Original research

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

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

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/thoraxjnl-2021-217487).

For numbered affiliations see end of article.

Correspondence to
Professor Adrian R Martineau and Professor Adrian R Martineau, Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London E1 2AT, UK; a.martineau@qmul.ac.uk, a.martineau@qmul.ac.uk

HH, MT, SOS and ARM contributed equally.

Received 19 April 2021
Accepted 9 September 2021

To cite: Holt H, Talaei M, Greenig M, *et al*. *Thorax* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/thoraxjnl-2021-217487

ABSTRACT

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 1 May 2020 to 5 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 ORs (aORs) for associations between potential risk factors and odds of COVID-19.

Results We recorded 446 incident cases of COVID-19 in 15 227 participants (2.9%). Increased odds of developing COVID-19 were independently associated with Asian/Asian British versus white ethnicity (aOR 2.28, 95% CI 1.33 to 3.91), household overcrowding (aOR per additional 0.5 people/bedroom 1.26, 1.11 to 1.43), any versus no visits to/from other households in previous week (aOR 1.31, 1.06 to 1.62), number of visits to indoor public places (aOR per extra visit per week 1.05, 1.02 to 1.09), frontline occupation excluding health/social care versus no frontline occupation (aOR 1.49, 1.12 to 1.98) and raised body mass index (BMI) (aOR 1.50 (1.19 to 1.89) for BMI 25.0–30.0 kg/m^2 and 1.39 (1.06 to 1.84) for BMI >30.0 kg/m^2 versus BMI <25.0 kg/m^2). Atopic disease was independently associated with decreased odds (aOR 0.75, 0.59 to 0.97). No independent associations were seen for age, sex, other medical conditions, diet or micronutrient supplement use.

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

Trial registration number ClinicalTrials.gov Registry (NCT04330599).

Key messages

What is the key question?

► How do demographic, socioeconomic, lifestyle, dietary, pharmacological and comorbidity factors relate to the risk of developing COVID-19 in the general adult population of the UK?

What is the bottom line?

► After rigorous adjustment for factors influencing exposure to SARS-CoV-2, Asian/Asian British ethnicity and raised body mass index were associated with increased risk of developing COVID-19, while atopic disease was associated with decreased risk; no associations were seen for age, sex or other underlying medical conditions.

Why read on?

► This large, population-based prospective study shows that there is limited overlap between risk factors for developing COVID-19 versus those for intensive care unit admission and death as reported in hospitalised cohorts.

INTRODUCTION

COVID-19 has taken a heavy toll on the health of populations globally.[1–3] Risk factors for severe and fatal disease are well recognised, and include male sex, black or Asian ethnic origin, obesity, deprivation and a range of comorbidities including diabetes mellitus, cardiovascular disease, COPD and hypertension.[4 5] Characterisation of risks for milder disease has been relatively neglected, but is important, both from a public health perspective (since it drives transmission to individuals at risk of severe disease), and from a biological perspective (since understanding susceptibility factors can provide insights into pathogenesis).

There is growing evidence from population-based studies to suggest that at least some risk factors for

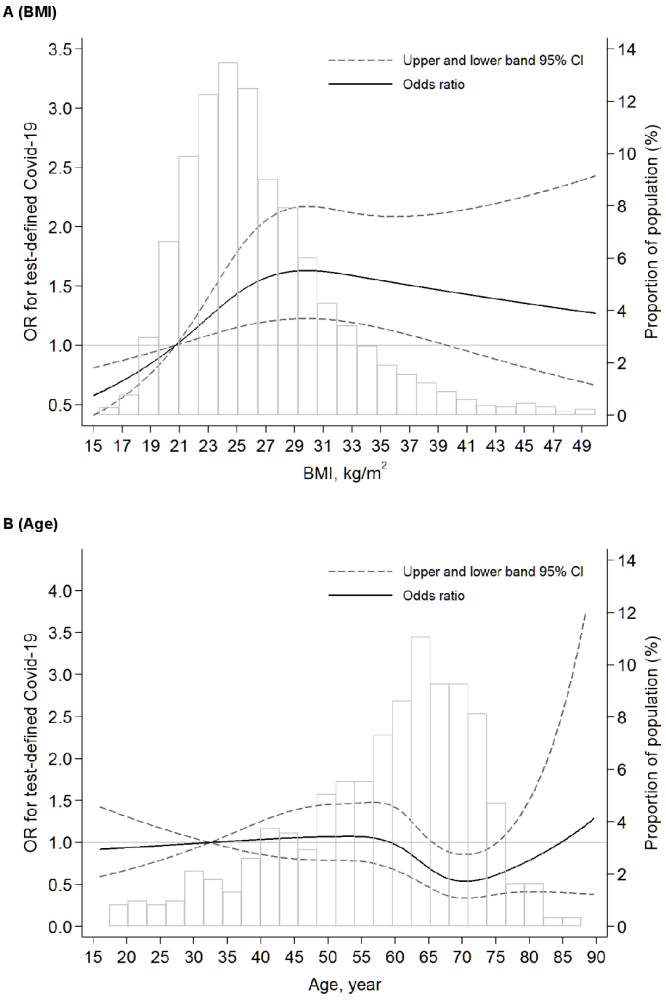

Figure 1 Dose–response relationship between body mass index (BMI) (A) and age (B) and risk of incident test-defined COVID-19 using restricted cubic spline analysis. The multivariable models were mutually adjusted for all factors presented in table 5.

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 USA and the UK have reported that risk of COVID-19 is higher in younger versus 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 versus 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 seropositivity 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 socioeconomic 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. Sufficient incident cases of test-confirmed COVID-19 have now accumulated to allow us to evaluate how a comprehensive panel of demographic, socioeconomic, lifestyle, dietary, pharmacological and comorbidity factors relate to the risk of developing COVID-19.

METHODS

Study design, setting and participants

COVIDENCE UK is a prospective longitudinal population-based observational study of coronavirus disease in the UK population (www.qmul.ac.uk/covidence). Inclusion criteria are age 16 years or more and residence in the UK at the point of enrolment; there are no exclusion criteria. Participants were invited via a national media campaign to complete an online baseline questionnaire to capture information on potential symptoms of COVID-19 experienced since 1 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 1 May 2020, and this paper reports findings of analysis of data collected up to 5 February 2021.

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 *et al*,[12] based on age, sex and self-reported loss of smell/taste, significant/severe persistent cough, severe fatigue and skipped meals (see online supplemental appendix for further details). This outcome was included in order to address potential underascertainment 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] Test-confirmed COVID-19 and symptom-defined probable COVID-19 were analysed as separate outcomes (ie, they were not combined). In order to minimise the potential for reverse causality to explain associations observed, outcomes occurring within 30 days of enrolment were excluded. At enrolment, participants were asked to provide details of potential symptoms of COVID-19 experienced since 1 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 online questionnaire capturing information about their sociodemographic characteristics, type of occupation, lifestyle, weight, height, long-standing medical conditions, medication use, vaccination status, diet and supplemental micronutrient intake (for baseline questionnaire, see online supplemental table 1). Monthly online follow-up questionnaires (online supplemental

