ISARIC (International Severe Acute Respiratory and Emerging Infections Consortium)

A global federation of clinical research networks, providing a proficient, coordinated, and agile research response to outbreak-prone infectious disease

Analysis Plan for ISARIC International Collaborative Covid-19 Adult Follow up Analysis

Please complete the following sections:

Title of proposed research
Characterization of long term Covid-19 outcomes in different populations: ISARIC Global Collaborative Adult Covid-19 Follow up analysis
Version: (Date: Day/Month/Year)
6 December 2021
Working Group Chairs(name, ORCID ID, email, institution, country)
Dr. Janet Scott, ISARIC4C UK, MRC- University of Glasgow Center for Virus Research, UK Email: janet.scott@glasgow.ac.uk Orcid ID: 0000-0001-8030-5223 Dr. Luis Felipe Reyes, Universidad de la Sabana, Chía, Colombia; University of Oxford, UK. Email: luisreyve@unisabana.edu.co / luis.reyes@wolfson.ox.ac.uk Orcid ID: 0000-0003-1172-6539 Prof. Daniel Munblit, Sechenov University, Russia E-mail: daniel.munblit08@imperial.ac.uk Orcid ID: 0000-0001-9652-6856 Dr. Louise Sigfrid, ISARIC Global, ISARIC4C UK, University of Oxford, UK E-mail: louise.sigfrid@ndm.ox.ac.uk Orcid ID: 0000-0003-2764-1177

[1]Working group data management and statistical leads(name, ORCID ID, email, institution, country)

Luis Felipe Reyes, Unversidad de la Sabana, Chía, Colombia; University of Oxford, UK.
Email: luisreyve@unisabana.edu.co / luis.reyes@wolfson.ox.ac.uk
Orcid ID: 0000-0003-1172-6539

Esteban Garcia, Universidad de La Sabana, Chía, Colombia.
E-mail: jegargax@gmail.com

Daniel Munblit, Sechenov University, Russia
E-mail: daniel.munblit08@imperial.ac.uk
Orcid ID: 0000-0001-9652-6856

Александр Суворов, Sechenov University, Russia
E-mail: suvorov_a_yu_1@staff.sechenov.ru

Ekatherina Pazukina, Sechenov University, Russia
E-mail: Pazukhina@gmail.com

Christiana Kartsonaki, University of Oxford, UK
Email: christiana.kartsonaki@dph.ox.ac.uk

Bronner P. Gonçalves, University of Oxford, UK
Email: bronner.goncalves@ndm.ox.ac.uk

ISARIC Global Covid-19 adult follow up analysis working group
(name, ORCID ID, email, institution, country)

Murray Dryden, South Africa, National Institute for Communicable Diseases
murrayd@nicd.ac.za
0000-0001-6975-9398

Waasila Jassat, South Africa, National Institute for Communicable Diseases
waasilaj@nicd.ac.za
0000-0003-4279-3056

Jan Cato Holter, Norway, Oslo University Hospital
jacaho@ous-hf.no
0000-0003-1618-5022

Anders Benjamin Kildal, Norway, Department of Anesthesiology and Intensive Care, University Hospital of North Norway, Tromsø, Norway
anders.benjamin.kildal@unn.no
0000-0002-1319-6511

Arne Søraas, Norway, The Norwegian Corona Cohort, Department of Microbiology
Oslo University Hospital

[1]

arne@meg.no
0000-0003-1622-591X

Anders B. Nygaard, Norway, The Norwegian Corona Cohort, Department of Microbiology
Oslo University Hospital
anders.b.nygaard@gmail.com
0000-0003-1922-0751

Merete Ellingjord-Dale, The Norwegian Corona Cohort, Department of Microbiology, Oslo University Hospital
mellingjord@hotmail.com/merell@ous-hf.no
0000-0003-2758-0140

Prof. Simone Piva, Italy, Department of Anesthesia, Critical Care and Emergency, Spedali Civili University Hospital, Brescia
Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia
Interdepartmental Research Center “Alessandra Bono” on LOng Term Outcome (LOTO) in Survivors of Critical Illness, University of Brescia
simone.piva@unibs.it
0000-0002-5483-8007

Fernando Bozza, Brazil, Fundação Oswaldo Cruz: Rio de Janeiro
bozza.fernando@gmail.com
0000-0003-4878-0256

