The Lancet Infectious Diseases

Association of age, gender, deprivation, urbanicity, ethnicity, and smoking with a positive test for COVID-19 in an English primary care surveillance network: cross sectional study of the first 500 cases

--Manuscript Draft--

Manuscript Number:	THELANCETID-D-20-02537
Article Type:	Article (Original Research)
Keywords:	General practice; medical record systems; computerized; sentinel surveillance; coronavirus; infections; pandemics
Corresponding Author:	Simon de Lusignan UNITED KINGDOM
First Author:	Simon de Lusignan
Order of Authors:	Simon de Lusignan
	Jienchi Dorward, PhD
	Ana Correa, Msc
	Nicholas Jones, Msc
	Oluwafunmi Akinyemi, PhD
	Gayatri Amirthalingam, FRCP
	Nick Andrews, PhD
	Rachel Byford, BSc
	Gavin Dabrera, FRCP
	Alex Elliot, PhD
	Joanna Ellis, PhD
	filipa Ferreira, PhD
	Jamie LopezBernal, FRCP
	Cecilia Okusi, MRES
	Mary Ramsay, PhD
	Julian Sherlock, BA
	Gillian Smith, MBBS FFPH
	Gary Howsam, FRCP
	Maria Zambon
	Richard Hobbs, MD
	Mark Joy, PhD
	John Williams, FRCP
Manuscript Region of Origin:	UNITED KINGDOM
Abstract:	Background There are few epidemiological studies of community cases in the current coronavirus-2019 (COVID-19) pandemic. We report on the first 500 COVID-19 cases identified through United Kingdom primary care surveillance and describe risk factors for testing COVID-19 positive. Methods The Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance

	Centre (RSC), is a nationally representative primary care sentinel network sharing pseudonymised data, including virological test data for COVID-19. We used multivariable logistic regression models with multiple imputation to identify risk factors for positive COVID-19 tests within this surveillance programme. Findings We identified 3,802 COVID-19 results between 28/01/20 and 04/04/2020, 587 were positive. Greater odds of testing COVID-19 positive included: working-age people (40-64years) and older age, (≥75 years) versus 0-17 year olds (adjusted odds ratio [aOR] 5.26, 95%CI:3.26-8.49 and 5.17,95%CI:2.99-8.92, respectively); male gender (aOR 1.56, 95%CI:1.28-1.90); black and mixed ethnicity compared with white (aOR 4.55, 95%CI:2.55-8.10 and 1.84 95%CO:1.1-3.14, respectively)); urban compared with rural areas (aOR 4.58, 95%CI:3.57-5.88); people with chronic kidney disease (CKD) (aOR 1.88, 95%CI:1.29-2.75) and increasing body mass index (aOR 1.02, 95%CI:1.00-1.03). People in the least deprived deprivation quintile had lower odds of a positive test (aOR 0.49 95%CI:0.36-0.65) as did current smokers (aOR 0.53, 95%CI:0.38-0.74). Interpretation A positive COVID-19 test result in primary care was associated with similar risk factors for severe outcomes seen in hospital settings, with the exception of smoking. We provide early evidence of potential sociodemographic factors associated with a positive test, including ethnicity, deprivation, population density, and CKD. Funding Public Health England, Wellcome Trust.

Association of age, gender, deprivation, urbanicity, ethnicity, and smoking with a positive test for COVID-19 in an English primary care surveillance network: cross sectional study of the first 500 cases

Authors

Simon de Lusignan,[1,2]
Professor of Primary Care and Clinical Informatics
Director Royal College of General Practitioners Research and Surveillance Centre
simon.delusignan@phc.ox.ac.uk

Jienchi Dorward[1,6]
Clinical PhD Fellow
jienchi.dorward@phc.ox.ac.uk

Ana Correa[4,5]
Visiting Researcher
ana.correa@surrey.ac.uk

Nicholas Jones[1]
Research Fellow
nicholas.jones2@hmc.ox.ac.uk

Oluwafunmi Akinyemi[1]
Research Fellow
oluwafunmi.akinyemi@phc.ox.ac.uk

Gayatri Amirthalingam[3]
Consultant Epidemiologist
Gayatri.Amirthalingam@phe.gov.uk

Nick Andrews[3]
Senior Statistician
Nick.Andrews@phe.gov.uk

Rachel Byford[1]
Senior SQL Developer
rachel.byford@phc.ox.ac.uk

Gavin Dabrera[3]
Consultant in Public Health, Acute Respiratory Infection Service
Gavin.Dabrera@phe.gov.uk

Alex Elliot[3]
Consultant Epidemiologist
Alex.Elliot@phe.gov.uk

Joanna Ellis[3]

Clinical Scientist
Joanna.Ellis@phe.gov.uk

Filipa Ferreira[1]
Senior Project Manager
filipa.ferreira@phc.ox.ac.uk

Jamie LopezBernal[3]
Consultant Epidemiologist
Head of Flu
Jamie.LopezBernal2@phe.gov.uk

Cecilia Okusi[1]
Data Scientist
cecilia.okusi@phc.ox.ac.uk

Mary Ramsay[3]
Head of Immunisation
Mary.Ramsay@phe.gov.uk

Julian Sherlock[1]
SQL Developer
julian.sherlock@phc.ox.ac.uk

Gillian Smith[3]
Consultant Epidemiologist
Gillian.Smith@phe.gov.uk

Gary Howsam[2]
Vice Chair RCGP
Gary.Howsam@rcgp.org.uk

Maria Zambon[3]
Deputy Director National Infection Service and Director, Reference Microbiology Services
Maria.Zambon@phe.gov.uk