Table 1 Selected cohort characteristics

	Total	Test-confirmed COVID-19	
		No	Yes
n (%)	15 227 (86.7)	14 781 (97.1)	446 (2.9)
Age, n (%)			
16–29.99 years	576 (3.8)	550 (3.7)	26 (5.8)
30–39.99 years	988 (6.5)	941 (6.4)	47 (10.5)
40–49.99 years	1787 (11.7)	1693 (11.5)	94 (21.1)
50–59.99 years	3361 (22.1)	3226 (21.8)	135 (30.3)
60–69.99 years	5105 (33.5)	5008 (33.9)	97 (21.7)
≥70.00 years	3410 (22.4)	3363 (22.8)	47 (10.5)
Age, years	59.4±13.4	59.5±13.4	53.3±13.2
Sex, n (%)			
Female	10 630 (69.8)	10 301 (69.7)	329 (73.8)
Male	4597 (30.2)	4480 (30.3)	117 (26.2)
Ethnicity, n (%)			
White	14 449 (94.9)	14 046 (95.0)	403 (90.4)
Mixed/multiple/other ethnic groups	401 (2.6)	382 (2.6)	19 (4.3)
South Asian	281 (1.8)	263 (1.8)	18 (4.0)
Black/African/ Caribbean/black British	96 (0.6)	90 (0.6)	6 (1.3)
Country of residence, n (%)			
England	13 463 (88.4)	13 052 (88.3)	411 (92.2)
Northern Ireland	294 (1.9)	281 (1.9)	13 (2.9)
Scotland	915 (6.0)	910 (6.2)	5 (1.1)
Wales	555 (3.6)	538 (3.6)	17 (3.8)
Household income sufficient for basic needs, n (%)			
Yes	14 209 (93.3)	13 793 (93.3)	416 (93.3)
Mostly/sometimes/no	1016 (6.7)	986 (6.7)	30 (6.7)
Housing, n (%)			
Owns own home	9326 (61.3)	9137 (61.8)	189 (42.4)
Mortgage	3743 (24.6)	3570 (24.2)	173 (38.8)
Privately renting	1069 (7.0)	1029 (7.0)	40 (9.0)
Renting from council	477 (3.1)	461 (3.1)	16 (3.6)
Others	610 (4.0)	582 (3.9)	28 (6.3)
Number of people per bedroom, n (%)			
≤0.50	5836 (38.6)	5738 (39.1)	98 (22.3)
>0.50–0.99	4179 (27.6)	4069 (27.7)	110 (25.0)
1.00–1.99	4766 (31.5)	4553 (31.0)	213 (48.4)
≥2.00	333 (2.2)	314 (2.1)	19 (4.3)
Highest educational level attained, n (%)			
Primary/secondary	1649 (10.8)	1598 (10.8)	51 (11.4)
Higher/further (A levels)	2233 (14.7)	2164 (14.7)	69 (15.5)
College	6720 (44.2)	6523 (44.2)	197 (44.2)
Postgraduate	4610 (30.3)	4481 (30.3)	129 (28.9)
Occupational status, n (%)			
Employed	5349 (35.1)	5091 (34.4)	258 (57.8)
Self-employed	1369 (9.0)	1330 (9.0)	39 (8.7)
Retired	6951 (45.6)	6856 (46.4)	95 (21.3)
Furloughed	365 (2.4)	348 (2.4)	17 (3.8)
Unemployed	267 (1.8)	254 (1.7)	13 (2.9)
Other	926 (6.1)	902 (6.1)	24 (5.4)
Frontline worker, n (%)			
No	12 474 (82.0)	12 198 (82.6)	276 (61.9)
Other frontline worker	1572 (10.3)	1492 (10.1)	80 (17.9)
Health or social care worker	1167 (7.7)	1077 (7.3)	90 (20.2)
Body mass index, n (%)			
<25 kg/m^2	7431 (48.9)	7254 (49.2)	177 (39.8)
25–30 kg/m^2	4829 (31.8)	4661 (31.6)	168 (37.8)
>30 kg/m^2	2929 (19.3)	2829 (19.2)	100 (22.5)
Self-reported general health, n (%)			
Excellent	3071 (20.2)	2982 (20.2)	89 (20.0)
Very good	6024 (39.6)	5850 (39.6)	174 (39.0)
Good	4048 (26.6)	3928 (26.6)	120 (26.9)
Fair	1633 (10.7)	1581 (10.7)	52 (11.7)
Poor	449 (2.9)	438 (3.0)	11 (2.5)
Tobacco smoking history, n (%)			
Never smoker	8529 (56.0)	8282 (56.0)	247 (55.4)
Ex-smoker	5862 (38.5)	5697 (38.5)	165 (37.0)
Current smoker	836 (5.5)	802 (5.4)	34 (7.6)
Alcohol consumption in week prior to questionnaire completion, n (%)			
None	4271 (28.1)	4152 (28.1)	119 (26.7)
1–7 units	5406 (35.5)	5244 (35.5)	162 (36.3)
8–14 units	2961 (19.4)	2876 (19.5)	85 (19.1)
15–21 units	1417 (9.3)	1366 (9.2)	51 (11.4)
22–28 units	657 (4.3)	640 (4.3)	17 (3.8)
>28 units	513 (3.4)	501 (3.4)	12 (2.7)
COVID-19 vaccine during follow-up, n (%)			
No	14 529 (85.7)	12 675 (85.8)	365 (81.8)
Yes	2416 (14.3)	2106 (14.2)	81 (18.2)

table 2) captured incident test-confirmed COVID-19 and potential symptoms of COVID-19.

Sample size and statistical methods

The sample size calculation and full details of statistical methods are presented in the online supplemental appendix. 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 ORs and 95% CIs for potential determinants of COVID-19 risk, first in a crude model, then in minimally adjusted and fully adjusted models. Correction for multiple comparisons was not applied, on the grounds that we were testing a priori hypotheses for all risk factors investigated.[14] We conducted sensitivity analysis for unmeasured

Table 2 Sociodemographic, occupational and lifestyle factors and risk of test-confirmed COVID-19: crude and minimally adjusted ORs

