medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

David Simons*[1,2], Olga Perski*[3], Lion Shahab[3], Jamie Brown[3] & Robin Bailey[2,4]

[1] Centre for Emerging, Endemic and Exotic Diseases, RVC, UK

[2] Hospital for Tropical Diseases, London, UK

[3] Department of Behavioural Science and Health, University College London, UK

[4] London School of Hygiene and Tropical Medicine, UK

*These authors contributed equally to this work

Corresponding author: David Simons, dsimons19@rvc.ac.uk

Abstract word count: 242

Manuscript word count: 3885

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Background: It is unclear whether smoking increases the risk of COVID-19 hospitalisation. We examined i) the association of smoking status with hospitalisation for COVID-19 compared with hospitalisation for other respiratory viral infections a year previous; and ii) concordance between smoking status recorded on the electronic health record (EHR) and the contemporaneous medical notes.

Methods: This case-control study enrolled adult patients (446 cases and 211 controls) at a single National Health Service trust in London, UK. The outcome variable was type of hospitalisation (COVID-19 vs. another respiratory virus a year previous). The exposure variable was smoking status (never/former/current smoker). Logistic regression analyses adjusted for age, sex, socioeconomic position and comorbidities were performed. The study protocol and analyses were pre-registered in April 2020 on the Open Science Framework.

Results: Current smokers had lower odds of being hospitalised with COVID-19 compared with other respiratory viruses a year previous (OR_{adj}=0.55, 95% CI=0.31-0.96, *p*=.04). There was no significant association among former smokers (OR_{adj}=1.08, 95% CI=0.72-1.65, *p*=.70). Smoking status recorded on the EHR (compared with the contemporaneous medical notes) was incorrectly recorded for 168 (79.6%) controls ($\chi^2(3)=256.5$, *p*=<0.001) and 60 cases (13.5%) ($\chi^2(3)=34.2$, *p*=<0.001).

Conclusions: In a single UK hospital trust, current smokers had reduced odds of being hospitalised with COVID-19 compared with other respiratory viruses a year previous, although it is unclear whether this association is causal. Targeted post-discharge recording of smoking status may account for the greater EHR-medical notes concordance observed in cases compared with controls.

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

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

1. Introduction

COVID-19 is a respiratory disease caused by the SARS-CoV-2 virus. There are in excess of 118 million confirmed COVID-19 cases globally, with over 2.6 million deaths reported (Johns Hopkins Coronavirus Resource Center, 2021). Large age and sex differences in case severity and mortality have been observed (Guan et al., 2020), with hypertension, diabetes and obesity identified as important risk factors (Fang et al., 2020). There are *a priori* reasons to believe that current smokers are at increased risk of contracting COVID-19 and experiencing greater disease severity once infected. SARS-CoV-2 enters epithelial cells through the ACE-2 receptor (Hoffmann et al., 2020). Evidence suggests that gene expression and subsequent ACE-2 receptor levels are elevated in the airway and oral epithelium of current smokers (Brake et al., 2020; Cai, 2020), potentially making smokers vulnerable to contracting SARS-CoV-2. Other studies, however, show that smoking downregulates the ACE-2 receptor (Oakes et al., 2018). In addition, smoking involves repeated hand-to-mouth movements, which may mean smokers are more likely to contract respiratory viruses such as SARS-CoV-2 (Simons et al., 2020). Early data from the ongoing pandemic have not provided clear evidence for an association of smoking status with COVID-19 outcomes, with a living review and unadjusted Bayesian meta-analysis of over 60 studies indicating that current compared with never smokers may be at reduced risk of SARS-CoV-2 infection, while former smokers are at increased risk of hospitalisation, disease severity and in-hospital mortality compared with never smokers (Simons et al., 2021).

Most studies to date have been limited by the lack of appropriate controls, poor recording of smoking status and insufficient adjustment for relevant covariates. Many studies relied on routine electronic health records (EHRs) to obtain data on demographic characteristics, comorbidities and smoking status. This is problematic, as previous research suggests that data on smoking status obtained via EHRs tend to be incomplete or outright inaccurate, with implausible longitudinal changes observed (Polubriaginof et al., 2018). As hospitalised populations differ by age and sex from the general population (Secondary Care Analytical Team, 2020), comparisons of current and former smoking prevalence in hospitalised and non-hospitalised populations are likely biased. There is therefore a need for alternative study designs with relevant comparator groups and adjustment for covariates to better understand the association of smoking status with COVID-19 disease outcomes.