Richard Hobbs[1]
Head of Department / Digital Health Lead
Richard.Hobbs@phc.ox.ac.uk

Mark Joy[1]
Senior Research Fellow
Mark.joy@phc.ox.ac.uk

John Williams[1]
Senior Research Fellow
John.williams@phc.ox.ac.uk

Correspondence to:

Professor Simon de Lusignan; simon.delusignan@phc.ox.ac.uk

[1]Nuffield Department of Primary Care Health Sciences,
University of Oxford
Eagle House, Walton Well Road, Jericho
Oxford, Ox2 6ED, UK

[2] Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC)
Royal College of General Practitioners
30 Euston Square
London, UK

[3] Public Health England
Colindale
London

[4]Institute for Global Health
University College London
London, UK

[5]Section of Clinical Medicine
University of Surrey
Guildford, UK

[6]Centre for the AIDS Programme of Research in South Africa (CAPRISA)
University of KwaZulu–Natal
Durban, South Africa

Word count: Abstract 251/250, Full Text 3544/3500

Abstract

Background
There are few epidemiological studies of community cases in the current coronavirus-2019 (COVID-19) pandemic. We report on the first 500 COVID-19 cases identified through United Kingdom primary care surveillance and describe risk factors for testing COVID-19 positive.

Methods
The Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC), is a nationally representative primary care sentinel network sharing pseudonymised data, including virological test data for COVID-19. We used multivariable logistic regression models with multiple imputation to identify risk factors for positive COVID-19 tests within this surveillance programme.

Findings
We identified 3,802 COVID-19 results between 28/01/20 and 04/04/2020, 587 were positive. Greater odds of testing COVID-19 positive included: working-age people (40-64years) and older age, (≥75 years) versus 0-17 year olds (adjusted odds ratio [aOR] 5.26, 95%CI:3.26-8.49 and 5.17,95%CI:2.99-8.92, respectively); male gender (aOR 1.56, 95%CI:1.28-1.90); black and mixed ethnicity compared with white (aOR 4.55, 95%CI:2.55-8.10 and 1.84 95%CO:1.1-3.14, respectively)); urban compared with rural areas (aOR 4.58, 95%CI:3.57-5.88); people with chronic kidney disease (CKD) (aOR 1.88, 95%CI:1.29-2.75) and increasing body mass index (aOR 1.02, 95%CI:1.00-1.03). People in the least deprived deprivation quintile had lower odds of a positive test (aOR 0.49 95%CI:0.36-0.65) as did current smokers (aOR 0.53, 95%CI:0.38-0.74).

Interpretation
A positive COVID-19 test result in primary care was associated with similar risk factors for severe outcomes seen in hospital settings, with the exception of smoking. We provide early evidence of potential sociodemographic factors associated with a positive test, including ethnicity, deprivation, population density, and CKD.

Funding
Public Health England, Wellcome Trust.

Introduction

The world is in the grip of a pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus, which causes coronavirus-2019 (COVID-19) disease.[1] The first imported cases were detected in the United Kingdom (UK) in late January, and community transmission began at the end of that month.[2] Initial reports from China, Italy and Spain have described clinical characteristics of people diagnosed with COVID-19 and risk factors for hospitalisation and death.[3] These studies report age, male gender, cardiovascular disease, hypertension and diabetes as being associated with an increased risk of severe disease.[4] However, most research to date has been conducted in hospital settings and the generalisability of these initial results to the much larger population of people who develop COVID-19 in community settings is uncertain. There is an urgent need to establish risk factors for COVID-19 in the broader community population to inform primary care management, public health measures and more personalised advice to patient groups.[5]

The Oxford-Royal College of General Practitioners (RCGP) Research and Surveillance Centre (RSC) is one of the longest established primary care sentinel networks globally. It is recruited to be nationally representative and includes over 500 urban and non-urban participating general practices, covering a population of over 4 million people. [6,7,8] The RCGP RSC works in close collaboration with Public Health England (PHE) on national surveillance of communicable diseases such as influenza[9] and assessing vaccine effectiveness,[10,11] including during the 2009 pandemic.[12] The network has adapted for COVID-19 surveillance by enlarging threefold to provide better coverage and introducing self-swabbing at home to reduce risks of disease transmission.[13,14]

We aimed to identify demographic and clinical risk factors for testing positive for COVID-19 within this primary care surveillance programme.

Methods

Study design and participants

We performed a cross-sectional study of patients in the Ox-RCGP RSC network who were tested for COVID-19 between 24th January and 4h April 2020. Since January 2020, RSC practices have submitted nasopharyngeal swabs to PHE for COVID-19 testing from patients presenting with symptoms of influenza or respiratory infections. Reverse transcriptase polymerase chain reaction (RT PCR) testing for the SARS-CoV-2 virus was performed at the PHE Colindale Laboratory. The RCGP RSC also extracts pseudonymized sociodemographic and clinical data from computerised medical records, including laboratory test results, and allows estimation of household size,[15] deprivation and rural-urban classification.[16] We included cases either from this PHE surveillance testing, through contact tracing, or primary care and hospital testing through routine NHS services.