		Number (%) with incident test-confirmed COVID-19	Total number	Crude OR (95% CI)	Minimally adjusted OR (95% CI)*
Age, years	16–29.99	26 (4.5)	15 227	1.00	1.00
	30–39.99	47 (4.8)		1.06 (0.65 to 1.73)	1.08 (0.65 to 1.78)
	40–49.99	94 (5.3)		1.17 (0.75 to 1.83)	1.24 (0.79 to 1.95)
	50–59.99	135 (4.0)		0.89 (0.58 to 1.36)	0.96 (0.62 to 1.49)
	60–69.99	97 (1.9)		0.41 (0.26 to 0.64)	**0.49 (0.31 to 0.77)**
	≥70.00	47 (1.4)		0.30 (0.18 to 0.48)	**0.39 (0.24 to 0.64)**
Sex	Female	329 (3.1)	15 227	1.00	1.00
	Male	117 (2.5)		0.82 (0.66 to 1.01)	1.08 (0.86 to 1.34)
Ethnicity	Asian/Asian British	18 (6.4)	15 227	2.39 (1.46 to 3.88)	**1.99 (1.20 to 3.30)**
	Black/African/Caribbean/black British	6 (6.3)		2.32 (1.01 to 5.34)	1.98 (0.84 to 4.65)
	Mixed/multiple/other ethnic groups	19 (4.7)		1.73 (1.08 to 2.78)	1.40 (0.86 to 2.28)
	White	403 (2.8)		1.00	1.00
Highest educational level attained	Primary/secondary	51 (3.1)	15 212	1.00	1.00
	Higher/further (A levels)	69 (3.1)		1.00 (0.69 to 1.44)	0.86 (0.59 to 1.25)
	College	197 (2.9)		0.95 (0.69 to 1.29)	**0.75 (0.54 to 1.03)**
	Postgraduate	129 (2.8)		0.90 (0.65 to 1.25)	**0.62 (0.44 to 0.87)**
Index of Multiple Deprivation rank, quartiles	Q1 (most deprived)	131 (3.6)	15 152	1.30 (1.01 to 1.69)	1.22 (0.93 to 1.60)
	Q2	108 (2.8)		1.02 (0.78 to 1.34)	0.97 (0.73 to 1.28)
	Q3	97 (2.5)		0.90 (0.68 to 1.19)	0.89 (0.67 to 1.19)
	Q4 (least deprived)	107 (2.8)		1.00	1.00
Household income sufficient for basic needs	Yes	416 (2.9)	15 225	1.00	1.00
	Mostly/sometimes/no	30 (3.0)		1.01 (0.69 to 1.47)	0.85 (0.58 to 1.25)
Claiming universal credit	No	425 (2.9)	15 178	1.00	1.00
	Yes	19 (4.7)		1.65 (1.03 to 2.64)	1.25 (0.77 to 2.03)
Housing	Owns own home	189 (2.0)	15 225	1.00	1.00
	Mortgage	173 (4.6)		2.34 (1.90 to 2.89)	**1.34 (1.04 to 1.74)**
	Privately renting	40 (3.7)		1.88 (1.33 to 2.66)	1.17 (0.79 to 1.73)
	Renting from council	16 (3.4)		1.68 (1.00 to 2.82)	1.20 (0.70 to 2.07)
	Others	28 (4.6)		2.33 (1.55 to 3.49)	**1.54 (0.95 to 2.49)**
Frontline worker	No	276 (2.2)	15 213	1.00	1.00
	Health or social care worker	90 (7.7)		3.69 (2.89 to 4.72)	**1.68 (1.28 to 2.21)**
	Other frontline worker	80 (5.1)		2.37 (1.84 to 3.06)	**1.85 (1.42 to 2.42)**
Shielding	No	423 (3.0)	15 171	1.00	1.00
	Yes	23 (2.0)		0.64 (0.42 to 0.98)	**0.63 (0.41 to 0.97)**
Number of people per bedroom	≥0.50	98 (1.7)	15 114	1.00	1.00
	>0.50–0.99	110 (2.6)		1.58 (1.20 to 2.08)	**1.34 (1.01 to 1.78)**
	1–1.99	213 (4.5)		2.74 (2.15 to 3.49)	**1.91 (1.46 to 2.50)**
	≥2.00	19 (5.7)		3.54 (2.14 to 5.87)	**2.41 (1.40 to 4.13)**
Any visits to/from other households in week prior to questionnaire completion	No	202 (2.7)	15 217	1.00	1.00
	Yes	244 (3.2)		1.21 (1.00 to 1.46)	**1.46 (1.19 to 1.80)**
Preschool children (0–4 years old) at home with participant	No	416 (2.8)	15 191	1.00	1.00
	Yes	30 (5.9)		2.15 (1.47 to 3.15)	1.19 (0.77 to 1.84)
Schoolchildren (5–15 years old) at home with participant	No	351 (2.6)	15 182	1.00	1.00
	Yes	95 (5.5)		2.18 (1.73 to 2.75)	**1.42 (1.08 to 1.87)**
Working-age adult (16–64 years old) at home with participant	No	117 (1.6)	15 183	1.00	1.00
	Yes	329 (4.2)		2.74 (2.21 to 3.39)	**1.86 (1.45 to 2.41)**
Number of different generations in household	Living alone	46 (1.6)	15 227	1.00	1.00
	Single generation	230 (2.8)		1.78 (1.29 to 2.45)	**1.66 (1.19 to 2.30)**
	Two generations	164 (4.1)		2.62 (1.88 to 3.64)	**2.08 (1.48 to 2.94)**
	Three generations	6 (5.0)		3.27 (1.37 to 7.83)	**2.54 (1.04 to 6.23)**

Continued

Table 2 Continued

		Number (%) with incident test-confirmed COVID-19	Total number	Crude OR (95% CI)	Minimally adjusted OR (95% CI)*
Cat at home	No	344 (2.9)	15 215	1.00	1.00
	Yes	102 (3.1)		1.06 (0.85 to 1.33)	0.88 (0.70 to 1.12)
Dog at home	No	317 (2.7)	15 215	1.00	1.00
	Yes	129 (3.6)		1.33 (1.08 to 1.64)	**1.21 (0.97 to 1.50)**
Number of public transport journeys in week prior to questionnaire completion	0	393 (2.9)	15 175	1.00	1.00
	1–5	35 (3.0)		1.02 (0.72 to 1.45)	1.12 (0.78 to 1.61)
	≥6	17 (3.7)		1.29 (0.79 to 2.12)	1.37 (0.82 to 2.28)
Travel to work/study in week prior to questionnaire completion	No	37 (2.5)	15 042	1.00	1.00
	Yes	318 (3.8)		1.54 (1.09 to 2.18)	**1.71 (1.20 to 2.44)**
	Not currently working or studying	88 (1.7)		0.68 (0.46 to 1.00)	1.20 (0.79 to 1.81)
Number of visits to shops and other indoor public places in week prior to questionnaire completion, quartiles	Q1	33 (1.5)	15 207	1.00	1.00
	Q2	151 (3.0)		2.05 (1.40 to 2.99)	**2.23 (1.51 to 3.29)**
	Q3	118 (3.3)		2.25 (1.52 to 3.32)	**2.77 (1.85 to 4.14)**
	Q4	144 (3.2)		2.16 (1.48 to 3.17)	**3.38 (2.25 to 5.07)**
Tobacco smoking status	Never smoker	247 (2.9)	15 227	1.00	1.00
	Ex-smoker	165 (2.8)		0.97 (0.80 to 1.19)	1.09 (0.89 to 1.35)
	Current smoker	34 (4.1)		1.42 (0.99 to 2.05)	1.24 (0.85 to 1.82)
Regular exposure to environmental tobacco smoke at home or in car	No	433 (2.9)	15 226	1.00	1.00
	Yes	13 (3.9)		1.35 (0.77 to 2.37)	1.17 (0.66 to 2.09)
Vaping status	Never vaper	410 (2.9)	15 185	1.00	1.00
	Ex-vaper	16 (3.2)		1.12 (0.67 to 1.86)	0.89 (0.53 to 1.50)
	Current vaper	19 (4.0)		1.40 (0.88 to 2.24)	1.04 (0.64 to 1.69)
Alcohol consumption in week prior to questionnaire completion, units	None	119 (2.8)	15 225	1.00	1.00
	1–7	162 (3.0)		1.08 (0.85 to 1.37)	1.08 (0.85 to 1.38)
	8–14	85 (2.9)		1.03 (0.78 to 1.37)	1.07 (0.80 to 1.43)
	15–21	51 (3.6)		1.30 (0.93 to 1.82)	**1.40 (0.99 to 1.98)**
	22–28	17 (2.6)		0.93 (0.55 to 1.55)	1.10 (0.65 to 1.87)
	>28	12 (2.3)		0.84 (0.46 to 1.52)	0.95 (0.52 to 1.76)
Hours of vigorous physical exercise in week prior to questionnaire completion†	0	171 (2.9)	15 176	1.00	1.00
	1–3	163 (2.9)		0.98 (0.79 to 1.22)	0.95 (0.75 to 1.18)
	≥4	109 (2.9)		0.99 (0.78 to 1.27)	1.03 (0.80 to 1.32)
Hours of light physical exercise in week prior to questionnaire completion‡	0–4	189 (3.6)	15 192	1.00	1.00
	5–9	140 (2.8)		0.76 (0.61 to 0.95)	0.87 (0.69 to 1.10)
	≥10	117 (2.4)		0.64 (0.51 to 0.81)	0.89 (0.70 to 1.14)
Hours of lower impact physical activity in week prior to questionnaire completion§	0	274 (3.1)	15 176	1.00	1.00
	1	88 (3.1)		0.97 (0.76 to 1.24)	0.90 (0.70 to 1.16)
	≥2	83 (2.3)		0.73 (0.57 to 0.94)	**0.76 (0.59 to 0.98)**
Estimated average hours of sleep per night in month prior to questionnaire completion¶	≤6	41 (3.1)	15 223	1.22 (0.86 to 1.73)	1.15 (0.80 to 1.64)
	7	113 (3.1)		1.20 (0.94 to 1.53)	**1.24 (0.97 to 1.59)**
	8	159 (2.6)		1.00	1.00
	≥9	133 (3.3)		1.28 (1.02 to 1.62)	**1.26 (0.99 to 1.60)**

