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

Preamble .................................................................................................................. Sonia Petrone, Editor 159
Data Science in a Time of Crisis: Lessons from the Pandemic
.................................................................................................................. Chiara Sabatti and John M. Chambers 160

Papers

Statistical Challenges in Tracking the Evolution of SARS-CoV-2
.................................................................................................................. Lorenzo Cappello, Jaehee Kim, Sijan Liu and Julia A. Palacios 162
Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality
.... George Nicholson, Marta Blangiardo, Mark Briers, Peter J. Diggle, Tor Erlend Fjelde, Hong Ge, Robert J. B. Goudie, Radka Jersakova, Ruairidh E. King, Brieuc C. L. Lehmann, Ann-Marie Mallon, Tullia Padellini, Yee Whye Teh, Chris Holmes and Sylvia Richardson 183
Real-Time Estimation of COVID-19 Infections: Deconvolution and Sensor Fusion
.................................................................................................................. Maria Jahja, Andrew Chin and Ryan J. Tibshirani 207
Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic
.................................................................................................................. Saskia Comess, Hannah Wang, Susan Holmes and Claire Donnat 229
Learning and Predicting from Dynamic Models for COVID-19 Patient Monitoring
.................................................................................................................. Zitong Wang, Mary Grace Bowring, Antony Rosen, Brian Garibaldi, Scott Zeger and Akihiko Nishimura 251

Perspectives

Seven Principles for Rapid-Response Data Science: Lessons Learned from Covid-19
Forecasting .......................................................................................................... Bin Yu and Chandan Singh 266
Being a Public Health Statistician During a Global Pandemic ........ Bhramar Mukherjee 270
Lessons Learned from the COVID-19 Pandemic: A Statistician’s Reflection ...... Xihong Lin 278
Data, Science, and Global Disasters ................................................................. John M. Chambers 284
Special Issue on Data Science in a Time of Crisis: Lessons from the Pandemic

Preamble
Sonia Petrone, Editor

Data Science in a Time of Crisis: Lessons from the Pandemic
Chiara Sabatti and John M. Chambers

Papers

Statistical Challenges in Tracking the Evolution of SARS-CoV-2
Lorenzo Cappello, Jaehee Kim, Sifan Liu and Julia A. Palacios

Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality
George Nicholson, Marta Blangiardo, Mark Briers, Peter J. Diggle, Tor Erlend Fjelde, Hong Ge, Robert J. B. Goudie, Radka Jersakova, Ruairidh E. King, Brieuc C. L. Lehmman, Ann-Marie Mallon, Tullia Padellini, Yee Whye Teh, Chris Holmes and Sylvia Richardson

Real-Time Estimation of COVID-19 Infections: Deconvolution and Sensor Fusion
Maria Jahja, Andrew Chin and Ryan J. Tibshirani

Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic
Saskia Comess, Hannah Wang, Susan Holmes and Clare Donnat

Learning and Predicting from Dynamic Models for COVID-19 Patient Monitoring
Zitong Wang, Mary Grace Bowering, Antony Rosen, Brian Garibaldi, Scott Zeger and Akhiko Nishimura

Perspectives

Seven Principles for Rapid-Response Data Science: Lessons Learned from Covid-19 Forecasting
Bin Yu and Chandan Singh

Being a Public Health Statistician During a Global Pandemic
Bhramar Mukherjee

Lessons Learned from the COVID-19 Pandemic: A Statistician’s Reflection
Xihong Lin

Data, Science, and Global Disasters
John M. Chambers
Preamble

Sonia Petrone, Editor
Data Science in a Time of Crisis: Lessons from the Pandemic

Chiara Sabatti and John M. Chambers

REFERENCES


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).
Statistical Challenges in Tracking the Evolution of SARS-CoV-2

Lorenzo Cappello, Jaehee Kim, Sifan Liu and Julia A. Palacios

Abstract. Genomic surveillance of SARS-CoV-2 has been instrumental in tracking the spread and evolution of the virus during the pandemic. The availability of SARS-CoV-2 molecular sequences isolated from infected individuals, coupled with phylodynamic methods, have provided insights into the origin of the virus, its evolutionary rate, the timing of introductions, the patterns of transmission, and the rise of novel variants that have spread through populations. Despite enormous global efforts of governments, laboratories, and researchers to collect and sequence molecular data, many challenges remain in analyzing and interpreting the data collected. Here, we describe the models and methods currently used to monitor the spread of SARS-CoV-2, discuss long-standing and new statistical challenges, and propose a method for tracking the rise of novel variants during the epidemic.

