<u>**README – Input_data**</u>

When extracting the input data for a model, we have a few recommendations that are detailed here. These are designed to the make the modelling consistent with the NHS sitrep that trusts in England are requested to submit on a daily basis. Alternatively, the user can make their own adjustments/assumptions, in order to make the model consistent with their internal data monitoring procedures.

Patient cohort

When selecting our patient cohort, we wish to select all patients with a positive COVID test associated with their hospital stay. To do this, we select all patients who are admitted to hospital within 14 days of their first positive COVID test **or** are tested during their hospital stay. This cohort can be separated into two branches: those tested before admission and those tested after admission. For individuals in the first cohort, we will count their COVID stay from their date of admission to hospital. For individuals in the second cohort, we will count their COVID stay from the date of their first positive test. Since we are counting by date of test, there will be a slightly lag over the last couple of days whilst patients are awaiting their test results. Therefore, we recommend removing all admissions from the last two days, and starting the modelling of the last included day. Since we are counting the second group from date of test, we need to ensure any ICU episodes are consistent with this start date. That is, we are only interested in ICU events that occur after they test positive or ICU events during which they tested positive.

In pseudo code, we wish to process the data as follows:

Extract all COVID patients with a PCR positive swab date either up to 14 days before admissions or during their hospital stay:

(AdmissionDate - SwabDate < 14) OR (SwabDate > AdmissionDate AND SwabDate < DischargeDate)

For individuals in the first group *"(AdmissionDate - SwabDate < 14)",* leave the data entries unchanged.

For individuals in the second group *"(SwabDate >= AdmissionDate AND SwabDate < DischargeDate)"*, change *AdmissionDate = SwabDate*. We will then only be interested in critical care stays that occur after *SwabDate* or overlap *SwabDate*, so that they are associated with the COVID stay.

For each patient extracted, we then require:

UniqueIdentifier; AdjustedAdmissionTime; CriticalCareStartTime; CriticalCareEndTime; DischargeTime; DateOfDeath; DateOfBirth

Where data is not available, e.g. *DateOfDeath* or *CriticalCareStartTime*, a black or "NULL" entry is recommended. For the input dates, we require the format **yyyy-mm-dd hh:mm:ss**.

An example SQL query is provided, see **Input_data.sql**. This extracts the data, though currently we amend the date formats and remove the last two days post extraction, so these parts are not included in the SQL.