Subgroup differences in COVID-19-related death in 17 million adults in England over time: an observational cohort study using the OpenSAFELY platform

Protocol

Background

During the peak of the first wave of the COVID-19 pandemic in April 2020, the Office for National Statistics (ONS) reported age-adjusted mortality rates for death due to COVID-19 of 626 and 495 per 100.000 people in England and Wales, respectively [1]. Despite high infection rates in December 2021, age-adjusted mortality rates have dropped to 56 and 59 in England and Wales, respectively. The decline in mortality rates is partly explained by measures such as social distancing, masking, shielding, improved clinical management and the roll out of a national vaccination programme.

An early report in June 2020 including 17 million patients in the UK, showed that COVID-19-related death was associated with being male, older age, deprivation, ethnicity, diabetes, severe asthma and various other medical conditions [2]. There is clear evidence that COVID-19-related death rates have dropped for the overall population, but it is largely unknown 1) to what extent mortality rates have changed over time in different subgroups of the population; and 2) if patterns in mortality rates over time are similar across different subgroups.

This study sets out to describe subgroup differences in COVID-19 related death rates over time (February 2020 - December 2021) and describe COVID-19 related death risk in different subgroups of the population in the three pandemic waves. Subgroups are based on age category, sex, bmi category, ethnicity, deprivation and comorbidities.

Methods

Study design and study population

A retrospective cohort study will be carried out within OpenSAFELY, a data analytics platform in England created to address urgent COVID-19-related questions. All data will be linked, stored and analysed securely within the OpenSAFELY platform: https://opensafely.org/. From 1st February 2020 - 31st December 2021, 22 monthly cohorts will be extracted of individuals aged 18 years or older registered with a TPP practice on the 1st of every month with at least one year of continuous GP registration prior to this date, to ensure that baseline

data could be adequately captured. We will exclude people with missing data on age, sex, or index of multiple deprivation, since these are likely to indicate poor data quality.
In addition, from 1st February 2020 - 31st December 2021, three cohorts will be extracted covering the three pandemic waves in England (March 23, 2020 - May 30, 2020 (wave 1) [3]; September 7, 2020 - April 24, 2021 (wave 2) [3]; and May 28, 2021 - December 14, 2021 (wave 3, delta)). The start date of the third pandemic wave (May 28, 2021) has been determined by the day that reported reproduction rates and growth rates in England were above 1 and positive, respectively (1.0-1.1 and 0-3, respectively) [4] and the end date (14 December 2021) was chosen as Omicron became dominant between 14 and 15 December [5]. Inclusion and exclusion criteria for these three cohorts will be identical to those listed for the 22 monthly cohorts above.

Outcomes

The outcome of interest is COVID-19-related mortality captured by the Office for National Statistics (ONS). COVID-19 deaths were defined as having an underlying or secondary cause of death listed as COVID-19 (ICD-10 codes U07.1 or U07.2).

Covariates

Covariates considered in the analysis will include health conditions listed in UK guidance on higher risk groups [6]; other common conditions that may cause immunodeficiency inherently or through medication, and other postulated risk factors for severe outcomes among COVID-19 cases. We will include age (grouped as 18-39, 40-49, 50-59, 60-69, 70-79 and ≥80 years for descriptive analysis), sex, ethnicity (White, Mixed, South Asian, Black, Other, categories from the UK census), body mass index (BMI) (categorised as not obese, class I [body mass index 30-34.9kg/m2], II [35-39.9kg/m2], III [≥40kg/m2]), smoking status (never, former, current), index of multiple deprivation quintile (derived from the patient's postcode at lower super output area level). We will also consider the following comorbidities: diagnosed hypertension, chronic respiratory diseases other than asthma, asthma (categorised as with or without recent use of oral steroids), chronic heart disease, diabetes (categorised according to the most recent glycated haemoglobin (HbA1c) recorded in the 15 months prior to the first day of the monthly cohort or pandemic wave), non-haematological and haematological cancer (both categorised by recency of diagnosis, <1, 1-4.9, ≥5 years), reduced kidney function (categorised by estimated glomerular filtration rate derived from the most recent serum creatinine measure (30-<60, 15-<30, <15 mL/min/1.73m2 or a record of dialysis), chronic liver disease, stroke, dementia, other neurological disease (motor neurone disease, myasthenia gravis, multiple sclerosis, Parkinson's disease, cerebral palsy, quadriplegia or hemiplegia, and progressive cerebellar disease), organ transplant, asplenia (splenectomy or a spleen dysfunction, including sickle cell disease), rheumatoid arthritis/lupus/psoriasis, and other immunosuppressive conditions (permanent immunodeficiency ever diagnosed, or aplastic anaemia or temporary immunodeficiency recorded within the last year). The Sustainability and Transformation Partnership (STP, an NHS administrative region) of the patient's general practice was included as an additional adjustment for geographical variation in infection rates across the country.

Missing data

Missing data will arise in BMI, smoking and ethnicity. Comorbidities will be coded as code present or not, thus no missing data will arise. There is likely to be some misclassification in ICD codes. In the primary analysis, those with missing BMI will be assumed to be non-obese and those with missing smoking information will be assumed to be non-smokers on the assumption that both obesity and smoking would be likely to be recorded if present. A sensitivity analysis will be run among those with complete BMI and smoking data only.

Statistical methods

Descriptive statistics

Patient numbers for wave 1, 2 and 3 will be depicted in a flowchart. Sociodemographic characteristics and clinical characteristics at baseline will be summarised using descriptive statistics, stratified by wave.