Key words and phrases: Phylodynamics, genetic epidemiology, coalescent, Bayesian nonparametrics, birth-death processes, SIR models.

REFERENCES


Lorenzo Cappello is Assistant Professor, Departments of Economics and Business, Universitat Pompeu Fabra, 08005, Spain (e-mail: lorenzo.cappello@upf.edu). Jaehee Kim is Assistant Professor, Department of Computational Biology, Cornell University, Ithaca, New York 14853, USA (e-mail: jaehae.kim@cornell.edu). Sifan Liu is a Ph.D. student, Department of Statistics, Stanford University, Stanford, California 94305, USA (e-mail: sfliu@stanford.edu). Julia A. Palacios is Assistant Professor, Departments of Statistics and Biomedical Data Sciences, Stanford University, Stanford, California 94305, USA (e-mail: julia.pr@stanford.edu).


Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality

George Nicholson, Marta Blangiardo, Mark Briers, Peter J. Diggle, Tor Erlend Fjelde, Hong Ge, Robert J. B. Goudie, Radka Jersakova, Ruairidh E. King, Brieuc C. L. Lehmann, Ann-Marie Mallon, Tullia Padellini, Yee Whye Teh, Chris Holmes and Sylvia Richardson

Abstract. We present interoperability as a guiding framework for statistical modelling to assist policy makers asking multiple questions using diverse datasets in the face of an evolving pandemic response. Interoperability provides an important set of principles for future pandemic preparedness, through the joint design and deployment of adaptable systems of statistical models for disease surveillance using probabilistic reasoning. We illustrate this through case studies for inferring and characterising spatial-temporal prevalence and reproduction numbers of SARS-CoV-2 infections in England.

Key words and phrases: Bayesian graphical models, Bayesian melding, COVID-19, evidence synthesis, interoperability, modularization, multi-source inference.

REFERENCES


[67] UK DATA SERVICE CENSUS. Available at https://ukdataservice.ac.uk/learning-hub/census/.


Real-Time Estimation of COVID-19 Infections: Deconvolution and Sensor Fusion

Maria Jahja, Andrew Chin and Ryan J. Tibshirani

Abstract. We propose, implement, and evaluate a method to estimate the daily number of new symptomatic COVID-19 infections, at the level of individual U.S. counties, by deconvolving daily reported COVID-19 case counts using an estimated symptom-onset-to-case-report delay distribution. Importantly, we focus on estimating infections in real-time (rather than retrospectively), which poses numerous challenges. To address these, we develop new methodology for both the distribution estimation and deconvolution steps, and we employ a sensor fusion layer (which fuses together predictions from models that are trained to track infections based on auxiliary surveillance streams) in order to improve accuracy and stability.

Key words and phrases: COVID-19, nowcasting, deconvolution, sensor fusion.

REFERENCES


Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. and Brilliant, L. (2009). Detecting in-


Abstract. Pooled testing offers an efficient solution to the unprecedented testing demands of the COVID-19 pandemic, despite their potentially lower sensitivity and increased costs to implementation in certain settings. Assessments of this trade-off typically assume the underlying infection statuses of pooled specimens to be independent and identically distributed. Yet, in the context of COVID-19, these assumptions are often violated: testing done on networks (housemates, spouses, co-workers) captures individuals with correlated infection statuses and risk, while infection risk varies substantially across time, place and individuals. Neglecting dependencies and heterogeneity may bias established optimality grids and induce a sub-optimal implementation of the procedure. As a lesson learned from this pandemic, this paper highlights the necessity of integrating field sampling information with statistical modeling to efficiently optimize pooled testing. Using real data, we show that (a) greater gains can be achieved at low logistical cost by exploiting natural correlations (nonindependence) between samples—allowing improvements in sensitivity and efficiency of up to 30% and 90%, respectively; and (b) these gains are robust despite substantial heterogeneity across pools (nonidentical). Our modeling results complement and extend the observations of Barak et al. (Sci. Transl. Med. 13 (2021) 1–8) who report an empirical sensitivity well beyond expectations. Finally, we provide an interactive tool for selecting an optimal pool size using contextual information.

