Current smoking and COVID-19 risk: results from a population symptom app in over 2.4 million people

Nicholas S Hopkinson,1 Niccolo Rossi,2 Julia El-Sayed_Moustafa,2 Anthony A Laverty,3 Jennifer K Quint,4 Maxim Freidin,5 Alessia Visconti,2 Ben Murray,6 Marc Modat,6 Sebastien Ourselin,7 Kerrin Small,2 Richard Davies,8 Jonathan Wolf,8 Tim D Spector,2 Claire J Steves,2 Mario Falchi2

ABSTRACT

Background The association between current tobacco smoking, the risk of developing symptomatic COVID-19 and the severity of illness is an important information gap.

Methods UK users of the Zoe COVID-19 Symptom Study app provided baseline data including demographics, anthropometrics, smoking status and medical conditions, and were asked to log their condition daily. Participants who reported that they did not feel physically normal were then asked by the app to complete a series of questions, including 14 potential COVID-19 symptoms and about hospital attendance. The main study outcome was the development of ‘classic’ symptoms of COVID-19 during the pandemic defined as fever, new persistent cough and breathlessness and their association with current smoking. The number of concurrent COVID-19 symptoms was used as a proxy for severity and the pattern of association between symptoms was also compared between smokers and non-smokers.

Results Between 24 March 2020 and 23 April 2020, data were available on 2 401 982 participants, mean (SD) age 43.6 (15.1) years, 63.3% female, overall smoking prevalence 11.0%. 834 437 (35%) participants reported being unwell and entered one or more related symptoms suggesting a diagnosis of COVID-19; classic symptoms including cardiovascular disease, diabetes and COPD.4–7

Interpretation These data are consistent with people who smoke being at an increased risk of developing symptomatic COVID-19.

INTRODUCTION

Although in many people infection with SARS-CoV-2 is asymptomatic, some develop symptomatic COVID-19 and in some individuals there is progressive lung involvement with respiratory failure and widespread systemic consequences.1–4 The risk of severe complications is higher in older people and those with long-term medical conditions including cardiovascular disease, diabetes and COPD.4–7

Tobacco smoking is a significant risk factor for both viral and bacterial infections of the respiratory system,8,9 with smokers five times as likely to develop influenza and twice as likely to have pneumonia.10 It may therefore be an important factor worsening the impact of SARS-CoV-2. There is evidence from case series that smoking is associated with more severe disease, a greater risk of intensive care unit admission and excess mortality in people with COVID-19 admitted to hospital.8,11–14 However, some reports have suggested that although smoking-related disease is common in patients with COVID-19, current smoking rates in hospitalised patients are lower than would be expected from population smoking prevalence. While some studies have looked specifically at current smoking,12–14 others combined current and ex-smokers,11 and smoking prevalence in these samples suggest data may have been incomplete. It is important to be able to distinguish the effect of current smoking from the impact of smoking-related disease.
on long-term conditions caused by former smoking, in order to
guide the appropriate prioritisation of smoking cessation as part of
the global response to COVID-19.

In order to establish the impact of current smoking more
accurately, we analysed data from a population of COVID-19
symptom reporting app, developed by Zoe Global in collabora-
tion with King’s College London. We hypothesised that current
smokers would be at increased risk of developing self-reported
symptoms suggesting that they had COVID-19 and would expe-
rience more symptoms reflecting a greater illness burden.

METHODS
Data and analysis
Adult members of the public were invited to download the Zoe
COVID-19 Symptom Study app after its launch on radio, TV and
via social media. On first use, the app records users’ self-reported
location, age and core health risk factors, including height and
weight, if they smoked with options; yes, not currently, never.
Participants were asked about common disease (eg, diabetes,
heart disease, lung disease) status, some medication use as well
as whether they thought that they already had COVID.15

With continued use, participants were asked to provide daily
updates on whether they have been tested for SARS-CoV-2 and the
results of these tests as well as if they ‘feel physically normal’. If
they reported that they did not, they were then invited by
the app to record the presence of 14 symptoms that have been
shown to be associated with COVID-1915 (ie, abdominal pain,
chest pain, delirium, diarrhoea, fatigue, fever, headache, hoarse
voice, loss of smell, persistent cough, shortness of breath, skipped
meals, sore throat, unusual muscle pains) and whether they had
attended hospital. Fatigue and shortness of breath were graded
by symptom severity, ranging from mild to severe.

Study population
The study population for this analysis comprised individuals
resident in the UK who were registered with the app and entered
data between 24 March and 23 April 2020. We collected all
symptoms declared by each participant during the period which
they engaged with the app. More details about user engagement
and the overall distribution of reported symptoms are shown
in online supplemental figures E1-2. While the response option
for the majority of symptoms was yes/no, fatigue and shortness
of breath were considered present if any level of severity was
reported. Where a possible symptom response was missing it was
reported as whether they thought that they already had COVID.15

For this analysis, people reporting their age as outside the
range 16–90 years or their body mass index (BMI) (kg/m2)
outside the range of 35°C–43°C or (iii) reported feeling unwell
outside 15–55 years were excluded (n=310 462). We further
declared themselves to be pregnant (n=8680).