Eligibility Criteria

We included patients who were registered at an RCGP RSC practice on the 30th September 2019, and with an entry in their medical record reporting a positive or negative test for COVID-19. We have developed a COVID-19 surveillance ontology to ensure consistency of case-definition and only included people with a coded positive test, not people with suspected disease.[17] Patients with codes in their medical records suggesting they have declined any form of data sharing were excluded (approximately 2.2% of the registered population).

Study variables

We included the following independent demographic variables: age, sex and ethnicity using an ontology to maximise case identification;[18] practice level deprivation using the English Index of Multiple Deprivation (IMD) quintiles (we combined the two most deprived quintiles because there was a low frequency of testing leading to sparse data in the most deprived quintile);[19] household size; and rural-urban classification. We included the following clinical variables, modified from those associated with increased vulnerability to influenza: body mass index (BMI); smoking status; pregnancy; hypertension; chronic kidney disease (CKD); coronary heart disease (CHD); chronic respiratory disease, asthma and chronic obstructive pulmonary disease; and diabetes type 1 and type 2. We grouped together malignancy and immunocompromised into a single group because of small numbers tested. Malignancy was identified from disease codes and we used records of prescriptions for prednisolone, and prescriptions of disease-modifying anti-rheumatic drugs (DMARDs) as surrogates for immunosuppression. The outcome variable was testing COVID-19 positive.

Statistical analysis

We used descriptive statistics and reported counts and proportions for categorical data, and measures of distribution for continuous data. We used a multivariate logistic regression to identify variables that were associated with a positive test of COVID-19 after multiple imputation of missing values. We included all variables in the multivariable model. We imputed missing data using the multiple imputation by chained equations method, with five imputed datasets and ten iterations[20]. For each variable, we specified a predictive mean matching model. Missing ethnicity observations were imputed using census data:[21] For each person with missing ethnicity in a given Lower Super Output Area (LSOA)[22] we randomly assigned an ethnic group, matching proportions of the ethnic group based on the census proportions. All analysis results were aggregated with Rubin's rule after appropriate transformation.[23] We checked the acceptability of the imputations by comparison of plots of the distribution of recorded and imputed values for all measurements. We used this method under the assumption that the missing observations for covariates were missing at random (MAR). All covariates in the logistic regression model (other than ethnicity) had fractions of missing data given in Table 1. We checked collinearity by measuring the variance inflationary factor (vif) for each covariate: all were deemed within acceptable bounds (the maximum vif was less than 2.0).

For all statistical analyses we used the open source software, R-version 3.5.3; we used the R library mice 3.4.0 for the multiple imputation routine.

Ethics

The data used for the analysis were pseudonymised at the point of extraction and encrypted prior to uploading to the Clinical Informatics Research Group secure server. Personal data were not identifiable during the analysis. The data extract was conducted as part of our surveillance work commissioned by PHE and approved under Regulation 3 of The Health Service (Control of Patient

Information) Regulations 2002. This study was approved by the RCGP RSC study approval committee and was classified as a study of "usual practice".[24] Therefore, no further ethical approval was required.

Results

Demographic characteristics

Between 24th January and 4th April 2020, we observed 587 patients with COVID-19 positive results and 3215 negative results in our surveillance programme. The first positive case presented on 30th January 2020, and the first 100 cases were passed on 17th March 2020. Overall, 2190 (57.6%) were male and 3456 (90.9) were of white ethnicity (Table 1). The median age of patients who had a test was 58.5 years for men (IQR 45-75) and 53 years old (IQR: 31-70) for women. Just over half (1986, 52.2%) lived in rural areas, and 1855 (48.8%) were ranked as least deprived in the IMD quintile 5. The most common clinical conditions were hypertension (1094, 28.8%) and chronic heart disease (600, 15.8%).

Table 1. Demographic Characteristics of Cohort (n, %), including missingness, N=3802

Variable	Level	N (%)
COVID-19 Test Result	*Negative*	3215 (84.6)
	Positive	587 (15.4)
	Missing	0
Age, years	*0-17*	*499 (13.1)*
	18-39	*666 (17.5)*
	40-64	*1316 (34.6)*
	65-74	*557 (14.7)*
	75+	*764 (20.1)*
	Missing	*0*
Sex	*Male*	*2190 (57.6)*
	Female	*1612 (42.4)*
	Missing	*0*
Ethnicity	*White*	*3456 (90.9)*
	Asian	*184 (4.8)*
	Black	*64 (1.68)*
	Mixed, Other	*98 (2.6)*
	Missing	*1014 (26.7)**
SES – Index of multiple Deprivation	*1&2 (most deprived)*	*668 (17.5)*
	3	*646 (17.0)*
	4	*633 (16.7)*
	5 (least deprived)	*1855 (48.8)*
	Missing	*0*

Variable	Level	N (%)
Household Size	*1*	*824 (21.7)*
	2	*1087 (28.6)*
	3	*660 (17.4)*
	4	*594 (15.6)*
	5-8	*408 (10.7)*
	9+	*135 (3.55)*
	Missing	*94 (2.47)*
Urban/Rural	*Rural*	*1986 (52.2)*
	Urban	*1816 (47.8)*
	Missing	*0*
Smoking Status	*Never smoked*	*1326 (34.9)*
	Active smoker	*501 (13.2)*
	Ex-smoker	*1469 (38.6)*
	Missing	*506 (13.3)*
Pregnancy	*No*	*3742 (98.4)*
	Yes	*60 (1.6)*
	Missing	*0*
Hypertension	*No*	*2708 (71.2)*
	Yes	*1094 (28.8)*
	Missing	*0*
CKD	*No*	*3595 (94.6)*
	Yes	*207 (5.4)*
	Missing	*0*
Diabetes	*No*	*3299 (86.8)*
	Yes	*503 (13.2)*
	Missing	*0*
CHD	*No*	*3202 (84.2)*
	Yes	*600 (15.8)*
	Missing	*0*
CRD	*No*	*3544 (3.2)*
	Yes	*258 (6.8)*
	Missing	*0*
Malignancy and Immuno-compromised	*No*	*3448 (90.7)*
	Yes	*354 (9.3)*
	Missing	*0*
BMI (continuous)	*Mean 27.2 (median 26.4) IQR: 23.0-30.1*	
	Missing	*586(15.4)*

