**Case fatality risk of the SARS-CoV-2 variant of concern B.1.1.7 in England**

**Version:** v1.1

**Date:** 10 May 2021

|  |  |  |
| --- | --- | --- |
| Version history | Date | Comment |
| 0.1 | 1 Feb 2021 | Initial draft created |
| 0.2 | 8 Feb 2021 | Refinements following comments received |
| 0.3 | 11 Feb 2021 | Addition of causal framework and DAG section. Decided to drop absolute risk difference models in favour of relative risk models. Absolute risk to be calculated from predicted values from relative risk models or logistic regression if convergence issues. |
| 0.4 | 19 Feb 2021 | Extended study period by one week |
| 1.0 | 22 Feb 2021 | Added references and finalised |
| 1.1 | 10 May 2021 | Addendum detailing further analyses of hospital admission and death given hospital admission |

**Contents**

[Background 3](#_Toc77257162)

[Objectives 3](#_Toc77257163)

[Study Design and Population 3](#_Toc77257164)

[Inclusion criteria 4](#_Toc77257165)

[Exclusion criteria 4](#_Toc77257166)

[Causal framework 4](#_Toc77257167)

[Study Measures 5](#_Toc77257168)

[Exposure 5](#_Toc77257169)

[Outcomes 5](#_Toc77257170)

[Covariates 5](#_Toc77257171)

[Statistical Methods 8](#_Toc77257172)

[Baseline characteristics 8](#_Toc77257173)

[28-day all-cause mortality 8](#_Toc77257174)

[Cox proportional hazards regression 8](#_Toc77257175)

[Covariate adjustment 8](#_Toc77257176)

[Defining regions 9](#_Toc77257177)

[*A priori* subgroup analyses 9](#_Toc77257178)

[Sensitivity and additional analyses 10](#_Toc77257179)

[Software and reproducibility 10](#_Toc77257180)

[Feasibility and power calculations 10](#_Toc77257181)

[Strengths and Limitations 10](#_Toc77257182)

[Risk models fail to converge 10](#_Toc77257183)

[Ethics 11](#_Toc77257184)

[Conflicts of Interests 11](#_Toc77257185)

[References 11](#_Toc77257186)

[Addendum: Analysis of hospital admissions data 12](#_Toc77257187)

## 

Background

The SARS-CoV-2 (COVID-19) variant of concern B.1.1.7 (VOC) was first identified in Kent, UK in autumn 2020. Early analysis suggests the VOC is more transmissible and it has since become the dominant strain throughout the UK. Only a small number of VOC cases are identified by whole-genome sequencing. Spike gene target failure (SGTF) has been adopted as a proxy for identifying VOC and has been shown to identify the VOC in more than 95% of cases during the period 16th November – 11th January.1

Studies using Public Health England (PHE) line listing data, hospital admissions and ONS death data have assessed the relative fatality of the VOC compared to the originally circulating viral strain (non-VOC) and have consistently demonstrated an increase in mortality associated with the VOC.2,3 Although these studies were able to account for age, sex, ethnicity, deprivation index, time period and geographical area, they were unable to account for comorbidities, which have been shown to be strongly associated with death among those diagnosed with COVID-19.4

### Objectives

To estimate the risk of death following confirmation of SARS-CoV-2 infection, comparing the risk among those infected with the VOC to those infected with the non-VOC accounting for both demographic factors and comorbidities. The risk of death will be quantified using the following methods:

1. A relative risk model, estimating relative and absolute 28-day all-cause mortality risk
2. A Cox Proportional hazards regression model, estimating a hazard ratio

Study Design and Population

We will use a cohort study design nested within the OpenSAFELY platform. Using test result data from the PHE Second Generation Surveillance System (SGSS), we will select all those people who are:

1. positive for SARS-CoV-2 based on PCR swab test results in the time window 16st November to 11th January and
2. have data on SGTF

The study will focus on the comparison between those with SGTF and those without. Inconclusive SGTF results are considered in the sensitivity and additional analyses section.

The primary analysis will focus on a 28-day all-cause mortality outcome. All-cause mortality will be determined from ONS death data and is expected to have full ascertainment of deaths with a 2-week delay. Therefore, the 28-day all-cause mortality analysis will include all individuals with at least 42-days follow-up from the date of COVID-19 diagnosis to the date of last ONS death data upload (28-days plus 14-days to account for the delay in the ONS death data).