Smoking and SARS-CoV-2 PCR test outcome using logistic regres-
sion, including age, sex and BMI as covariates. Statistical analyses were carried out using R (V3.6.0). The
Bonferroni-adjusted threshold for statistical significance was
calculated by dividing a conventional alpha value of 0.05 by the
number of tested groups and the effective number of indepen-
dent symptoms, which were estimated based on their pairwise
polychoric correlation coefficients (polychor R package, V0.7-10),
using the method by Li and Ji.17 The resulting p value threshold
was 0.05/4/11=1.14×10−4. Association p values below this
threshold were considered to be statistically significant.

Smoking association with COVID-19 symptom burden in
standard users
We further explored the effect of smoking on COVID-19
severity among standard users (ie, subjects neither tested for
SARS-CoV-2 nor self-reporting prior suspected COVID-19
infection at the time of registration) by examining associa-
tion between smoking status and the development of a greater
symptom burden. We took the number of reported COVID-19
symptoms as a proxy for disease severity, but did not attempt
to weight different symptoms. Tested logistic regression models
first addressed the impact of current smoking on the primary
end point defined as the presence of all the classic triad of symp-
toms (ie, fever, persistent cough and shortness of breath consist-
tent with COVID-19 caseness). In addition, we investigated the
presence of (i) >5 symptoms and (ii) >10 symptoms taking these
as arbitrary categories of increasing disease burden to the indi-
vidual. Age, sex and BMI were included as covariates.

Ethnicity was not included in these analyses, due to its limited
availability in the dataset. The small number of subjects with
missing smoking status (n=1313 (0.05%)) were considered non-
smokers. Sensitivity analyses using the same model but including
ethnicity as a covariate and including missing smoking status as
a category in its own right are reported in online supplemental
tables E1-2. In both analyses, the ORs resulting from the sensi-
tivity analysis were consistent with those from the primary
analysis, thus suggesting that our associations are valid across
different ethnicities and that treating subjects with missing
smoking status as non-smokers does not affect the association
outcome.

Smoking association with hospital attendance
Users of the app who reported that they had symptoms were
asked to report and then to update their location using the
options (i) at home not been to hospital for suspected COVID-19
symptoms, (ii) attending hospital with suspected COVID-19
symptoms, (iii) back from hospital I’d like to tell you about my
treatment, (iv) back from hospital I’ve already told you about
my treatment. We treated hospital attendance as an additional
proxy for COVID-19 symptom severity, and therefore tested

Statistical analysis
The main study outcome was the development of ‘classic’
symptoms of COVID-19 during the pandemic defined as fever,
new persistent cough and breathlessness and their association
with current smoking. The number of concurrent COVID-19
symptoms was used as a proxy for severity and the pattern of
association between symptoms was also compared between
smokers and non-smokers. We first investigated associations
between smoking status and individual COVID-19 symptoms
using logistic regression, including age, sex and BMI as covari-
ates. Statistical analyses were carried out using R (V3.6.0). The
Bonferroni-adjusted threshold for statistical significance was
calculated by dividing a conventional alpha value of 0.05 by the
number of tested groups and the effective number of indepen-
dent symptoms, which were estimated based on their pairwise
polychoric correlation coefficients (polychor R package, V0.7-10),
using the method by Li and Ji.17 The resulting p value threshold
was 0.05/4/11=1.14×10−4. Association p values below this
threshold were considered to be statistically significant.

Smoking association with hospital attendance
We also assessed the association between smoking status and a
positive SARS-CoV-2 PCR test outcome using logistic regres-
sion, including age, sex and BMI as covariates. As a sensitivity
analysis, we controlled for confounding effects of healthcare
working status on smoking association with SARS-CoV-2 posi-
tivity by testing the identical multivariate regression model
including healthcare worker status as a covariate.

Smoking and SARS-CoV-2 PCR result in tested subset
We also assessed the association between smoking status and a
positive SARS-CoV-2 PCR test outcome using logistic regres-
sion, including age, sex and BMI as covariates. As a sensitivity
analysis, we controlled for confounding effects of healthcare
working status on smoking association with SARS-CoV-2 posi-
tivity by testing the identical multivariate regression model
including healthcare worker status as a covariate.

Statistical analysis
The main study outcome was the development of ‘classic’
symptoms of COVID-19 during the pandemic defined as fever,
new persistent cough and breathlessness and their association
with current smoking. The number of concurrent COVID-19
symptoms was used as a proxy for severity and the pattern of
association between symptoms was also compared between
smokers and non-smokers. We first investigated associations
between smoking status and individual COVID-19 symptoms
using logistic regression, including age, sex and BMI as covari-
ates. Statistical analyses were carried out using R (V3.6.0). The
Bonferroni-adjusted threshold for statistical significance was
calculated by dividing a conventional alpha value of 0.05 by the
number of tested groups and the effective number of indepen-
dent symptoms, which were estimated based on their pairwise
polychoric correlation coefficients (polychor R package, V0.7-10),
using the method by Li and Ji.17 The resulting p value threshold
was 0.05/4/11=1.14×10−4. Association p values below this
threshold were considered to be statistically significant.
by logistic regression its association with smoking in individuals testing positive for SARS-CoV-2. Age, sex, BMI and presence of comorbidities (ie, diabetes, lung disease, kidney disease, cancer and heart disease) or comorbidity-related treatments (ie, blood pressure medications and immunosuppressants) were included as covariates in the model.