* Missing variables imputed by census data and inputted into table

Risk factors for testing COVID-19 positive

In univariable analysis the odds of testing positive for COVID-19 were higher amongst older people, and people of Black, Asian or Mixed/Other Ethnicity (Table 2). A higher odds of a positive test were also seen for people living in more deprived areas, urban areas and single households or households including nine or more people. Among clinical factors in the univariable analysis, CKD, malignancy or immunocompromise, diabetes, respiratory disease and hypertension were all associated with increased odds of a positive test for COVID-19 (Table 3). Active smoking was associated with decreased odds of a positive test.

In multivariable analysis, adjusted for all other variables in Tables 2 and 3, male gender remained independently associated with testing COVID-19 positive (adjusted OR [aOR] 1.46, 95%CI:1.17 to 1.81). Adults were at increased risk compared to children, but people in the age group 40 to 64 years were at greatest risk (aOR 5.26, 95%CI 3.26-8.49). Compared to people of white ethnicity, the adjusted odds of a positive test were over four times greater in people of black ethnicity (aOR 4.55, 95% 2.55-8.1) and remained high in people of Mixed/Other Ethnicity. Urban areas (aOR 4.58, 95%CI:3.57-5.88) were associated with an increased risk of a positive COVID-19 test, whereas those living in the least deprived areas (quintile 5) had a lower risk (aOR 0.49, 95%CI:0.36-0.65). The association with household size was no longer statistically significant.

Active smoking was linked to decreased odds of a positive COVID-19 test result (aOR 0.53, 95%CI 0.38-0.74). People with CKD were more likely to test positive for COVID-19 in the adjusted analysis (aOR 1.88, 95%CI:1.29-2.75) but there was no longer a statistically significant association between other chronic conditions and risk of a positive COVID-19 test in the multivariable analysis (Table 3). For every increase in one BMI unit, there was an increased aOR of 1.03 (95%CI:1.01-1.04) of having a positive test.

Table 2. Univariable and multivariable analysis of demographic risk factors for testing COVID-19 positive using logistic regression.

Variable	Level	COVID-19 Positivity n/N	(%)	Unadjusted OR (95% CI)	P-value	Adjusted OR (95% CI)	P-value
Age (years)	*0-17*	*23/499*	*(4.6)*	*1*		*1*	
	18-39	*84/666*	*(12.6)*	*2.99 (1.85-4.81)*	*<0.0001*	*2.86 (1.71-4.76)*	*0.0001*
	40-64	*243/1316*	*(18.4)*	*4.69 (3.02-7.28)*	*<0.0001*	*5.26 (3.26-8.49)*	*<0.0001*
	65-74	*88/557*	*(15.8)*	*3.88 (2.41-6.25)*	*<0.0001*	*4.35 (2.51-7.54)*	*<0.0001*
	75+	*149/764*	*(19.5)*	*5.01 (3.18-7.90)*	*<0.0001*	*5.16 (2.99-8.92)*	*<0.0001*
Sex	*Female*	*291/2190*	*(13.3)*	*1*		*1*	
	Male	*296/1612*	*(18.4)*	*1.47 (1.23-1.75)*	*<0.0001*	*1.56 (1.28-1.90)*	*<0.0001*
SES – Index of Multiple Deprivation Quintile	*1&2 (most deprived)*	*197/668*	*(29.5)*	*1*		*1*	
	3	*135/646*	*(26.4)*	*0.63 (0.49-0.81)*	*0.0004*	*1.13 (0.85-1.51)*	*0.4*
	4	*112/633*	*(17.7)*	*0.51 (0.39-0.67)*	*<0.0001*	*0.72 (0.54-0.97)*	*0.0325*
	5 (least deprived)	*143/1855*	*(7.7)*	*0.20 (0.16-0.25)*	*<0.0001*	*0.49 (0.36-0.65)*	*<0.0001*
Settlement / population density	*Rural*	*111/1986*	*(5.6)*	*1*		*1*	
	Urban	*476/1816*	*(26.2)*	*6.0 (4.82-7.46)*	*<0.0001*	*4.58 (3.57-5.88)*	*<0.0001*
Ethnicity	*White*	*475/3456*	*(13.7)*	*1*		*1*	
	Asian	*53/184*	*(28.8)*	*2.54 (1.82-3.54)*	*<0.0001*	*1.38 (0.95-2.0)*	*0.094*
	Black	*36/64*	*(56.3)*	*8.07 (4.88-13.35)*	*<0.0001*	*4.55 (2.55-8.1)*	*<0.0001*
	Mixed, Other	*23/98*	*(23.5)*	*1.92 (1.2-3.1)*	*0.0007*	*1.84 (1.1-3.14)*	*0.0257*
Household Size	*1*	*163/824*	*(19.7)*	*1*		*1*	
	2	*158/1087*	*(14.5)*	*0.69 (0.54-0.88)*	*0.002*	*0.92 (0.7-1.22)*	*0.57*
	3	*92/660*	*(13.9)*	*0.66 (0.5-0.87)*	*0.003*	*1.06 (0.77-1.47)*	*0.7*
	4	*70/594*	*(11.8)*	*0.54 (0.4-0.73)*	*<0.0001*	*1.0 (0.7-1.42)*	*0.99*
	5-8	*53/408*	*(12.9)*	*0.61 (0.43-0.85)*	*0.004*	*0.89 (0.6-1.31)*	*0.54*
	9-+	*35/135*	*(25.9)*	*1.42 (0.93-2.16)*	*0.1*	*1.23 (0.76-2.0)*	*0.4*