Raph Hamers, Indonesia, EOCRU Jakarta, Universities of Indonesia
Raph.hamers@eocru.org

Matteo Puntoni, Italy, Clinical Epidemiology - Research & Innovation Unit
University Hospital of Parma
mpuntoni@ao.pr.it
0000-0002-7908-0626

Giuseppe Maglietta, Italy, Clinical & Epidemiological Research Unit
University Hospital of Parma
gmaglietta@ao.pr.it
0000-0003-4497-1872

Anna Beltrame, Italy, Clinician, Department of Infectious Diseases, Tropical and Microbiology, IRCCS Sacro Cuore Don Calabria Hospital, Negrar di Valpolicella
anna.beltrame@sacrocuore.it
0000-0002-7499-5843

Prof. Nicola Latronico,Italy, Department of Anesthesia, Critical Care and Emergency, Spedali Civili University Hospital, Brescia
Department of Medical and Surgical Specialties, Radiological Sciences and Public Health, University of Brescia
Interdepartmental Research Center “Alessandra Bono” on LOng Term Outcome (LOTO) in Survivors of Critical Illness, University of Brescia

nicola.latronico@unibs.it
0000-0002-2521-5871

Roberta Meta, Gibraltar, St Bernard's Hospital
Roberta.Meta@gha.gi
0000-0001-9232-1052

Kyle Gomez, Gibraltar, St Bernard's Hospital
Kyle.Gomez@gha.gi

Natalie Wright Public Health UK
Natalie.Wright@phe.gov.uk

Allegra Chatterjee Public Health UK
Allegra.Chatterjee@phe.gov.uk

Adam Ali, Gibraltar, St Bernard's Hospital
adam.ali96@btinternet.com

Michael Edelstein, Israel, Bar-Ilan University
michael.edelstein@biu.ac.il
000-0002-7323-0806

Sung-Min Cho, Johns Hopkins
csungmi1@jhmi.edu
0000-0002-5132-0958

Alejandro Martín-Quirós
a.martinquiros@gmail.com
0000-0003-4630-7668

Karl Trygve Kalleberg, Norway, Age Labs
karltk@agelabs.com
0000-0003-4968-2295

This SAP was developed in August 2020 and published as part of the follow up protocol in BMJ Open in March 2021 (doi: 10.1136/bmjopen-2020-043887) (PMID: 33692181). The draft protocol and data collection form was circulated to partners in June 2020. Sites with capacity and interest in follow up at the time, formed the ISARIC Global Covid-19 follow up working group. The working group finalized the protocol and analysis plan, which is available on the www.isaric.org website. We now have a data set ready for analyses. In the intervening 13 months, a substantial amount has been published in this area. However, this is to our knowledge the first international, multisite study seeking to characterize long term Covid-19 sequalae and complications in different populations using a standardized data collection form.

In essence of the rapidly advancing field, and importance, we are aiming for a rapid initial analysis including those that were hospitalised during the acute phase,

followed by a second analysis including longer follow up time and cohorts that were hospitalised and community cohorts. The aim is to generate evidence on long term Covid-19 outcomes in different at risk populations, to inform clinical and public health management and further studies.

Introduction

This document details the initial analysis plan for publication. This analysis builds on data collected using the ISARIC/WHO Covid-19 Clinical Characterisation Protocol (CCP) and associated data collection forms already in operation, the Core and Rapid case report forms (CRFs). These CRFs were developed to standardise clinical data collection on patients admitted with suspected or confirmed Covid-19. These CRFs collect data on demographics, pre-existing comorbidities and risk factors, signs and symptoms experienced during the acute phase, and care and treatments received during hospitalization.

Recognising that post-acute Covid-19 sequelae and post-Covid condition affects people who were hospitalised and those who were not, for the follow up study we developed case report forms to enable inclusion of both community and hospitalised cohorts.

A subset of the global cohort was followed up after their acute illness, or in some sites linked to SARS-CoV-2 testing, to create the global follow up cohort. There are around 18,800 patients in the ISARIC database from 14 countries, as of 25th October 2021. This data will represent the global experience of the first 6 – 12 months of patient follow up.

Participatory Approach

All contributors to the ISARIC database are invited to participate in our follow up analysis through review and input on the statistical analysis plan, and contribution of data and statistical support for the planned first and second analysis. Please contact us on ncov@isaric.org if of interest. This is the first, collaborative follow up analysis, which will be followed by additional analysis including additional sites, patients and longer follow up time. The outputs of this work will be disseminated as widely as possible to inform patient care and public health policy, this will include a collaborative submission for publication in an international, peer-reviewed journal.