However, the selection of an appropriate comparator group is not straightforward. Ideally, controls should represent the underlying population from which cases emerged, both geographically and demographically (Grimes and Schulz, 2005). In the context of COVID-19 hospitalisation, disease severity and death, we therefore reasoned *a priori* that historical controls – i.e. patients hospitalised at the same trust with another respiratory viral infection (e.g. influenza) a year previous – would act as a useful comparator, as they represent a geographically matched population at risk of severe disease from a circulating respiratory

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

virus with a similar route of transmission (i.e. respiratory droplets and aerosols) and detection (i.e. laboratory-confirmed infection prior to or upon hospitalisation) (McCarthy and Giesecke, 1999). In addition, risk factors for hospitalisation with other respiratory viruses are similar to those for hospitalisation with COVID-19 (e.g. older age, comorbidities) (Falsey et al., 2014; Peralta et al., 2010).

In the present case-control study, we therefore aimed to examine i) the association of smoking status with hospitalisation for COVID-19 compared with hospitalisation for other respiratory viral infections (e.g. influenza, respiratory syncytial virus) a year previous at a single UK hospital trust; and ii) whether there is discordance between smoking status recorded on the summary EHR and within the contemporaneous medical notes. As current smoking in April 2020 (when our study protocol was registered) was *a priori* expected to be associated with an increased risk of COVID-19 hospitalisation (Alqahtani et al., 2020; Simons et al., 2020), and with the association expected to be of a similar magnitude to that observed for other respiratory viruses, we opted for a non-inferiority design to test the hypothesis that the proportion of current smokers in patients hospitalised with COVID-19 is similar to that in patients hospitalised with other respiratory viral infections a year previous.

2. Methods

2.1. Study design

This was an observational case-control study with historical controls, performed at a single National Health Service (NHS) hospital trust (comprising two hospital sites) in London, UK. The study protocol and analysis plan were pre-registered on the Open Science Framework in April 2020 (https://bit.ly/3kFYh06). The pre-registered protocol stipulated a non-inferiority design (i.e. a one-tailed statistical test) to maximise statistical power to detect a significantly lower proportion of current smokers (i.e. <10%) among patients hospitalised with COVID-19 compared with patients hospitalised with another respiratory viral infection a year previous (i.e. 20%). The protocol was amended after data collection but prior to statistical analysis in September 2020 (https://osf.io/ezfqs/) to implement a traditional case-control design (i.e. a two-tailed statistical test), as a delay in study approval meant that the number of eligible cases and controls exceeded our expectations, providing sufficient power for a two-tailed test. We had also planned to compare current smoking in cases with age- and sex-matched London prevalence, with data obtained from the representative Annual Population Survey. However, following external review on an earlier manuscript draft, we decided against presenting data from this comparison due to smoking rates in hospitalised populations typically being greater than in the general population (Benowitz et al., 2009). A completed STROBE checklist is available in Supplementary file 1. Ethical approval was obtained from a local research committee and approved by the NHS Health Research Authority (IRAS_282704). The requirement for informed consent was waived due to the nature of the study.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

2.2. Eligibility criteria

2.2.1. Inclusion criteria

- Cases
    1. Consecutive patients admitted to an adult hospital ward (i.e. 18+ years) between 1st March 2020 and 26th August 2020 (the date on which data were obtained);
    2. Diagnosis of COVID-19 on or within 5 days of hospital admission, identified via associated International Classification of Diseases version 10 (ICD-10) codes (World Health Organisation, 2019). This temporal boundary was set to prevent inclusion of patents with nosocomial (hospital-acquired) infection and allowed for a delay of 3 days in requesting a COVID-19 test and 2 days for receiving and reporting the results on the EHR. The median incubation time for COVID-19 is estimated at 5.1 days (95% CI = 4.5-5.8) (Lauer et al., 2020). We sought to exclude individuals with nosocomial COVID-19 infection as they are a different population (e.g. older, more frail) compared with those infected in the community and subsequently requiring hospitalisation.

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

2.2.2. Exclusion criteria

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

2.3. Measures

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

2.3.1. Outcome variable

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

2.3.2. Exposure variable

Smoking status (i.e. current, former, never) was obtained from the summary EHR or the medical notes. A number of cases were recorded as 'non-smokers' without distinguishing between 'former smokers' and 'never smokers'. For the primary analysis, patients categorised as a 'non-smoker' were treated as 'never smokers'. Where possible, information on use of smokeless tobacco, waterpipe and/or alternative nicotine products (e.g. e-cigarettes) was extracted. The joint first authors searched within the contemporaneous medical records for free-text entries of smoking status. The most recently available record of smoking status, obtained from either the summary EHR or the medical notes, was extracted. Where available, data on pack-year history of smoking were extracted.

