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 COVID-19 Patients

Please complete the following sections:

Title of proposed research
A tool-box for epidemic research: the ISARIC Clinical Characterisation group on strategies to simplify and improve collaboration-fuelled research during health emergencies
Version: (Date: Day/Month/Year)
Version: 1 - November 19, 2021
Working Group co-chairs (name, institution, country, ORCID ID)
Joaquin Baruch, 0000-0002-0806-3183, ISARIC, UK joaquinbaruch2@gmail.com Bharath Kumar Tirupakuzhi Vijayaraghavan, ORCID: 0000-0002-1801-0667, Apollo Hospitals, Chennai, India bharath@icuconsultants.com Christiana Kartsonaki, 0000-0002-3981-3418, University of Oxford, UK christiana.kartsonaki@dph.ox.ac.uk Trokon Omarley Yeabah, National Public Health Institute, Liberia trokonyeabah@gmail.com Piero Olliaro, 0000-0003-1426-3680, University of Oxford, UK piero.olliaro@ndm.ox.ac.uk Laura Merson, 0000-0002-4168-1960, University of Oxford, UK, laura.merson@ndm.ox.ac.uk Bronner P. Gonçalves, University of Oxford, UK bronner.goncalves@ndm.ox.ac.uk

Amanda Rojek, University of Oxford, UK
amanda.rojek@gmail.com

Jake Dunning, University of Oxford, UK
jake.dunning@ndm.ox.ac.uk

Introduction

A novel disease outbreak. What is the clinical presentation? What is the case fatality ratio among hospitalized patients? Are healthcare workers being infected at higher rates than the general population? These and other critical questions were being asked by clinicians and governments during the evolving COVID-19 outbreak in the early months of 2020.

As clinicians and governments ask these questions, epidemiologists and statisticians will think about the study design and data required to answer them. Standardised data collection can be a big part of our solution by allowing researchers to compare and aggregate collected data. These data are typically managed through clinical research to answer specific questions, but clinical research usually lacks the agility required to have an impact during a health emergency. Data governance structures, case report forms and approved protocols prepared before the start of an outbreak can help to tackle some of these challenges and avoid delays.

Several initiatives have emerged in response to earlier epidemics and pandemics for the early detection and data collection during disease threats (e.g., ISARIC, Global.Health, European SARI-Network, others). For example, in 2013, ISARIC and the World Health Organization (WHO) implemented the standardised Clinical Characterisation Protocol (CCP) (Dunning JW, Merson L, et al., 2014) to rapidly transform clinical data into research evidence during health emergencies using a pre-approved protocol and a standardised data collection form. The specific focus on standardization in these programmes is aimed at circumventing the fractured and poorly coordinated efforts of early outbreaks – leading to data that was incomparable between hospitals or districts, let alone countries. The pre-approval process allows for data collection during the critical early phases of an outbreak – where case numbers may be high, but significant uncertainty exists.

As a result, the ISARIC data have produced a vast amount of evidence for decision-making (https://isaric.org/research/covid-19-clinical-research-resources/evidence-reports/). Although other data sources such as electronic health records exist in some high-income countries, a key focus of ISARIC's data collection and analysis is in low- and middle-income countries, using a designed data collection tool. Through its scale and breadth, these data allow characterisation of disease presentation and progression in different populations; identification of regional differences in treatment or outcome; and assessment of associations between relatively uncommon risk factors (e.g. specific comorbidities,) and outcomes. However, almost two years into this pandemic, we need to ask ourselves: what can we learn from our response to COVID-19 to be better prepared to answer those basic questions next time?

How much data should we collect? Which specific questions and variables should be prioritized? Undoubtedly, data collection is costly. There is a trade-off between the number of variables collected and the time and effort required. In novel outbreaks, while very early data can be wrong (e.g. exaggerated fatality rates), data from very large studies that take longer can be 'old news' very quickly and add little to what we already learned from faster, smaller studies. Similarly, there is an ethical argument for only collecting essential clinical research data during a pandemic when health services and staff are stretched for resources; effective prioritisation of research and clinical data collection is critical. To answer these questions, we propose to use the ISARIC COVID-19 database as an example to define potential targets and objectives to optimise the creation and use of a large clinical database.

Participatory Approach

All contributors to the ISARIC database are invited to participate in this analysis through review and input on the statistical analysis plan and resulting publication. The outputs of this work will be disseminated as widely as possible to inform patient care and public health policy; this will include submission for publication in an international, peer-reviewed journal. ISARIC aims to include the names of all those who contribute data as cited contributors of this publication, subject to review of the final manuscript per the ISARIC publication policy.

