# HEALTH BURDEN OF COVID-19 AND HEALTHCARE RESOURCE UTILIZATION IN ENGLAND

## An observational retrospective cohort study describing clinical outcomes and utilisation of healthcare resources among persons with COVID-19 in England, stratified by infection severity and selected comorbidities

19 October 19, 2021

## Study Synopsis

| | |
|---|---|
| **Background** | The novel coronavirus disease 2019 (COVID-19) that has caused a pandemic has placed an enormous burden on the healthcare systems. In December 2020, a nationwide vaccination programme began in the United Kingdom (UK). While vaccinations have greatly reduced the incidence of COVID-19 infection, they have not been a panacea, as populations exist who may not be eligible owing to medical contraindications or who demonstrate suboptimal response due to prior/current treatment(s) and/or underlying disease. The magnitude of these various populations for whom vaccination against COVID-19 may prove suboptimal is not well-understood; similarly, as COVID-19 is still a new phenomenon, relatively little is known concerning the impact of COVID-19 infection. Therefore, this study aims to describe the incidence of COVID-19 infection during both the pre-vaccination and vaccination (including vaccine boosters) periods and associated healthcare resource utilisation (HCRU) and cost, stratified by age, severity, and selected comorbidities. This study additionally will describe the populations in England who are potentially ineligible for COVID-19 vaccines or who are at risk of COVID-19 infection following vaccination. |
| **Objectives** | **Core Study Objectives**<br>1 To estimate the size of populations (pre-defined) in England who potentially are ineligible for vaccine or are at risk of COVID-19 infection following vaccination.<br>2 To estimate incidence of COVID-19, by age group, disease severity, and selected comorbidities.<br>3 To estimate incidence of long COVID-19 syndrome, by age, disease severity, and selected comorbidities.<br>4 To describe patterns of HCRU and cost associated with an episode of COVID-19, stratified by age, selected comorbidities, disease severity and the occurrence (vs. absence) of long COVID-19 syndrome.<br>**Exploratory Objectives**<br>1. To identify patients who developed COVID-19 infection after vaccination stratified by the number of doses received and explore potential risk factors therefor.<br>2. To develop a prediction model to identify risk profiles associated with COVID-19 infection after the deployment of vaccination campaign in England, and to estimate the prevalence of subgroups consistent with each identified risk profile. |
| **Study Design** | This is an observational retrospective cohort study using a series of datasets in England, accessible through National Health Services (NHS) Digital. The study period will start on 1 September 2015 and end on the time of data extraction (i.e., the latest day of data overlap across required datasets). The period that will be used to identify COVID-19 infections will start on 1 September 2020 and end 12 weeks prior to the end of the study period. Patient baseline characteristics will be assessed up to a 5-year period immediately before 1 September 2020 (i.e., 1 September 2015 – 31 |

| | |
|---|---|
| | August 2020). The pre-vaccination period will be defined as beginning on 1 September 2020 and ending on 7 December 2020. The vaccination period will be defined as beginning on 8 December 2020 and ending on the last day of the study period. |
| **Study Population** | **Inclusion Criteria**<br><br>This study will include all subjects in England who were active in the datasets described below for at least one day during the COVID-19 infection identification period.<br><br>Data sets:<br><br>• General Practice Extraction Service data for pandemic planning and research (GDPPR)<br>• COVID-19 Second Generation Surveillance System (SGSS) from Pillar 1 and Pillar 2<br>• COVID-19 Vaccination Status Data<br>• COVID-19 Hospitalisation in England Surveillance System (CHESS)<br>• Hospital Episode Statistics (HES)<br>• Uncurated Low Latency Hospital Data Sets for APC, OP and ACC<br>• SUS Payment by Result (SUS PbR)<br>• Electronic Prescribing and Medicines Administration (EPMA) data in Secondary Care for COVID-19<br>• NHS Business Service Authority (BSA)<br>• Civil registration (deaths)<br><br>**Exclusion Criteria**<br><br>Subjects will be excluded from this study if they have less than 12 months of baseline data prior to 1 September 2020. |
| **Outcome Measures** | **COVID-19 Infections**<br>A COVID-19 infection will be defined based on the presence of either:<br>• A positive COVID-19 PCR test, **OR**<br>• A medical encounter with a COVID-19 diagnosis code in either the primary or secondary care setting<br><br>**Long COVID-19**<br>A patient will be considered to have long COVID-19 if (s)he: |

