

1 **The association of smoking status with hospitalisation for COVID-19 compared with other**  
2 **respiratory viruses a year previous: A case-control study at a single UK National Health**  
3 **Service trust**

4 David Simons\*<sup>1,2</sup>, Olga Perski\*<sup>3</sup>, Lion Shahab<sup>3</sup>, Jamie Brown<sup>3</sup> and Robin Bailey<sup>2,4</sup>

5 <sup>1</sup> Centre for Emerging, Endemic and Exotic Diseases, RVC, UK

6 <sup>2</sup> Hospital for Tropical Diseases, London, UK

7 <sup>3</sup> Department of Behavioural Science and Health, University College London, UK

8 <sup>4</sup> London School of Hygiene and Tropical Medicine, UK

9 \*These authors contributed equally to this work

10 Abstract word count: 250

11 Manuscript word count: 3317

12 Key words: tobacco, smoking, COVID-19, case-control study, hospitalisation

13

14 **Abstract**

15

16 **Background:** It is unclear whether smoking increases the risk of COVID-19 hospitalisation. We  
17 aimed to i) examine the association of smoking status with hospitalisation for COVID-19 compared  
18 with hospitalisation for other respiratory virus infections a year previous; ii) compare current  
19 smoking in cases with age- and sex-matched London prevalence; and iii) examine concordance  
20 between smoking status recorded on the electronic health record (EHR) and the medical notes.

21

22 **Methods:** This retrospective case-control study enrolled adult patients (446 cases and 211  
23 controls) at a single National Health Service trust in London, UK. London smoking prevalence was  
24 obtained from the representative Annual Population Survey. The outcome variable was type of  
25 hospitalisation (COVID-19 vs. another respiratory virus). The exposure variable was smoking  
26 status (never/former/current smoker). Logistic regression analyses adjusted for age, sex,

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

27 socioeconomic position and comorbidities were performed. The study protocol and analyses were  
28 pre-registered on the [Open Science Framework](#).

29

30 **Findings:** Patients hospitalised with COVID-19 had lower odds of being current smokers than  
31 patients admitted with other respiratory viruses (OR<sub>adj</sub>=0.55, 95% CI=0.31-0.96,  $p=.04$ ). Odds were  
32 equivocal for former smokers (OR<sub>adj</sub>=1.08, 95% CI=0.72-1.65,  $p=.70$ ). Current smoking in cases  
33 was significantly lower than expected from London prevalence (9.4% vs. 12.9%,  $p=.02$ ). Smoking  
34 status recorded on the EHR deviated significantly from that recorded within the medical notes  
35 ( $\chi^2(3)=226.7$ ,  $p<.001$ ).

36

37 **Interpretation:** In a single hospital trust in the UK, patients hospitalised with COVID-19 had  
38 reduced odds of being current smokers compared with patients admitted with other respiratory  
39 viruses a year previous.

40 **Funding:** UK BBSRC, Cancer Research UK, UKPRP.

41

## 42 INTRODUCTION

43

44 COVID-19 is a respiratory disease caused by the SARS-CoV-2 virus. There are in excess of 50  
45 million confirmed COVID-19 cases globally, with over one million deaths reported[1]. Large age  
46 and sex differences in case severity and mortality have been observed[2], with hypertension,  
47 diabetes and obesity identified as important risk factors[3]. There are *a priori* reasons to believe  
48 that current smokers are at increased risk of contracting COVID-19 and experiencing greater  
49 disease severity once infected. SARS-CoV-2 enters epithelial cells through the ACE-2 receptor[4].  
50 Evidence suggests that gene expression and subsequent ACE-2 receptor levels are elevated in  
51 the airway and oral epithelium of current smokers[5, 6], potentially making smokers vulnerable to  
52 contracting SARS-CoV-2. The regular hand-to-mouth movements involved in cigarette smoking  
53 may also increase SARS-CoV-2 transmission. Other studies, however, show that smoking  
54 downregulates the ACE-2 receptor[7]. These uncertainties notwithstanding, both former and

55 current smoking increase the risk of other respiratory viral[8] and bacterial[9] infections and are  
56 associated with worse outcomes once infected. For example, a large observational study in the UK  
57 found that former and current smokers had increased odds of being hospitalised with community-  
58 acquired pneumonia compared with never smokers[10]. However, early data from the ongoing  
59 pandemic have not provided clear evidence for an independent association of smoking status with  
60 COVID-19 outcomes[11, 12].

61

62 A living evidence review of observational and experimental studies examining the association of  
63 smoking status with COVID-19 infection, hospitalisation, disease severity and mortality has found  
64 that, among 279 studies, current smoking (not adjusted for age, sex or comorbidities) was  
65 generally lower than what would be expected from national smoking prevalence[12]. In addition,  
66 current smokers were at reduced risk of testing positive for SARS-CoV-2 compared with never  
67 smokers, and former smokers were at increased risk of hospitalisation, disease severity and in-  
68 hospital mortality compared with never smokers. However, the majority of included studies were  
69 limited by the lack of appropriate controls, poor recording of smoking status and insufficient  
70 adjustment for relevant covariates. Many studies relied on routine electronic health records (EHRs)  
71 to obtain individual-level data on demographic characteristics, comorbidities and smoking status.  
72 Previous research suggests that data on smoking status obtained via EHRs tend to be incomplete  
73 or outright inaccurate, with implausible longitudinal changes observed[13]. In addition, as  
74 hospitalised populations differ by age and sex from the general population, comparisons of current  
75 and former smoking prevalence in hospitalised and non-hospitalised populations are likely limited.  
76 There is hence an urgent need for alternative study designs with relevant comparator groups (e.g.  
77 populations hospitalised with a similar type of respiratory viral infection) and adjustment for  
78 covariates to better understand the association of smoking status with COVID-19 disease  
79 outcomes.

80

81 We therefore aimed to i) examine, in a retrospective case-control study conducted at a single UK  
82 hospital trust, the association of smoking status with hospitalisation for COVID-19 compared with  
83 hospitalisation for other respiratory virus infections (e.g. influenza, respiratory syncytial virus) a

84 year previous; ii) compare current smoking prevalence in the case population with age- and sex-  
85 matched London prevalence, with estimates obtained from the nationally representative Annual  
86 Population Survey; and iii) examine whether there is discordance between smoking status  
87 recorded on the summary EHR and within the contemporaneous medical notes.