Table 3. Univariable and multivariable analysis of clinical risk factors for testing COVID-19 positive using logistic regression.

Smoking Status	*Never smoked*	*219/1326*	*(16.5)*	*1*		*1*	
	Active smoker	*58/501*	*(11.6)*	*0.66 (0.49-0.90)*	*0.009*	*0.53 (0.38-0.74)*	*0.0002*
	Ex-smoker	*274/1469*	*(18.7)*	*1.56 (0.95-1.41)*	*0.14*	*0.92 (0.73-1.15)*	*0.46*
Pregnancy	*No*	*583/3742*	*(15.6)*	*1*		*1*	
	Yes	*4/60*	*(6.7)*	*0.39 (0.14-1.1)*	*0.07*	*0.67 (0.23-1.97)*	*0.47*
Hypertension	*No*	*378/2708*	*(14.0)*	*1*		*1*	
	Yes	*209/1094*	*(19.1)*	*1.46 (1.21-1.75)*	*<0.0001*	*0.90 (0.70-1.15)*	*0.4*
CKD	*No*	*519/3595*	*(14.4)*	*1*		*1*	
	Yes	*68/207*	*(32.9)*	*2.9 (2.14-3.93)*	*<0.0001*	*1.88 (1.29-2.75)*	*0.001*
Diabetes	*No*	*473/3299*	*(14.3)*	*1*		*1*	
	Yes	*114/503*	*(22.6)*	*1.75 (1.39-2.2)*	*<0.0001*	*1.02 (0.78-1.35)*	*0.86*
CHD	*No*	*451/3202*	*(14.1)*	*1*		*1*	
	Yes	*136/600*	*(22.7)*	*1.79 (1.44-2.2)*	*<0.0001*	*1.21 (0.92-1.6)*	*0.18*
CRD	*No*	*529/3544*	*(14.9)*	*1*		*1*	
	Yes	*58/258*	*(22.5)*	*1.65 (1.22-2.25)*	*0.001*	*0.99 (0.69-1.44)*	*0.97*
Malignancy and	*No*	*504/3448*	*(14.6)*	*1*		*1*	
Immunocompromised	*Yes*	*83/354*	*(23.4)*	*1.79 (1.37-2.33)*	*<0.0001*	*1.12 (0.82-1.52)*	*0.48*
BMI (continuous)			*1.03 (1.01-1.04)*		*<0.0001*	*1.02 (1.0-1.03)*	*0.04*

Discussion

We report one of the first and largest cross-sectional analysis of a community sample of people tested for COVID-19. We found increasing age, male sex, deprivation, urban location, black and mixed/other ethnicity were associated with an increased risk of a positive COVID-19 test. Current smoking was linked to decreased odds of a positive test. CKD and increased BMI were the only clinical factors independently associated with a positive test.

Evidence before this study

A literature review suggests that COVID-19 has affected more men than women, and principally those in the 30 to 65 year-old age band, with roughly half of cases over 50 years old.[25] We found a similar increased risk of a positive COVID-19 test in men, and in people above 40 years old, particularly among those in the second half of their working life (40 to 65 years) and those aged 75 years or older.

COVID-19 transmission is known to be associated with high population density due to increased social mixing,[26] which is consistent with the higher odds of a positive test in urban areas in our primary care surveillance network. While social deprivation has been associated with increased risk of other respiratory infections,[27] we are not aware of evidence demonstrating this for COVID-19. We found a dose response relationship between decreasing deprivation and decreased odds of a positive test, independent of household size, urban location and smoking. Other factors which we did not measure, such as employment, may have contributed to this effect and should be explored.

Some ethnic differences have been reported, that suggest Asian and Black ethnicities may be associated with higher risk of disease or more severe infection, or both.[28] For example, among 3370 people admitted to intensive care in the UK with confirmed COVID-19 and ethnicity data, 402 (11.9%) were black, 486 Asian (14.4%) and 2236 (66.4%) white, compared to respective national figures of 3.4%, 6.8% and 80%.[29] These results did not adjust for potential sociodemographic or clinical confounders. While overall numbers of Black, Asian and Minority Ethnic people were small in our study, we found an association between black or mixed/other ethnicity and the odds of a positive COVID-19 test result, which remained significant after adjusting for important comorbidities such as hypertension and diabetes, the prevalence of which is increased in Black and Asian ethnic groups.[30]

Systematic reviews have shown that people with COVID-19, who have chronic co-morbidities such as hypertension, diabetes and cardiovascular disease are at a higher risk of progressing to severe COVID-19 disease.[31,32] There is little evidence regarding the impact of these conditions on risk of SARS-CoV-2 infection, and we found that chronic kidney disease and increased BMI were associated with testing positive for COVID-19. Risk factors for infection and progression to severe disease may be different. There has been controversy over smoking and COVID-19, with most of the data focussing on the impact of smoking on disease severity and progression, rather than risk of testing COVID-19 positive. We found that active smoking was associated with lower odds of having a positive test result. We speculate whether active smoking might have an impact on the viral test results rather than on disease process. The association between smoking and COVID-19 merits further investigation, particularly as there is a risk our data might encourage people to smoke. This would be unwarranted given the well documented harms to overall health from smoking, the potential for smoking to increase COVID-19 disease severity,[33] and the preliminary nature of our findings.