**RESULTS**

Overall, 2 401 982 UK residents, a majority female (63.3%), entered data via the app between 24 March 2020 and 23 April 2020 and were eligible for analysis. Median (IQR) per-user number of login days was 5 (2–14), and median number of days between first and last engagement with the app was 14 (1–26) (online supplemental figure E1). During this period, 834 437 individuals (35%) reported that they ‘did not feel physically normal’ and entered at least one COVID-19-related symptom. We classified the 2 401 982 participants into four groups (table 1; figure 1): (i) tested SARS-CoV-2 positive (SC2P; n=7123); (ii) tested SARS-CoV-2 negative (SC2N; n=16 759); (iii) subjects self-reporting COVID-19 (SC2S; n=157 406) who believed, based on symptoms, that they already had COVID-19 at the time of registration; (iv) the remaining 2 221 088 were defined as ‘standard users’. The distribution of symptoms in these four categories is presented in figure 2.

Individuals reporting hospital attendance reported experiencing a higher number of co-occurring symptoms than those who did not attend hospital (effect size (ES)=2.53; SE=0.02; p<2.22×10⁻11).

### Table 1: Characteristics of study subjects by category

<table>
<thead>
<tr>
<th>Demographics</th>
<th>All users</th>
<th>Self-reported COVID-19</th>
<th>SARS-CoV-2 tested</th>
<th>P value</th>
<th>Standard users</th>
</tr>
</thead>
<tbody>
<tr>
<td>N (%)</td>
<td>2 401 982</td>
<td>157 406 (6.5)</td>
<td>26 918 (1.1)</td>
<td>16 759 (0.7)</td>
<td>7123 (0.3)</td>
</tr>
<tr>
<td>Sex (% F)</td>
<td>63.3</td>
<td>64.39</td>
<td>68.81</td>
<td>68.41</td>
<td>67.82</td>
</tr>
<tr>
<td>Age (years)</td>
<td>43.63±15.01 (32–55)</td>
<td>42.5±13.26 (32–52)</td>
<td>42.35±13.23 (32–52)</td>
<td>43.2±13.51 (33–53)</td>
<td>41.1±12.69 (31–50)</td>
</tr>
<tr>
<td>Healthcare worker (%)</td>
<td>83 120 (3.46)</td>
<td>10 533 (6.69)</td>
<td>5385 (20.01)</td>
<td>2391 (14.27)</td>
<td>2218 (13.14)</td>
</tr>
<tr>
<td>Ethnicity (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>6014 (2.1)</td>
<td>395 (3.02)</td>
<td>120 (4.72)</td>
<td>51 (3.63)</td>
<td>50 (3.69)</td>
</tr>
<tr>
<td>Black</td>
<td>1700 (0.59)</td>
<td>123 (0.94)</td>
<td>27 (1.06)</td>
<td>15 (1.07)</td>
<td>10 (1.28)</td>
</tr>
<tr>
<td>Chinese</td>
<td>1040 (0.36)</td>
<td>52 (0.4)</td>
<td>16 (0.63)</td>
<td>11 (0.78)</td>
<td>4 (0.51)</td>
</tr>
<tr>
<td>Middle East</td>
<td>1219 (0.43)</td>
<td>81 (0.62)</td>
<td>16 (0.63)</td>
<td>6 (0.43)</td>
<td>7 (0.89)</td>
</tr>
<tr>
<td>Mixed</td>
<td>4260 (1.49)</td>
<td>272 (2.08)</td>
<td>28 (1.10)</td>
<td>14 (1.00)</td>
<td>5 (0.64)</td>
</tr>
<tr>
<td>White</td>
<td>272 537 (95.04)</td>
<td>12 142 (92.94)</td>
<td>2335 (91.86)</td>
<td>1309 (93.10)</td>
<td>707 (90.29)</td>
</tr>
<tr>
<td>N/A</td>
<td>2 115 212 (88.06)</td>
<td>144 341 (91.70)</td>
<td>24 376 (90.56)</td>
<td>15 353 (91.61)</td>
<td>6 340 (89.01)</td>
</tr>
</tbody>
</table>

### Smoking prevalence (% all; % F; % M)*

| Overall | 240 872; 146 010; 94 697 (10.96; 10.35; 12.02) | | | | |
| Age (years) | | | | | |
| 16–24 | 30 510; 20 167; 10 313 (13.79; 13.02; 15.55) | | | | |
| 25–44 | 135 199; 82 509; 52 594 (13.11; 12.07; 15.14) | | | | |
| 45–64 | 66 741; 39 238; 27 504 (8.96; 8.46; 9.49) | | | | |
| >65 | 8423; 4096; 4322 (4.21; 3.78; 4.71) | | | | |