## 88 **METHODS**

89

### 90 **Study design**

91

92 This was an observational, retrospective, case-control study performed at a single National Health  
93 Service (NHS) hospital trust (comprising two hospital sites) in London, UK. The study protocol and  
94 analysis plan were pre-registered on the Open Science Framework (<https://bit.ly/3kFYh06>). The  
95 pre-registered protocol stipulated a non-inferiority design (i.e. a one-tailed statistical test) to  
96 maximise statistical power to detect a significantly lower proportion of current smokers among  
97 patients hospitalised with COVID-19 compared with patients hospitalised with another respiratory  
98 viral infection a year previous. The protocol was amended after data collection but prior to  
99 statistical analysis (<https://osf.io/ezfqs/>) to implement a traditional case-control design (i.e. a two-  
100 tailed statistical test), as a delay in study approval meant that the number of eligible cases and  
101 controls exceeded our expectations. Hence, the obtained sample size was deemed sufficient for a  
102 two-tailed test. A completed STROBE checklist is available in the additional file. This amended  
103 study was designed to test the hypothesis that current smoking is underrepresented in patients  
104 hospitalised with COVID-19 compared with those hospitalised with other respiratory viruses a year  
105 previous. Ethical approval was obtained from a local research committee and the NHS Health  
106 Research Authority (IRAS\_282704). The requirement for informed consent was waived due to the  
107 nature of the study.

108

109 For comparison, aggregated data on current, former and never smoking prevalence for London  
110 was obtained from the Office for National Statistics (ONS) Annual Population Survey[14].

111

112 A sample size calculation, updated after data collection but prior to data analysis  
113 (<https://osf.io/ezfqs/>), indicated that 363 cases and 109 controls would provide 80% power to  
114 detect a 10% reduction in current smoking prevalence in cases compared with controls (i.e. 10% in  
115 cases and 20% in controls) with alpha set to 5%. We included all cases from 1<sup>st</sup> March 2020 to the  
116 26<sup>th</sup> August 2020 (the date on which data were obtained) and all controls from the 1<sup>st</sup> January 2019  
117 and the 31<sup>st</sup> December 2019.

118

## 119 **Eligibility criteria**

120

### 121 *Inclusion criteria*

#### 122 • Cases

- 123 1. Consecutive patients admitted to an adult hospital ward (i.e. 18+ years) between 1<sup>st</sup>  
124 March 2020 and 26<sup>th</sup> August 2020 (the date on which data were obtained);
- 125 2. Diagnosis of COVID-19 on or within 5 days of hospital admission, identified via  
126 associated International Classification of Diseases version 10 (ICD-10) codes[15].  
127 This temporal boundary was set to prevent inclusion of patents with nosocomial  
128 (hospital-acquired) infection and allowed for a delay of 3 days in requesting a  
129 COVID-19 test and 2 days for receiving and reporting the results on the EHR. The  
130 median incubation time for COVID-19 is estimated at 5.1 days (95% CI = 4.5-  
131 5.8)[16]. We sought to exclude individuals with nosocomial COVID-19 infection as  
132 they are a different population (e.g. older, more frail) compared with those infected  
133 in the community and subsequently requiring hospitalisation.

134

#### 135 • Controls

- 136 1. Consecutive patients admitted to an adult hospital ward (i.e. 18+ years) between 1<sup>st</sup>  
137 January 2019 and 31<sup>st</sup> December 2019;
- 138 2. Diagnosis of a viral respiratory infection (e.g. influenza, parainfluenza) on or within 5  
139 days of admission, identified via ICD-10 codes. Patients hospitalised with

140 respiratory viruses other than COVID-19 were deemed appropriate controls due to  
141 the similar mechanism for transmission (i.e. respiratory droplets and aerosols)[17].  
142 In addition, risk factors for hospitalisation with other respiratory viruses are similar to  
143 those for hospitalisation with COVID-19 (e.g. age, comorbidities)[18, 19].

#### 144 *Exclusion criteria*

- 145 • Cases and controls
- 146 1. No record of smoking status on the summary EHR or within the medical notes;
- 147 2. A primary diagnosis of infectious exacerbation of COPD due to the strong causal  
148 association of COPD with current and former smoking.

#### 149 **Measures**

150

151 Data on demographic and smoking characteristics were collected from the summary EHR or the  
152 medical notes. In the UK, the summary EHR is produced at the point of an individual's first  
153 interaction with a specific NHS hospital trust. Further information is added to the summary EHR  
154 following subsequent interactions with the hospital trust. The medical notes include  
155 contemporaneous clinical notes, General Practitioner referral letters and outpatient clinic letters,  
156 and are updated more frequently than the summary EHR.

157

#### 158 *Outcome variable*

159

160 The outcome of interest was the type of hospital admission (i.e. with COVID-19 vs. other  
161 respiratory viral infections).

162

#### 163 *Exposure variable*

164

165 Smoking status (i.e. current, former, never) was obtained from the summary EHR or the medical  
166 notes. A number of cases were recorded as 'non-smokers' without distinguishing between 'former

167 smokers' and 'never smokers'. For the primary analysis, patients categorised as a 'non-smoker'  
168 were treated as 'never smokers'. Where possible, information on use of smokeless tobacco,  
169 waterpipe and/or alternative nicotine products (e.g. e-cigarettes) was extracted. The joint first  
170 authors searched within the contemporaneous medical records for free-text entries of smoking  
171 status. The most recently available record of smoking status, obtained from either the summary  
172 EHR or the medical notes, was extracted. Where available, data on pack-year history of smoking  
173 were extracted.

174

### 175 *Covariates*

176

177 Covariates included age, sex, ethnicity, socioeconomic position (with post codes linked by the  
178 research team to the Index of Multiple Deprivation (IMD)[20]) and comorbidities (classified by  
179 organ system, including cardiac, metabolic and respiratory diseases). Medical conditions not  
180 expected to be strongly associated with COVID-19 hospitalisation were not considered in the  
181 analyses (e.g. sciatica and fibromyalgia; see the additional file). Age was treated as a continuous  
182 variable in the primary analysis, with banded age groups (i.e. 18-29 years, 30-44 years, 45-59  
183 years, 60-74 years, 75-89 years and > 90 years) used in exploratory analyses. The IMD was  
184 categorised as quintiles to reduce the impact of sparse data.

185

### 186 **Data analysis**

187

188 All analyses were conducted in R version 4.0.2.[21] Descriptive statistics for cases and controls are  
189 reported. To explore differences between cases and controls, Pearson's Chi-square tests,  
190 Cochran-Armitage tests for trend and ANOVAs were used, as appropriate.

191

192 To examine the association of former and current smoking with hospitalisation for COVID-19  
193 compared with hospitalisation for other respiratory virus infections, unadjusted and adjusted  
194 generalised linear models with a binomial distribution and logit link function were performed. We  
195 report odds ratios (ORs), 95% confidence intervals (CIs) and *p*-values. Two sensitivity analyses

196 were performed. First, those recorded as ‘non-smokers’ were removed from the analysis. Second,  
197 those excluded from the analytic sample due to missing data on smoking status (see section above  
198 on ‘Exclusion criteria’) were included and coded as i) ‘never smokers’ and then as ii) ‘current  
199 smokers’.