Minimally adjusted ORs (95% CI) with a p value of <0.10 are emboldened.
*Adjusted for age, sex, duration of participation and test frequency.
†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 (eg, vigorous digging, landscaping), swimming, jogging, aerobics, football, tennis, cycling and gym workout, reported to the nearest hour.
‡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 and walking to work, reported to the nearest hour.
§Defined as exercise to improve flexibility or core strength such as yoga, tai chi or Pilates, reported to the nearest hour.
¶Reported to the nearest hour.

confounding by estimating E-values[15] using the 'evalue' package in Stata.[16] Two other sensitivity analyses were also 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. We also performed an exploratory analysis to determine whether COVID-19 risk differed for participants with atopic versus non-atopic asthma endotypes; this was conducted on the basis of evidence suggesting that decreased expression of *ACE2*, the gene encoding the SARS-CoV-2 receptor, has been reported in people with asthma who have high levels of allergic sensitisation.[17]

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

		Number (%) with incident test-confirmed COVID-19	Total number	Crude OR (95% CI)	Minimally adjusted OR (95% CI)*
Underlying conditions					
Self-rated general health	Excellent	89 (2.9)	15225	1.00	1.00
	Very good	174 (2.9)		1.00 (0.77 to 1.29)	1.06 (0.81 to 1.38)
	Good	120 (3.0)		1.02 (0.77 to 1.35)	1.03 (0.77 to 1.37)
	Fair	52 (3.2)		1.10 (0.78 to 1.56)	0.99 (0.70 to 1.42)
	Poor	11 (2.4)		0.84 (0.45 to 1.59)	0.75 (0.40 to 1.43)
Self-rated anxiety or depression	No	316 (2.8)	15216	1.00	1.00
	Yes	130 (3.2)		1.13 (0.92 to 1.39)	0.91 (0.73 to 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 to 1.83)	**1.58 (1.27 to 1.97)**
	>30	100 (3.4)		1.45 (1.13 to 1.86)	**1.37 (1.06 to 1.77)**
Asthma	No	384 (3.0)	15227	1.00	1.00
	Yes	62 (2.5)		0.82 (0.62 to 1.07)	**0.67 (0.51 to 0.88)**
Atopic disease†	No	349 (3.1)	15227	1.00	1.00
	Yes	97 (2.5)		0.80 (0.64 to 1.01)	**0.65 (0.51 to 0.82)**
COPD	No	438 (2.9)	15227	1.00	1.00
	Yes	8 (2.5)		0.84 (0.41 to 1.70)	1.14 (0.55 to 2.34)
Diabetic status	No diabetes	415 (3.0)	15206	1.00	1.00
	Pre-diabetes	14 (3.1)		1.06 (0.62 to 1.82)	1.38 (0.79 to 2.41)
	Type 1 diabetes	3 (2.8)		0.96 (0.30 to 3.02)	0.81 (0.25 to 2.62)
	Type 2 diabetes	13 (2.1)		0.70 (0.40 to 1.22)	0.92 (0.52 to 1.62)
Heart disease‡	No	432 (3.0)	15227	1.00	1.00
	Yes	14 (2.4)		0.80 (0.47 to 1.38)	1.28 (0.73 to 2.25)
Arterial disease§	No	431 (3.0)	15227	1.00	1.00
	Yes	15 (1.9)		0.63 (0.38 to 1.07)	0.91 (0.53 to 1.57)
Hypertension	No	378 (3.2)	15227	1.00	1.00
	Yes	68 (2.1)		0.64 (0.50 to 0.84)	0.88 (0.67 to 1.17)
Kidney disease	No	441 (3.0)	15227	1.00	1.00
	Yes	5 (1.6)		0.54 (0.22 to 1.31)	0.62 (0.25 to 1.52)
Major neurological conditions¶	No	440 (3.0)	15227	1.00	1.00
	Yes	6 (1.5)		0.50 (0.22 to 1.12)	0.62 (0.27 to 1.41)
Cancer	Never	410 (3.0)	15227	1.00	1.00
	Past (cured or in remission)	33 (2.6)		0.87 (0.61 to 1.25)	1.12 (0.77 to 1.63)
	Present (active treatment in progress)	3 (2.2)		0.72 (0.23 to 2.27)	0.89 (0.28 to 2.87)
Immunodeficiency**	No	446 (2.9)	15140	1.00	1.00
	Yes	0 (0.0)		Not applicable	Not applicable
Periodontitis††	No	240 (2.7)	15227	1.00	1.00
	Yes	129 (3.3)		1.20 (0.96 to 1.49)	1.30 (1.04 to 1.62)
	Missing	77 (3.05)		1.12 (0.86 to 1.45)	1.13 (0.86 to 1.49)
Medications					
Statins	No	398 (3.2)	15227	1.00	1.00
	Yes	48 (1.8)		0.57 (0.42 to 0.77)	0.93 (0.67 to 1.29)
ACE inhibitors	No	415 (3.0)	15227	1.00	1.00
	Yes	31 (2.1)		0.68 (0.47 to 0.99)	0.90 (0.62 to 1.32)
Proton pump inhibitors	No	395 (3.0)	15227	1.00	1.00
	Yes	51 (2.4)		0.80 (0.60 to 1.08)	0.88 (0.65 to 1.20)
Inhaled corticosteroids	No	410 (3.0)	15227	1.00	1.00
	Yes	36 (2.5)		0.83 (0.59 to 1.18)	**0.68 (0.48 to 0.97)**
Systemic immunosuppressants	No	437 (3.0)	15227	1.00	1.00
	Yes	9 (1.3)		0.43 (0.22 to 0.84)	**0.39 (0.20 to 0.77)**
Selective serotonin reuptake inhibitors (SSRIs)	No	401 (2.8)	15227	1.00	1.00
	Yes	45 (4.2)		1.50 (1.09 to 2.05)	1.19 (0.86 to 1.64)