### Comorbidities (%)

| Diabetes | 2.84 | 25 | 2.95 | 2.76 | 3.26 | 2.86 |
| Heart disease | 2.03 | 1.68 | 2.11 | 2.33 | 1.69 | 2.06 |
| Lung disease | 12.2 | 13.38 | 14.69 | 15.01 | 13.38 | 12.09 |
| Kidney disease | 0.59 | 0.57 | 0.9 | 0.95 | 0.83 | 0.59 |
| Cancer | 1.21 | 0.67 | 1.75 | 2.29 | 0.94 | 1.26 |

### Hospital attendance

| N (%) | 12 093 (1.46) | 2746 (2.84) | 2725 (13.49) | 1250 (12.16) | 1148 (18.35) | 7208 (1.00) |

*Smoking prevalence in the UK (adult smoking habits in the UK: 2018, Office for National Statistics). Overall: 14.7% (16.5% of men; 13% of women). By age group: 18–24: 16.8%; 25–34: 19.2%; 35–44: 16.4%; 45–54: 16.8%; 55–64: 14.5%; 65+: 7.9. Age and BMI are given as means±SD and first, third IQR (in brackets). Standard users refers to individuals who did not believe that they already had COVID-19 when registering with the app and had not been tested for SARS-CoV-2. Significant differences in study variables between tested positive and tested negative subjects are shown (P). Categorical variables (ie, gender, healthcare worker and smoking) were compared using Pearson’s χ² test, while age was compared using Wilcoxon’s test. Association of p-values with BMI is from linear regression adjusting for age and sex. The ‘self-reported COVID-19’ group includes 1395 and 1767 individuals who reported having tested positive and negative for SARS-CoV-2, respectively. BMI, body mass index.
Among standard users, current smokers were more likely to develop the classic triad of symptoms suggesting a diagnosis of COVID-19 (ie, fever, persistent cough and shortness of breath; OR (95% CI)=1.14 (1.10 to 1.18); p=5.81×10^{-13}) and to meet a higher threshold of symptom burden, than non-smokers (>5 symptoms OR=1.29 (1.26 to 1.31); p=8.00×10^{-187}; >10 symptoms OR=1.50 (1.42 to 1.58); p=3.46×10^{-48}; figure 3).

Smoking was associated with increased overall COVID-19 symptom burden in all categories (standard users $\beta=2.93$, SE=0.07, p<2.22×10^{-16}; SC2P $\beta=3.28$, SE=1.10, p=2.79×10^{-3}; SC2N $\beta=1.95$, SE=0.78, p=1.22×10^{-2}; figure 4). Smoking was also associated with higher symptom burden in SC2S and, more importantly, SC2P (OR (95% CI) of presenting with >10 symptoms 1.33 (1.22 to 1.45); p=1.91×10^{-10} and 1.42 (1.09 to 1.83); p=7.13×10^{-3}, respectively).

Although smokers had more symptoms than non-smokers, we did not observe any specific pattern of co-occurring symptoms distinguishing smokers from non-smokers, with symptom correlation matrices showing significant intercorrelation in all four user categories (lowest Mantel’s test $r^2=0.54; p=0.001; p<2.22×10^{-16}; SC2P \neq 3.28, SE=1.10, p=2.79×10^{-3}; SC2N \neq 1.95, SE=0.78, p=1.22×10^{-2};$ figure 4).

**Figure 1** Flow diagram of the progress through phases of the analysis. BMI, body mass index; SC2N, SARS-CoV-2 negative; SC2P, SARS-CoV-2 positive; SC2S, self-reporting COVID-19.

**Figure 2** COVID-19 symptom distribution. Bar plot showing the proportion of individuals reporting COVID-19 symptoms within user categories. All symptoms except delirium and sore throat were more frequently observed among subjects who reported testing positive for SARS-CoV-2 ($\chi^2 p=1.27×10^{-6}$). The ‘self-reported COVID-19’ group includes 1295 and 1767 individuals who reported having tested positive and negative for SARS-CoV-2, respectively.
Smoking was strongly associated with having most COVID-19 symptoms among standard users and SC2S (table 2).

Healthcare workers were defined as those who responded positively that they worked in a hospital, with the elderly or in a community care setting. The association between smoking and increased symptom burden remained significant in all groups but SC2N when healthcare workers were excluded from the analysis (standard users ES=2.96, SE=0.07, p<2.22×10−16; SC2S ES=2.97, SE=0.22, p<2.22×10−16; SC2P ES=3.96, SE=1.34, p=3.05×10−3; SC2N ES=1.43, SE=0.85, p=0.09; online supplemental figure E6).