Research Plan

Summary of Research Objectives
The primary objective of this study is to characterise the physical and psychosocial consequences in patients with Long Covid (post-acute Covid-19). Secondary objectives include estimating the risk of and risk factors for post-Covid-19 medical sequalae, psychosocial consequences.
Proposed Target Population
Inclusion criteria • People aged 18 years and older.

- Laboratory confirmed SARS-CoV-2 infection or physician confirmed Covid-19
- At least 1-month post-onset of Covid-19, or post-SARS-CoV-2 positive test.
- Person (or family member/carer for patients who lack capacity) consent to participate.

Inclusion of vulnerable participants

The data collection surveys and validated tools are developed for anyone who fit the inclusion criteria, including pregnant women, elderly and those who are immunosuppressed. A paediatric study is planned separately.

Clinical Questions/Descriptive Analyses

- What are the medium and long-term impact of SARS-CoV-2 infection on physical health?
- What are the medium and long-term impacts of SARS-CoV-2 infection on psychosocial health and functioning?
- What are the risk factors (including vaccination) for developing post-acute SARS-CoV-2 sequelae?

Planned Statistical Analyses, Methodology and Representation

General Framework

Using the data, we will test for differences in outcomes across important demographic groups (age categories, sex, ethnicity, socioeconomic factors, comorbidities), specific exposures (mild, moderate, severe acute Covid -19, critical care admission, ventilation) and initial clinical presentation, and basic medical management. We plan to use this platform to conduct analyses to address public health and scientific questions which may arise. Given these requirements, new questions we have not specified within this protocol may arise. For the post-hospitalisation cohorts, the data collected through the follow up module will be linked with data on demographics, comorbidities, clinical characteristics, care and treatments collected using the ISARIC/WHO Core- or RAPID Covid-19 CRF. Fields contained within the CRFs will be combined and if an area of interest is found, the maximal amount of data will be used to investigate this to maintain sample size and power. For the non-hospitalised cohorts/community cohorts the acute data documented in the Tier 1 Freestanding survey module will be incorporated in the analysis. The plan below presents our guiding statistical framework. Analyses will be developed concurrently with data collection using statistical coding script. The data documented at the follow up assessments include, self-reported recovery, persistent and ongoing symptoms not existing prior to Covid-19 onset, complications, breathlessness (MRC dyspnea scale), fatigue (Visual analogue scale), impact on daily activities and functioning, disability assessment (UN/Washington disability score), impact on occupation, sickness

absence, and socioeconomic deprivation data (Supplemental file 1). The follow up assessments are carried out using a range of methods for wide dissemination and uptake during the pandemic. People will be identified either during a hospital admission with Covid-19 or when presenting in a health centre and/or linked to SARS-CoV-2 diagnostic testing and asked for consent to participate following local ethical regulations and approvals. Sites may use one or a combination of methods for assessing consenting participants, including, by telephone, online link, by postal survey or in-clinic assessment.

Data will be summarised first by using summary statistics. Histograms for each variable will be plotted. Categorical variables will be explored using frequencies and percentages, with differences in disease severity and treatment groups tested for using Chi-square tests or Fisher's exact test where cell counts are under five. Numeric variables will be summarized using means and standard deviations and medians and 25^{th} and 75^{th} percentiles. Differences in numeric variables by disease severity and treatment will be assessed using linear regression and t- and F-tests.

Outcomes will broadly be of one of three types: 1) binary event data (for presence or absence of outcome of interest), 2) change over time (for continuous or ordinal data) or as 3) time-to-event data (for patients with serial follow up assessments). We will calculate changes in the frequencies of different symptoms and other outcome variables over time and use these changes over time as outcomes to assess the association of treatments or exposures with them. To explore incidence of outcomes cumulative incidence curves will be plotted and associations of explanatory variables with outcomes will be assessed using Cox regression. Individuals will be followed up from illness onset or discharge after acute illness to readmission.