For Cox proportional hazards analysis all individuals will be included. Follow-up will be censored two weeks prior to the date of ONS death data upload for those without a documented date of death.

### Inclusion criteria

* A positive SARS-CoV-2 PCR swab test result in SGSS within the window 16th November to 11th January
* Data available on SGTF in SGSS.
* Registered with a primary care practice using The Phoenix Partnership (TPP) software on the date of COVID-19 diagnosis, with at least one year of continuous GP registration.

### Exclusion criteria

* Missing age, sex, or index of multiple deprivation, as these are indicators of poor data quality.
* COVID-19 diagnoses prior to the diagnosis in the study time window (based on either a positive test for SARS-CoV-2 in SGSS data or a diagnosis for COVID-19 in primary care).
* Receipt of vaccination against COVID-19 prior to diagnosis in the study time window.

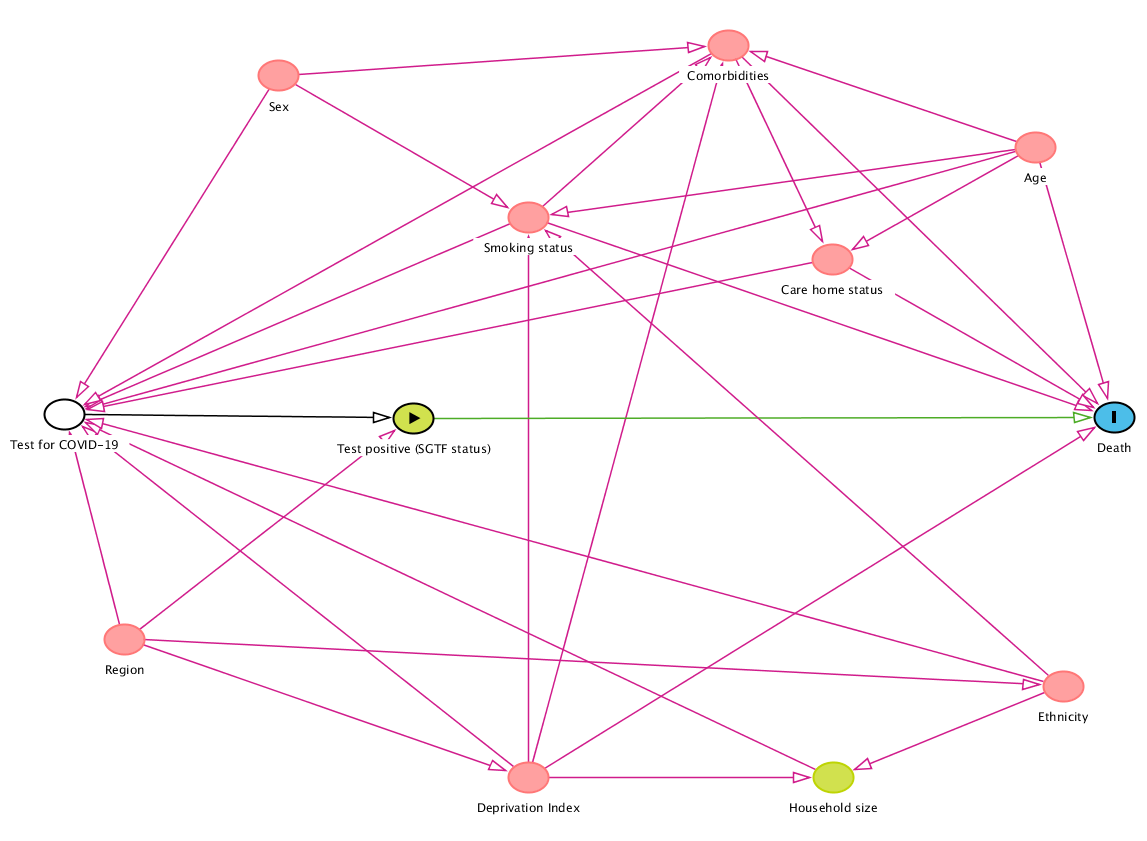
### 

### Causal framework

The motivation for adjusting for demographics and comorbidities is not that they impact on the variant of COVID-19 infection *per se*, but are likely to be associated with the upstream process of getting a test (e.g. test-seeking behaviour, ability to access testing facilities). Therefore, adjustment attempts to correct for imbalances between the VOC and non-VOC exposure groups with respect to factors associated with getting a test. With the study population defined by SARS-CoV-2 positive test and SGTF data available, the minimum sufficient adjustment set implied by **Figure 1** is:

Age, care home status, comorbidities, deprivation index, smoking status.