Impact of smoking in the SARS-CoV-2 tested subgroup

The group who reported that they had undergone testing at this relatively early stage in the pandemic (n=26,918 10.1% of app users) differed in a number of ways from the general study population and testing was non-random so their data cannot be used to make inferences about the risk to the general population (table 1). Twenty per cent of those tested were healthcare workers (HCW) compared with 3.4% in the whole study population and 31% of those testing positive were HCW. Rates of comorbidities were higher in tested individuals compared with

Figure 3  Effect of current smoking on risk of presenting with COVID-19 symptoms. The plot shows the OR (95% CI) for smokers from the standard users category of presenting with COVID-19 classic symptoms (ie, all three of cough, fever and breathlessness) or a higher symptom burden (ie, >5 or >10 symptoms).

Figure 4  Effect of current smoking on COVID-19 symptom burden. For each user category, the bar plot shows the relative abundance of individuals by total number of COVID-19 symptoms reported, among non-smokers (grey) and smokers (black). We report the regression effect size, along with its SE and p value when testing the association between smoking status and COVID-19 symptom burden. Current smoking was associated with increased overall symptom burden in all categories. NA, not available.

the general population. Smoking rates among tested individuals were lower compared with all users of the app (8.9% vs 11.0%).

The average number of symptoms reported by subjects who reported being tested positive for SARS-CoV-2 (mean (IQR) = 6.16 (4–9)), was markedly higher than SC2N users (3.22 (2–6); Wilcoxon’s test p < 2.22 × 10−6), and more importantly, than standard users (1.12 (0–1); p < 2.2 × 10−16; online supplemental figure E3). The incidence of all COVID-19-related symptoms, with the exception of delirium and sore throat, was higher in the SC2P group as compared with the other categories (χ² test p < 1.27 × 10−8; figure 2). Of note, among subjects who tested positive for SARS-CoV-2, current smokers were more likely to develop delirium (OR (95% CI) = 1.83 (1.47 to 2.27); p = 4.38 × 10−3) and abdominal pain (OR = 1.43 (1.15 to 1.77); p = 1.07 × 10−3), compared with non-smokers.

Smoking rates were lower in those who had tested positive for SARS-CoV-2 compared with those testing negative (7.4%) vs 9.3%) as were rates of cancer, heart disease and lung disease. In the tested subpopulation, smoking was negatively associated with the risk of having a positive PCR for SARS-CoV-2 infection (OR (95% CI) = 0.73 (0.65 to 0.81); p = 9.49 × 10−5). This association remained significant when accounting for healthcare working status by including it as covariate in the model (OR (95% CI) = 0.70 (0.63–0.79); p = 6.82 × 10−10), suggesting that the association between smoking status and SARS-CoV-2 test outcome is not confounded by the higher number of healthcare workers being tested for SARS-CoV-2 nor the lower smoking prevalence among healthcare workers.

Despite this apparently protective effect of smoking in the tested group, the symptom burden was higher in smokers who tested positive for SARS-CoV-2 as compared with non-smokers. Smoking was associated with higher symptom burden in those testing positive (OR (95% CI) of presenting with >10 symptoms 1.42 (1.09 to 1.83); p = 7.13 × 10−4). In addition, current smokers who tested positive had a higher risk of attending hospital due to COVID-19 compared with non-smokers (OR 2.11 (1.41 to 3.11); p = 1.99 × 10−5). This association remained significant when participants reporting the presence of comorbidities were removed from the analysis (OR 1.87 (1.15 to 2.95); p = 9.02 × 10−5) or when accounting for healthcare working status by including it as covariate in the model (OR 2.08 (1.37 to 3.12); p = 4.76 × 10−5).

### DISCUSSION

The main finding from this large prospective population cohort is that current smoking was associated with a substantially increased risk of developing self-reported symptoms suggestive of COVID-19, as well as a greater symptom burden, indicating an impact of smoking on disease severity. Among those who reported being tested positive for SARS-CoV-2, smokers had a higher symptom burden and were more likely to need to attend hospital, compared with non-smokers.

### Significance of findings

The COVID-19 pandemic continues to threaten to overwhelm healthcare systems and measures to contain it impose significant burdens on populations.18-19 There is a need for strategies to mitigate this, which include preventing or delaying transmission of the virus to ‘flatten the curve’, as well as any measures that might reduce the severity of the condition. In addition, steps should be taken to increase health system capacity and reduce health system demand from other sources. The present data suggest that smoking cessation should be considered as an element in strategies to address COVID-19, as smoking increases both the likelihood of symptomatic disease defined according to the presence of ‘classic’ symptoms of fever, cough and breathlessness and the severity of disease, defined based on the number of symptoms. Combined with this, a reduction in population smoking rates is likely to reduce the health system burden from other conditions that require hospitalisation such as acute vascular events and exacerbations of lung disease20 as well as improving resilience by reducing overall sickness absence among key workers.20

It has been reported that smoking is associated with upregulation of ACE2,\(^{21-23}\) the receptor for the SARS-CoV-2 virus\(^ {24}\) in the lung, although a recent meta-analysis suggests divergent effects with upregulation in epithelial cells and down-regulation in alveolar type 2 cells.\(^ {25}\) The situation is further complicated by the possibility that internalisation of ACE2 due to viral infection leads to unopposed ACE inhibitor activity and high angiotensin II levels, contributing to endothelial damage and the coagulopathy and microthrombosis seen in severe COVID-19.\(^ {1, 2}\) Smoking itself causes vascular endothelial damage,\(^ {26}\) a prominent feature in the pathophysiology of severe COVID-19.