200

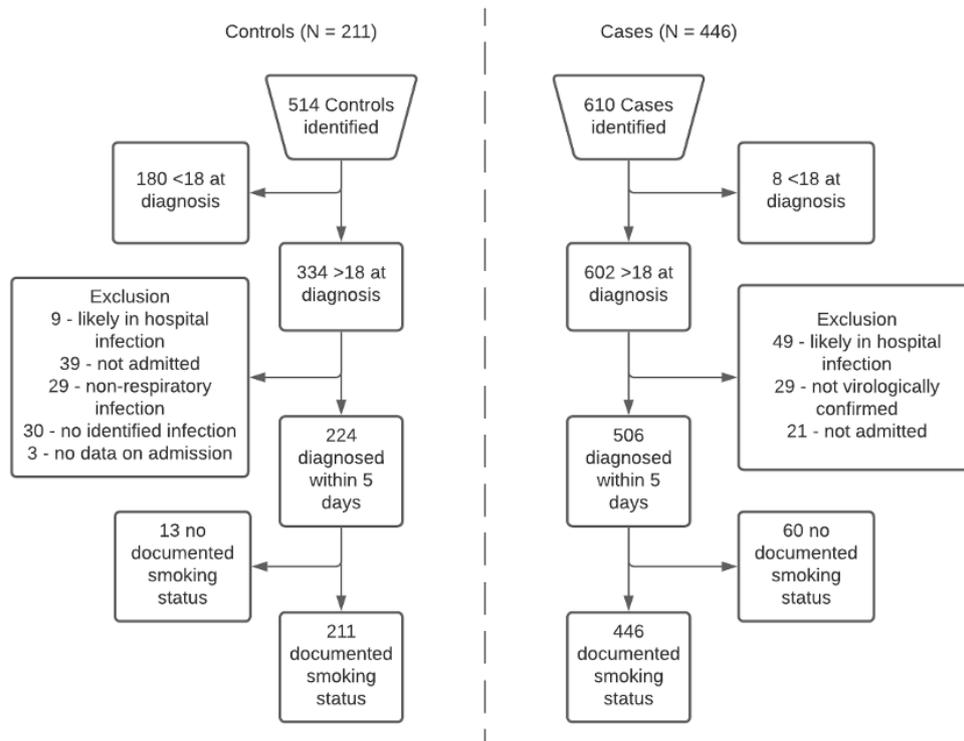
201 We then compared age-stratified current smoking prevalence in the case population with overall  
202 London prevalence stratified by age and weighted by the sex distribution of the case population.

203

204 To examine the concordance between smoking status recorded on the summary EHR and within  
205 the contemporaneous medical notes, Pearson’s Chi-squared tests were performed for the entire  
206 sample, and then separately for cases and controls.

## 207 **RESULTS**

208 A total of 610 potential cases and 514 potential controls were identified. A total of 446 cases and  
209 211 controls were included in the analytic sample (see Figure 1). Thirteen potential controls and 60  
210 potential cases were excluded due to not having a record of documented smoking status. This was  
211 likely due to these patients having no prior contact with the selected NHS foundation trust. Notably,  
212 37 (62%) of potential cases that were excluded because of missing smoking status did not survive  
213 to hospital discharge, with no in-hospital mortality in potential controls, which suggests that data  
214 were not missing at random.



215

216 *Figure 1.* Eligibility flow diagram for controls (left hand side) and cases (right hand side).

217

218 Compared with controls, cases were more likely to be male (55% vs. 35.9%) and older (64.9 years  
 219 vs 62.5 years) (see Table 1). Approximately 10% of cases and controls had missing data for  
 220 ethnicity. Compared with cases, controls were more likely to be admitted from more deprived areas  
 221 (IMD quintiles 1 and 2) (41.8% vs. 32.9%,  $p < 0.001$ ) and have pre-existing metabolic (30.3% vs  
 222 13.3%) and cardiac comorbidities (53.4% vs 30.3%). A significantly larger proportion of cases  
 223 compared with controls did not survive to discharge (28.7% vs. 4.3%). Among 128 cases not  
 224 surviving to discharge, 53 (41.4%) were never smokers, 63 (49.2%) were former smokers and 12  
 225 (9.4%) were current smokers. For patients who survived to discharge, the median length of  
 226 hospital stay for cases and controls was 9 (IQR = 4-18) and 4 (IQR = 2-9) days, respectively (see  
 227 Table 1).

228

229 Table 1. Demographic and smoking characteristics of cases and controls.

	Controls (N=211)	Cases (N=446)	Total (N=657)	p value
<b>Female sex</b>	116 (55.0%)	160 (35.9%)	276 (42.0%)	< 0.001
<b>Age in years (continuous)</b>				0.019
- Median (IQR)	62.5 (48.5-75.8)	64.9 (52.4-76.2)	64.58 (51.3-76.1)	
<b>Age in years (banded)</b>				0.007
- 18-29	18 (8.5%)	11 (2.5%)	29 (4.4%)	
- 30-44	22 (10.4%)	37 (8.3%)	59 (9.0%)	
- 45-59	46 (21.8%)	113 (25.3%)	159 (24.2%)	
- 60-74	67 (31.8%)	162 (36.3%)	229 (34.9%)	
- 75+	58 (27.5%)	123 (27.6%)	181 (27.5%)	
<b>Ethnicity</b>				0.243
- Black British, Black African or Black Caribbean	19 (9.0%)	63 (14.1%)	82 (12.5%)	
- White British or White Other	116 (55.0%)	220 (49.3%)	336 (51.1%)	
- Not stated	26 (12.3%)	43 (9.6%)	69 (10.5%)	
- Asian British or Asian	28 (13.3%)	68 (15.2%)	96 (14.6%)	
- Other	22 (10.4%)	52 (11.7%)	74 (11.3%)	
<b>IMD quintile</b>				0.06
- Missing*	5 (2.4%)	5 (1.1%)	10 (1.5%)	
- 1 - Most deprived	21 (10.0%)	26 (5.8%)	47 (7.2%)	
- 2	67 (31.8%)	121 (27.1%)	188 (28.6%)	
- 3	42 (19.9%)	122 (27.4%)	164 (25.0%)	
- 4	38 (18.0%)	98 (22.0%)	136 (20.7%)	
- 5 - Least deprived	38 (18.0%)	74 (16.6%)	112 (17.0%)	
<b>Most recent record of smoking status (summary EHR or medical notes)</b>				0.012
- Never smoker	108 (51.2%)	232 (52.0%)	340 (51.8%)	
- Former smoker	67 (31.8%)	172 (38.6%)	239 (36.4%)	
- Current smoker	36 (17.1%)	42 (9.4%)	78 (11.9%)	
<b>Record of pack-year smoking history in former and current smokers</b>	36 (35%)	91 (42.5%)	127 (40%)	
- Pack-year smoking history in current smokers, median (IQR)	30 (22.5-45)	30 (6-45)	30 (16.8-45.5)	
- Pack-year smoking history in former smokers, median (IQR)	20 (10-40)	25 (10-40)	25 (10-40)	
<b>Recorded smoking status on summary EHR</b>				< 0.001
- Never smoker	23 (10.9%)	233 (52.2%)	256 (39.0%)	
- Former smoker	22 (10.4%)	152 (34.1%)	174 (26.5%)	
- Current smoker	7 (3.3%)	30 (6.7%)	37 (5.6%)	
- Unknown	159 (75.4%)	31 (7.0%)	190 (28.9%)	
Survival to discharge	202 (95.7%)	318 (71.3%)	520 (79.1%)	< 0.001
<b>Length of hospital admission for survivors (days)</b>				< 0.001
- Median (IQR)	4 (2-9)	9 (4-18)	7 (3-14)	
Cancer (current or past)	66 (31.3%)	90 (20.2%)	156 (23.7%)	0.002
Auto-immune disease (present)	12 (5.7%)	16 (3.6%)	28 (4.3%)	0.213
Metabolic disease (present)	28 (13.3%)	135 (30.3%)	163 (24.8%)	< 0.001
Haematological disease (present)	3 (1.4%)	4 (0.9%)	7 (1.1%)	0.541
Cardiac disease (present)	64 (30.3%)	238 (53.4%)	302 (46.0%)	< 0.001
Neurological disease (present)	24 (11.4%)	54 (12.1%)	78 (11.9%)	0.786
Respiratory disease (present)	54 (25.6%)	91 (20.4%)	145 (22.1%)	0.134