| | |
|---|---|
| | • Has a long COVID-19 diagnosis code during either the pre- or vaccination periods (SNOMED codes: 1325181000000106 or 1325161000000102); **OR**<br><br>• Has an episode of COVID-19 infection lasting (i.e., end date – start date +1) ≥4 weeks.<br><br>**COVID-19 Severity**<br>Clinical signs and symptoms used in WHO's definitions of COVID-19 severity are typically not available in the electronic healthcare datasets. Algorithms based on diagnosis codes, procedure code and usage of healthcare resources (e.g., admission into intensive care unit) will be developed to classify disease severity in line with WHO's clinical guidelines.<br><br>**HCRU**<br>o Inpatient visits<br>o Hospital length of stay<br>o Days mechanically ventilated<br>o Days in intensive care<br>o Outpatient visits<br>o Specialist outpatient visits<br>o Primary care visits (GP or nurse)<br>o Primary care referrals<br>o Prescriptions/dispenses<br>o Procedures<br>o Surgeries<br>o Diagnostic scans<br>o Cost of care |
| **Statistical Methods** | **Core Objective 1**<br>The number (N) and percentage (%) of various subgroups in England who potentially are ineligible for vaccines or who are at increased risk of COVID-19 infection following vaccination will be reported.<br><br>**Core Objective 2**<br>Overall incidence of COVID-19 (per 100 person-months) will be estimated using the total number of COVID-19 episodes identified during the relevant time period as the numerator, and the total person-months at risk during that time as the denominator. Analyses also will be conducted that stratify the population by age, COVID-19 severity, and selected key comorbidities. Incidence rates with corresponding 95% confidence intervals (CIs) will be reported.<br><br>**Core Objective 3**<br>The total number (N) and percentage (%) of unique patients identified who develop long COVID-19 during the pre- and vaccination periods will be summarised. The |

| | |
|---|---|
| | incidence of long COVID-19 (per 100 person-months) will be estimated. Stratified analyses by age, COVID-19 severity and selected key comorbidities will be included. Incidence rates with corresponding 95% CIs will be reported.<br><br>**Core Objective 4**<br>HCRU and costs incurred during COVID-19 episodes will be described for each resource of interest separately (e.g., GP visits, outpatient visits, hospitalisations). The rate of utilisation for each resource will be calculated on a per-patient and per-patient per-COVID-19 episode basis, respectively. Analyses will be conducted for all patients who develop COVID-19, as well as stratified by age, COVID-19 severity, and selected key comorbidities. A sub-analysis also will be conducted to focus on HCRU incurred by long COVID-19 patients.<br><br>**Exploratory Objectives**<br>proportional-hazard regression models will be fit to explore risk factors for predicting populations at potential risk of COVID-19 infection. Estimated relative risks and corresponding 95% CIs will be reported. Analysis will be stratified by the number of doses received (i.e., “one dose”, “both doses”, or “both doses and booster” [if information on vaccine boosters is available at the time of data provision]).<br><br>In order to identify risk profiles that predict COVID-19 infection, a data cube for machine learning will be built, and data for 80% of subjects will be used to develop and train the predictive model. Random forest (RF), gradient boosting machine (GBM) and support vector machine (SVM) models will be implemented in the training set and model performance assessed in the 20% of remaining subjects withheld from the training set (i.e., the test dataset). If these models do not produce robust results, an Artificial neural network (ANN) model may be fitted to improve model performance.<br><br>**Sensitivity Analysis**<br>It is expected that sensitivity analyses will be conducted using alternative algorithms to define COVID-19 severity based on data availability in real world data. Additionally, post-COVID-19 syndrome (i.e., symptoms lasting for 12 weeks or more) will be used as an alternative outcome for long COVID-19 (i.e., symptoms lasting for 4 weeks or more) and repeat analysis for objective 3. |