In addition to an increase in individual susceptibility to developing COVID-19 following viral infection, smoking might also be expected to increase the risk of infection because of the repeated hand to mouth movements that smoking entails. The increase in severity that we found associated with smoking suggests that the whole curve of disease is shifted as opposed to just the chance of infection. Importantly, although the number and degree of symptoms was higher in smokers the pattern of symptoms did not differ importantly, indicating that the phenomenon that we observed was not caused by the presence of symptoms related directly to smoking itself.

Data were also available for a limited number of individuals who reported the results of swab tests performed to identify the presence of SARS-CoV-2. Only 1% of the study population had undergone testing and this group was heavily selected because of testing policies which focused on healthcare workers and others interacting with healthcare, in particular, patients tested who may have been attending healthcare settings for other non-COVID-19-related conditions. As numerous health conditions are smoking-related, this would tend to increase the exposure of smokers without COVID-19 to testing. As a consequence, extrapolation from this subgroup to population risk is inappropriate. Smoking prevalence was lower in the tested group than in the whole study cohort, 8.9% vs 11.0%, both values substantially below the UK population average and smoking prevalence was lower in those testing positive than those testing negative (7.4% vs 9.3%). This raises several possibilities. First, it may be that smoking does reduce susceptibility to infection with SARS-CoV-2, although the likelihood of developing symptoms and severity of disease in current smokers remains higher, according both to the whole population (2.4 million app users) data and the fact that smokers testing positive were more likely to report >10 symptoms and to need to have attended hospital. Of note, the OPEN-Safely study which found that in a large general practice population, adjusted for age and sex, current smokers had an HR of 1.14 (95% CI 1.05 to1.23) for COVID-19-related death compared with never smokers.\(^ {7}\) Alternatively, this finding may represent sampling bias. Among those tested, a negative PCR result was also associated with a higher rate of heart disease, cancer and respiratory disease, all factors associated with worse outcomes in people with COVID-19 but unlikely to be protective against infection. The use of swab testing for PCR diagnosis of viral infection has limitations with a significant false negative rate. It is possible that non-random differences in the way that tests were conducted (eg, nasopharyngeal or oropharyngeal swabs) or the temporal relation of testing to the development of symptoms (eg, likely more rapid for healthcare workers) could have influenced results.\(^ {27, 28}\) In due course, the question of relative infectivity may be answered based either on surveillance screening or widespread antibody testing.

**Methodological issues**

The triad of ‘classic’ symptoms we used as the primary end point was based on Public Health England guidance. We chose the presence of 5 or 10 symptoms arbitrarily to illustrate the ‘burden’ of COVID-19 on the basis that a greater number of symptoms was likely to represent more severe illness. All symptoms were given equal weight. Attending hospital in the context of potential COVID-19 was taken as a marker of disease severity. Of note, the questions were framed in such a way that hospital attendance with symptoms was clearly distinguished from being at a hospital because of work.

The study uses the presence of self-reported symptoms to define the occurrence and severity of COVID-19. Although it is possible that some of these represented conditions other than COVID-19, we have previously reported concordance between reported symptoms SARS-CoV-2 test results.\(^ {19}\) Additionally, we observe in this larger dataset an increased symptom burden among tested positive users compared with both tested negative and untested subjects. Importantly, the pattern of correlation between symptoms did not differ between smokers and non-smokers suggesting that the increased symptom burden was not directly related to pre-existing smoking-related symptoms.

Behavioural factors or individual prior expectations and beliefs may have influenced reporting. Thus smokers may have felt themselves to be at a higher risk of developing COVID-19 making them more likely to report symptoms, thus exaggerating the apparent impact of smoking. Alternatively, non-smokers being in general healthier, might be expected to be more sensitive to changes in their well-being making them more likely to report symptoms, thus underestimating the impact of smoking.

Participants in the study were self-selected, having chosen to register to use the app. In addition, the app was only available on a smartphone raising issues of digital access and digital literacy, so caution is needed about extrapolation to the entire population, in particular older or poorer sections of the population. The overall self-reported smoking rate among app users was 11.0% and lower in all age and sex strata compared with the general UK population. This may be due to sampling bias, reflecting the demographics of this sample, as smoking rates are lower in wealthier groups\(^ {29}\) who may also have been more likely to be able to make use of a smartphone app. Smokers tended to report via the app for a shorter period, with fewer days between first and last engagement (online supplemental figure E4), which may have led to an underestimate of the effect of smoking on the development of symptoms. In addition, there were relatively fewer people in the older age groups who are most susceptible to COVID-19. Smoking status, being based on contemporary self-report by participants is likely to have been captured more accurately by the app used in the present study than in studies based on routine, historical administrative records and those where there is substantial missing data.