Strengths and limitations

To our knowledge, this is the first study to report risk factors for testing positive for COVID-19 among a well described cohort of patients in a community setting. The RCGP RSC is an established network of sentinel practices, meaning clinicians are experienced in undertaking surveillance research. The same approach to sample collection and analysis was used in all patients. Practices have been provided with coding ontologies in order to standardise reporting.

This study is based on a cohort of symptomatic people who self-presented to a primary care sentinel practice. Although health care is free at the point of access in the UK, results may in part reflect which groups of patients were more likely to present for assessment. Our study population contained a high proportion of people living in rural and least deprived areas, suggesting there may be some disparities in testing based on socioeconomic status. This could bias our results, as if people living in more deprived areas were only likely to present to practices when more severely unwell, those who did present could be more likely to be COVID-positive. In less deprived areas, there may be less barriers to presenting, meaning people may present with less severe symptoms, and therefore be more likely to test negative. Population-based surveys are therefore required to ensure consistent levels of testing across sub-groups.

Whilst RT PCR testing is the current gold standard for SARS-CoV-2 diagnosis, reports suggest theses assays are approximately 75% sensitive, meaning some COVID-19 cases would have been missed. Also, the sentinel network changed from in-practice nasopharyngeal to self-swabbing, using a previously tested approach.[34] Testing, coding and the likelihood of infection being travel-related have all changed over time and are potential confounders. It was not possible to assess the impact of these factors on positive tests in this cross-sectional analysis. Missing data for smoking, ethnicity, and BMI may also have impacted our results, although we corrected for this with multiple imputation techniques.

Implications for practice and future research

Further data are needed to establish the epidemiology of COVID-19, particularly in relation to emerging factors such as ethnicity, deprivation and population density. Population-based surveys could help to reduce selection bias and ensure adequate inclusion of different population sub-groups. Our data from primary care may be able to help inform the impact of social distancing and lockdown on particular groups; it should also be able to inform about rates of hospitalisation and death as the pandemic unfolds.

Conclusions

Sentinel network data provides some insights into the epidemiology of COVID-19. Notwithstanding the limitations of the scale of this study and that it is based on routine data; primary care provides insights into which groups are more likely to test positive for COVID-19. Our finding about cigarette smoking should not be taken as permission to carry on. Increasing age, male gender, black and mixed ethnicity, and increased population density are all associated with an increased risk of a positive test.

Research in context

Evidence before this study

We conducted a focused search of PubMed, Medline and Trip database from inception to 14.4.20 to look for community-based studies describing the epidemiology of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) virus or the associated illness, coronavirus-2019 (COVID-19), using the terms "(COVID-19 or 2019-nCoV or SARS-CoV-2) AND (primary care or general practice or family practice or community)". We found no relevant studies.

Hospital-based studies have reported increasing age, male gender and certain co-morbidities, such as hypertension and diabetes, are associated with an increased risk of COVID-19 and with more severe disease. The generalisability of these initial results to the much larger population of people who develop COVID-19 in community settings is currently uncertain.

Added value of this study

We performed a cross-sectional study of patients with a positive COVID-19 test code in the Oxford Royal College of General Practitioners Research Surveillance Centre network between 24th January and 4h April 2020. We observed 587 patients with positive and 3215 with negative results.

Increasing age, population density, male sex and black ethnicity were associated with an increased risk of a positive COVID-19 test. Chronic kidney disease and increasing body mass index were the only clinical factors associated with a positive test. Current smokers have a lower rate of positive tests.

People in the least deprived socioeconomic group, those of white ethnicity and people living in rural areas were more likely to have a negative test result than their comparative subgroups, suggesting there may currently be disparities in testing.

Implications of all the available evidence

Our findings suggest risk factors for developing COVID-19 in the community are similar to those which lead to hospital admission, with males, people aged over 40 years and those living in urban areas most at risk. Research is needed into reasons for potential disparities in testing, ethnic variations in COVID-19 incidence and the risk to smokers.

Acknowledgements:

The participating practices and patients for providing the data for this cohort. Collaboration with the GP computer system suppliers – EMIS, TPP and InPractice Systems; Wellbeing systems; Public Health England, Wellcome Trust and our other funders and collaborators. Funders have no role in the writing of the manuscript or the decision to submit it for publication.

Contributors:

SdeL conceived the study with MZ and wrote the first full draft and led subsequent revisions. AC did the first analysis with input and help from MJ and wrote an initial draft manuscript. JS with input from RB created the first and revised data output from the database. MJ conducted the final statistical analysis. JD and NJ and other authors helped interpret the analysis, and contributed substantially to the write up. CO additionally assisted with the data presentation. All other authors contributed substantially to ensuring the effective availability of data to support this paper, interpretation and analysis of this data to the manuscript and its revisions.