Renal disease (present)	12 (5.7%)	37 (8.3%)	49 (7.5%)	0.235
HIV (present)	2 (0.9%)	7 (1.6%)	9 (1.4%)	0.522
No relevant comorbidities	43 (20.4%)	85 (19.1%)	128 (19.5%)	0.690

230 \* IMD is not available for individuals with home addresses outside of England.

231

232 Cases and controls were predominantly admitted from North central and North East central London  
233 (see additional file figure 1). The number of cases admitted from peripheral locations was greater  
234 than in controls and represents transfer of inpatients from other hospitals and diversion of patients  
235 that would otherwise have attended local hospitals due to bed pressures. The Chi-square test for  
236 trend found inconclusive evidence for any difference in socioeconomic position between cases and  
237 controls,  $\chi^2(3) = 8.93$ ,  $p = 0.06$  (see additional file figure 2).

238

### 239 **Association of smoking status with type of hospitalisation**

240

241 The prevalence of former smoking was higher in cases compared with controls (38.6% vs. 31.8%).  
242 Current smoking prevalence was lower in cases compared with controls (9.4% vs. 17.1%). A single  
243 patient from the case cohort was recorded as a dual cigarette and e-cigarette user. Two patients,  
244 one from each cohort, were recorded as dual cigarette and shisha/waterpipe users. Pack-year  
245 history of smoking was only recorded for 40% of patients with a smoking history (see Table 1).

246

247 In the univariable analysis, patients hospitalised with COVID-19 had reduced odds of being a  
248 current smoker compared with those admitted with other respiratory viruses (OR = 0.52, 95% CI =  
249 0.31-0.86,  $p = 0.01$ ). The odds for former smokers were equivocal (OR = 1.16, 95% CI = 0.81-  
250 1.68,  $p = 0.43$ ).

251 In the multivariable analysis adjusted for sex, age, socioeconomic position and comorbidities,  
252 patients hospitalised with COVID-19 had reduced odds of being a current smoker compared with  
253 those admitted with other respiratory viruses (OR = 0.55, 95% CI = 0.31-0.96,  $p = 0.04$ ). The odds  
254 for former smokers were equivocal (OR = 1.08, 95% C.I. = 0.72-1.61,  $p = 0.70$ ).

255

## 256 **Sensitivity analyses**

257 First, in a sensitivity analysis with patients recorded as 'non-smokers' excluded from the sample  
258 (leaving 398 cases and 159 controls), patients hospitalised with COVID-19 had reduced odds of  
259 being a current smoker compared with those admitted with other respiratory viruses (OR = 0.41,  
260 95% CI = 0.22-0.74,  $p = 0.03$ ). The odds for former smokers were equivocal (OR = 0.78, 95% C.I.  
261 = 0.49-1.23,  $p = 0.28$ ).

262 Second, in a sensitivity analysis with those with missing data on smoking status ( $n = 73$ ) treated as  
263 'never smokers' (resulting in 506 cases and 224 controls), patients hospitalised with COVID-19 had  
264 reduced odds of being a current smoker compared with those admitted with other respiratory  
265 viruses (OR = 0.51, 95% CI = 0.30-0.88,  $p = 0.01$ ). Next, when those with missing data on smoking  
266 status were treated as 'current smokers', patients hospitalised with COVID-19 had equivocal odds  
267 of being a current smoker compared with those admitted with other respiratory viruses (OR = 0.94,  
268 95% CI = 0.61-1.46,  $p = 0.80$ ).

269

## 270 **Comparison with London smoking prevalence**

271 Compared with age-stratified London prevalence weighted by sex, similar current smoking  
272 prevalence in cases were observed in the 30-44 years, 60-74 years and 75+ years age groups  
273 (see Table 2). Current smokers were overrepresented in the 18-29 years age group and  
274 underrepresented in the 45-59 years age group. Former smokers were consistently  
275 overrepresented in cases compared with age-stratified London prevalence weighted by sex, except  
276 for those aged 18-29 years.

277

278

279 *Table 2. Smoking prevalence in cases and controls compared with London prevalence.*

Age	Controls			Cases			London*		
	Current smoker	Former smoker	Never smoker	Current smoker	Former smoker	Never smoker	Current smoker	Former smoker	Never smoker
18-29	3 (15%)	2 (10%)	15 (75%)	6 (50%)	0 (0%)	6 (50%)	17%	7.3%	75.7%
30-44	7 (30.4%)	1 (4.3%)	15 (65.2%)	6 (13.3%)	11 (24.4%)	28 (62.2%)	14.7%	17.2%	68%
45-59	9 (18%)	14 (28%)	27 (54%)	5 (4.5%)	35 (31.2%)	72 (64.3%)	14.5%	23.3%	62.2%
60-74	13 (21%)	22 (35.5%)	27 (43.5%)	15 (9.5%)	60 (38%)	83 (52.5%)	11.6%	32.5%	56%
75+	4 (7.1%)	28 (50%)	24 (42.9%)	10 (8.4%)	66 (55.5%)	43 (36.1%)	6.8%	37.3%	56%