**Figure 1 Causal framework DAG**



Study Measures

### Exposure

SGTF on SARS-CoV-2 PCR swab test from SGSS data, referred to as the VOC exposure group. The comparator group being SARS-CoV-2 diagnoses without SGTF in SGSS data, referred to as the non-VOC group.

### Outcomes

Death from any cause.

### Covariates

Age, sex, deprivation index, ethnicity, smoking status, household size.

Region, defined by middle layer super output area (MSOA) from patient post code, or NHS England region.

Rural and urban location classification, and care home status.

Epidemiological week of the positive test.

Comorbidities: obesity, hypertension, chronic respiratory diseases other than asthma, chronic heart disease, diabetes, non-haematological and haematological cancer, reduced kidney function, chronic liver disease, stroke, dementia, other neurological disease, organ transplant, asplenia, rheumatoid arthritis/lupus, psoriasis, and other immunosuppressive conditions.

1. **Study measures**

|  |  |
| --- | --- |
| Exposures | OpenCodelists Definition |
| SGTF | PHE Second Generation Surveillance System (SGSS) |
| Outcomes |  |
| Death | All-cause registered deaths, from ONS |
| Covariates |  |
| Ethnicity | <https://codelists.opensafely.org/codelist/opensafely/ethnicity/> |
| Region | MSOA and STP are extracted from patient post code  UTLA regions are found using these look-up tables:  <https://geoportal.statistics.gov.uk/datasets/lower-layer-super-output-area-2011-to-upper-tier-local-authorities-2019-lookup-in-england-and-wales->  <https://geoportal.statistics.gov.uk/datasets/9f4c270148014f20bf24abff9a7aef62_0> |
| Comorbidities | <https://codelists.opensafely.org/codelist/opensafely/aplastic-anaemia/>  <https://codelists.opensafely.org/codelist/opensafely/asplenia/>  <https://codelists.opensafely.org/codelist/opensafely/asthma-diagnosis/>  <https://codelists.opensafely.org/codelist/opensafely/asthma-inhaler-salbutamol-medication/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/asthma-inhaler-steroid-medication/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/asthma-oral-prednisolone-medication/2020-04-27/>  <https://codelists.opensafely.org/codelist/opensafely/bone-marrow-transplant/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/cancer-excluding-lung-and-haematological/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/chemotherapy-or-radiotherapy-updated/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/chronic-cardiac-disease/2020-04-08/>  <https://codelists.opensafely.org/codelist/opensafely/chronic-respiratory-disease/2020-04-10/>  <https://codelists.opensafely.org/codelist/opensafely/chronic-liver-disease/2020-06-02/>  <https://codelists.opensafely.org/codelist/opensafely/dementia/2020-04-22/>  <https://codelists.opensafely.org/codelist/opensafely/diabetes/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/chronic-kidney-disease/2020-04-14/>  <https://codelists.opensafely.org/codelist/opensafely/gi-bleed-or-ulcer/2020-04-08/>  <https://codelists.opensafely.org/codelist/opensafely/haematological-cancer/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/hiv/2020-07-13/>  <https://codelists.opensafely.org/codelist/opensafely/hypertension/2020-04-28/>  <https://codelists.opensafely.org/codelist/opensafely/inflammatory-bowel-disease/2020-04-07/>  <https://codelists.opensafely.org/codelist/opensafely/lung-cancer/2020-04-15/>  <https://codelists.opensafely.org/codelist/opensafely/other-neurological-conditions/2020-06-02/>  <https://codelists.opensafely.org/codelist/opensafely/permanent-immunosuppression/2020-06-02/>  <https://codelists.opensafely.org/codelist/opensafely/ra-sle-psoriasis/2020-04-14/>  <https://codelists.opensafely.org/codelist/opensafely/sickle-cell-disease/2020-04-14/>  <https://codelists.opensafely.org/codelist/opensafely/smoking-clear/2020-04-29/>  <https://codelists.opensafely.org/codelist/opensafely/smoking-unclear/2020-04-29/>  <https://codelists.opensafely.org/codelist/opensafely/solid-organ-transplantation/2020-04-10/>  <https://codelists.opensafely.org/codelist/opensafely/stroke-updated/2020-06-02/>  <https://codelists.opensafely.org/codelist/opensafely/temporary-immunosuppression/2020-04-24/> |

