COVID-19: STATISTICS IN ACTION

Prof Sylvia Richardson CBE
Director
MRC Biostatistics Unit
Cambridge Institute of Public Health
University of Cambridge

President-Elect
Royal Statistical Society

Little was known about the epidemiology of infection by COVID-19 in early February 2020, but some of the characteristics of the SARS-CoV-2 transmission that had been described in China, namely high infectivity with early estimates of R in the order of 3 and suspicion that transmission could take place before the development of symptoms, were already an early warning of the potential for worldwide spread. Even less was known about the disease presentation and clinical management of the fraction of COVID-19 infected individuals who became severely ill with complications requiring intensive care treatment, and at significant risk of mortality. From the outset, it became apparent that the combination of epidemiological characteristics and clinical manifestations of COVID-19 could lead to a pandemic and required an unprecedented and urgent scientific collaborative effort.

There were - and still are - many unanswered questions regarding COVID-19. Statistics is contributing to provide evidence on many aspects of COVID-19, evidence which in turn provides a sound basis for policy decisions. The range of questions that are tackled straddles from basic science to public health, from understanding the immune and inflammation response to the virus to quantifying the overall disease burden. Progress is being made by matching each scientific question with appropriate data sources, purposely designed or routinely collected, and by using statistical approaches which are tailored to the type of data and question. To illustrate the productive melding of statistics and science that has taken place at pace since March 2020, I will draw on the experience of the MRC Biostatistics Unit (BSU) and the breadth of COVID-19 related projects that the BSU COVID-19 Working Group is engaged in.

MELDING OF STATISTICS AND SCIENCE TO TACKLE COVID-19

Engagement of statisticians has been most effective when it has been able to build on an existing network of trusted collaborations. Not only this has facilitated rapid access to relevant data sources, but it has also ensured that the much-needed dialog between analysts and scientific researchers can flow immediately. Through our long-standing collaboration with Public Health England (PHE) led by Daniela De Angelis, we were able to set-up quickly an agreement enabling the BSU Covid-19 Working Group to have access to hospital records of infected patients from the hospital surveillance systems. Crucially, we were able to report back on data quality, missing data and inconsistencies, and to discuss the interpretation of any results so that these were as robust as possible. Without an established network of collaboration, detailed understanding and critical appraisal of the data collection are difficult.

Existing collaborations have also been the basis for setting up new data collection protocols at speed. Our collaboration with clinical teams in Intensive Care Units (ICU) at Addenbrooke's and our previous work on understanding Electronic Care Records (ECR) were the basis of a COVID-19 ICU project, aimed at understanding how to target care to patients in most need. This project was approved quickly in March, and first ECR data were extracted within a month. Despite this fast start, full access to ECR data on a safe haven where powerful data science tools could be deployed was only operational mid-June as there were many regulatory barriers to satisfy. Finding an adequate balance between the much-needed rapid access to data and the importance of data protection has been raised by the current crisis and a fruitful topic for further discussion.

The breadth of questions raised simultaneously by this unprecedented pandemic is reflected in the diversity of statistical approaches that are being utilised to try to answer them. Much progress has been made recently on multidimensional methods for precision medicine, going away from one-at-a-time analyses of each biomarker towards integrative analyses of whole panels of biomarkers to get a deeper understanding of coordinated responses to external stimulus. Such integrative analyses will be key to understand the observed heterogeneity of the immune

Acknowledgment: I would like to acknowledge the BSU COVID-19 Working Group for many stimulating discussions and for their stellar and dedicated work.

BSU COVID-19 Working Group: Daniela De Angelis, Thomas Jaki, Sylvia Richardson, Brian Tom; Joshua Blake, Paul Birrell, Robert Goudie, Christopher Jackson, Peter Kirwan, Kevin Kunzmann, Anne Presanis, Pantelis Samartsidis, Shaun Seaman, Helene Ruffieux, Martin Wiegand.