2.3.3. Covariates

Covariates included age, sex, ethnicity, socioeconomic position (SEP; with post codes linked by the research team to the Index of Multiple Deprivation (IMD) (Department for Communities and Local Government, 2019)) and comorbidities (classified by organ system, including cardiac, metabolic and respiratory diseases). Medical conditions not expected to be strongly associated with COVID-19 hospitalisation were not considered in the analyses (e.g. sciatica and fibromyalgia; see Supplementary file 1). Age was treated as a continuous variable in the primary analysis, with banded age groups (i.e. 18-29 years, 30-44 years, 45-59 years, 60-74 years, 75-89 years and > 90 years) used in exploratory analyses. The IMD was categorised as quintiles to reduce the impact of sparse data.

2.4. Data analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

All analyses were conducted in R version 4.0.2.(R Core Team, 2020) Descriptive statistics for cases and controls are reported. To explore differences between cases and controls, Pearson's Chi-square tests, Cochran-Armitage tests for trend and ANOVAs were used, as appropriate.

To examine the association of former and current smoking with hospitalisation for COVID-19 compared with hospitalisation for other respiratory viral infections, unadjusted and two different adjusted generalised linear models with a binomial distribution and logit link function were performed. The first model adjusted for age, sex and SEP, with a second model adjusting for age, sex, SEP and comorbidities. We report odds ratios (ORs), 95% confidence intervals (CIs) and *p*-values. Two sensitivity analyses were subsequently performed. First, those recorded as 'non-smokers' were removed from the analysis. Second, those excluded from the analytic sample due to missing data on smoking status (see section above on 'Exclusion criteria') were included and coded as i) 'never smokers' and then as ii) 'current smokers' to assess the robustness of the associations.

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

3. Results

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

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

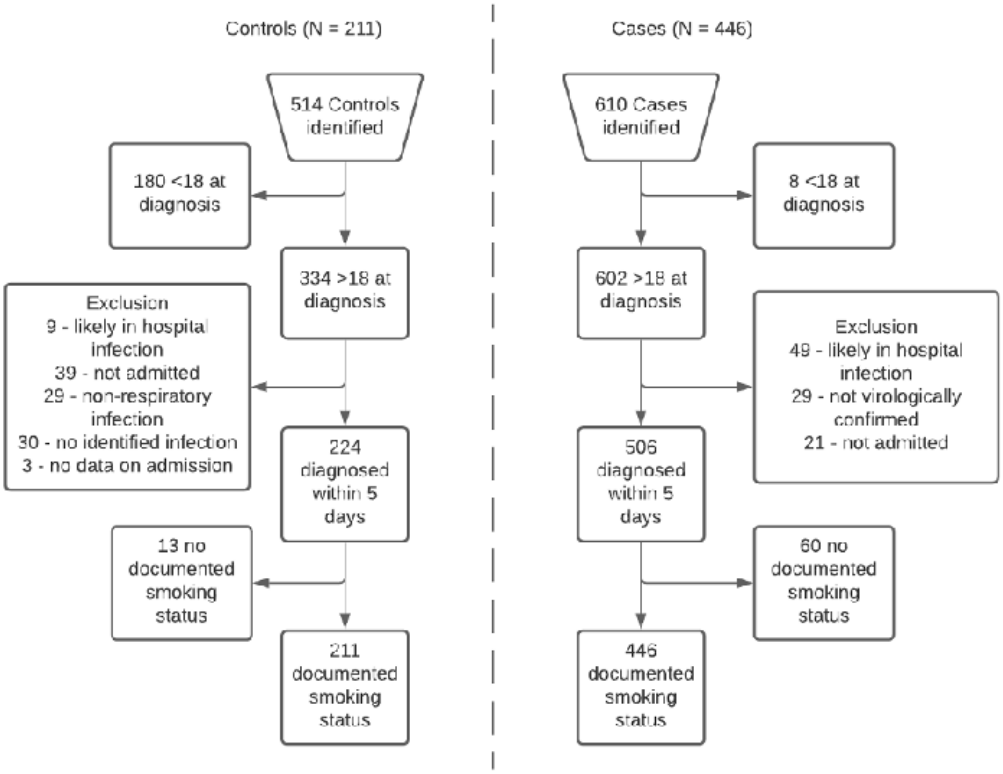

Figure 1. Eligibility flow diagram for controls (left hand side) and cases (right hand side).

