5-15 years) at home with participant	No	1.00	--
	Yes	1.16 (0.86-1.55)	
Dog at home	No	1.00	--
	Yes	1.19 (0.94-1.49)	
Shielding	No	1.00	--
	Yes	0.98 (0.59-1.61)	
Any visits to/from other households in week prior to questionnaire completion	No	1.00	--
	Yes	1.33 (1.07-1.64)	
Number of visits to shops and other indoor public places in week prior to questionnaire completion[3]	Q1	1.00	<0.001
	Q2	1.98 (1.30-3.03)	
	Q3	2.22 (1.43-3.46)	
	Q4	2.62 (1.67-4.10)	
Travel to work/study in week prior to questionnaire completion	No	1.00	--
	Yes	1.36 (0.93-1.99)	
	Not currently working or studying	1.32 (0.86-2.04)	
Frontline worker	No	1.00	--
	Health or social care worker	1.31 (0.97-1.76)	
	Other frontline worker	1.49 (1.12-1.98)	
Number of portions of fruit, vegetables, and salad intake per day in week prior to questionnaire completion, quartiles	Q1	1.00	0.29
	Q2	1.01 (0.75-1.36)	
	Q3	1.04 (0.74-1.46)	
	Q4	0.86 (0.63-1.18)	
Taking vitamin D supplement	No	1.00	
	Yes	0.93 (0.75-1.16)	
Hours of lower-impact physical activity in week prior to questionnaire completion[4]	0	1.00	0.21
	1	1.00 (0.77-1.30)	
	≥2	0.82 (0.62-1.08)	
Estimated average hours of sleep per night in month prior to questionnaire completion[5]	≤6	1.16 (0.80-1.68)	0.30[6]
	7	1.13 (0.87-1.47)	
	8	1.00	
	≥9	1.27 (0.99-1.64)	
Alcohol consumption in week prior to questionnaire completion, units	None	1.00	0.38
	1-7	1.08 (0.83-1.41)	
	8-14	1.08 (0.79-1.47)	
	15-21	1.42 (0.98-2.04)	
	22-28	1.22 (0.71-2.10)	

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

Characteristic	Categories	Fully-adjusted odds ratio (95% CI)[1]	P for trend
	>28	0.85 (0.44-1.64)	
Body mass index, kg/m^2	<25	1.00	0.006
	25-30	1.51 (1.20-1.90)	
	>30	1.38 (1.05-1.82)	
Asthma	No	1.00	--
	Yes	0.88 (0.61-1.28)	
Atopic disease[7]	No	1.00	--
	Yes	0.76 (0.59-0.98)	
Systemic immunosuppressants	No	1.00	--
	Yes	0.43 (0.19-0.95)	
Inhaled corticosteroids	No	1.00	--
	Yes	0.92 (0.50-1.68)	
Inhaled bronchodilators[8]	No	1.00	--
	Yes	1.01 (0.58-1.76)	
BCG-vaccinated	No	1.00	--
	Yes	1.44 (1.00-2.05)	
	Unsure	1.31 (0.81-2.11)	
Duration of participation, days		1.00 (1.00-1.00)	--
Swab test frequency		1.62 (1.52-1.73)	--

Abbreviation: BCG, *M. bovis* Bacillus Calmette Guérin

1, adjusted for age, sex, duration of participation, test frequency, ethnicity, highest educational level attained, IMD rank, hours of sleep per night, housing, number of people per bedroom, presence of schoolchildren at home, dog at home, shielding, visits to/from other households, visits to shops and other indoor places, travel to work or study, frontline worker status, fruit, vegetable and salad intake, supplemental vitamin D intake, low-impact physical activity, alcohol intake, body mass index, history of asthma, history of atopic disease, use of systemic immunosuppressants, use of inhaled corticosteroids, use of inhaled bronchodilators and BCG vaccination status.