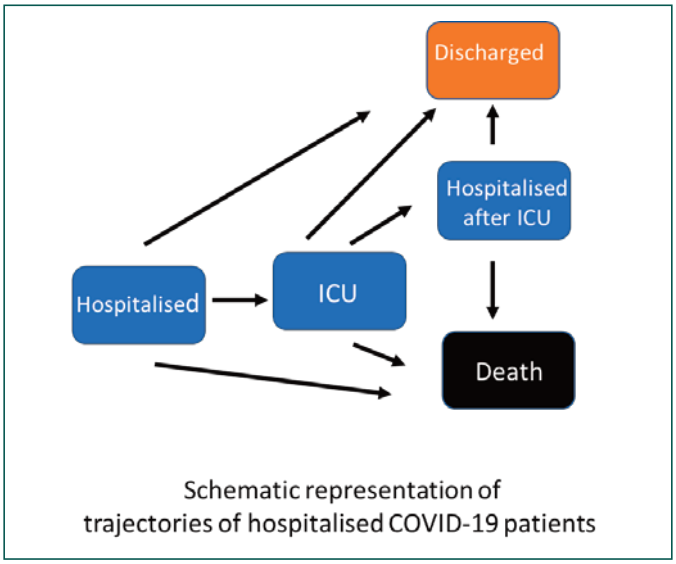
response to COVID-19. In collaboration with colleagues in the Cambridge Institute of Therapeutic Immunology & Infectious Disease, we are involved in a study of the different phases of the immune and inflammation responses to infection by SARS-CoV-2. We will use clustering and other integrative analysis tools to find coordinated modules of dysregulation in severe patients. Characterising subgroups of patients with similar immune responses is the prelude to better target treatment for each patient.

One important challenge since the beginning of the epidemic has been robust treatment evaluation since there was no known treatment for COVID-19 infections. Early on, there were many reports from small studies with inadequate designs, which created confusion. A recommendation from WHO and UK NERVTAG to evaluate whether existing treatments could be repurposed to treat COVID-19 was the impetus for the community to set-up a large Randomised Clinical Trial (RCT) with multiple arms and flexible design. The RECOVERY trial was conceived with chief investigators from Oxford University, and supported by a steering group which includes a BSU Lead. In view of the urgency and the much-needed flexibility, an adaptive design was chosen. Adaptive design is a framework that goes beyond classical RCT designs involving two groups, treated versus control. It includes features such as the ability to compare several arms to a common control arm, to introduce new treatments and secondary randomisation during the trial, to stop treatment at interim analyses, and to carry out dose finding. Crucially, this increased flexibility is not at the cost of the integrity and the validity of the results which is maintained. Such methodological underpinning had been previously developed by researchers at the BSU, working together with a network of trial methodologists to establish statistical properties of a variety of adaptive designs. Thanks to its design, the RECOVERY trial, which started on the 19th March, has already being able to report on three interim analyses with immediate impact on the clinical care of patients.

For a minority of infected individuals, the viral phase is followed by an excessive inflammatory response, which can have severe consequences in particular on the lung, creating pneumonia and acute respiratory distress syndrome, with fatal outcome for some. It is important to study the trajectory of hospitalised patients to better understand the severity burden that COVID-19 imposes on the health system and to inform the general population. After hospital admission, patients can follow a number of trajectories (see diagram) including admission to ICU, readmission to a ward after ICU, discharge or death. It is particularly useful to estimate the probabilities of transition between these different states as well as the length of stay in ICU or other wards. This can be done through the framework of multistate models.

Typically, such analyses are carried out using hospital records data, which has been collected for different purposes such as hospital management or audit. Currently, the main sources of data are the COVID-19 Hospitalisation in England Surveillance System (CHESS), which has been set up by PHE, and the Covid-19 Clinical Information Network (CO-CIN). As these observational data sources are not purpose built, it is important to consider carefully issues related to inconsistencies, missing data, censoring, and population selection as these will influence the results and their interpretation. As is good practice in any analysis of observational data, care must be taken to assess the sensitivity of the analyses to these issues.

A final and beautiful example of melding between statistics and science on COVID-19 is the modelling work carried out by Daniela De Angelis and her BSU-PHE team to reconstruct the evolution of the pandemic in the UK. This is fully detailed in a separate article to which I refer the readers.



Schematic representation of trajectories of hospitalised COVID-19 patients

THE ROLE OF STATISTICS

In summary, as statisticians, our role is to produce evidence from data and quantify uncertainty. This needs to be done in a principled, transparent and interpretable way so that policy makers are fully aware of the assumptions underpinning the analyses and can make informed decisions. Having access to good-quality data is paramount and greatly facilitated by long-term multi-disciplinary collaborations. There has been a huge mobilisation of the scientific community on Covid-19; statisticians are involved in the whole spectrum of projects, making use of a large portfolio of statistical methods.

Exploiting and repurposing routine data collection are certainly useful. The UK is internationally known for its strength in health data science made possible through initiatives like Health Data Research UK, and this has been of major benefit to the current crisis. But observational data comes with its limitations and needs careful analysis as well as consideration of potential sources of biases. The latter can be avoided by having purposely designed and well-conducted studies. Hence, being nimble in setting-up quickly such studies to tackle emerging health threats is essential.

Significant strides have been made but lessons can be learnt to better prepare for the future. These include planning comprehensive, well-designed and aligned data streams covering multi-facet surveillance and involving all the relevant disciplines, designing methods for triangulating evidence in the context of surveillance, and carrying out a constant evaluation of operational systems and policies. □