Statistical Methods

### Baseline characteristics

Participant characteristics, including all covariates listed above, will be described at baseline (the date of positive SARS-CoV-2 test), comparing the two exposure groups (VOC and non-VOC). Continuous variables will be summarised by the mean and standard deviation and compared with a t-test, or median and interquartile range and Wilcoxon signed-rank test, as appropriate. Categorical variables will be summarized by the number and proportion in each group (n (%)) and compared with a chi-square test.

The median time-to-death and interquartile range of those who die will be presented by exposure.

The proportion of SARS-CoV-2 positive tests with SGTF, identifying the VOC, will be plotted over the study period by NHS England region and descriptively compared to PHE data for the whole population.

### 28-day all-cause mortality

Case fatality risk will be calculated at 28-days post SARS-CoV-2 positive test result. Therefore, only those with 28-days of follow-up or a date of death within this window will be included in this analysis.

The relative risk for VOC cases vs. non-VOC will be calculated from a generalised linear regression model with binomial distribution and log link function. Absolute risk will be estimated by the predicted risk from this model.

**POST ANALYSIS NOTE:** there were model convergence issues when working on the risk scale. As per the limitations section, 28-day risk was therefore calculated from a logistic regression model.

### Cox proportional hazards regression

The relative hazard of death for SGTF cases vs. non-SGTF will be calculated from a Cox proportional hazards regression model, with no requirement for 28-days follow-up time. Follow-up will be censored at the earliest of two weeks prior to the date of ONS death data upload or 7 days prior to receipt of a COVID-19 vaccination.

The hazard of death following a SARS-CoV-2 positive test result is expected to vary considerably between regions in England over time. Consequently, adjustment for region is unlikely to satisfy the proportional hazards assumption of a Cox model. To account for this variability, we will stratify the analysis on region, allowing a separate baseline hazard to be estimated for each region, but with covariate effects estimated over the full population – a stratified Cox PH model. The definition of regions is discussed below.

### Covariate adjustment

Unadjusted, demographically adjusted, and fully adjusted estimates will be presented for each analysis.

Demographically adjusted models will include adjustment for the following covariates: Age will be included as a cubic spline term. Ethnicity will be grouped into five categories. The primary analysis will exclude patients with missing ethnicity. Sex, deprivation index, household size, and type of residence will be included as categorical terms.

Epidemiological week of the baseline SARS-CoV-2 positive test will be included as a categorical variable.

Region will be defined by UTLA, unless data sparsity prevents this level of granularity. In which case region will be defined by STP, or aggregated geographical areas defined by NHS England region. Rural or urban location classification will be included as a categorical variable with 5 levels in line with other work.

Fully adjusted models will additionally adjust for patient comorbidities, smoking status, and obesity status. Comorbidities will be aggregated into a categorical term taking values none, 1, and 2 or more. In line with previous work on the risk of death from COVID-19 on the OpenSAFELY platform. For smoking and obesity, missing values will be categorised as never smoked and no evidence of obesity, in line with previous OpenSAFELY studies.4,5

The causal framework indicates both these adjustment sets result in a causal estimate of the effect of the VOC on mortality. For comparison, we will also fit a model using the minimum sufficient adjustment set implied by the causal DAG (Age, comorbidities, deprivation index, smoking status).

### Defining regions

Regional stratification will be a key consideration due to variability in the incidence of COVID-19 outcomes over time. Regions will be defined using patient middle super output area (MSOA) codes derived from patient post codes. MSOA data will be aggregated into upper tier local authority areas (UTLA) which will be the primary definition of regions for analysis.

Sensitivity analysis will define region by sustainability and transformation partnership (STP) areas, also defined from patient post codes. This analysis will assess the impact of regional definitions on the estimated risk and hazard of death for SGTF cases vs. non-SGTF.

Should data sparsity preclude regional adjustment at the UTLA and STP level, aggregated geographical areas defined by NHS England region will be used instead.

### *A priori* subgroup analyses

Case fatality relative and absolute risk will be estimated in subgroups of *a priori* interest, after adjustment for confounding. Differences in risk in these subgroups will be formally tested with a likelihood ratio test for an interaction with SGTF exposure status.