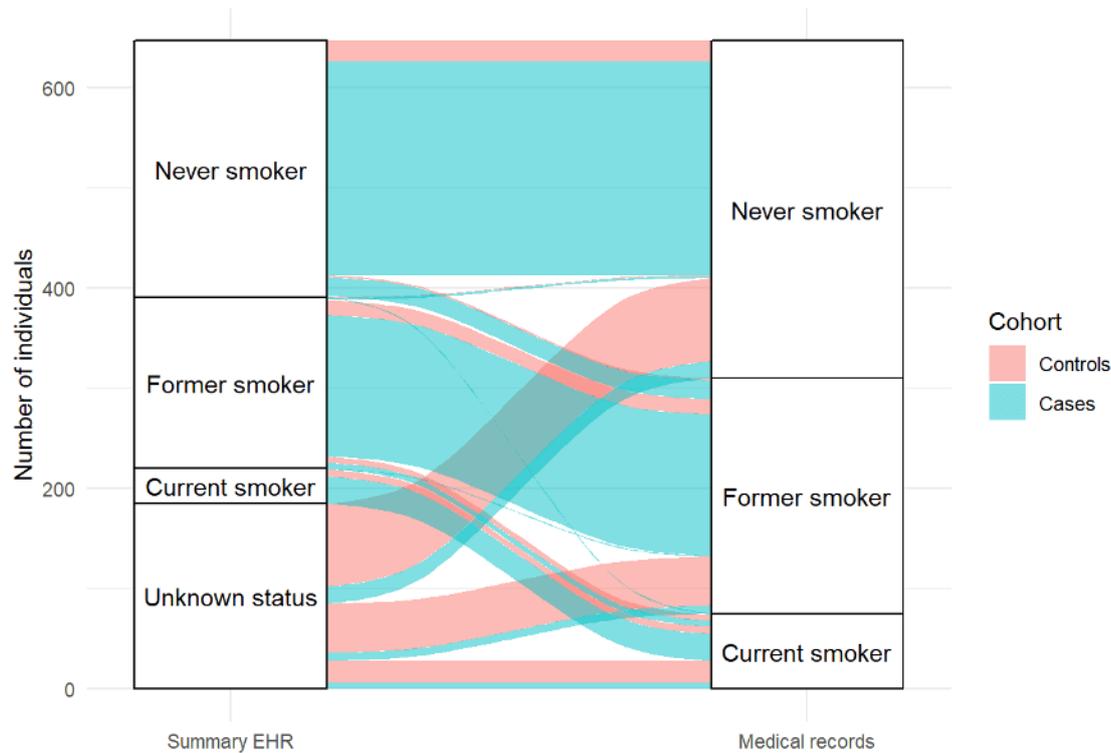
280 \* Age-stratified London prevalence weighted by the sex distribution of the case population (see  
 281 additional file table 1 for weightings).

282

283 **Concordance of smoking status recorded on the summary EHR and the medical notes**

284 Controls were more likely to have no record of smoking status on the summary EHR compared  
 285 with cases (75.4% vs. 7%) (see Figure 2). Smoking status on the summary EHR was incorrectly  
 286 recorded for 168 controls and 60 cases ( $\chi^2(3) = 226.7, p = < 0.001$ ). In cases, six current smokers  
 287 were misclassified as former smokers, one current smoker as a never smoker and six current  
 288 smokers had no record of smoking status on the summary EHR. In controls, six current smokers  
 289 were misclassified as former smokers and 23 current smokers had no record of smoking status on  
 290 the summary EHR. There was greater discordance between smoking status recorded on the  
 291 summary EHR and within the medical notes in controls ( $\chi^2(3) = 256.5, p = < 0.001$ ) than in cases  
 292 ( $\chi^2(3) = 34.2, p = < 0.001$ ).

293



294

295 *Figure 2. Concordance between smoking status recorded on the summary EHR and the medical*  
296 *notes for controls (red) and cases (blue).*

297

## 298 **DISCUSSION**

299

300 This observational, retrospective, case-control study with patients admitted to a single UK hospital  
301 trust found a lower proportion of current smokers in cases hospitalised with COVID-19 during the  
302 first phase of the pandemic compared with controls hospitalised with other respiratory viral  
303 infections a year previous. The observed reduction in current smoking prevalence was robust to  
304 adjustment for sex, age, socioeconomic position and comorbidities. The low prevalence of current  
305 smoking in cases appeared driven by current smokers being underrepresented in those aged 45-  
306 59 years, with similar to expected current smoking prevalence in the other age groups. It should be  
307 noted that cases aged 45-59 years were more likely to reside in less deprived areas than controls.  
308 As smoking status is strongly associated with socioeconomic position[14], this may partly explain  
309 the lower smoking prevalence in this age group.

310

311 Further, we found that smoking status is typically poorly recorded in the summary EHR. This was  
312 more prominent in controls than cases – a difference that is likely explained by the observation that

313 COVID-19 patients were followed up by the respiratory medicine team after discharge as part of a  
314 COVID-19 follow-up clinic where they specifically asked about smoking status[22]. The observed  
315 discrepancy between smoking status recorded on summary EHRs and the contemporaneous  
316 medical notes is a concern, particularly for studies relying solely on EHRs as the source of  
317 information on smoking status.

318

### 319 **Strengths and limitations**

320

321 To our knowledge, this is one of few studies specifically designed to examine the association  
322 between smoking status and hospitalisation with COVID-19. It was further strengthened by the  
323 inclusion of a control population hospitalised with other respiratory viruses, a comparison with local  
324 smoking prevalence derived from a nationally representative survey and an assessment of the  
325 quality of data on smoking status gleaned from summary EHRs.

326

327 This study also had a number of limitations. First, current smoking is expected *a priori* to be  
328 associated with hospitalisation for non-COVID-19 respiratory viruses[23]. Hence, the use of this  
329 control population may have biased towards the null any association between smoking status and  
330 type of hospitalisation due to greater smoking prevalence in controls compared with the general  
331 population at risk of COVID-19[24]. To minimise any impact of this on the interpretation of results,  
332 we included a community-level comparator of overall London smoking prevalence. Second, the  
333 IMD was used to capture the socioeconomic position of patients, calculated for a spatial boundary  
334 referred to as a Lower Layer Super Output Area[20], which are geographic areas of populations of  
335 ~1500 individuals, which can then be linked to post codes. In densely populated and  
336 socioeconomically diverse areas such as central London, the ability of the IMD to capture  
337 socioeconomic position may be limited[25]. Third, the control population was selected to reflect  
338 commonalities in the transmission of respiratory viruses. However, emerging evidence suggests  
339 that COVID-19 has a significantly different pathological process compared with other respiratory  
340 viruses. For example, mortality rates from COVID-19 differ widely from those due to epidemic  
341 influenza[26]. Further, mortality rates for COVID-19 differ between and within countries and over

342 the course of the pandemic, at least partly due to improving knowledge and treatment of the  
343 disease[27]. In addition, the risk of superadded bacterial infection in COVID-19 patients appears  
344 lower than for those with influenza[28]] and COVID-19 patients have increased rates of  
345 coagulopathies compared with other respiratory viruses[29]. Taken together, these emerging  
346 observations may limit any direct comparison of risk profiles in cases and controls. Fourth, a  
347 history of current or past cancer was high in both groups at greater than 20% and was significantly  
348 greater in controls compared with cases. This reflects a bias in the population that regularly  
349 interacts with the selected NHS hospital trust, which is a specialist cancer referral centre. We  
350 visualised the geographic regions where patients were admitted from to examine any systemic  
351 differences between cases and controls, and caution that the differing catchment areas of the two  
352 cohorts may have led to important differences in the underlying populations. In addition, during the  
353 peak of the first wave of the pandemic in the UK, many cases were transferred across hospital  
354 sites due to bed pressures.