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

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

	Controls (N=211)	Cases (N=446)	Total (N=657)	*p* value
Female sex	116 (55.0%)	160 (35.9%)	276 (42.0%)	< 0.001
Age in years (continuous)				0.019
- Median (IQR)	62.5 (48.5-75.8)	64.9 (52.4-76.2)	64.58 (51.3-76.1)	
Age in years (banded)				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%)	
Ethnicity				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%)	
IMD quintile				0.06
- Missing [a]	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%)	
Most recent record of smoking status (summary EHR or medical notes)				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%)	
Record of pack-year smoking history in former and current smokers	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)	

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

- Pack-year smoking history in former smokers, median (IQR)	20 (10-40)	25 (10-40)	25 (10-40)	
Recorded smoking status on summary EHR				< 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
Length of hospital admission for survivors (days)				< 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

[a] IMD is not available for individuals with home addresses outside of England.

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

3.1. Association of smoking status with type of hospitalisation

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

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

In the multivariable analysis adjusted for sex, age and SEP, current smokers had reduced odds of being hospitalised with COVID-19 compared with other respiratory viruses a year previous (OR = 0.48, 95% CI = 0.28-0.83, $p < 0.01$). There was no significant association among former smokers (OR = 0.90, 95% CI = 0.61-1.34, $p = 0.61$). Results were not materially altered when also adjusting for relevant comorbidities (current smokers, OR = 0.55, 95% CI = 0.31-0.96, $p = 0.04$; former smokers, OR = 1.08, 95% CI = 0.72-1.61, $p = 0.70$).

3.2. Sensitivity analyses

First, in a sensitivity analysis with patients recorded as 'non-smokers' excluded from the sample (leaving 398 cases and 159 controls), current smokers had reduced odds of being hospitalised with COVID-19 compared with other respiratory viruses a year previous (OR = 0.41, 95% CI = 0.22-0.74, $p = 0.03$). There was no significant association among former smokers (OR = 0.78, 95% C.I. = 0.49-1.23, $p = 0.28$).

Second, in a sensitivity analysis with those with missing data on smoking status (n = 73) treated as 'never smokers' (resulting in 506 cases and 224 controls), patients hospitalised with COVID-19 had reduced odds

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

of being a current smoker compared with those admitted with other respiratory viruses (OR = 0.51, 95% CI = 0.30-0.88, p = 0.01). Next, when those with missing data on smoking status were treated as 'current smokers', there was no significant association between current smoking and hospitalisation with COVID-19 (OR = 0.94, 95% CI = 0.61-1.46, p = 0.80).

3.3. Concordance of smoking status recorded on the summary EHR and the medical notes

Controls were more likely to have no record of smoking status on the summary EHR compared with cases (75.4% vs. 7%) (see Figure 2). However, smoking status could be ascertained from the contemporaneous medical notes for all included cases and controls. Smoking status on the summary EHR (including 'unknown' status) was incorrectly recorded for 168 (79.6%) controls and 60 cases (13.5%) ($\chi^2(3) = 226.7$, p = < 0.001). In cases, six current smokers were misclassified as former smokers, one current smoker as a never smoker and six current smokers had no record of smoking status on the summary EHR. In controls, six current smokers were misclassified as former smokers and 23 current smokers had no record of smoking status on the summary EHR. There was greater discordance between smoking status recorded on the summary EHR and within the contemporaneous medical notes in controls ($\chi^2(3) = 256.5$, p = < 0.001) than in cases ($\chi^2(3) = 34.2$, p = < 0.001).

