Containing COVID, part 1 | First things first: the difficulty of building an evidence base

*Unprecedented cooperation meant that scientists quickly built a picture of the novel coronavirus. Yet the evidence they accumulated should not be taken at face value. **Linda Hantrais (LSE)** and **Susanne MacGregor (LSHTM)** introduce a series about the challenges of learning from statistical datasets and other scientific evidence.*

In the months after the pandemic began, teams of biostatisticians, virologists, epidemiologists, clinicians and statistical modellers, collaborating across national boundaries, built on the knowledge acquired from previous pandemics to develop the scientific evidence base for COVID-19. This collaboration resulted in thousands of scientific papers, often posted pre-publication and subject to validation, reporting early evidence about the virus.

The impact of the COVID pandemic was found to differ from previous epidemics and global crises in several ways — not least due to the over-arching characteristics of modernity, such as ageing populations and the growth of science and medicine, as well as overall Western economic prosperity and global inequality. Initial evidence indicated that COVID was a novel and virulent strain of coronavirus against which populations appeared to have no natural immunity, and for which no known cure or proven therapeutic strategies existed. Unlike earlier pandemics and global crises, older people with co-morbidities were found to be more likely to die from the disease. Comparisons with death rates in the Spanish flu pandemic in 1918 and both world wars show how the impact of COVID deaths fell on very different age groups.

Published statistics should not be taken at face value

Using the best data available while acknowledging uncertainty and gaps in the evidence base, statisticians developed interactive online tools to track the spread of the pandemic in real-time (nowcasts), to identify cumulative trends in the numbers of infections and deaths, and to predict (forecast) how it might progress. Nowcasters and forecasters invited challenges to their interpretations of the data and were exposed to harsh criticisms by scientific colleagues, politicians who claimed to have based life-changing decisions on the 'best scientific evidence available', and by an unforgiving public when events proved predictions to be wrong.

Based on global datasets showing the total number of reported deaths since the beginning of the pandemic, by late April 2021 the US, Brazil, Mexico, India, the UK, Italy, Russia and France were all recording over 100,000 COVID-19-related deaths. When analysed in relation to population size, the cumulative figures tell a different story: 41 countries reported over 1,000 deaths per million inhabitants, rising to over 2,000 per million in 11 countries, with Hungary and Czechia in 1st and 3rd positions in the rank order. According to this indicator, the US was in 16th position, Brazil in 14th, Mexico 21st, India 119th, the UK 13th, Italy 12th, Russia 58th and France 23rd. Vietnam, Taiwan, Tanzania and New Zealand remained near to the bottom of the list for both indicators. A year earlier when Europe was at the epicentre of the first wave of the pandemic, Hungary and Czechia and three other Central and Eastern European member states, which had moved into the 2,000 per million bracket in April 2021, had been among the countries displaying the lowest death rates per million inhabitants in the EU27, suggesting that they had not drawn lessons from their earlier experience. The low ranking per million inhabitants for India could be attributed both to the relatively small number of COVID deaths during the first wave and to under-reporting on a massive scale in the second wave.

One of the first lessons to be drawn from COVID datasets over time and space, as presented in international comparative tables, was that published statistics should not be taken at face value. With hindsight, it was clear that all but the most circumscribed lessons drawn from data about causal links between death rates and national policy responses should have been treated with extreme caution. Before any meaningful lessons could be learnt, the raw figures needed to be contextualised not only in relation to population size but also with reference to country-specific data collection methods and resources. Even then, assessing, interpreting and understanding the wider implications of publicly available data would never be a straightforward process. It requires detailed scrutiny of a multitude of potential contributing factors encompassing epidemiology and epistemology. In combination with social and biological forces, these factors create what has been aptly described as a 'causal patchwork'.

For scientists seeking reliable evidence from their analysis of the first year of the pandemic to support more widely applicable lessons, COVID continued to pose unprecedented challenges. In subsequent posts, we assemble international evidence from a range of disciplines to consider what lessons, if any, may be gleaned by unpicking the different threads in the patchwork and by adding new pieces. Scientists, politicians and the wider public faced a steep learning curve. We explore both 'known unknowns' and 'unknown unknowns' in searching for causal explanations and in recognising the limits of scientific knowledge, as well as the opportunities that ignorance provides for further investigation.

This post represents the views of the authors and not those of the COVID-19 blog, nor LSE.

Date originally posted: 2021-04-29

Permalink: https://blogs.lse.ac.uk/covid19/2021/04/29/containing-covid-part-1-first-things-first-the-difficulty-of-building-an-evidence-base/

Blog homepage: https://blogs.lse.ac.uk/covid19/