355

### 356 **Implications for policy and practice**

357

358 COVID-19 will continue to place a large burden on healthcare services in the UK and  
359 internationally over the coming months and years. To mitigate against this, multiple non-  
360 pharmacological interventions are being implemented to reduce the intensity of demand on acute  
361 and intensive services. Irrespective of any direct link between smoking and COVID-19 disease  
362 outcomes, smoking is a significant cause for healthcare demand globally. We have argued  
363 elsewhere for the need to ramp up smoking cessation support to reduce the current and future  
364 burden on healthcare and social services[30].

365

### 366 **Avenues for future research**

367

368 To better understand the role of smoking as a risk factor for COVID-19 hospitalisation, longitudinal,  
369 representative population-level studies from multiple sites are required. A large, multinational  
370 clinical cohort study (the International Severe Acute Respiratory and emerging Infection

371 Consortium; ISARIC) collects smoking status but this is likely obtained from summary EHRs.

372 Purposeful acquisition of accurate smoking status is required to characterise the role of smoking in  
373 COVID-19 disease outcomes.

374

## 375 **Conclusion**

376

377 In a single hospital trust in the UK, patients hospitalised with COVID-19 had reduced odds of being  
378 current smokers compared with patients admitted with other respiratory viruses a year previous. In  
379 the 45-59 year age group, lower smoking prevalence than would be expected based on age- and  
380 sex-matched local smoking prevalence was observed. Smoking status was poorly recorded, with  
381 high observed discordance between smoking status recorded on the summary EHR and the  
382 medical notes.

383

## 384 **DATA SHARING**

385 Anonymised and de-identified individual-level data are available upon request and approval from  
386 the NHS foundation trust. Requests should be directed to the corresponding author. Alternatively,  
387 data can be requested directly from the Biomedical Research Centre Clinical and Research  
388 Informatics Unit at UCL/UCLH.

## 389 **CONTRIBUTERS**

390 DS, OP, LS, JB and RB conceived of and designed the study. DS and RB acquired the study data.  
391 DS and OP extracted and analysed the data. DS, OP, LS, JB contributed to the interpretation of  
392 the data and to the drafting of the manuscript. All authors have read and approved the manuscript.

## 393 **DECLARATION OF INTERESTS**

394 DS, OP and RB have no conflicts of interest to declare. LS has received a research grant and  
395 honoraria for a talk and travel expenses from manufacturers of smoking cessation medications  
396 (Pfizer and Johnson & Johnson). JB has received unrestricted research funding to study smoking

397 cessation from companies who manufacture smoking cessation medications. All authors declare  
398 no financial links with tobacco companies or e-cigarette manufacturers or their representatives.

### 399 **ACKNOWLEDGEMENTS**

400 DS is supported by a PhD studentship from the UK Biotechnology and Biological Sciences  
401 Research Council [BB/M009513/1]. OP receives salary support from Cancer Research UK  
402 (C1417/A22962). JB, LS, & OP are members of SPECTRUM, a UK Prevention Research  
403 Partnership Consortium (MR/S037519/1). UKPRP is an initiative funded by the UK Research and  
404 Innovation Councils, the Department of Health and Social Care (England) and the UK devolved  
405 administrations, and leading health research charities.

### 406 **ROLE OF FUNDERS**

407 We gratefully acknowledge all funding listed above. The views expressed are those of the authors  
408 and not necessarily those of the funders.

### 409 **ETHICAL APPROVAL**

410 Ethical approval was obtained from a local research committee and the NHS Health Research  
411 Authority (IRAS\_282704). The requirement for informed consent was waived due to the nature of  
412 the study.

413

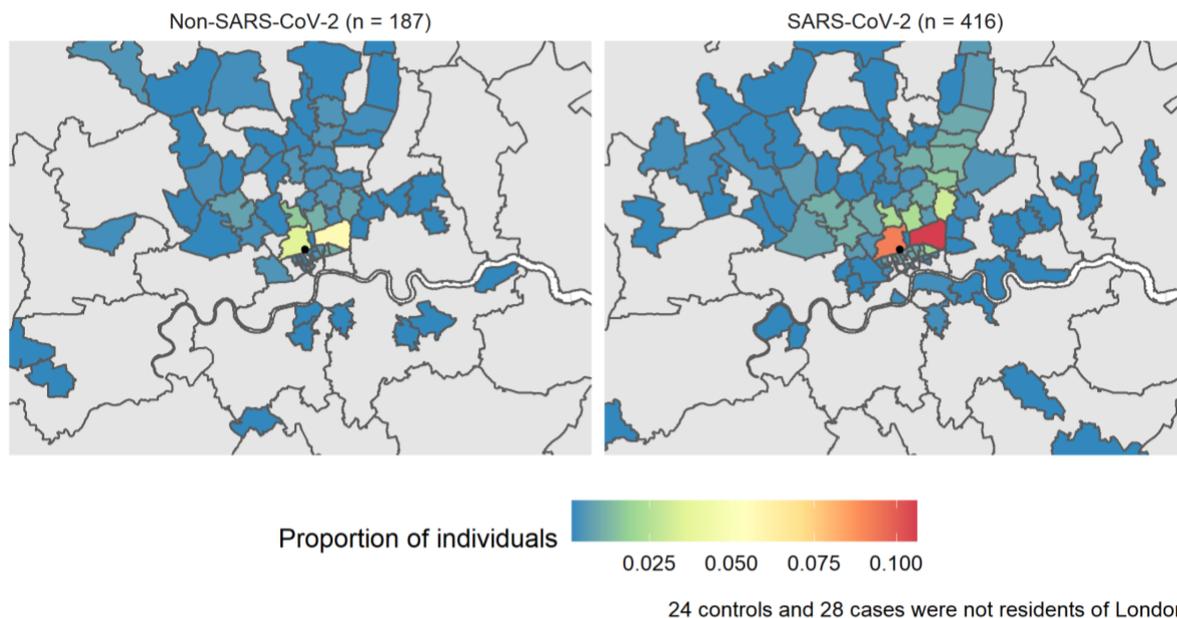
### 414 **REFERENCES**

- 415 1. COVID-19 Map. Johns Hopkins Coronavirus Resource Center.  
416 <https://coronavirus.jhu.edu/map.html>. Accessed 1 Oct 2020.
- 417 2. Guan W, Liang W, Zhao Y, Liang H, Chen Z, Li Y, et al. Comorbidity and its impact on 1590  
418 patients with COVID-19 in China: a nationwide analysis. *Eur Respir J.* 2020;55.  
419 doi:10.1183/13993003.00547-2020.
- 420 3. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at  
421 increased risk for COVID-19 infection? *Lancet Respir Med.* 2020;:0.
- 422 4. Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2  
423 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease  
424 Inhibitor. *Cell.* 2020;181:271-280.e8.
- 425 5. Brake SJ, Barnsley K, Lu W, McAlinden KD, Eapen MS, Sohal SS. Smoking Upregulates  
426 Angiotensin-Converting Enzyme-2 Receptor: A Potential Adhesion Site for Novel Coronavirus  
427 SARS-CoV-2 (Covid-19). *J Clin Med.* 2020;9, Page 841:841.