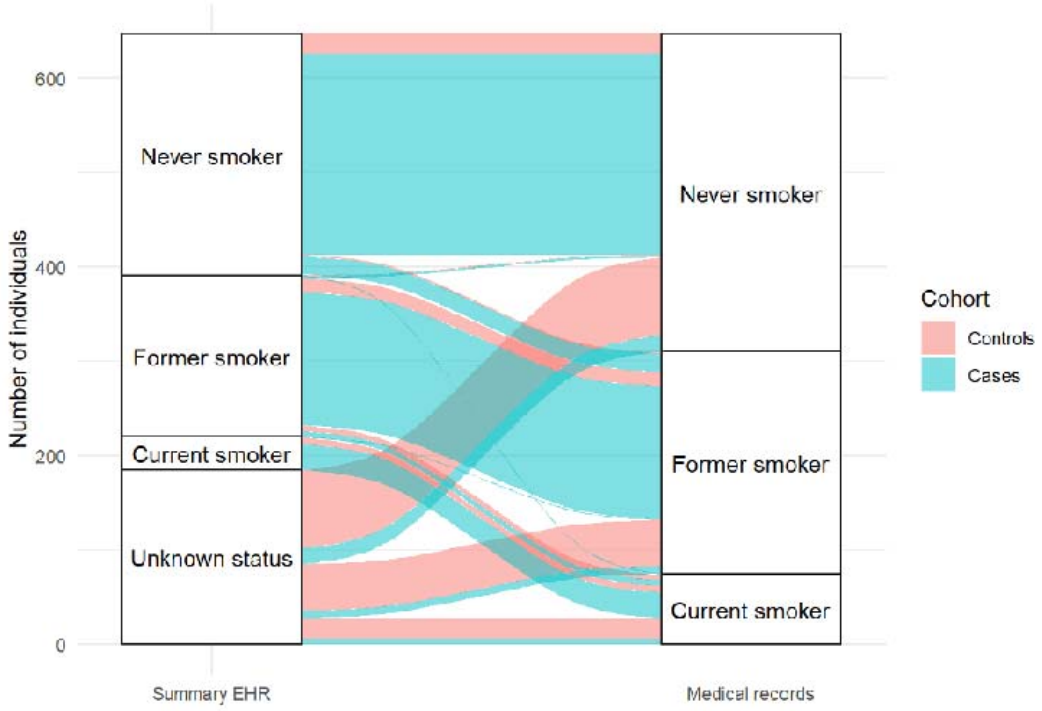

Figure 2. Concordance between smoking status recorded on the summary EHR and the medical notes for controls (red) and cases (blue).

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

4. Discussion

This observational case-control study with patients admitted to a single UK hospital trust found a lower proportion of current smokers in cases hospitalised with COVID-19 during the first phase of the pandemic compared with controls hospitalised with other respiratory viral infections a year previous. Further, we found that smoking status is typically poorly recorded in the summary EHR. This was more prominent in controls than cases – a difference that is likely explained by the observation that COVID-19 patients were followed up by the respiratory medicine team after discharge as part of a COVID-19 follow-up clinic where they specifically asked about smoking status (Mandal et al., 2020). The observed discrepancy between smoking status recorded on summary EHRs and the contemporaneous medical notes is a concern, particularly for studies relying solely on EHRs as the source of information on smoking status.

4.1. Strengths and limitations

To our knowledge, this is one of few studies specifically designed to examine the association between smoking status and hospitalisation with COVID-19. It was further strengthened by an assessment of the quality of data on smoking status gleaned from summary EHRs.

However, this study has several important limitations, the majority of which pertain to the selection of the controls. First, current smoking is expected *a priori* to be associated with hospitalisation for non-COVID-19 respiratory viruses (Stämpfli and Anderson, 2009). Ideally, hospital-based case-control studies should avoid selecting a control disease which is associated with the exposure of interest (i.e. smoking status) (Vandenbroucke and Pearce, 2012). However, to our knowledge, there is no other control disease with a similar route of acquisition and mechanism for hospitalisation/severe disease that is not *a priori* also associated with smoking status. The greater smoking prevalence in controls compared with the general population from which the cases emerged (Vandenbroucke and Pearce, 2012) therefore likely contributes to the significantly reduced odds of current smoking in our cases.

Second, the risk profile for controls likely differs from cases in that there is prior immunity to other respiratory viruses (e.g. influenza, respiratory syncytial virus), with no prior immunity in the population to SARS-CoV-2.

Third, we selected the controls on the basis of sharing a similar route of transmission and risk factors for hospitalisation as cases. However, at the time of writing (March 2021), we now suspect that COVID-19 differs from other respiratory viruses in several ways. For example, COVID-19 gains cell entry via the ACE-2 receptor (Hoffmann et al., 2020) – with unknown receptor binding in flu (Killingley and Nguyen-Van-Tam,

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2013) – and appears to display less fomite and physical contact transmission than flu (Ben-Shmuel et al., 2020). In addition, emerging evidence suggests that COVID-19 has a significantly different pathological process compared with other respiratory viruses. For example, mortality rates from COVID-19 differ widely from those due to epidemic influenza (Office for National Statistics, 2020a). Although we currently do not know the importance of these factors, taken together, emerging observations may mean that direct comparison of risk profiles in cases and controls is limited.

Fourth, while no known behavioural restrictions were implemented during the control period, London was under lockdown restrictions from March to July 2020, which likely impacted the risk of viral exposure in cases (Davies et al., 2020). This may further have impacted the different risk profiles of controls and cases beyond the adjustments made in this analysis for sex, age and SEP.