Continued

Table 3 Continued

		Number (%) with incident test-confirmed COVID-19	Total number	Crude OR (95% CI)	Minimally adjusted OR (95% CI)*
Non-SSRI antidepressants	No	424 (2.9)	15 227	1.00	1.00
	Yes	22 (3.4)		1.19 (0.77 to 1.84)	1.17 (0.75 to 1.83)
Angiotensin receptor blockers	No	427 (3.0)	15 227	1.00	1.00
	Yes	19 (2.2)		0.74 (0.46 to 1.17)	1.01 (0.62 to 1.63)
Vitamin K antagonists	No	444 (2.9)	15 227	1.00	1.00
	Yes	2 (1.8)		0.61 (0.15 to 2.46)	0.73 (0.18 to 3.06)
Beta-blockers	No	424 (3.0)	15 227	1.00	1.00
	Yes	22 (2.0)		0.67 (0.43 to 1.03)	0.91 (0.59 to 1.43)
Thiazides	No	437 (3.0)	15 227	1.00	1.00
	Yes	9 (1.8)		0.61 (0.31 to 1.19)	0.81 (0.41 to 1.60)
H2-receptor antagonists	No	442 (2.9)	15 227	1.00	1.00
	Yes	4 (4.6)		1.60 (0.59 to 4.39)	1.36 (0.48 to 3.83)
Calcium channel blockers	No	417 (3.0)	15 227	1.00	1.00
	Yes	29 (2.0)		0.64 (0.44 to 0.93)	0.91 (0.61 to 1.36)
Inhaled bronchodilators‡‡	No	14 (2.0)	15 227	1.00	1.00
	Yes	411 (3.0)		0.84 (0.59 to 1.20)	**0.54 (0.31 to 0.94)**
Non-steroidal anti-inflammatory drugs	No	421 (3.0)	15 227	1.00	1.00
	Yes	25 (2.0)		0.66 (0.44 to 0.99)	0.84 (0.55 to 1.28)
Sodium-glucose co-transporter-2 inhibitors	No	444 (2.9)	15 227	1.00	1.00
	Yes	2 (2.5)		0.86 (0.21 to 3.51)	0.95 (0.23 to 4.00)
Anti-platelet drugs	No	422 (3.0)	15 227	1.00	1.00
	Yes	24 (2.5)		0.83 (0.55 to 1.26)	1.29 (0.83 to 1.99)
Sex hormone therapy	No	401 (2.8)	15 227	1.00	1.00
	Yes	45 (4.0)		1.44 (1.05 to 1.97)	1.19 (0.85 to 1.65)
Paracetamol	No	430 (2.9)	15 227	1.00	1.00
	Yes	16 (2.5)		0.84 (0.51 to 1.39)	0.89 (0.53 to 1.50)
Metformin	No	434 (2.9)	15 227	1.00	1.00
	Yes	12 (2.8)		0.94 (0.52 to 1.67)	1.28 (0.71 to 2.31)
Bisphosphonates	No	441 (2.9)	15 227	1.00	1.00
	Yes	5 (1.9)		0.65 (0.27 to 1.58)	0.76 (0.31 to 1.90)
Vaccinations					
BCG-vaccinated	No	40 (2.0)	15 197	1.00	1.00
	Yes	366 (3.1)		1.55 (1.12 to 2.16)	1.36 (0.97 to 1.90)
	Unsure	38 (2.5)		1.23 (0.79 to 1.93)	1.26 (0.80 to 2.00)
MMR-vaccinated	No	131 (2.6)	11 451	1.00	1.00
	Yes	226 (3.5)		1.38 (1.11 to 1.71)	1.04 (0.81 to 1.32)

Minimally adjusted ORs (95% CI) with a p value of <0.10 are bold.
*Adjusted for age, sex, duration of participation and test frequency.
†Defined by atopic eczema/dermatitis and/or hayfever/allergic rhinitis.
‡Defined as coronary artery disease or heart failure.
§Defined as ischaemic heart disease, peripheral vascular disease or cerebrovascular disease.
¶Defined as stroke, transient ischaemic attack, dementia, Parkinson's disease, multiple sclerosis or motor neuron disease.
**Defined as HIV, primary immune deficiency or other immunodeficiency.
††Defined as being present if participant answered 'yes' to any of the three questions relating to gum disease/dental health listed in online supplemental table 2.
‡‡Defined as β-2 adrenoceptor agonists or anticholinergics.
MMR, measles, mumps and rubella.

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 2 November 2020, we excluded those who were identified as already having had test-confirmed and/or symptom-defined probable COVID-19 (n=1477). Of the remaining 16 081 participants, 15 227 completed at least one subsequent monthly follow-up questionnaire at least 30 days after enrolment and were included in this prospective analysis; 14 348 completed the final follow-up questionnaire on or before 5 February 2021, giving a retention rate of 89.2% (online supplemental figure 1). Selected baseline characteristics of participants included in the prospective analysis are presented in table 1; their characteristics are compared with those who were excluded from this analysis in online supplemental table 3. Mean age of those contributing data to prospective analyses was 59.4 years (range 16.0–94.4 years), 69.8% were female and 94.9% identified their ethnic origin as white.

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

		Number (%) with incident test-confirmed COVID-19	Total number	Crude OR (95% CI)	Minimally adjusted OR (95% CI)*
Diet					
Dietary restrictions	None	419 (2.9)	15227	1.00	1.00
	Vegetarian†	22 (3.2)		1.11 (0.72 to 1.72)	0.89 (0.57 to 1.40)
	Vegan‡	5 (2.2)		0.75 (0.31 to 1.83)	0.60 (0.24 to 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 to 1.17)	0.90 (0.68 to 1.20)
	Q3	87 (3.0)		0.85 (0.62 to 1.16)	0.89 (0.64 to 1.22)
	Q4	120 (2.4)		0.66 (0.49 to 0.88)	**0.73 (0.54 to 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 to 1.48)	1.18 (0.91 to 1.53)
	3–5	99 (2.7)		0.96 (0.73 to 1.27)	1.06 (0.80 to 1.41)
	≥6	91 (2.9)		1.03 (0.78 to 1.37)	1.24 (0.93 to 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 to 1.15)	0.99 (0.73 to 1.36)
	Q3	111 (2.7)		0.71 (0.53 to 0.93)	0.88 (0.66 to 1.18)
	Q4	162 (2.6)		0.70 (0.54 to 0.90)	0.95 (0.72 to 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 to 0.90)	0.86 (0.69 to 1.08)
	2	71 (2.4)		0.64 (0.49 to 0.85)	0.80 (0.60 to 1.06)
	≥3	32 (2.1)		0.57 (0.39 to 0.83)	0.76 (0.52 to 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 to 1.04)	0.99 (0.78 to 1.26)
	2	92 (2.6)		0.76 (0.58 to 1.00)	0.95 (0.72 to 1.26)
	≥3	53 (3.3)		0.98 (0.71 to 1.35)	1.16 (0.83 to 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 to 1.32)	1.03 (0.76 to 1.41)
	Q3	116 (2.5)		0.84 (0.64 to 1.10)	0.81 (0.62 to 1.07)
	Q4	149 (3.4)		1.14 (0.88 to 1.47)	0.99 (0.76 to 1.29)
Micronutrient supplement use					
Multivitamin supplement	No	331 (2.7)	15227	1.00	1.00
	Yes	115 (3.6)		1.33 (1.07 to 1.65)	1.12 (0.89 to 1.40)
Vitamin A supplement	No	445 (2.9)	15227	1.00	1.00
	Yes	1 (1.1)		0.38 (0.05 to 2.73)	0.43 (0.06 to 3.15)
Vitamin C supplement	No	397 (2.9)	15227	1.00	1.00
	Yes	49 (3.1)		1.09 (0.80 to 1.47)	1.03 (0.76 to 1.41)
Vitamin D supplement	No	305 (3.2)	15227	1.00	1.00
	Yes	141 (2.5)		0.80 (0.65 to 0.98)	**0.80 (0.65 to 0.99)**
Zinc supplement	No	425 (2.9)	15227	1.00	1.00
	Yes	21 (2.8)		0.97 (0.62 to 1.51)	0.93 (0.58 to 1.46)
Iron supplement	No	423 (2.9)	15227	1.00	1.00
	Yes	23 (4.6)		1.63 (1.06 to 2.51)	1.27 (0.82 to 1.98)
Probiotic supplement	No	426 (3.0)	15227	1.00	1.00
	Yes	20 (2.2)		0.72 (0.46 to 1.13)	0.69 (0.44 to 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 to 1.25)	0.98 (0.71 to 1.35)
Cod liver oil supplement	No	409 (2.9)	15227	1.00	1.00
	Yes	37 (3.0)		1.03 (0.73 to 1.45)	1.29 (0.91 to 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 to 1.63)	0.93 (0.43 to 2.02)
Selenium supplement	No	445 (3.0)	15227	1.00	1.00
	Yes	1 (0.6)		0.21 (0.03 to 1.47)	0.20 (0.03 to 1.46)

