**Phenotype definitions for analyses**

**6th april 2020 - version 1**

Working group: Kenneth Baillie, Mark Daly, Lea Davis, Andrea Ganna, Rachel Liao, Sulggi Lee, James Priest, Alessandra Renieri, Vijay Sankaran, David van Heel

**RATIONALE:**Identification of human genetic factors impacting the COVID19 pandemic represents a simple and scalable set of analyses which holds the potential to rapidly offer important information. Within individual cohorts, we propose establishing a shared set of phenotypes to be combined with genomic data for standard GWAS and further meta-analysis by the analysis group. First, we define a minimal set of analysis phenotypes, which we suggest to be the main analysis deliverable from the participating groups. Second, we provide a number of phenotypes that are less likely to be widely available across groups, but they are still valuable for downstream meta-analytic efforts.

We first provide a definition of SARS-CoV-2 infection.

Definition of SARS-CoV-2 infection:

1. Laboratory confirmation of SARS-CoV-2 infection (RNA and/or serology based)

**OR**

1. ICD/administrative/EHR coding-based definition of SARS-CoV-2 infection where large-scale & rapid clinical testing unavailable (see **Appendix 1** for codes).

*Note: Case definition of SARS-CoV-2 infection may change as testing strategies are*

*rapidly evolving*

**MINIMAL ANALYSIS**

Our minimal analysis set focuses on disease severity only among SARS-CoV-2 infected individuals.

**Analysis 1** (name: ANA1):

* **Inclusion criteria**: SARS-CoV-2 infected individuals with a clinical window of 15-90 days from diagnosis \*.
* **Cases:** death OR respiratory support (intubation, CPAP, CPN or BiPAP, See **Appendix 1** for codes).
* **Controls**: alive individuals without respiratory support at any point since diagnosis.

**Analysis 2** (name: ANA2):

* **Inclusion criteria**: SARS-CoV-2 infected individuals with a clinical window of 15-90 days from diagnosis \*.
* **Cases:** hospitalized (in-patient) individuals
* **Controls**: non-hospitalized individuals

**EXTENDED ANALYSES**

Extended analyses try to accommodate different analyses that depend on the specific study design or collection approach.

**Analysis 3-mortality** (phenotype name: ANA3)**:**

* **Inclusion criteria**: SARS-CoV-2 infected individuals with a clinical window of 15-90 days from diagnosis \*.
* **Cases:** individuals that died
* **Controls**: alive individuals

**Analysis 4-severity ordinal**  (phenotype name: ANA4)**:**

* **Inclusion criteria**: hospitalized SARS-CoV-2 infected individuals with a clinical window of 15-90 days from diagnosis \*.
* **Continuous outcome:** 3 level scale (mild,severe,critical) of COVID19 disease severity based on <https://jamanetwork.com/journals/jama/fullarticle/2762130>. Mild: nonpneumonia and mild pneumonia. Severe: dyspnea, respiratory frequency ≥30/min, blood oxygen saturation ≤93%, partial pressure of arterial oxygen to fraction of inspired oxygen ratio <300, and/or lung infiltrates >50% within 24 to 48 hours. Critical: respiratory failure, septic shock, and/or multiple organ dysfunction or failure. *Adaptations of this scale are possible within each study*. Studies will need to report the exact criteria used to define: mild,severe,critical.

**Analysis 5-susceptibility**  (phenotype name: ANA5)**:**

* **Inclusion criteria**: everyone.
* **Cases:** individuals diagnosed with COVID-19.
* **Controls:** individualswithout documented diagnosis of COVID-19.

**Analysis 6-susceptibility-severe**  (phenotype name: ANA6)**:**

* **Inclusion criteria**: everyone.
* **Cases:** individuals diagnosed with COVID-19 and hospitalized.
* **Controls:** individualswithout documented diagnosis of COVID-19.

**Analysis 7-flu symptoms**  (phenotype name: ANA7)**:**

* **Inclusion criteria**: everyone.
* **Cases:** individuals reportingflu-related symptoms (definition is cohort-specific and based on questionnaire data, see **Appendix 2** for an example) during the COVID-19 pandemic period.
* **Controls**: individuals reporting no flu-related symptoms

\* we suggest to include only COVID-19 patients that are followed long enough to allow a reliable classification in “cases” and “controls”. Thus, we suggest that a time window between 15 and 90 days from the COVID-19 diagnosis is used to decide if the patient should be assigned to the case or control group.

**APPENDIX 1**:

Diagnostic codes for COVID-19 severity (provided by Lea Davis and Julia Sealock)

<https://drive.google.com/file/d/1ck0ABYZ6oYnMStoYnGpnA7n1W6wcY3_6/view>

**APPENDIX 2:**

Cases can be defined if they experienced at least two of the following complaints:

* Headache
* Dizziness
* Pain in the chest or heart area
* Lower back pain
* Nausea or an upset stomach
* Sore muscles
* Difficulty breathing
* Feeling very warm sometimes, then feeling very cold again
* A numb or tingling sensation somewhere in your body
* A lump in your throat
* Feeling limp somewhere
* Feeling heavy in the arms or legs
* Shortness of breath
* Pain when breathing
* Running nose
* Sore throat
* Cough without mucus
* Coughing with mucus
* Fever (38 degrees or more)
* Diarrhea or abdominal pain
* Loss of smell or taste
* Red, sore or itchy eyes
* Sneezing