To address the main study aims of characterizing the long-term impact of SARS-CoV-2 infection on physical and psychosocial health, we will first provide simple summaries of incidence and second, characterise which patients are at risk of developing these. We will present description of baseline characteristics in patients with follow-up at different times and also compare baseline characteristics for patients who were included in the analysis and those who did not accept to participate. Median and interquartile range time since acute presentation admission and discharge will be presented. To assess the associations of demographics, disease severity, and treatments with complications, functional impairment or quality of life, we will fit linear (for numeric outcomes) and logistic (for binary outcomes) regression models, adjusting for potential confounders. To quantify associations between patient characteristics and acute clinical presentation and the development of persistent complications, including functional impairment or reduced quality of life, we will use linear (for continuous outcomes)

or logistic (for binary outcomes) regression models, adjusting for potential confounders.

Analyses will be conducted within country/site and if appropriate combined using random effect models. Potential confounders will be selected on the basis of clinical plausibility.

For binary event data, multilevel logistic regression will be used, and estimates presented as odds ratios with 95% confidence intervals. For continuous data, linear regression will be used, and estimates presented as model coefficients with 95% confidence intervals. Also, we will compare baseline characteristics (e.g. acute severity) to make analyses per groups and determine factors associated with the development of long covid. Moreover, we will study time-to-event (i.e., post-Covid condition) when appropriate and feasible according to the data available.

P values < 0.05 will be interpreted as some evidence against the null hypothesis. Analyses will be conducted in secure R (R Foundation for Statistical Computing, Vienna, AUT), SPSS 29 (IBM, California, USA) or STATA (StataCorp LLC, TX, USA) environments.

Specific research questions

What are the medium and longer-term impact of SARS-CoV-2 infection on physical health?

Population:

Hospitalised cohort: Adults, hospitalised with Covid-19, or diagnosed with Covid-19 during hospital admission, diagnosed by laboratory testing or highly suspected (clinically confirmed) infection in absence of positive test.

Non-hospitalised cohort: Adults, diagnosed with SARS-CoV-2 infection who were not hospitalised during the acute phase, diagnosed by laboratory testing or highly suspected (clinically confirmed) infection in absence of positive test (e.g. settings with limited access to diagnostic testing during the peak of the pandemic).

Exposures: Different patient characteristics, demographic factors (age, sex, ethnicity), pre-existing co-morbidities and risk factors, duration of illness, Covid-19 immunisation, clinical characteristics (symptoms), severity and clinical management during the acute phase (hospitalisation, critical care, empirical treatment), hospitalisation duration, in-hospital complications, calendar time. In sites where data from both people with history of hospitalisation due to acute COVID-19 and from those with evidence of infection without hospitalization (i.e. mild and asymptomatic cases), these will also be compared in unadjusted and adjusted/multivariable analyses.

Outcome: Persistent or new symptoms and complications, impact assessed using the UN/ Washington Disability Score, MRC Dyspnoea Scale, EQ5D5L, Fatigue VAS scale, occupational status. Outcomes should be studied at defined timepoints and not combined inappropriately (i.e. in-hospital outcomes should not be compared with out of hospital outcome data). Timing of comparison should be within specific follow up assessment windows (3 to 6 months) as below. First assessment at an interval around three months post-onset.

Statistical approach:

Incidence of each sequelae or complication at defined time intervals (i.e. at 1- 3 months, > 3 – 6 months, > 9 – 12 months). Continuous variables (e.g. EQ-5D-5L) will be reported using median (with IQR) or mean (with SD) values following the Euroqol analysis guidance. We will calculate changes in summary EQ5D-5L index and estimates for individual EQ5D-5L dimensions. Where multiple measurements over time are available, we will use visualisation and modelling of changes over time.

Statistical output:

- Table of frequency of symptom/sequelae or complication at given time points in patients with Covid-19
- Identification and prevalence of specific cluster of sequelae, including symptoms and impact on functioning
- Stratified by age, sex, ethnicity, comorbidities, BMI, and severity during the acute phase (whether patient received critical care or invasive mechanical ventilation)
- For longitudinal data, graphical representation over time
- Comparison across submitting countries, by geographical region, world bank or human development index.

What are the medium and long-term impacts of SARS-CoV-2 infection on psychosocial health and functioning?

Population: Same as above.

Exposure: Same as above *Outcome:* EQ-5D – 5L questions in the psychosocial health domain, correlated with pre-existing risk factors, comorbidities, acute severity, management and clusters of persistent sequelae. Identification of clusters of sequelae and impact on functioning using random forest analysis and other unsupervised models.