- 428 6. Cai G. Bulk and Single-Cell Transcriptomics Identify Tobacco-Use Disparity in Lung Gene  
429 Expression of ACE2, the Receptor of 2019-nCov. preprint. LIFE SCIENCES; 2020.  
430 doi:10.20944/preprints202002.0051.v3;  
431 <http://web.archive.org/web/20200925104805/https://www.preprints.org/manuscript/202002.0051/v3>  
432 .
- 433 7. Oakes JM, Fuchs RM, Gardner JD, Lazartigues E, Yue X. Nicotine and the renin-angiotensin  
434 system. *Am J Physiol - Regul Integr Comp Physiol*. 2018;315:895–906.
- 435 8. Abadom TR, Smith AD, Tempia S, Madhi SA, Cohen C, Cohen AL. Risk factors associated with  
436 hospitalisation for influenza-associated severe acute respiratory illness in South Africa: A case-  
437 population study. *Vaccine*. 2016;34:5649–55.
- 438 9. Feldman C, Anderson R. Cigarette smoking and mechanisms of susceptibility to infections of the  
439 respiratory tract and other organ systems. *J Infect*. 2013;67:169–84.
- 440 10. Millett ERC, De Stavola BL, Quint JK, Smeeth L, Thomas SL. Risk factors for hospital  
441 admission in the 28 days following a community-acquired pneumonia diagnosis in older adults, and  
442 their contribution to increasing hospitalisation rates over time: A cohort study. *BMJ Open*.  
443 2015;5:008737.
- 444 11. Vardavas CI, Nikitara K. COVID-19 and smoking: A systematic review of the evidence. *Tob  
445 Induc Dis*. 2020;18:20.
- 446 12. Simons D, Shahab L, Brown J, Perski O. The association of smoking status with SARS-CoV-2  
447 infection, hospitalisation and mortality from COVID-19: A living rapid evidence review with  
448 Bayesian meta-analyses (version 9). *Qeios*. 2020. doi:10.32388/UJR2AW.10.
- 449 13. Polubriaginof F, Salmasian H, Albert DA, Vawdrey DK. Challenges with Collecting Smoking  
450 Status in Electronic Health Records. *AMIA Annu Symp Proc*. 2018;2017:1392–400.
- 451 14. ONS. Adult smoking habits in the UK: 2019. Office for National Statistics; 2020.  
452 [https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpecta  
453 ncies/bulletins/adultsmokinghabitsingreatbritain/2019](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/adultsmokinghabitsingreatbritain/2019). Accessed 10 Jul 2020.
- 454 15. ICD-10 Version:2019. <https://icd.who.int/browse10/2019/en#/U07.1>. Accessed 28 Oct 2020.
- 455 16. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation Period of  
456 Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and  
457 Application. *Ann Intern Med*. 2020;172:577–82.
- 458 17. McCarthy N, Giesecke J. Case-case comparisons to study causation of common infectious  
459 diseases. *Int J Epidemiol*. 1999;28:764–8.
- 460 18. Peralta PS-O, Cortes-García M, Vicente-Herrero M, Castrillo-Villamandos C, Arias-Bohigas P,  
461 Amo IP, et al. Risk factors for disease severity among hospitalised patients with 2009 pandemic  
462 influenza A (H1N1) in Spain, April – December 2009. *Eurosurveillance*. 2010;15:19667.
- 463 19. Falsey AR, McElhaney JE, Beran J, van Essen GA, Duval X, Esen M, et al. Respiratory  
464 Syncytial Virus and Other Respiratory Viral Infections in Older Adults With Moderate to Severe  
465 Influenza-like Illness. *J Infect Dis*. 2014;209:1873–81.
- 466 20. The English Indices of Deprivation 2019. :31.
- 467 21. R Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R  
468 Foundation for Statistical Computing; 2020. <https://www.R-project.org/>.

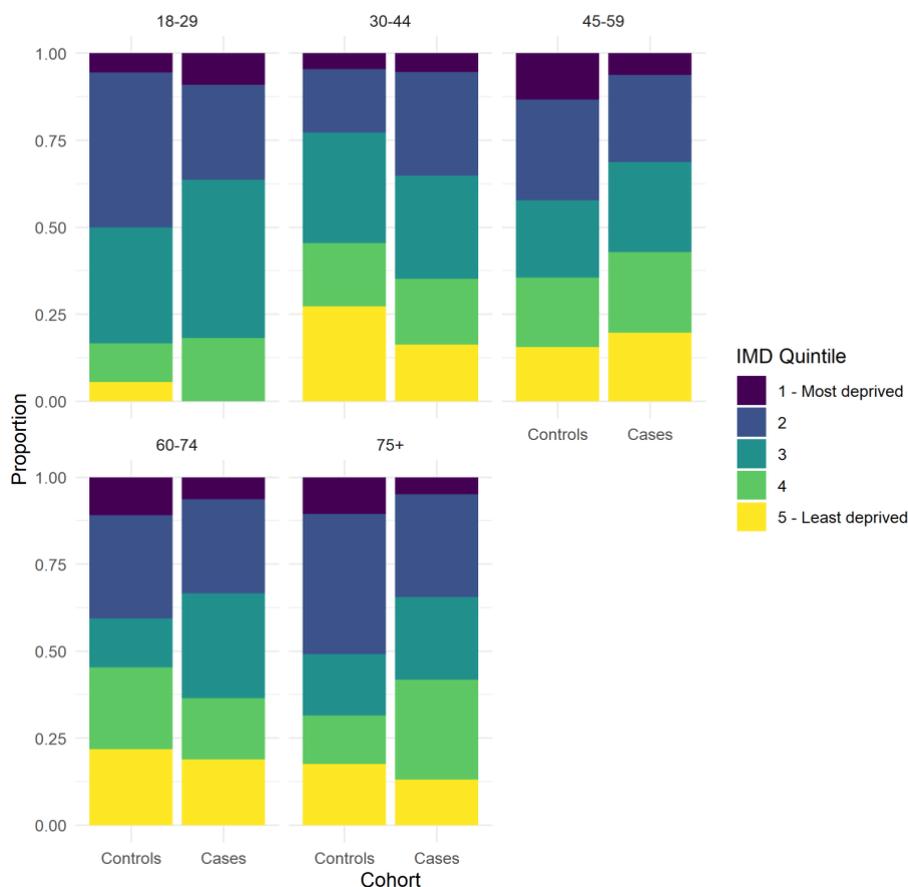
- 469 22. Mandal S, Barnett J, Brill SE, Brown JS, Denny EK, Hare SS, et al. 'Long-COVID': a cross-  
470 sectional study of persisting symptoms, biomarker and imaging abnormalities following  
471 hospitalisation for COVID-19. *Thorax*. 2020. doi:10.1136/thoraxjnl-2020-215818.
- 472 23. Stämpfli MR, Anderson GP. How cigarette smoke skews immune responses to promote  
473 infection, lung disease and cancer. *Nat Rev Immunol*. 2009;9:377–84.
- 474 24. Vandenbroucke JP, Pearce N. Case–control studies: basic concepts. *Int J Epidemiol*.  
475 2012;41:1480–9.
- 476 25. Deas I, Robson B, Wong C, Bradford M. Measuring Neighbourhood Deprivation: A Critique of  
477 the Index of Multiple Deprivation. *Environ Plan C Gov Policy*. 2003;21:883–903.
- 478 26. ONS. Deaths due to coronavirus (COVID-19) compared with deaths from influenza and  
479 pneumonia, England and Wales.  
480 <https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsduetocoronaviruscovid19comparedwithdeathsfrominfluenzaandpneumoniaenglandandwales/deathsoccurringbetween1januaryand31august2020>. Accessed 29 Oct 2020.  
482
- 483 27. Levin AT, Hanage WP, Owusu-Boaitey N, Cochran KB, Walsh SP, Meyerowitz-Katz G.  
484 ASSESSING THE AGE SPECIFICITY OF INFECTION FATALITY RATES FOR COVID-19:  
485 SYSTEMATIC REVIEW, META-ANALYSIS, AND PUBLIC POLICY IMPLICATIONS. medRxiv.  
486 2020;:2020.07.23.20160895.
- 487 28. Lansbury L, Lim B, Baskaran V, Lim WS. Co-infections in people with COVID-19: a systematic  
488 review and meta-analysis. *J Infect*. 2020;81:266–75.
- 489 29. Klok FA, Kruip MJHA, van der Meer NJM, Arbous MS, Gommers DAMPJ, Kant KM, et al.  
490 Incidence of thrombotic complications in critically ill ICU patients with COVID-19. *Thromb Res*.  
491 2020;191:145–7.
- 492 30. Simons D, Perski O, Brown J. Covid-19: The role of smoking cessation during respiratory virus  
493 epidemics. *The BMJ*. 2020. <https://blogs.bmj.com/bmj/2020/03/20/covid-19-the-role-of-smoking-cessation-during-respiratory-virus-epidemics/>.  
494
- 495
- 496