2, adjusted odds ratio per additional 0.5 people per bedroom 1.26 (95% CI 1.11 to 1.43)

3, adjusted odds ratio per extra visit per week 1.05 (95% CI 1.01 to 1.09)

4, defined as exercise to improve flexibility or core strength such as yoga, tai chi or pilates, reported to the nearest hour

5, reported to nearest hour

6, only applies to sleeping categories below 8 hours

7, defined by atopic eczema/dermatitis and/or hayfever/allergic rhinitis

8, defined as β-2-adrenoceptor agonists or anticholinergics

medRxiv preprint doi: https://doi.org/10.1101/2021.03.27.21254452; this version posted March 30, 2021. 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-NC-ND 4.0 International license .

References

1. Aburto JM, Kashyap R, Schöley J, et al. Estimating the burden of COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: A population-level analysis. *MedRxiv* 2020.https://doi.org/10.1101/2020.07.16.20155077.
2. Miller IF, Becker AD, Grenfell BT, Metcalf CJE. Disease and healthcare burden of COVID-19 in the United States. *Nat Med* 2020; **26**(8): 1212-7
3. Pifarre IAH, Acosta E, Lopez-Casasnovas G, et al. Years of life lost to COVID-19 in 81 countries. *Scientific Reports* 2021; **11**(1): 3504
4. Clift AK, Coupland CAC, Keogh RH, et al. Living risk prediction algorithm (QCOVID) for risk of hospital admission and mortality from coronavirus 19 in adults: national derivation and validation cohort study. *BMJ* 2020; **371**: m3731
5. Williamson EJ, Walker AJ, Bhaskaran K, et al. Factors associated with COVID-19-related death using OpenSAFELY. *Nature* 2020; **584**(7821): 430-6
6. Venkatesan P. The changing demographics of COVID-19. *The Lancet Respiratory Medicine* 2020; **8**(12): e95
7. Ward H, Atchison C, Whitaker M, et al. Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults. *MedRxiv* 2020.https://doi.org/10.1101/2020.08.12.20173690.
8. Stringhini S, Wisniak A, Piumatti G, et al. Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland (SEROCoV-POP): a population-based study. *Lancet* 2020; **396**(10247): 313-9
9. Bhala N, Curry G, Martineau AR, Agyemang C, Bhopal R. Sharpening the global focus on ethnicity and race in the time of COVID-19. *Lancet* 2020; **395**(10238): 1673-6
10. Martineau AR, Forouhi NG. Vitamin D for COVID-19: a case to answer? *The Lancet Diabetes & Endocrinology* 2020; **8**(9): 735-6
11. Hamer M, Kivimaki M, Gale CR, Batty GD. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: A community-based cohort study of 387,109 adults in UK. *Brain, Behavior, and Immunity* 2020; **87**: 184-7
12. Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020; **26**(7): 1037-40
13. Griffith GJ, Morris TT, Tudball MJ, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature Communications* 2020; **11**(1): 5749
14. Cramér H. Mathematical Methods of Statistics. Princeton: Princeton University Press; 1946.
15. Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990; **1**(1): 43-6.
16. The OpenSAFELY Collaborative, Mathur R, Rentsch CT, et al. Ethnic differences in COVID-19 infection, hospitalisation, and mortality: an OpenSAFELY analysis of 17 million adults in England. *MedRxiv* 2020.https://doi.org/10.1101/2020.09.22.20198754.
17. Popkin BM, Du S, Green WD, et al. Individuals with obesity and COVID-19: A global perspective on the epidemiology and biological relationships. *Obes Rev* 2020; **21**(11): e13128
18. Ho FK, Celis-Morales CA, Gray SR, et al. Modifiable and non-modifiable risk factors for COVID-19, and comparison to risk factors for influenza and pneumonia: results from a UK Biobank prospective cohort study. 2020; **10:e040402**.10.1136/bmjopen-2020-040402.
19. Jackson DJ, Busse WW, Bacharier LB, et al. Association of respiratory allergy, asthma, and expression of the SARS-CoV-2 receptor ACE2. *J Allergy Clin Immunol* 2020; **146**(1): 203-6 e3