The subgroups of interest to be assessed are:

* Age group (<50; 50-64; 65-74; 75-84; 85+)
* Ethnicity (in 5 categories)
* Comorbidity status (none, 1, 2+)
* Deprivation index (deciles of deprivation)
* Epidemiological week of positive SARS-CoV-2 test (each week of the study period, fortnightly if data are sparse)
* NHS England region (East, London, South East, South West, Midlands, North East and Yorkshire, North West)

### Sensitivity and additional analyses

40-day all-cause mortality

Differential time from positive SARS-CoV-2 test to death by SGTF exposure status has the potential to bias the analysis of risk. An additional analysis will consider an increase in the risk period to 40-days to assess the sensitivity of the findings to the risk period definition.

Inconclusive SGTF results

SGTF flags will be inconclusive in some cases. SGTF data are expected to take the values yes, no, maybe, unknown. The primary analysis will focus on the comparison of the yes group (VOC) with the no group (non-VOC). In additional analysis the risk of death in the maybe and unknown groups will also be quantified and compared to that of the VOC and non-VOC exposure groups.

Multiple imputation of missing ethnicity

Previous work in OpenSAFELY has identified that ethnicity data are missing for up to one quarter of all patients. The primary analysis will use the complete case set with regards to ethnicity. This sensitivity analysis will assess the impact of excluding records with missing ethnicity by imputing missing ethnicity using multiple imputation based on all variables included in the full adjustment set.

### Software and reproducibility

Data management will be performed using Python and Google BigQuery, with analysis carried out using Stata 16.1 / Python. Code for data management and analysis as well as codelists archived online <https://github.com/opensafely/sgtf-cfr-research>.

### 

### Feasibility and power calculations

To be assessed when SGTF data are available.

Strengths and Limitations

### Risk models fail to converge

Although the size of our study population will likely be considerable with a large number of deaths, it’s possible that the models of risk may fail to converge due to data sparsity in some covariate subsets. If this is the case we will revert to a logistic regression approach, estimating the log odds ratio. Inference on the risk of death will then be performed by converting predicted odds to estimates of absolute risk.

### Non-random availability of SGTF data

Although the fact that we adjust for factors associated with getting tested should help account for possible non-random availability of SGTF data, we will also compare the characteristics of people included in the study (who all have SGTF data) with those not included in the study due to lack of SGTF data. This will help us assess whether those with SGTF are representative of all those tested during the time period of the study, and allow us to discuss the implications of this in our write-up as necessary.

Ethics

This study was approved by the Health Research Authority (REC reference 20/LO/0651) and by the LSHTM Ethics Board (reference 21863).

Conflicts of Interests

None to declare

References

1. Chand M. Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01 (PDF). *Public Health England PHE* 2020.

2. NERVTAG. NERVTAG paper on COVID-19 variant of concern B.1.1.7: Paper from the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) on new coronavirus (COVID-19) variant B.1.1.7. <https://www.gov.uk/government/publications/nervtag-paper-on-covid-19-variant-of-concern-b117>.

3. Iacobucci G. Covid-19: New UK variant may be linked to increased death rate, early data indicate. British Medical Journal Publishing Group; 2021.

4. Williamson E, Walker AJ, Bhaskaran K, et al. OpenSAFELY: factors associated with COVID-19-related hospital death in the linked electronic health records of 17 million adult NHS patients. *MedRxiv* 2020.  
5. Bhaskaran K, Bacon S, Evans S, et al. Factors associated with deaths due to COVID-19 versus other causes: population-based cohort analysis of UK primary care data and linked national death registrations within the OpenSAFELY platform. medRxiv 2021: 2021.01.15.21249756.

Addendum: Analysis of hospital admissions data

Expanding on the initial analysis of case fatality, analysis of the intermediate patient pathway outcomes of hospital admission, death given hospital admission, and death given ICU admission, we will use Secondary Uses Statistics (SUS) data to define hospital admission and intensive care unit (ICU) admission.

These analyses will follow the same methodology as described in the Statistical Methods section above (page 8), namely logistic regression to estimate 28-day risk and stratified Cox proportional hazards regression to estimate relative hazards. Covariate adjustment will be the same as detailed previously, with unadjusted, demographically adjusted, and fully adjusted estimates presented for each analysis.