Minimally adjusted ORs (95% CI) with a p value of <0.10 are emboldened.
*Adjusted for age, sex, duration of participation and test frequency.
†Defined as excluding meat and fish, but not eggs or cow's milk, from the diet.
‡Defined as excluding meat, fish, eggs and cow's milk from the diet.

Table 5 Independent risk factors for test-confirmed COVID-19: final multivariable model (n=14 556)

Characteristic	Categories	Fully adjusted OR (95% CI)*	P for trend	E_{est} (E_{CI})
Age, years	16–29.99	1.00	0.08	
	30–39.99	1.20 (0.68 to 2.12)		
	40–49.99	1.20 (0.69 to 2.10)		
	50–59.99	1.12 (0.65 to 1.93)		
	60–69.99	0.80 (0.44 to 1.44)		
	≥70.00	0.74 (0.39 to 1.41)		
Sex	Female	1.00	—	
	Male	0.97 (0.76 to 1.23)		
Ethnicity	Asian/Asian British	2.28 (1.33 to 3.91)	—	3.99 (1.99)
	Black/African/Caribbean/black British	1.84 (0.70 to 4.82)		
	Mixed/multiple/other ethnic groups	1.53 (0.92 to 2.53)		
	White	1.00		
Highest educational level attained	Primary/secondary	1.00	0.08	
	Higher/further (A levels)	0.90 (0.61 to 1.33)		
	College	0.82 (0.58 to 1.15)		
	Postgraduate	0.74 (0.51 to 1.07)		
Index of Multiple Deprivation (IMD) rank, quartiles	Q1 (most deprived)	1.11 (0.84 to 1.48)	0.44	
	Q2	0.90 (0.67 to 1.20)		
	Q3	0.87 (0.65 to 1.16)		
	Q4 (least deprived)	1.00		
Housing	Owns own home	1.00	—	
	Mortgage	1.19 (0.91 to 1.56)		
	Privately renting	0.96 (0.63 to 1.46)		
	Renting from council	0.96 (0.53 to 1.74)		
	Others	1.34 (0.81 to 2.21)		
Number of people per bedroom†	≤0.50	1.00	<0.001	
	>0.50–0.99	1.24 (0.93 to 1.66)		
	1.00–1.99	1.67 (1.25 to 2.23)		2.73 (1.81)
	≥2.00	2.04 (1.15 to 3.60)		3.52 (1.43)
Schoolchildren (aged 5–15 years) at home with participant	No	1.00	—	
	Yes	1.16 (0.86 to 1.56)		
Dog at home	No	1.00	—	
	Yes	1.17 (0.93 to 1.47)		
Shielding	No	1.00	—	
	Yes	1.04 (0.64 to 1.68)		
Any visits to/from other households in week prior to questionnaire completion	No	1.00	—	
	Yes	1.31 (1.06 to 1.62)		1.95 (1.31)
Number of visits to shops and other indoor public places in week prior to questionnaire completion‡	Q1	1.00	<0.001	
	Q2	1.96 (1.29 to 2.98)		3.33 (1.90)
	Q3	2.21 (1.43 to 3.43)		3.84 (2.21)
	Q4	2.63 (1.69 to 4.10)		4.70 (2.77)
Travel to work/study in week prior to questionnaire completion	No	1.00	—	
	Yes	1.36 (0.93 to 1.99)		
	Not currently working or studying	1.27 (0.82 to 1.95)		
Frontline worker	No	1.00	—	
	Health or social care worker	1.35 (1.00 to 1.82)		2.04 (1.02)
	Other frontline worker	1.49 (1.12 to 1.98)		2.34 (1.49)
Number of portions of fruit, vegetables and salad intake per day in week prior to questionnaire completion, quartiles	Q1	1.00	0.31	
	Q2	1.00 (0.75 to 1.35)		
	Q3	1.04 (0.74 to 1.45)		
	Q4	0.86 (0.63 to 1.19)		

Continued

Table 5 Continued

Characteristic	Categories	Fully adjusted OR (95% CI)*	P for trend	E_{est} (E_{CI})
Taking vitamin D supplement	No	1.00		
	Yes	0.93 (0.75 to 1.16)		
Hours of lower impact physical activity in week prior to questionnaire completion§	0	1.00	0.22	
	1	1.01 (0.78 to 1.31)		
	≥2	0.83 (0.63 to 1.09)		
Estimated average hours of sleep per night in month prior to questionnaire completion¶	≤6	1.14 (0.79 to 1.65)	0.31**	
	7	1.14 (0.88 to 1.48)		
	8	1.00		
	≥9	1.29 (1.01 to 1.66)		1.90 (1.11)
Alcohol consumption in week prior to questionnaire completion, units	None	1.00	0.39	
	1–7	1.11 (0.85 to 1.44)		
	8–14	1.08 (0.79 to 1.47)		
	15–21	1.45 (1.01 to 2.08)		
	22–28	1.23 (0.72 to 2.11)		
	>28	0.83 (0.43 to 1.60)		
Body mass index, kg/m^2	<25	1.00	0.004	
	25–30	1.50 (1.19 to 1.89)		2.37 (1.67)
	>30	1.39 (1.06 to 1.84)		2.13 (1.31)
Asthma	No	1.00	—	
	Yes	0.82 (0.55 to 1.24)		
Atopic disease††	No	1.00	—	
	Yes	0.75 (0.59 to 0.97)		2.00 (1.21)
Periodontitis‡‡	No	1.00	—	
	Yes	1.20 (0.95 to 1.52)		
	Missing	1.03 (0.77 to 1.37)		
Systemic immunosuppressants	No	1.00	—	
	Yes	0.47 (0.22 to 0.99)		3.68 (1.11)
Inhaled corticosteroids	No	1.00	—	
	Yes	1.01 (0.55 to 1.84)		
Inhaled bronchodilators§§	No	1.00	—	
	Yes	1.07 (0.61 to 1.88)		
BCG-vaccinated	No	1.00	—	
	Yes	1.38 (0.97 to 1.96)		
	Unsure	1.24 (0.78 to 1.99)		
Duration of participation, days		1.00 (1.00 to 1.00)	—	
Swab test frequency		1.62 (1.52 to 1.72)	—	

