Statistical Science
2022, Vol. 37, No. 2, 160–161
https://doi.org/10.1214/22-STS372IN

Data Science in a Time of Crisis: Lessons from the Pandemic

Chiara Sabatti and John M. Chambers

The exceptional shock of the COVID-19 pandemic has brought about an equally exceptional scientific response, over a wide range of disciplines and with a spirit of collaboration and mutual support.

This issue of *Statistical Science* contains five reports on research related to the pandemic and four perspectives on lessons learned and thoughts for the future. We hope the issue will contribute to ongoing exploration and discussion from the special perspective of the *Statistical Science* community, particularly in the context of avoiding or managing future crises.

As with any substantial data science challenge, three fundamental aspects are involved: *data*—obtaining, organizing and validating the information relevant to the study; *analysis*—models, summaries and other computations motivated by the scientific questions; and *communication*—valid information derived from the study in a form helpful for the scientific community, for decision makers or for the public.

The COVID-19 pandemic has challenged the scientific response in all aspects. Data on the prevalence and effects of the disease fall short because of the rapidity of the spread, the wide range of symptoms and especially from the global context. The accuracy, consistency and simple availability of the relevant data suffer. Future fixes will need improvements in policy but statistical techniques can to some extent compensate. Pooled testing can use models to improve accuracy (Comess et al. [3]) and analytic techniques can compensate for time delays (Jahja, Chin and Tibshirani [4]).

The relevant science ranges from the most basic genomic understanding of the virus to all the social, economic and other human effects of the pandemic. At the genomic level, powerful models are available but with challenges at the limits of analysis and computing, as discussed by Cappello et al. [1].

Communication of the scientific insights is arguably the challenge needing the greatest novelty of approach. For the pandemic, this has to encompass a wide variety of listeners: those directly involved in the fight; those involved in formulating policy; and the whole population affected by the results. Wang et al. [8] present analysis and graphical techniques to aid patient monitoring in the crucial hospital context. Nicholson et al. [7] introduce the concept of *interoperability* as a framework for communication between policy makers and statistical analysis.

The technical papers are supplemented by four perspectives. Yu and Singh [9] distill seven principles from their experience with the pandemic. Mukherjee [6] and Lin [5] recount their experiences, in India and with the early Wuhan data respectively, and go on to present reflections with implications for dealing with future epidemics. Chambers [2] considers how science might respond to a broad range of still graver threats facing us, using the COVID-19 experience and the example of Bell Labs research as reference points.

Any single journal issue can only report a very small fraction of the dramatic surge in activity. To the credit of the profession, many statisticians engaged directly with the analysis and modeling of data related to the pandemic. They provided support for local, national and international health organizations. Presentation of their analysis helped the public to interpret information and assisted authorities in designing policies. The vast majority of the researchers involved in these efforts had not specialized in such data before spring 2020.

Our goal for this special issue was to highlight a few of these contributions. In no sense could we hope to find a "representative" sample, let alone a "best" selection. Given the desire to produce the issue in a reasonable time frame, and considering the heavy commitments of potential authors, we used a thoroughly *ad hoc* search process. A first targeted request for contributions was sent out in early autumn 2020. We also sought suggestions for additional contributions, leading to further invitations. This very unsystematic sampling is inevitably biased, notably towards United States activities, for which we apologize. The international perspectives in the papers by Nicholson et al., Mukherjee and Lin are particularly appreciated.

Inevitably also a number of technical topics are conspicuously absent; for example, studies related to vaccines or to the health disparities exposed by the pandemic. Much other research has taken place and still more challenges exist for future investigation. The perspective by Mukherjee has a valuable summary of important areas of statistical research.

Chiara Sabatti is Professor, Biomedical Data Science and of Statistics, Stanford University, Stanford, California 94305-2070, USA (e-mail: sabatti@stanford.edu). John M. Chambers is Adjunct Professor in Statistics and Senior Advisor, Data Science, Stanford University, Stanford, California 94305-2070, USA (e-mail: jmc4@stanford.edu).