Fifth, the selection of historical controls may mean that there are non-trivial differences in smoking status between controls and cases due to a declining trend in London smoking prevalence (Office for National Statistics, 2020b). However, as a single year was used for the selection of controls, and there was no large change in national smoking prevalence from 2019 to 2020 in England (Brown, 2020), we expect any impact of the time-varying exposure to be minimal. We considered using a contemporaneous control (i.e. patients hospitalised with other respiratory viral infections in 2020), which would have mitigated against this potential bias. However, due to factors such as reduced national and international travel, physical distancing, increased hand hygiene and potential viral dominance by COVID-19, exposure to and hospitalisation with other respiratory viruses has been substantially reduced in 2020 (GOV.UK, 2020), which would have limited the sample size for controls.

Sixth, a history of current or past cancer was high in both groups at greater than 20% and was significantly greater in controls compared with cases. This reflects a bias in the population that regularly interacts with the selected NHS hospital trust, which is a specialist cancer referral centre. We visualised the geographic regions where patients were admitted from to examine any systemic differences between cases and controls, and caution that the differing catchment areas of the two cohorts may have led to important differences in the underlying populations. In addition, during the peak of the first wave of the pandemic in the UK, many cases were transferred across hospital sites due to bed pressures.

Finally, there was a greater proportion of cases than controls with no record of smoking status on either the summary EHR or within the contemporaneous medical notes (thus, they were not eligible for inclusion), with patients with missing data having unusually high mortality. It is plausible that many of them were smokers. We attempted to mitigate against this by conducting several sensitivity analyses, with the results largely remaining robust. However, when those excluded from the cohort due to missing data on smoking

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

status were treated as current smokers, there was no significant association between current smoking and hospitalisation for COVID-19.

Despite these limitations, alternative designs were impracticable or would have had different limitations. In the future, the current study can be considered alongside findings across multiple such alternative methodological approaches, each with different sources of bias, to triangulate on the extent to which associations between smoking and COVID-19 are causal.

4.2. Implications for policy and practice

COVID-19 will continue to place a large burden on healthcare services in the UK and internationally over the coming months and years. To mitigate against this, multiple non-pharmacological interventions are being implemented to reduce the intensity of demand on acute and intensive services. Irrespective of any direct link between smoking and COVID-19 disease outcomes, smoking is a significant cause for healthcare demand globally. We have argued elsewhere for the need to ramp up smoking cessation support to reduce the current and future burden on healthcare and social services (Simons et al., 2020).

4.3. Avenues for future research

The selection of appropriate controls in hospital-based case-control studies is very challenging for a novel respiratory virus such as COVID-19 (which means we converged on a hybrid approach, combining elements from hospital-based case series and case-control designs with historical controls). We recommend the use of representative population-studies with data from multiple sites and with purposeful acquisition of smoking status to better understand the role of smoking as a potential risk or protective factor for COVID-19 hospitalisation and disease severity.

5. Conclusion

In a single hospital trust in the UK, current smokers had reduced odds of being hospitalised with COVID-19 compared with other respiratory viruses a year previous, although we caution against interpreting this as a causal association. Smoking status was poorly recorded, with high observed discordance between smoking status recorded on the summary EHR and the contemporaneous medical notes.

Data sharing

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

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

Contributors

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

Conflict of interest

DS, OP and RB have no conflicts of interest to declare. LS has received a research grant and honoraria for a talk and travel expenses from manufacturers of smoking cessation medications (Pfizer and Johnson & Johnson). JB has received unrestricted research funding to study smoking cessation from companies who manufacture smoking cessation medications. All authors declare no financial links with tobacco companies or e-cigarette manufacturers or their representatives.

Acknowledgements

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

Role of funding source

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

Ethical approval

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

This study was approved by the UCL/UCLH Joint Research Office Research Strategy Group and UCLH Data Access Committee. Approval to conduct research limited to pseudonymised patient data was provided by the NHS Health Research Authority (IRAS_282704). The requirement for informed consent was waived due to the nature of the study.

6. References

Alqahtani, J.S., Oyelade, T., Aldhahir, A.M., Alghamdi, S.M., Almehmadi, M., Alqahtani, A.S., Quaderi, S., Mandal, S., Hurst, J.R., 2020. Prevalence, Severity and Mortality associated with COPD and Smoking in patients with COVID-19: A Rapid Systematic Review and Meta-Analysis. PLoS ONE 15. https://doi.org/10.1371/journal.pone.0233147

Benowitz, N.L., Schultz, K.E., Haller, C.A., Wu, A.H.B., Dains, K.M., Jacob, P., 2009. Prevalence of smoking assessed biochemically in an urban public hospital: a rationale for routine cotinine screening. Am. J. Epidemiol. 170, 885–891. https://doi.org/10.1093/aje/kwp215