497 **Additional File**



498

499 **Figure 1.** Primary location of residence of controls (L) and cases (R) enrolled in the study. The location of the study site  
500 is marked by a black dot. n.b. visualised region is limited to central London.



501

502 **Figure 2.** IMD quintile for admitted controls and cases for each age group.

503

504 **Table 1.** Weightings matching London APS data to case population

APS age band	Sex	n	Proportion
18-29	Male	10	0.83
	Female	2	0.17
30-44	Male	28	0.62
	Female	17	0.38
45-59	Male	79	0.71
	Female	33	0.29
60-74	Male	101	0.64
	Female	57	0.36
75+	Male	68	0.57
	Female	51	0.43

505

506 **ICD-10 code list:**

- 507
- Cases
    - U07.1 - COVID-19, virus identified
  - 509 • Controls
    - J10.1 - Influenza with other respiratory manifestations, seasonal influenza virus identified
    - J10.0 - Influenza with pneumonia, seasonal influenza virus identified
    - J11.0 - Influenza with pneumonia, virus not identified
    - J12.2 - Parainfluenza virus pneumonia
    - J12.0 - Adenoviral pneumonia
    - J10.8 - Influenza with other manifestations, seasonal influenza virus identified
    - J11.1 - Influenza with other respiratory manifestations, virus not identified
    - J12.8 - Other viral pneumonia
    - J12.9 - Viral pneumonia, unspecified
    - J12.3 - Human metapneumovirus
    - J20.8 - Acute bronchitis due to other specified organisms
    - J21.8 - Acute bronchiolitis due to other specified organisms

522 **Comorbidity code list:**

- 523
- Auto-immune
    - Myasthenia gravis
    - Systemic Lupus Erythematosus
    - Crohn's disease
    - Ulcerative colitis
    - Granulomatosis with Polyangiitis
    - Sjogren's syndrome
    - Cold agglutinin disease
    - Grave's disease
    - Rheumatoid arthritis
  - 534 • Cancers
    - All cancers
    - All carcinomas
    - All myelodysplasias
    - Waldenstrom's macroglobulinaemia
    - All myelomas
    - All lymphomas
    - All leukaemias
    - Gliomas
    - Sarcomas
    - Melanomas
  - 545 • Cardiac

- 546           ○ Hypertension
- 547           ○ Ischaemic heart disease
- 548           ○ Coronary artery disease
- 549           ○ Atrial fibrillation
- 550           ○ Congestive cardiac failure
- 551           ○ All other heart failure
- 552           ○ Non-ST elevation myocardial infarction
- 553           ○ Myocarditis
- 554           ○ History of coronary artery bypass graft
- 555           ○ Aortic stenosis
- 556           ○ Other valvular abnormalities
- 557           ○ Presence of a pacemaker/implantable cardiac defibrillator
- 558           ○ All cardiomyopathies
- 559           ○ Structural cardiac abnormalities
- 560       • Haematological
  - 561           ○ Sickle cell disease
  - 562           ○ All thalassaemias
  - 563           ○ Thrombotic thrombocytopenia purpura
  - 564           ○ Anaemia
- 565       • Metabolic
  - 566           ○ Obesity
  - 567           ○ Type-1 Diabetes Mellitus
  - 568           ○ Type-2 Diabetes Mellitus
  - 569           ○ Syndrome of inappropriate ADH
  - 570           ○ Hypothyroidism
  - 571           ○ Addison's disease
  - 572           ○ Pituitary hormone abnormalities
  - 573           ○ Chronic pancreatitis
- 574       • Neurological
  - 575           ○ History of acute stroke
  - 576           ○ Parkinson's disease
  - 577           ○ History of transient ischaemic attack
  - 578           ○ Multiple sclerosis
  - 579           ○ Epilepsy
  - 580           ○ Cerebral Palsy
  - 581           ○ All dementias
  - 582           ○ Autism
  - 583           ○ Ataxic disorders
- 584       • No relevant comorbidities
  - 585           ○ Nil
  - 586           ○ Alcohol excess
  - 587           ○ Achalasia
  - 588           ○ Sciatica
  - 589           ○ Substance misuse
  - 590           ○ Benign prostatic hypertrophy
- 591       • Psychiatric
  - 592           ○ Schizophrenic disorders
  - 593           ○ Anorexia nervosa
- 594       • Renal
  - 595           ○ Chronic kidney disease
  - 596           ○ History of acute kidney disease
  - 597           ○ Renal transplant
- 598       • Respiratory
  - 599           ○ Obstructive sleep apnoea
  - 600           ○ Pulmonary hypertension
  - 601           ○ Chronic obstructive pulmonary disease
  - 602           ○ Asthma
  - 603           ○ Pulmonary fibrosis
  - 604           ○ Bronchiectasis
  - 605           ○ Interstitial lung disease
  - 606           ○ Emphysema
  - 607           ○ History of or current tuberculosis
  - 608           ○ Sarcoidosis

609

- History of pulmonary embolus