The COVID-19 pandemic has exacerbated socioeconomic inequalities and ethnicity is emerging as a crucial area in its development.\(^ {7}\) Unfortunately, ethnicity was not included in the original set of app registration data. This omission was addressed so that this is now being collected, but unfortunately for the period covered by this analysis it was only available for 12% of individuals. Although in the sensitivity analysis in this subgroup our results seem to hold across different ethnicities, further experiments would be needed to better characterise any possible interaction between smoking and symptoms in different ethnic groups. The availability of more detailed socioeconomic data as well as more detailed occupational history would have allowed

---


\(^ {2}\) Data were also available for a limited number of individuals who reported the results of swab tests performed to identify the presence of SARS-CoV-2. Only 1% of the study population had undergone testing and this group was heavily selected because of testing policies which focused on healthcare workers and others interacting with healthcare, in particular, patients tested who may have been attending healthcare settings for other non-COVID-19-related conditions. As numerous health conditions are smoking-related, this would tend to increase the exposure of smokers without COVID-19 to testing. As a consequence, extrapolation from this subgroup to population risk is inappropriate. Smoking prevalence was lower in the tested group than in the whole study cohort, 8.9% vs 11.0%, both values substantially below the UK population average and smoking prevalence was lower in those testing positive than those testing negative (7.4% vs 9.3%). This raises several possibilities. First, it may be that smoking does reduce susceptibility to infection with SARS-CoV-2, although the likelihood of developing symptoms and severity of disease in current smokers remains higher, according both to the whole population (2.4 million app users) data and the fact that smokers testing positive were more likely to report >10 symptoms and to need to have attended hospital. Of note, the OPEN-Safely study which found that in a large general practice population, adjusted for age and sex, current smokers had an HR of 1.14 (95% CI 1.05 to1.23) for COVID-19-related death compared with never smokers.\(^ {7}\) Alternatively, this finding may represent sampling bias. Among those tested, a negative PCR result was also associated with a higher rate of heart disease, cancer and respiratory disease, all factors associated with worse outcomes in people with COVID-19 but unlikely to be protective against infection. The use of swab testing for PCR diagnosis of viral infection has limitations with a significant false negative rate. It is possible that non-random differences in the way that tests were conducted (eg, nasopharyngeal or oropharyngeal swabs) or the temporal relation of testing to the development of symptoms (eg, likely more rapid for healthcare workers) could have influenced results.\(^ {27, 28}\) In due course, the question of relative infectivity may be answered based either on surveillance screening or widespread antibody testing.

**Methodological issues**

The triad of ‘classic’ symptoms we used as the primary end point was based on Public Health England guidance. We chose the presence of 5 or 10 symptoms arbitrarily to illustrate the ‘burden’ of COVID-19 on the basis that a greater number of symptoms was likely to represent more severe illness. All symptoms were given equal weight. Attending hospital in the context of potential COVID-19 was taken as a marker of disease severity. Of note, the questions were framed in such a way that hospital attendance with symptoms was clearly distinguished from being at a hospital because of work.

The study uses the presence of self-reported symptoms to define the occurrence and severity of COVID-19. Although it is possible that some of these represented conditions other than COVID-19, we have previously reported concordance between reported symptoms SARS-CoV-2 test results.\(^ {19}\) Additionally, we observe in this larger dataset an increased symptom burden among tested positive users compared with both tested negative and untested subjects. Importantly, the pattern of correlation between symptoms did not differ between smokers and non-smokers suggesting that the increased symptom burden was not directly related to pre-existing smoking-related symptoms.

Behavioural factors or individual prior expectations and beliefs may have influenced reporting. Thus smokers may have felt themselves to be at a higher risk of developing COVID-19 making them more likely to report symptoms, thus exaggerating the apparent impact of smoking. Alternatively, non-smokers being in general healthier, might be expected to be more sensitive to changes in their well-being making them more likely to report symptoms, thus underestimating the impact of smoking.

Participants in the study were self-selected, having chosen to register to use the app. In addition, the app was only available on a smartphone raising issues of digital access and digital literacy, so caution is needed about extrapolation to the entire population, in particular older or poorer sections of the population. The overall self-reported smoking rate among app users was 11.0% and lower in all age and sex strata compared with the general UK population. This may be due to sampling bias, reflecting the demographics of this sample, as smoking rates are lower in wealthier groups\(^ {29}\) who may also have been more likely to be able to make use of a smartphone app. Smokers tended to report via the app for a shorter period, with fewer days between first and last engagement (online supplemental figure E4), which may have led to an underestimate of the effect of smoking on the development of symptoms. In addition, there were relatively fewer people in the older age groups who are most susceptible to COVID-19. Smoking status, being based on contemporary self-report by participants is likely to have been captured more accurately by the app used in the present study than in studies based on routine, historical administrative records and those where there is substantial missing data.