Ben-Shmuel, A., Brosh-Nissimov, T., Glinert, I., Bar-David, E., Sittner, A., Poni, R., Cohen, R., Achdout, H., Tamir, H., Yahalom-Ronen, Y., Politi, B., Melamed, S., Vitner, E., Cherry, L., Israeli, O., Beth-Din, A., Paran, N., Israely, T., Yitzhaki, S., Levy, H., Weiss, S., 2020. Detection and infectivity potential of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) environmental contamination in isolation units and quarantine facilities. Clin. Microbiol. Infect. 26, 1658–1662. https://doi.org/10.1016/j.cmi.2020.09.004

Brake, S.J., Barnsley, K., Lu, W., McAlinden, K.D., Eapen, M.S., Sohal, S.S., 2020. Smoking Upregulates Angiotensin-Converting Enzyme-2 Receptor: A Potential Adhesion Site for Novel Coronavirus SARS-CoV-2 (Covid-19). J Clin Med 9, Page 841, 841. https://doi.org/10.3390/jcm9030841

Brown, J., 2020. The Smoking Toolkit Study, 2020. Google Docs.

Cai, G., 2020. Bulk and Single-Cell Transcriptomics Identify Tobacco-Use Disparity in Lung Gene Expression of ACE2, the Receptor of 2019-nCov. https://doi.org/10.20944/preprints202002.0051.v3; http://web.archive.org/web/20200925104805/https://www.preprints.org/manuscript/202002.0051/v3

Davies, N.G., Kucharski, A.J., Eggo, R.M., Gimma, A., Edmunds, W.J., Jombart, T., O'Reilly, K., Endo, A., Hellewell, J., Nightingale, E.S., Quilty, B.J., Jarvis, C.I., Russell, T.W., Klepac, P., Bosse, N.I., Funk, S., Abbott, S., Medley, G.F., Gibbs, H., Pearson, C.A.B., Flasche, S., Jit, M., Clifford, S., Prem, K., Diamond, C., Emery, J., Deol, A.K., Procter, S.R., Zandvoort, K. van, Sun, Y.F., Munday, J.D., Rosello, A., Auzenbergs, M., Knight, G., Houben, R.M.G.J., Liu, Y., 2020. Effects of non-pharmaceutical interventions on COVID-19 cases, deaths, and demand for hospital services in the UK: a modelling study. Lancet Public Health 5, e375–e385. https://doi.org/10.1016/S2468-2667(20)30133-X

Department for Communities and Local Government, 2019. The English Indices of Deprivation 2019 31.

Falsey, A.R., McElhaney, J.E., Beran, J., van Essen, G.A., Duval, X., Esen, M., Galtier, F., Gervais, P., Hwang, S.-J., Kremsner, P., Launay, O., Leroux-Roels, G., McNeil, S.A., Nowakowski, A., Richardus, J.H., Ruiz-Palacios, G., St Rose, S., Devaster, J.-M., Oostvogels, L., Durviaux, S., Taylor, S., 2014. Respiratory Syncytial Virus and Other Respiratory Viral Infections in Older Adults With Moderate to Severe Influenza-like Illness. J. Infect. Dis. 209, 1873–1881. https://doi.org/10.1093/infdis/jit839

Fang, L., Karakiulakis, G., Roth, M., 2020. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? Lancet Respir Med 0. https://doi.org/10.1016/s2213-2600(20)30116-8.

GOV.UK, 2020. Weekly national flu and COVID-19 surveillance reports published [WWW Document]. GOV.UK. URL https://www.gov.uk/government/news/weekly-national-flu-and-covid-19-surveillance-reports-published (accessed 3.11.21).

Grimes, D.A., Schulz, K.F., 2005. Compared to what? Finding controls for case-control studies. The Lancet 365, 1429–1433. https://doi.org/10.1016/S0140-6736(05)66379-9

Guan, W., Liang, W., Zhao, Y., Liang, H., Chen, Zi-sheng, Li, Y., Liu, X., Chen, R., Tang, C., Wang, T., Ou, C., Li, L., Chen, P., Sang, L., Wang, W., Li, J., Li, C., Ou, L., Cheng, B., Xiong, S., Ni, Z., Xiang, J., Hu, Yu, Liu, L., Shan, H., Lei, C., Peng, Y., Wei, L., Liu, Y., Hu, Ya-hua, Peng, P., Wang, J., Liu, J., Chen, Zhong, Li,

medRxiv preprint doi: https://doi.org/10.1101/2020.11.26.20238469; this version posted March 20, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

G., Zheng, Z., Qiu, S., Luo, J., Ye, C., Zhu, S., Cheng, L., Ye, F., Li, S., Zheng, J., Zhang, N., Zhong, N., He, J., 2020. Comorbidity and its impact on 1590 patients with COVID-19 in China: a nationwide analysis. Eur. Respir. J. 55. https://doi.org/10.1183/13993003.00547-2020

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.-H., Nitsche, A., Müller, M.A., Drosten, C., Pöhlmann, S., 2020. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell 181, 271-280.e8. https://doi.org/10.1016/j.cell.2020.02.052

Johns Hopkins Coronavirus Resource Center, 2021. COVID-19 Map [WWW Document]. Johns Hopkins Coronavirus Resour. Cent. URL https://coronavirus.jhu.edu/map.html (accessed 3.11.21).