*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.
†Adjusted OR per additional 0.5 people per bedroom 1.26 (95% CI 1.11 to 1.43).
‡Adjusted OR per extra visit per week 1.05 (95% CI 1.02 to 1.09).
§Defined as exercise to improve flexibility or core strength such as yoga, tai chi or Pilates, reported to the nearest hour.
¶Reported to the nearest hour.
**Only applies to sleeping categories below 8 hours.
††Defined by atopic eczema/dermatitis and/or hayfever/allergic rhinitis.
‡‡Defined as being present if participant answered 'yes' to any of the three questions relating to gum disease/dental health listed in online supplemental table 2.
§§Defined as β-2 adrenoceptor agonists or anticholinergics.
E_{CI}, E-value for 95% CI; E_{est}, E-value for point estimate.

The geographical distribution of COVIDENCE UK participants aligned closely with that of incident COVID-19 in the UK (online supplemental figure 2).

A total of 446 participants experienced at least one episode of PCR-confirmed 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 (aORs) for associations between risk of test-confirmed COVID-19 and sociodemographic, occupational and lifestyle factors (table 2); long-standing 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 odds of COVID-19 with $p<0.10$: Asian/Asian British versus white ethnic origin, housing type (paying mortgage and 'other' vs owning own home), frontline versus

non-frontline worker status, household overcrowding (>0.5 vs ≤0.5 people per bedroom), any visit versus no visits to/from other households in the previous week, presence versus absence of schoolchildren and working-age adults in the household, living with others versus living alone, periodontitis, having versus not having a dog at home, any versus 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 versus no history of BCG vaccination. The following factors associated with decreased odds of COVID-19 with p<0.10 after minimal adjustment: age ≥60 vs 16–29.99 years, education to college or postgraduate level versus primary/secondary level, shielding versus non-shielding, low-impact physical activity (≥2 vs 0 hours/week), presence versus absence of asthma diagnosis, presence versus absence of atopic disease (defined by atopic eczema/dermatitis and/or hayfever/allergic rhinitis), use versus 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 versus 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 (online supplemental figure 3) revealed a high degree of collinearity between the number of working-age adults in the household, multigenerational households and number of people per bedroom. Since household overcrowding (as indicated by number of people per bedroom) was deemed to relate most closely to SARS-CoV-2 exposure risk from a clinical/epidemiological perspective, 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, Index of Multiple Deprivation 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 odds of developing COVID-19 were independently associated with Asian/Asian British versus white ethnicity (aOR 2.28, 95% CI 1.33 to 3.91), household overcrowding (aOR per additional 0.5 people/bedroom 1.26, 1.11 to 1.43), any versus no visits to/from other households in previous week (aOR 1.31, 1.06 to 1.62), number of visits to indoor public places (aOR per extra visit per week 1.05, 1.02 to 1.09), frontline occupation outside health/social care versus no frontline occupation (aOR 1.49, 1.12 to 1.98), raised BMI (aOR 1.50 (1.19 to 1.89) for BMI 25.0–30.0 kg/m^2 and 1.39 (1.06 to 1.84) for BMI >30.0 kg/m^2 versus BMI <25.0 kg/m^2) and sleep >9 hours per night (aOR 1.29, 1.01 to 1.66). Lower odds of test-confirmed COVID-19 were independently associated with history of atopic disease (aOR 0.75, 0.59 to 0.97) and taking systemic immunosuppressants (aOR 0.47, 0.22 to 0.99). Restricted cubic spline analysis showed non-linear associations between BMI (P for non-linearity 0.009) and age (P for non-linearity 0.02) and incident test-defined COVID-19 (figure 1).

We performed two sensitivity analyses to explore the robustness of the multivariable results. The first (online supplemental table 4) excluded 3202 participants who received one or more doses of COVID-19 vaccine before the date of the data download (5 February 2021). The second (online supplemental table 5) 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. An exploratory analysis to determine whether COVID-19 risk differed for participants with atopic versus non-atopic asthma endotypes (as defined by the presence or absence of atopic eczema/dermatitis and/or hayfever/allergic rhinitis) showed reduced odds of COVID-19 for participants with atopic asthma (aOR 0.62, 0.41 to 0.93), but not for those with non-atopic asthma (aOR 0.88, 0.60 to 1.30), as compared with participants without atopic disease or asthma. These effect estimates did not materially change after further adjustment for inhaled corticosteroids (online supplemental table 6).

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 6035 COVIDENCE UK participants entering the prospective analysis who had one or more tests for COVID-19 during the follow-up period, this case definition had sensitivity and specificity for test-confirmed COVID-19 of 0.47 and 0.97, respectively, with an area under the receiver operating characteristic curve of 0.72 (95% CI 0.69 to 0.74; online supplemental table 7). Potential risk factors associating with probable symptom-defined COVID-19 with p<0.10 in a minimally adjusted model (ie, adjusting for age, sex and duration of participation) were included in the multivariable model presented in online supplemental table 8. Increased risk of probable symptom-defined COVID-19 was independently associated with Asian/Asian British versus white ethnicity, housing type (having a mortgage vs home ownership), household overcrowding (>1.0 vs≤0.5 people per bedroom), health or social care occupation, use of cod liver oil supplements, poorer self-assessed general health, periodontitis 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. Associations with household overcrowding and visits to indoor public places showed dose–response relationships, strengthening causal inference. 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 long-standing conditions and medications.

This study sheds new light on the degree of overlap between risk factors for developing COVID-19 (irrespective of severity) versus 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 18] 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 for further research to investigate social and biological factors that might explain ethnic disparities in risk of developing COVID-19, including vitamin D deficiency.[9] 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 19 20]

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, where cases were predominantly mild (93.1% non-hospitalised). Our finding of no association between kidney disease and susceptibility to COVID-19 contrasts with that of de Lusignan *et al*,[21] who reported such an association in a study that was conducted earlier in the UK pandemic when testing was limited to those with more severe COVID-19 illness presenting to hospital. The bias resulting from focusing testing on more severe disease in that study may have contributed to the different findings in our study, which was conducted over a later period when testing was more widely available. In contrast with other studies,[22 23] we found no association between intake of micronutrient supplements and protection against COVID-19: this may reflect a false-negative result from our study (arising due to a relative lack of power to detect modest protective effects), or a false-positive result from other studies arising as a result of less rigorous adjustment for potential socioeconomic confounding.