For analysis of hospital admission following a positive SARS-CoV-2 test the causal diagram is the same as shown in **Figure 1**, but with hospital admission defined as the outcome rather than death. The causal diagram is the same as the previous analysis as the processes are all identical, the endpoint is just earlier along the patient pathway (**Figure A1**).

For analysis of death given hospital admission and death given ICU admission we introduce another condition on the study population, i.e. admission to hospital or ICU. As hospital admission and ICU admission lies on the causal pathway between testing positive for SARS-CoV-2 and death for most people these analyses no longer estimate the total causal effect of SGTF on death. Though they are suitable to estimate the direct causal effect of SGTF on death among these study populations, i.e. the causal effect of SGTF on death given hospital admission and the causal effect of SGTF on death given ICU admission (**Figure A2 and Figure A3**).

Figure A1 Causal framework DAG, outcome: Hospital admission

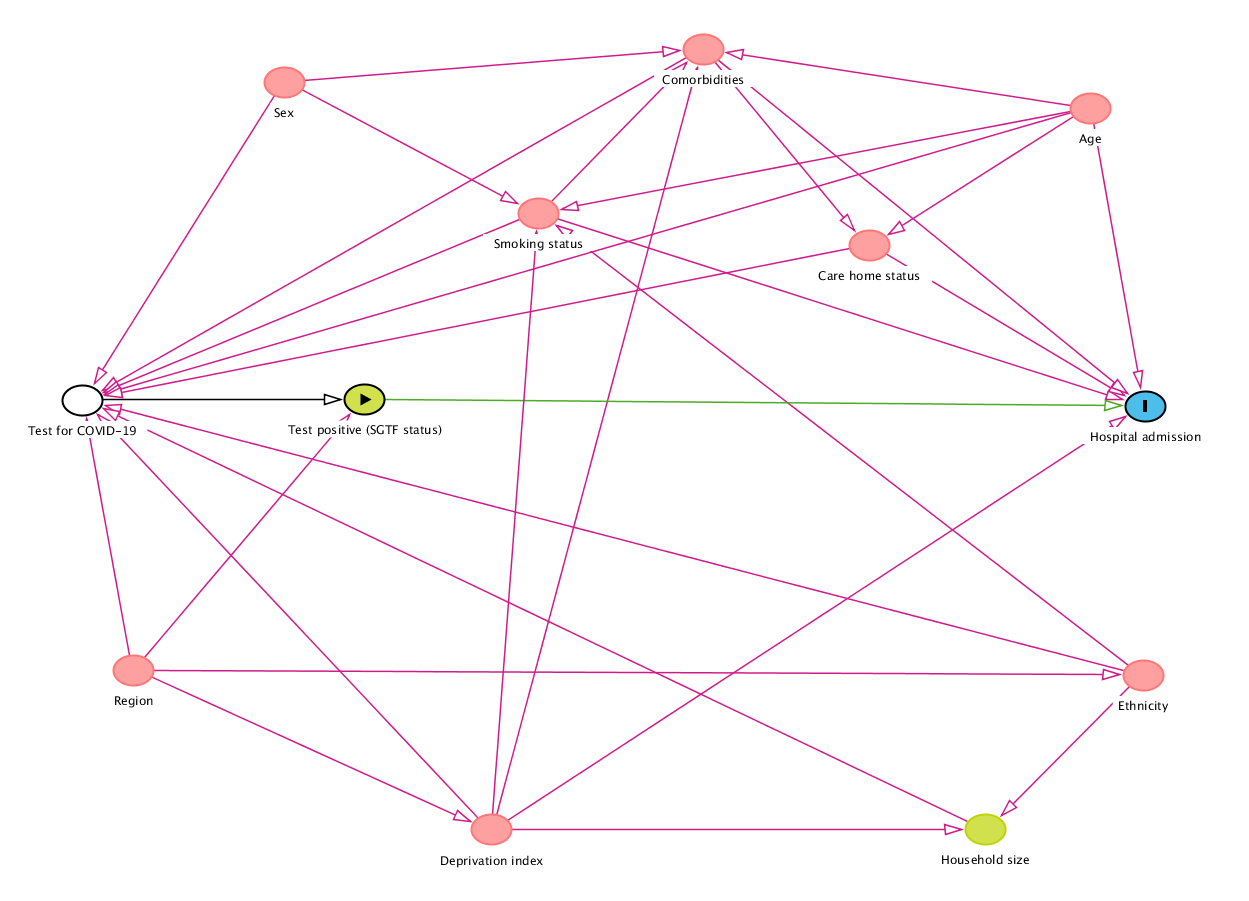


Figure A2 Causal framework DAG, outcome: Death given hospital admission

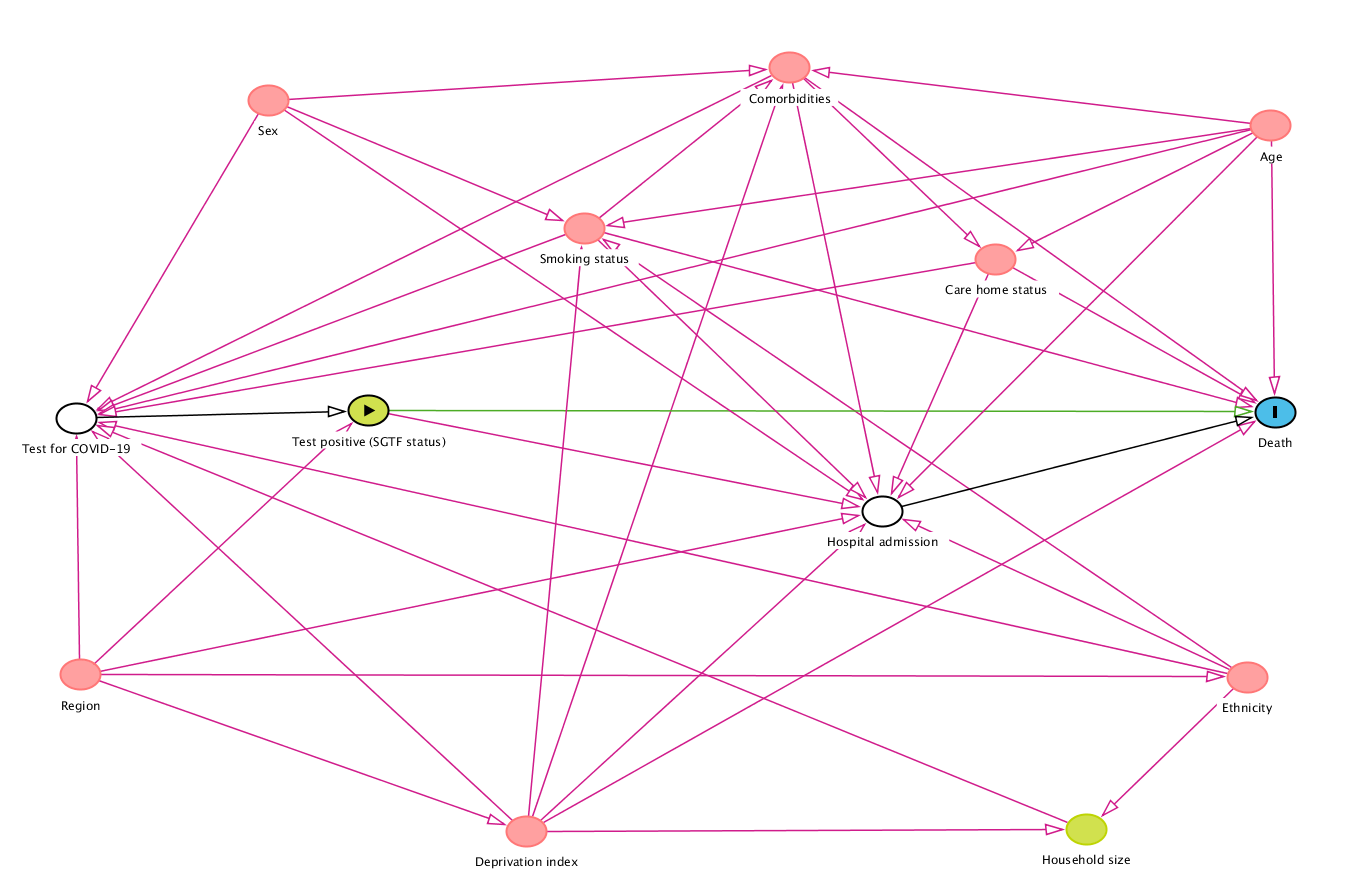


Figure A2 Causal framework DAG, outcome: Death given intensive care unit (ICU) admission