Killingley, B., Nguyen-Van-Tam, J., 2013. Routes of influenza transmission. Influenza Other Respir. Viruses 7 Suppl 2, 42–51. https://doi.org/10.1111/irv.12080

Lauer, S.A., Grantz, K.H., Bi, Q., Jones, F.K., Zheng, Q., Meredith, H.R., Azman, A.S., Reich, N.G., Lessler, J., 2020. The Incubation Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed Cases: Estimation and Application. Ann. Intern. Med. 172, 577–582. https://doi.org/10.7326/M20-0504

Mandal, S., Barnett, J., Brill, S.E., Brown, J.S., Denneny, E.K., Hare, S.S., Heightman, M., Hillman, T.E., Jacob, J., Jarvis, H.C., Lipman, M.C.I., Naidu, S.B., Nair, A., Porter, J.C., Tomlinson, G.S., Hurst, J.R., 2020. 'Long-COVID': a cross-sectional study of persisting symptoms, biomarker and imaging abnormalities following hospitalisation for COVID-19. Thorax. https://doi.org/10.1136/thoraxjnl-2020-215818

McCarthy, N., Giesecke, J., 1999. Case-case comparisons to study causation of common infectious diseases. Int. J. Epidemiol. 28, 764–768. https://doi.org/10.1093/ije/28.4.764

Oakes, J.M., Fuchs, R.M., Gardner, J.D., Lazartigues, E., Yue, X., 2018. Nicotine and the renin-angiotensin system. Am J Physiol - Regul Integr Comp Physiol 315, 895–906. https://doi.org/10.1152/ajpregu.00099.2018

Office for National Statistics, 2020a. Deaths due to coronavirus (COVID-19) compared with deaths from influenza and pneumonia, England and Wales [WWW Document]. URL https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/deathsduetocoronaviruscovid19comparedwithdeathsfrominfluenzaandpneumoniaenglandandwales/deathsoccurringbetween1januaryand31august2020 (accessed 10.29.20).

Office for National Statistics, 2020b. Adult smoking habits in the UK: 2019, Annual Population Survey. Office for National Statistics.

Peralta, P.S.-O., Cortes-García, M., Vicente-Herrero, M., Castrillo-Villamandos, C., Arias-Bohigas, P., Amo, I.P., Sierra-Moros, M.J., Spain, on behalf of the S.G. for N.I.A.V.I. and C.T. in, 2010. Risk factors for disease severity among hospitalised patients with 2009 pandemic influenza A (H1N1) in Spain, April – December 2009. Eurosurveillance 15, 19667. https://doi.org/10.2807/ese.15.38.19667-en

Polubriaginof, F., Salmasian, H., Albert, D.A., Vawdrey, D.K., 2018. Challenges with Collecting Smoking Status in Electronic Health Records. AMIA. Annu. Symp. Proc. 2017, 1392–1400.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Secondary Care Analytical Team, 2020. Hospital Admitted Patient Care Activity 2019-20, Hospital Admitted Patient Care Activity.

Simons, D., Perski, O., Brown, J., 2020. Covid-19: The role of smoking cessation during respiratory virus epidemics. The BMJ.

Simons, D., Shahab, L., Brown, J., Perski, O., 2021. The association of smoking status with SARS-CoV-2 infection, hospitalisation and mortality from COVID-19: A living rapid evidence review with Bayesian meta-analyses (version 11). Qeios. https://doi.org/10.32388/UJR2AW.13

Stämpfli, M.R., Anderson, G.P., 2009. How cigarette smoke skews immune responses to promote infection, lung disease and cancer. Nat. Rev. Immunol. 9, 377–384. https://doi.org/10.1038/nri2530

Vandenbroucke, J.P., Pearce, N., 2012. Case–control studies: basic concepts. Int. J. Epidemiol. 41, 1480–1489. https://doi.org/10.1093/ije/dys147

World Health Organisation, 2019. ICD-10 [WWW Document]. URL https://icd.who.int/browse10/2019/en#/U07.1 (accessed 10.28.20).