Editorial: Data Science in Action in Response to the Outbreak of COVID-19

Dean Follmann[1], Peter X. K. Song[2], Hansheng Wang[3], and Jun Yan[4]

[1]*Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland, USA*
[2]*Department of Biostatistics, University of Michigan, Ann Arbor, Michigan, USA*
[3]*Guanghua School of Management, Peking University, Beijing, China*
[4]*Department of Statistics, University of Connecticut, Storrs, Connecticut, USA*

This special issue of the *Journal of Data Science* on "Data Science in Action in Response to the Outbreak of COVID-19" was designed to provide a timely platform for data scientists fighting the COVID-19 pandemic to communicate their research works. This was the first special issue of this kind from statistical journals on this topic. The idea was initiated and formed near the end of February, 2020, when the COVID-19 outbreak was close to an end in China. Data scientists along with experts from other substantive fields have been actively taking up various challenges arising from the pandemic. This was a perfect opportunity to demonstrate the critical impact of statistics and data science in a global emergency. Our hope was to attract contributions from any branch of data science, including mathematics, computer science, statistics, epidemiology, economics, engineering, and environmental sciences, as long as they were relevant to the pandemic of COVID-19.

As co-guest-editors of the special issue, we were excited about the positive response from colleagues working on various types of COVID-19 data. This issue contains eight papers, two of which are accompanied with discussions and rejoinders, covering a wide range of methods and applied topics on data science research about COVID-19. The two discussion papers both extend the classic susceptible-infectious-removed (SIR) model in epidemiology. Wang et al. incorporate time-varying quarantine measures, with a companion R package `eSIR` which facilitates assessment of the effects of different levels of social distancing interventions. The R package has been widely used. Sun et al. allow tracking the effective reproductivity of the epidemic and assessment of the control measures through a varying coefficient SIR model. Zhang et al. study the transmissibility of COVID-19 through meta analyses on the basic reproduction number, the incubation time, the serial interval, and the epidemic doubling time. Cai et al. report risk factors affecting COVID-19 fatality using an adaptive LASSO penalized slice inverse regression. Dong et al. estimate the basic reproduction number of COVID-19 with a meta analysis and applied it to evaluating the effectiveness of isolation measures in different countries. Wong assesses the impacts of mutations to the structure of COVID spike protein via sequential Monte Carlo. Charvadeh and Yi presents a case study on data visualization and descriptive analysis for the epidemiological characteristics of COVID-19. Chen and Ting give a comprehensive review of design considerations for vaccine trials with a special focus on COVID-19. Reproducibility is one of our highlights: all the papers have their code/data available through public repositories or online supplementary materials.

Some of the papers have already had a notable impact on COVID-19 data science research and practice since their online appearance as early as April, 2020. For example, the eSIR method has been used to predict the COVID-19 pandemic in India, leading to a nationwide intervention policy issued by the Indian government (Ray et al., 2020). Three authors, Peter Song, Song Xi Chen, and Grace Yi, presented parts of their works in a webinar series "Data Science in Action in Response to the Outbreak of COVID-19", which ran weekly from April 17 to July 24, 2020. The series was co-organized by the *Journal of Data Science* and the American

Statistical Association (ASA) Section on Statistical Learning and Data Science (SLDS), with co-sponsorship from a few sister societies and ASA sections. Over 1000 people registered. The recordings and slides from the most of the webinars have been publicly available at the University of Connecticut Statistical Data Science Lab (https://statds.org/events/webinar_dsa2020/schedule.html). The release of this special issue will stimulate further contributions to the battle with COVID-19 from the data science community.

Time is of the essence. We are grateful to the authors, discussants, and reviewers for their swift action in writing/revising, commenting, and reviewing, without which the timely publication of this special issue would be impossible. The editorial assistant Dr. Jing Zhou of Renmin University of China managed the manuscripts manually during the transition to a new editorial system. Dr. Wenjie Wang of Eli Lilly developed a new LaTeX template that gives the journal a much needed fresh look. Gradaute students from Dr. Jun Yan's group at the University of Connecticut typeset the articles. Their tremendous efforts are deeply appreciated.

The outbreak continues and we are continuing accepting submissions to the next special issue on "Data Science in Action in Response to the Outbreak of COVID-19". To speed up communications in Chinese, we have a special issue in Chinese with English abstracts which is also accepting submissions. We look forward to receiving more high quality contributions that advance the research in data science in response to COVID-19.

References

Ray D, Salvatore M, Bhattacharyya R, Wang L, Du J, Mohammed S, et al. (2020). Predictions, role of interventions and effects of a historic national lockdown in India's response to the the COVID-19 pandemic: Data science call to arms. *Harvard Data Science Review*. Forthcoming, https://doi.org/10.1162/99608f92.60e08ed5.