Declaration of interests:

The authors have no competing interests. SdeL is the Director of the Oxford RCGP RSC, RB, JS, FF, EK and GH are part funded by PHE; and CO and AC by a Wellcome Biomedical resources grant (212763/Z/18/Z). JD is funded by Wellcome Trust (216421/Z/19/Z)

Funding:

Public Health England provides the core funding for RCGP RSC, no specific funding was provided for this analysis.

Data sharing statement:

The RCGP RSC data set can be accessed by researchers, approval is on a project-by-project basis (www.rcgp.org.uk/rsc). Ethical approval by an NHS Research Ethics Committee is needed before any data release/other appropriate approval. Researchers wishing to directly analyse the patient-level pseudonymised data will be required to complete information governance training and work on the data from the secure servers at the University of Surrey. Patient-level data cannot be taken out of the secure network. We encourage interested researchers to attend the short courses on how to analyse primary-care data/RCGP RSC data offered twice a year.

References

[1] Lai, C. C., Shih, T. P., Ko, W. C., Tang, H. J., & Hsueh, P. R. (2020). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. International journal of antimicrobial agents, 105924.

[2] Kinross, P., Suetens, C., Dias, J. G., Alexakis, L., Wijermans, A., Colzani, E., & Monnet, D. L. (2020). Rapidly increasing cumulative incidence of coronavirus disease (COVID-19) in the European Union/European Economic Area and the United Kingdom, 1 January to 15 March 2020. Eurosurveillance, 25(11).

[3] Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. JAMA. 2020 Feb 24. doi: 10.1001/jama.2020.2648.

[4] Zhou F, Yu T, Du R, Fan G, Liu Y, Liu Z, Xiang J, Wang Y, Song B, Gu X, Guan L, Wei Y, Li H, Wu X, Xu J, Tu S, Zhang Y, Chen H, Cao B. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. Lancet. 2020 Mar 28;395(10229):1054-1062. doi: 10.1016/S0140-6736(20)30566-3.

[5] Lipsitch M, Swerdlow DL, Finelli L. Defining the Epidemiology of Covid-19 - Studies Needed. N Engl J Med. 2020 Mar 26;382(13):1194-1196. doi:10.1056/NEJMp2002125.

[6] Fleming, DM, Miles J. The representativeness of sentinel practice networks. J Public Health (Oxf). 2010 Mar;32(1):90-6.

[7] Correa A, Hinton W, McGovern A, van Vlymen J, Yonova I, Jones S, de Lusignan S. Royal College of General Practitioners Research and Surveillance Centre (RCGP RSC) sentinel network: a cohort profile. BMJ Open. 2016 Apr 20;6(4):e011092.

[8] de Lusignan S, Correa A, Smith GE, Yonova I, Pebody R, Ferreira F, Elliot AJ, Fleming D. RCGP Research and Surveillance Centre: 50 years' surveillance of influenza, infections, and respiratory conditions. Br J Gen Pract. 2017 Oct;67(663):440-441. doi: 10.3399/bjgp17X692645.

[9] Fleming DM. Weekly Returns Service of the Royal College of General Practitioners. Commun Dis Public Health 1999;2(2):96 –100

[10] Pebody RG, Whitaker H, Ellis J, Andrews N, Marques DFP, Cottrell S, Reynolds AJ, Gunson R, Thompson C, Galiano M, Lackenby A, Robertson C, O'Doherty MG, Owens K, Yonova I, Shepherd SJ, Moore C, Johnston J, Donati M, McMenamin J, Lusignan S, Zambon M. End of season influenza vaccine effectiveness in primary care in adults and children in the United Kingdom in 2018/19. Vaccine. 2020 Jan 16;38(3):489-497. doi: 10.1016/j.vaccine.2019.10.071.

[11] Pebody RG, Warburton F, Andrews N, Sinnathamby M, Yonova I, Reynolds A, Robertson C, Cottrell S, Sartaj M, Gunson R, Donati M, Moore C, Ellis J, de Lusignan S, McMenamin J, Zambon M. Uptake and effectiveness of influenza vaccine in those aged 65 years and older in the United Kingdom, influenza seasons 2010/11 to 2016/17. Euro Surveill. 2018 Sep;23(39). doi:10.2807/1560-7917.ES.2018.23.39.1800092

[12] Fleming DM, Durnall H. Ten lessons for the next influenza pandemic-an English perspective: a personal reflection based on community surveillance data. Hum Vaccin Immunother. 2012 Jan;8(1):138-45. doi: 10.4161/hv.8.1.18808.

[13] de Lusignan S, Borrow R, Tripathy M, Linley E, Zambon M, Hoschler K, Ferreira F, Andrews N, Yonova I, Hriskova M, Rafi I, Pebody R. Serological surveillance of influenza in an English sentinel network: pilot study protocol. BMJ Open. 2019 Mar 8;9(3):e024285. doi: 10.1136/bmjopen-2018-024285.

[14] de Lusignan S, Lopez Bernal J, Zambon M, Akinyemi O, Amirthalingam G, Andrews N, Borrow R, Byford R, Charlett A, Dabrera G, Ellis J, Elliot AJ, Feher M, Ferreira F, Krajenbrink E, Leach J, Linley E, Liyanage H, Okusi C, Ramsay M, Smith G, Sherlock J, Thomas N, Tripathy M, Williams J, Howsam G, Joy M, Hobbs R. Emergence of a Novel Coronavirus (COVID-19): Protocol for Extending Surveillance Used by the Royal College of General Practitioners Research and Surveillance Centre and Public Health England. JMIR Public Health Surveill. 2020 Apr 2;6(2):e18606. doi: 10.2196/18606.