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 risk for people with diabetes, heart disease or hypertension. While 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 gene encoding the SARS-CoV-2 receptor, which has been reported in people with both high levels of allergic sensitisation and asthma.[17]

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 underascertainment 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. However, the lack of direct swabbing surveillance and reliance on results of routine testing that will usually have been prompted by incident symptoms may have led to underascertainment of asymptomatic SARS-CoV-2 infection. A second issue relates to the self-selected nature of the cohort participants. Ethnic minorities, particularly people of black, African and Caribbean ethnic origin, were under-represented in the study; a lack of statistical power may explain why we did not confirm an increased risk of disease in these groups. People with limited internet access or with fewer digital skills are also less likely to have participated. However, lack of representativeness in a study population does not preclude identification of causal associations.[24] Third, as with any observational study, we cannot exclude the possibility that the associations we report may be explained by residual and/or unmeasured confounding. We sought to minimise this by capture of, and mutual adjustment for, a comprehensive panel of potential confounders. Calculation of E-values enabled us to determine how likely it was that our main findings might be 'explained away' by unmeasured/unknown confounding factors. For example, for an unknown confounder to fully explain the association between frequent visits to shops/indoor places and COVID-19 risk (OR 2.63 comparing top vs bottom quartile of exposure), it would need to be associated with both the outcome and the exposure (above and beyond the measured confounders) by an OR of nearly 5 or more (table 5); weaker confounding could not explain away the association. In contrast, for the association between >9 hours' sleep and COVID-19 risk (OR 1.29) to be explained away, an unknown confounder would need to be associated with the exposure and outcome by an OR of at least 1.90; this effect estimate is comparable with other risk factor associations we have found, and hence is more plausible for an unknown confounder. Accordingly, evidence for an unconfounded, causal association is considerably stronger for frequent visits to shops/indoor places than it is for prolonged sleep duration.

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.

Author affiliations
[1]Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK
[2]Jane Symons Media, London, UK
[3]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[4]Department of Health Services Research and Policy, London School of Hygiene and Tropical Medicine, London, UK
[5]National Heart and Lung Institute, Imperial College London, London, UK
[6]Edenfield Road Surgery, Rochdale, UK
[7]Department of Radiology, Salisbury District Hospital, Salisbury, UK
[8]School of Dentistry, Institute of Clinical Sciences, University of Birmingham, Birmingham, UK
[9]Population Data Science, Swansea University Medical School, Swansea, UK

[10]Centre for Public Health Research (NI), Queen's University Belfast, Belfast, UK
[11]Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh, UK

Twitter Nicholas S Hopkinson @COPDdoc and Aziz Sheikh @DrAzizSheikh

Acknowledgements We thank all the people who participated in the COVIDENCE UK Study, and the following organisations which supported study recruitment: Asthma UK, the British Heart Foundation, the British Lung Foundation, the British Obesity Society, Cancer Research UK, Diabetes UK, Future Publishing, 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, the Terence Higgins Trust and Vasculitis UK.

Contributors ARM wrote the study protocol, with input from HH, MT, CR, GB and SOS. HH, MT, JS, CR, KSY, MRD, KNT, SF, SI, AM, PEP, GL-J, TD, IC, DM, GD, RAL, CG, FK, AS, GB, SOS and ARM contributed to questionnaire development and design. HH coordinated and managed the study, with input from ARM, MT, JS and SOS. HH, JS, ARM, SOS, NSH, OT and BA supported recruitment. MT, HH, MG, MRD, KNT, SSR, AAK, SER, PJL and DJ 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.

Funding This study was supported by a grant from Barts Charity to ARM and CG (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).

Disclaimer The views expressed are those of the authors and not necessarily those of Barts Charity, BREATHE or Health Data Research UK.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The study was sponsored by Queen Mary University of London and approved by Leicester South Research Ethics Committee (ref 20/EM/0117).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Deidentified participant data are available from the corresponding author (a.martineau@qmul.ac.uk) upon reasonable request, subject to the terms of Research Ethics Committee approval.

ORCID iDs
Mohammad Talaei http://orcid.org/0000-0002-6901-3665
Sarah El Rifai http://orcid.org/0000-0002-9194-479X
Philippa J Lloyd http://orcid.org/0000-0002-6946-0780
Nicholas S Hopkinson http://orcid.org/0000-0003-3235-0454
Graham Lloyd-Jones http://orcid.org/0000-0001-8191-4389
Paul E Pfeffer http://orcid.org/0000-0003-0369-2885
Frank Kee http://orcid.org/0000-0002-0606-8167
Adrian R Martineau http://orcid.org/0000-0001-5387-1721http://orcid.org/0000-0001-5387-1721

REFERENCES

1 Aburto JM, Kashyap R, Schöley J, *et al*. Estimating the burden of the COVID-19 pandemic on mortality, life expectancy and lifespan inequality in England and Wales: a population-level analysis. *J Epidemiol Community Health* 2021;75:735–40.
2 Miller IF, Becker AD, Grenfell BT, *et al*. Disease and healthcare burden of COVID-19 in the United States. *Nat Med* 2020;26:1212–7.
3 Pifarré I Arolas H, Acosta E, López-Casasnovas G, *et al*. Years of life lost to COVID-19 in 81 countries. *Sci Rep* 2021;11:3504.
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;584:430–6.
6 Venkatesan P. The changing demographics of COVID-19. *Lancet Respir Med* 2020;8:e95.
7 Ward H, Atchison C, Whitaker M. Antibody prevalence for SARS-CoV-2 following the peak of the pandemic in England: REACT2 study in 100,000 adults. *MedRxiv* 2020.
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;396:313–9.
9 Bhala N, Curry G, Martineau AR, *et al*. Sharpening the global focus on ethnicity and race in the time of COVID-19. *Lancet* 2020;395:1673–6.
10 Martineau AR, Forouhi NG. Vitamin D for COVID-19: a case to answer? *Lancet Diabetes Endocrinol* 2020;8:735–6.
11 Hamer M, Kivimäki M, Gale CR, *et al*. Lifestyle risk factors, inflammatory mechanisms, and COVID-19 hospitalization: a community-based cohort study of 387,109 adults in UK. *Brain Behav Immun* 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;26:1037–40.
13 Griffith GJ, Morris TT, Tudball MJ, *et al*. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nat Commun* 2020;11:5749.
14 Rothman KJ. No adjustments are needed for multiple comparisons. *Epidemiology* 1990;1:43–6.
15 Ding P, VanderWeele TJ. Sensitivity analysis without assumptions. *Epidemiology* 2016;27:368–77.
16 Linden A, Mathur MB, VanderWeele TJ. Conducting sensitivity analysis for unmeasured confounding in observational studies using E-values: the evalue package. *Stata J* 2020;20:162–75.
17 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;146:203–6.
18 Mathur R, Rentsch CT, The OpenSAFELY Collaborative. Ethnic differences in COVID-19 infection, hospitalisation, and mortality: an OpenSAFELY analysis of 17 million adults in England. *MedRxiv* 2020.
19 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:e13128.
20 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. *BMJ Open* 2020;10:e040402.
21 de Lusignan S, Dorward J, Correa A, *et al*. Risk factors for SARS-CoV-2 among patients in the Oxford Royal College of general practitioners research and surveillance centre primary care network: a cross-sectional study. *Lancet Infect Dis* 2020;20:1034–42.
22 Ma H, Zhou T, Heianza Y, *et al*. Habitual use of vitamin D supplements and risk of coronavirus disease 2019 (COVID-19) infection: a prospective study in UK Biobank. *Am J Clin Nutr* 2021;113:1275–81.
23 Louca P, Murray B, Klaser K, *et al*. Modest effects of dietary supplements during the COVID-19 pandemic: insights from 445 850 users of the COVID-19 symptom study APP. *BMJ Nutr Prev Health* 2021;4:149–57.
24 Doll R, Hill AB. Mortality in relation to smoking: ten years' observations of British doctors. *Br Med J* 1964;1:1460–7.