Research Plan

Summary of Research Objectives
To propose sampling strategies to determine the following characteristics for an emerging infection: a) the clinical presentation or change of the clinical presentation, b) case fatality ratio (CFR) among hospitalised patients, c) risk factors and prognostic scores for severe outcomes (death, ICU admission, invasive mechanical ventilation, or composite outcome). To provide an evidence based approach for updating the ISARIC COVID-19 case report forms by evaluating the completion of these forms during the COVID-19 pandemic and the utility of the data in each variable.
Proposed Target Population
COVID-19 hospitalised patients
Clinical Questions/Descriptive Analyses

What are the best sampling strategies and minimum viable data required to answer the following questions?
 	
● Clinical presentation for COVID-19
  ○ When did we achieve enough sample size to clinically characterize COVID-19 patients?
  ○ What would be the best strategy to detect a change in clinical presentation? How could we use our learning from COVID-19 to improve for future outbreaks? What novel pathogenesis should we be ready to respond to (different presentations, different age groups)?

- How do case definitions from WHO, ECDC, PHE, and US CDC identify the patients in the ISARIC database over time and geographical region?

- Case fatality ratio(CFR)
  - What is COVID-19 CFR among hospitalised patients and what are the limitations of our data?
  - How does the representativeness of the data and geographical variability affect its estimation and what strategies could we use to minimize the impact of this variability?
  - When estimating a CFR, should we only incorporate patients from sites that enrol all patients or a random sample of them?
- Risk factors for severe outcomes
  - What is the required sample size and sampling strategy to detect a difference in the risk factor pattern for COVID-19 severe outcomes?
  - What is the impact of missing data on our risk factor analysis?
- How could we adapt our case report form based on the acquired knowledge (completeness and utility of data) over the last 2 years?

Planned Statistical Analyses, Methodology and Representation

Clinical Question	Planned statistical analysis	Planned representation in manuscript
Clinical presentation 1. When did we achieve a sufficient sample size to define COVID-19 clinical presentation? 2. What should be the sampling strategy necessary to monitor changes? 3. How do case definitions from WHO, ECDC, PHE, and US CDC identify the patients in the ISARIC database over time and geographical region?	1. Retrospective analysis of our dataset to evaluate the clinical presentation of COVID-19 over time. 2. Subset analyses among those sites that collect all cases. Re-sample patients within sites to determine the sampling strategies to detect a change in symptom presentation. Use different sampling strategies (random vs sequential) and sentinel sites within the dataset. Check for data `check-points` at different times during the pandemic.	Graphical representation of the timelines in which we achieved this sample size. Graphical or table representation of the different clinical presentations obtained by using different sampling strategies to detect a change in disease presentation. Graphical representation.

Case fatality ratio 1. Can we use a clinical platform to calculate CFRs? What are the limitations of our data to estimate hospital based CFRs. 2. How does this case fatality ratio differ if we take into account the different countries? Are those differences because of hospital admission policies, treatments, or admission into the ISARIC database? 3. Should we only incorporate patients from sites that enrol all patients or a random sample of them? 4. What would be the ideal study design to account for these factors?	1 and 2. Estimations of CFRs and estimation of their variability using HR. 3 and 4. Sample size estimations and evaluation of sampling strategies .	Graphical representations of CFRs by sample size and sampling strategies.
Risk factors for severe outcomes 1. What is the required sample size and what are the sampling strategies to detect a difference in the risk factor pattern for	1&3. Sample size estimations by simulating datasets with relationships (age, sex, comorbidities, site/country) similar to the one in the final data of ISARIC. Subset analysis among those sites that collect all	Graphical representation of sample size and sampling strategies required for risk factor analysis. Tables with HR or OR for the different risk factors by

COVID-19 severe outcomes? 2. What are the impacts of missing data on our risk factor analyses? 3. What volume and quality of data are required to generate and validate prognostic scores?	cases. Re-sample patients within sites to determine the sample size required to detect a change in risk factor presentation. Use different sampling strategies and sentinel sites within the dataset. 2. Assume that if data are missing, they are "no". Use of multiple imputation. Contrast the raw results, the ones using these strategies, and the ones using sampling strategies.	different sampling strategies.
How could we adapt our case report form based on the acquired knowledge (completeness and utility of data) over the last 2 years?	Descriptively evaluate data completeness in the entire case report form. Descriptively count the times that variables were used in different partner analysis and ISARIC reports. Qualitatively assess the usefulness of the collected variables based on the results from our first objective. Evaluation of the temporal patterns of data completeness. Evaluation of the geographical distribution (site or country level) of data completeness.	Graphical representations and tabulations.
Handling of Missing Data		

No missing data will be replaced, as the idea is to understand how missing data impacts our analyses.

Other Information/Timelines

Circulation of proposal and receipt of feedback from partners: December 20th

Analysis: December - February

Manuscript preparation and circulation for feedback: February - March

Publication: March - April 2022

References:

Dunning JW, Merson L, et al. **Open source clinical science for emerging infections.** *Lancet Infectious Diseases*. 2014 Jan;14(1):8-9. doi:10.1016/S1473-3099(13)70327-X