[15] de Lusignan S, Sherlock J, Ferreira F, O'Brien S, Joy M. Household presentation of acute gastroenteritis in a primary care sentinel network: retrospective database studies. BMC Public Health. 2020 Apr 5;20(1):445. doi: 10.1186/s12889-020-08525-8.

[16] de Lusignan S, McGee C, Webb R, Joy M, Byford R, Yonova I, Hriskova M, Matos Ferreira F, Elliot AJ, Smith G, Rafi I. Conurbation, Urban, and Rural Living as Determinants of Allergies and Infectious Diseases: Royal College of General Practitioners Research and Surveillance Centre Annual Report 2016-2017. JMIR Public Health Surveill. 2018 Nov 26;4(4):e11354. doi: 10.2196/11354.

[17] Bioportal. COVID-19 Surveillance Ontology. URL: https://bioportal.bioontology.org/ontologies/COVID19

[18] Tippu Z, Correa A, Liyanage H, Burleigh D, McGovern A, Van Vlymen J, Jones S, de Lusignan S. Ethnicity Recording in Primary Care Computerised Medical Record Systems: An Ontological Approach. J Innov Health Inform. 2017 Mar 14;23(4):920. doi: 10.14236/jhi.v23i4.920.

[19] Department for Communities and Local Government. The English Index of Multiple Deprivation (IMD) 2015 – Guidance. UK Government. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/464430/English_Index_of_Multiple_Deprivation_2015_-_Guidance.pdf

[20] Multivariate Imputation by Chained Equations in R, Journal of Statistical Software December 2011, Volume 45, Issue 3. http://www.jstatsoft.org/

[21] Office for National Statistics (2011): KS201EW: Ethnic group. https://www.nomisweb.co.uk/

[22] Office of National Statistics. Census geography: 2001 Output areas. URL: https://www.ons.gov.uk/methodology/geography/ukgeographies/censusgeography

[23] Mertens BJA, Banzato E, de Wreede LC. Construction and assessment of prediction rules for binary outcome in the presence of missing predictor data using multiple imputation and cross-validation: Methodological approach and data-based evaluation. Biom J. 2020 Feb 13. doi: 10.1002/bimj.201800289.

[24] Health Research Authority. Defining research table - October 2017. 2017. Available from: http://www.hra-decisiontools.org.uk/research/docs/DefiningResearchTable_Oct2017-1.pdf [accessed 6th February 2018].

[25] Yi Y, Lagniton PNP, Ye S, Li E, Xu RH. COVID-19: what has been learned and to be learned about the novel coronavirus disease. Int J Biol Sci. 2020 Mar 15;16(10):1753-1766. doi: 10.7150/ijbs.45134.

[26] Wang KW, Gao J, Wang H, Wu XL, Yuan QF, Guo FY, Zhang ZJ, Cheng Y. Epidemiology of 2019 novel coronavirus in Jiangsu Province, China after wartime control measures: A population-level retrospective study. Travel Med Infect Dis. 2020 Apr 5:101654. doi: 10.1016/j.tmaid.2020.101654.

[27] Smith S, Morbey R, de Lusignan S, Pebody RG, Smith GE, Elliot AJ. Investigating regional variation of respiratory infections in a general practice syndromic surveillance system. J Public Health (Oxf). 2020 Feb 2. pii: fdaa014. doi: 10.1093/pubmed/fdaa014.

[28] Intensive Care National Audit & Research Centre (ICNARC) Case Mix Programme Database. ICNARC report on COVID-19 in critical care 10 April 2020. URL: https://www.icnarc.org/DataServices/Attachments/Download/c31dd38d-d77b-ea11-9124-00505601089b

[29] UK Government. Ethnicity Facts and Figures. URL: https://www.ethnicity-facts-figures.service.gov.uk/uk-population-by-ethnicity/national-and-regional-populations/population-of-england-and-wales/latest

[30] Carson AP, Howard G, Burke GL, Shea S, Levitan EB, Muntner P. Ethnic differences in hypertension incidence among middle-aged and older adults: the multi-ethnic study of atherosclerosis. Hypertension. 2011 Jun;57(6):1101-7. doi:10.1161/HYPERTENSIONAHA.110.168005.

[31] Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, Ji R, Wang H, Wang Y, Zhou Y. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. Int J Infect Dis. 2020 Mar 12. pii:S1201-9712(20)30136-3. doi: 10.1016/j.ijid.2020.03.017.

[32] Li B, Yang J, Zhao F, Zhi L, Wang X, Liu L, Bi Z, Zhao Y. Prevalence and impact of cardiovascular metabolic diseases on COVID-19 in China. Clin Res Cardiol. 2020 Mar 11. doi: 10.1007/s00392-020-01626-9.

[33] Vardavas CI, Nikitara K. COVID-19 and smoking: A systematic review of the evidence. Tob Induc Dis. 2020 Mar 20;18:20. doi: 10.18332/tid/119324. eCollection 2020.

[34] Elliot AJ, Bermingham A, Charlett A, Lackenby A, Ellis J, Sadler C, Sebastianpillai P, Powers C, Foord D, Povey E, Evans B, Durnall H, Fleming DM, Brown D, Smith GE, Zambon M. Self-sampling for community respiratory illness: a new tool for national virological surveillance. Euro Surveill. 2015 Mar 12;20(10):21058.