The COVID-19 pandemic has exacerbated socioeconomic inequalities and ethnicity is emerging as a crucial area in its development.\(^ {7}\) Unfortunately, ethnicity was not included in the original set of app registration data. This omission was addressed so that this is now being collected, but unfortunately for the period covered by this analysis it was only available for 12% of individuals. Although in the sensitivity analysis in this subgroup our results seem to hold across different ethnicities, further experiments would be needed to better characterise any possible interaction between smoking and symptoms in different ethnic groups. The availability of more detailed socioeconomic data as well as more detailed occupational history would have allowed

---

further refinement of the effect of smoking on symptoms, particularly as smoking is one of the most important drivers of socio-economic gradients in life expectancy.10

We took the diagnosis of COVID-19 to require an individual to have symptoms. The data from the app do not, by definition, provide information about asymptomatic viral carriage which is acknowledged to be of considerable importance in the spread of SARS-CoV-2.11

Conclusion
Evidence that current smoking appears to increase individual and thus health system burden from COVID-19 is a strong argument for governments to accelerate rather than pause measures to deliver on tobacco control plans.12 13 These should include options such as polluter-pays levies on tobacco transnationals as they will continue, through the pandemic, to be highly profitable as well as enormously damaging to global health.14

Our results provide compelling evidence for an association between current smoking and individual risk from COVID-19, including self-reported symptom burden and risk of attending hospital. Smoking cessation should be incorporated into public health campaigns and other efforts to address the COVID-19 pandemic.

Author affiliations
1National Heart and Lung Institute, Imperial College London, London, UK
2The Department of Twin Research & Genetic Epidemiology, King’s College London, London, UK
3Department Primary Care and Public Health, Imperial College, London, UK
4Respiratory Epidemiology, Occupational Medicine and Public Health, Imperial College London, London, UK
5Department of Twin Research, King’s College London, London, UK

7Biomedical Engineering & Imaging Sciences, King’s College London, London, UK
8Zoe Global Limited, London, UK

Correction notice This article has been corrected since it was published Online First. Figures 2 and 4 were interchanged and figure legends for figures 2 and 3 were also listed with the incorrect figure.

Twitter Nicholas S Hopkinson @COPDdoc

Acknowledgements The authors would like to thank members of the public for taking the time to enter their information into the Zoe COVID Symptom Study app in order to improve understanding of and response to the COVID-19 pandemic.

Contributors NSH, CIS and MF conceived the analysis which was developed with input from all authors. Symptom data cleaning and analyses were performed by BM, MM, SO, NR and MF. NSH, NR and MF produced the first draft of the paper to which all authors contributed. All authors have reviewed and approved the final version. MF is the guarantor.

Funding This work was supported by Zoe Global Limited as well as grants from the Wellcome Trust (212904/2/18/Z) and the Medical Research Council (MRC)/British Heart Foundation Ancestry and Biological Informative Markers for Stratification of Hypertension (AIMHY; MR/M016560/1). TwinsUK is funded by the Wellcome Trust Medical Research Council, European Union, Chronic Disease Research Foundation (CDRF), Zoe Global Ltd and the National Institute for Health Research (NIHR)-funded BioResource, Clinical Research Facility and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. JSES, TDS and KSS acknowledge support from the Medical Research Council (MR/M004422/1). This work is supported by BREATHE - The Health Data Research Hub for Respiratory Health [MC_PC_190004]. BREATHE is funded through the UK Research and Innovation Industrial Strategy Challenge Fund and delivered through Health Data Research UK.

Competing interests NSH is Chair of Action on Smoking and Health and Medical Director of the British Lung Foundation. TDS is a consultant to Zoe Global Ltd (Zoe), who developed the app. JW and RD are employees of Zoe Global Ltd.

Patient consent for publication Not required.

Ethics approval The study has been approved by the King’s College London Research Ethics Committee REMAS-ID 18210, review reference LRS-19/20-18210 and all subscribers provided informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data may be obtained from a third party and are not publicly available. Anonymised research data will be shared with third parties via the centre for Health Data Research UK (HDRUK.ac.uk). Data updates can be found on https://COVID.joinzoe.com.

This article is made freely available for use in accordance with BMJ’s website terms and conditions for the duration of the covid-19 pandemic or until otherwise determined by BMJ. You may use, download and print the article for any lawful, non-commercial purpose (including text and data mining) provided that all copyright notices and trade marks are retained.

ORCID iDs
Nicholas S Hopkinson http://orcid.org/0000-0003-3235-0454
Anthony A Lavery http://orcid.org/0000-0003-1318-8439
Jennifer K Quint http://orcid.org/0000-0003-0149-4869
Richard Davies http://orcid.org/0000-0003-2050-3994

REFERENCES
22 Smith JC, Sheltzer JM. Cigarette smoke triggers the expansion of a subpopulation of respiratory epithelial cells that express the SARS-CoV-2 receptor ACE2. bioRxiv 2020.


32 Hopkinson NS. The path to a smoke-free England by 2030. BMJ 2020;368:m518.