Monthly mortality rates

A crude COVID-19 mortality rate will be calculated for the period between February 2020 and December 2021, for females and males separately. Subsequently, subgroup specific (based on the covariates discussed in 'covariates') monthly age-adjusted COVID-19 mortality rates will be calculated for the period between February 2020 and December 2021, for females and males separately. Monthly mortality rates will be calculated by (the number of deaths in a specific month in a specific subgroup: the total number of individuals followed during that month in that specific subgroup) x 100,000. To account for age differences in the various subgroups of the population, mortality rates will be directly standardised using the European Standard Population using five-year age bands. Crude rates and standardised rates (DSR) will be scaled to a consistent month length of 30 days. The estimated monthly mortality rates (y-axis) will be plotted over time (x-axis). Different graphs are used to visualise change in mortality rates by age category (1 graph, 6 lines), sex (1 graph, 2 lines), bmi category (1 graph, 6 lines), ethnicity (1 graph, 6 lines), IMD quintile (1 graph, 5 lines). For the comorbidities, each comorbidity is plotted in a separate graph, comparing the individuals with the comorbidity to individuals without the comorbidity (16 graphs, 2 lines per graph).
Comparative Mortality Figures (CMF) will be calculated by taking the DSR among subgroup xyz (e.g., asthmatic individuals) and dividing these by the DSR among individuals not in subgroup xyz (e.g., non asthmatic individuals), for females and males separately. Confidence intervals for the CMF will be calculated using standard approaches [7]. For the variables with more than 2 categories, the first category will be used as the reference, except for age, where 50-59 is used as the reference. For visualisation, CMF (y-axis) will be plotted over time (x-axis), identical to explained above for the cruden rates and DSR.

Modelling pandemic waves

Crude COVID-19 mortality incidence rates will be calculated for the overall population and then stratified by subgroup (based on the covariates described above) for waves 1, 2 and 3. To account for age and sex differences in the various subgroups of the population, mortality rates will be directly age- and sex-standardised to the European Standard Population using five-year age bands. Confidence intervals will be obtained taking the normal approximation

to the binomial distribution. Follow-up will begin on the first date of the wave and will end at the earliest occurrence of COVID-19-related death, death by other causes, or the end date of the wave.

For each of the covariates, a Cox cause-specific proportional hazards model will be fitted. Follow-up will begin on March 23, 2020, for wave 1, on Sept 7, 2020, for wave 2, and on May 28, 2021 for wave 3 and will end at the earliest occurrence of the outcome of interest, death, deregistration from a primary care practice, or the censoring date for the dataset capturing the outcome of interest (to May 30, 2020, (for wave 1) or to Sept 7, 2021 (for wave 2) or to Dec 14 2021 (for wave 3)). All analyses will be adjusted for age (using restricted cubic splines) and sex, and all models will be stratified by STP region to account for regional differences in infection rates. All analyses will be done separately for wave 1, wave 2 and wave 3. Proportional hazards assumptions will be assessed by testing for a zero slope in the scaled Schoenfeld residuals and graphical inspection of Kaplan-Meier plots.

Interpretation

This study aims to *describe* risks for COVID-19-related death in subgroups of the UK population over the time course of the three pandemic waves (March 2020 - December 2021). In addition, this study explores if subgroup specific mortality risks changed over time. Since this study is purely descriptive, this study does not explain *why* risks are higher/lower in certain subgroups nor does it explain why, compared to the first pandemic wave, risks may or may not have changed in subgroups.
The drivers of mortality risks are complex and often multidimensional. When individuals with e.g. a comorbidity have a higher risk of COVID-19-related death than individuals without the comorbidity, it is not known whether the comorbidity is the *cause* of the increased risk, nor is it known if targeting e.g. vaccination efforts to the individuals with the comorbidity will decrease COVID-19-related death risks in that group. For example, individuals with the comorbidity may in fact be vaccinated, but vaccines may be less effective in that group. Moreover, differences in infection rates in subgroups may explain observed differences in COVID-19-related death. There may be differences in infection rates between subgroups due to, for example, geographical clustering, occupation, household size, previous infection, vaccination, masking and social distancing. In fact, higher infection rates in the first pandemic wave may prevent COVID-19-related death in the second and third wave (if an increased mortality risk in the first wave is driven by high infection rates, this may cause mortality risk in the second wave to drop).
Future research should continue to explore *what* causes higher mortality risks in various subgroups of the population and if so, *why* mortality risks changed (in comparison to the first pandemic wave) in one subgroup while they have not changed in another subgroup.

References

[1] ‘Coronavirus (COVID-19) latest insights - Office for National Statistics’. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19latestinsights/deaths (accessed Mar. 02, 2022).

[2] E. J. Williamson *et al.*, ‘Factors associated with COVID-19-related death using OpenSAFELY’, *Nature*, vol. 584, no. 7821, Art. no. 7821, Aug. 2020, doi: 10.1038/s41586-020-2521-4.

[3] ‘Coronavirus (COVID-19) Infection Survey technical article: waves and lags of COVID-19 in England, June 2021’. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsurveytechnicalarticle/wavesandlagsofcovid19inenglandjune2021#waves-of-covid-19 (accessed Mar. 02, 2022).

[4] ‘The R value and growth rate’, *GOV.UK*. https://www.gov.uk/guidance/the-r-value-and-growth-rate (accessed Mar. 02, 2022).

[5] R. Paton, C. Overton, and T. Ward, ‘Infectious Disease Modelling Team’, p. 6.

[6] ‘Who is at high risk from coronavirus (COVID-19)’, *nhs.uk*, Mar. 01, 2021. https://www.nhs.uk/conditions/coronavirus-covid-19/people-at-higher-risk/who-is-at-high-risk-from-coronavirus/ (accessed Mar. 02, 2022).

[7] ‘Information on Public Health Observatory recommended methods’. https://www.scotpho.org.uk/media/1403/inphorm-6-final.pdf (accessed Mar. 02, 2022).