Statistical approach: Incidence of each sequela or complication at defined time intervals (i.e. pre-onset, at 1- 3 months, > 3 – 6 months, > 9 – 12 months assessment windows). For continuous variables median (with IQR) and mean

(with SD) values will be reported. Where multiple measurements exist in time (i.e. longitudinal data) are available, visualisation will be used.

Statistical output:

- Table of incidence of psychosocial health outcomes at given time points (pre-onset, and at follow up) in patients with Covid-19
- Identification of and prevalence of clusters of sequelae and impact on functioning
- Comparison across age, sex, ethnicity, pre-existing comorbidites and whether patient received critical care or invasive mechanical ventilation
- For longitudinal data, graphical representation over time
- Comparison across submitting countries, by world bank and human development index.

What are the risk factors for developing long term sequelae post SARS-CoV-2 infection?

Population: Same as above.

Exposure: Same as above.

Outcome: Cluster of persistent symptoms, increased breathlessness, new or worsened disability, low EQ-5D-5L scores.

Statistical approach: Use linear and logistic regression models to identify who is at highest risk of developing sequelae of Covid-19

Statistical output:

- Logistic regression models including explanatory variables such as age, sex, ethnicity, pre-existing comorbidities, socioeconomic deprivation and acute severity (critical care or advance respiratory support (e.g., invasive mechanical ventilation, non-invasive mechanical ventilation, or HFNC).
- Unsupervised clustering analyses of patients, symptoms and outcomes, with or without subsequent dimension reduction to include in models.

Handling of Missing Data

Handling missing data

Missing data should be studied using heat-map approaches to explore whether data may be missing at random which is considered highly unlikely. Where missing data exceed 30% in any given field, multiple imputation through chained equations may be used to impute the explanatory variables only. Outcomes should be included in the imputation, but not imputed themselves.

Handling unclear data

If data are not clear as to when collected or how measured or how the outcome is defined, the source data or investigator for that site / country should be contacted. If not clear, then data cannot be included.

Reporting

Analyses should be done on an intention to recruit approach i.e. if the patient is approached for follow-up and does not complete it, they should still be retained in the analysis. In addition to this, analyses should be repeated for those who attempted at least one field of follow-up. Study reporting should be in line with the applicable EQUATOR network guideline and checklist.

Other Information

The analysis plan was developed in August 2020. The first international, collaborative analysis is scheduled to be completed in early 2022. The results will be disseminated via peer-reviewed publication, and social media. The aim is to submit the first analysis for publication by January/February 2022, followed by a second analysis including longer follow up time, and community cohorts. Local investigators will disseminate the results using national media and social media channels as appropriate, for wide dissemination and translation of the findings into national policies as appropriate.

The ISARIC follow up protocol, data collection forms and analysis plans have been developed in collaboration with members of the Long Covid support group. We aim to involve the Long Covid support group members in translation of the key findings into lay terms, to provide a summary for dissemination to lay as well as professional audiences.

Associated publications

- Sigfrid, Cevik et al., What is the recovery rate and risk of long-term consequences following a diagnosis of Covid-19? A harmonised, global longitudinal observational study protocol, BMJ Open, 2020
- Michelen et al., Characterising long-term covid-19: a rapid living systematic review, BMJ Global Health, Oct. 2021
- Norton, Olliaro et al., Long COVID: tackling a multifaceted condition requires a multidisciplinary approach, The Lancet Infectious Diseases, May 2021
- Sigfrid, Drake et al., Long Covid in adults discharged from UK hospitals after Covid-19: A prospective, multicentre cohort study using the ISARIC WHO Clinical Characterisation Protocol, The Lancet Regional Health – Europe, 2021
- Osmanov et al., Risk factors for long covid in previously hospitalised children using the ISARIC Global follow-up protocol: A prospective cohort study, ERJ, 2021

- Buonsenso et al., Evidence of lung perfusion defects and ongoing inflammation in an adolescent with post-acute sequelae of SARS-CoV-2 infection, The Lancet Child and Adolescent health, July 2021
- Munblit, Sigfrid, et al., Setting Priorities to Address Research Gaps in Long-term COVID-19 Outcomes in Children, Jama Network, 2021
- Munblit D, Bobkova P et al., Incidence and risk factors for persistent symptoms in adults previously hospitalised for COVID-19., Clin Exp Allergy, 2021