Key words and phrases: COVID-19, pooled testing, correlations, heterogeneity.

REFERENCES


CAPS (2020). Pooled Testing: Guidance from the CAP’s Microbiology Committee. Available at www.cap.org/covid-19/pooled-
Learning and Predicting from Dynamic Models for COVID-19 Patient Monitoring

Zitong Wang, Mary Grace Bowring, Antony Rosen, Brian Garibaldi, Scott Zeger and Akihiko Nishimura

Abstract. COVID-19 has challenged health systems to learn how to learn. This paper describes the context, methods and challenges for learning to improve COVID-19 care at one academic health center. Challenges to learning include: (1) choosing a right clinical target; (2) designing methods for accurate predictions by borrowing strength from prior patients’ experiences; (3) communicating the methodology to clinicians so they understand and trust it; (4) communicating the predictions to the patient at the moment of clinical decision; and (5) continuously evaluating and revising the methods so they adapt to changing patients and clinical demands.

To illustrate these challenges, this paper contrasts two statistical modeling approaches—prospective longitudinal models in common use and retrospective analogues complementary in the COVID-19 context—for predicting future biomarker trajectories and major clinical events. The methods are applied to and validated on a cohort of 1678 patients who were hospitalized with COVID-19 during the early months of the pandemic. We emphasize graphical tools to promote physician learning and inform clinical decision making.

Key words and phrases: Longitudinal data analysis, prediction, inverse regression, decision support, statistical graphics.

REFERENCES

Zitong Wang is a PhD student, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA (e-mail: zwang238@jhmi.edu). Mary Grace Bowring is an MD/PhD student, Departments of Biomedical Engineering and Biostatistics, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (e-mail: mbowrin1@jhmi.edu). Antony Rosen is Mary Betty Stevens Professor of Medicine and Vice Dean for Research, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (e-mail: arosen@jhmi.edu). Brian Garibaldi is Associate Professor, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (e-mail: bgaribal@jhmi.edu). Scott L. Zeger (co-senior author) is John C. Malone Professor, Department of Biostatistics and Medicine, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA (e-mail: sz@jhu.edu). Akihiko (Aki) Nishimura (co-senior author) is Assistant Professor, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA (e-mail: aki.nishimura@jhu.edu).


Seven Principles for Rapid-Response Data Science: Lessons Learned from Covid-19 Forecasting

Bin Yu and Chandan Singh

Abstract. In this article, we take a step back to distill seven principles out of our experience in the spring of 2020, when our 12-person rapid-response team used skills of data science and beyond to help distribute 340,000+ units of Covid PPE. This process included tapping into domain knowledge of epidemiology and medical logistics chains, curating a relevant data repository, developing models for short-term county-level death forecasting in the US, and building a website for sharing visualization (an automated AI machine). The principles are described in the context of working with Response4Life, a then-new nonprofit organization, to illustrate their necessity. Many of these principles overlap with those in standard data-science teams, but an emphasis is put on dealing with problems that require rapid response, often resembling agile software development. The technical work from this rapid response project resulted in a paper (Altieri et al. (2021)); see also this interview for more background (Yu and Meng (2021)).

Key words and phrases: Coronavirus, forecasting, county-level, data-science.

REFERENCES


Being a Public Health Statistician During a Global Pandemic

Bhramar Mukherjee

Abstract. In this perspective, I first share some key lessons learned from the experience of modeling the transmission dynamics of SARS-CoV-2 in India since the beginning of the COVID-19 pandemic in 2020. Second, I discuss some interesting open problems related to COVID-19 where statisticians have a lot to contribute to in the coming years. Finally, I emphasize the need for having integrated and resilient public health data systems: good data coupled with good models are at the heart of effective policymaking.

Key words and phrases: COVID-19 pandemic, epidemiological models, forecasting, public health data systems, research infrastructure, transmission dynamics, vaccine effectiveness.

REFERENCES

[17] COVIDSURG COLLABORATIVE et al. (2021). Machine learning risk prediction of mortality for patients undergoing surgery...


[21] COV-IND-19 STUDY GROUP (2020). What can we learn from the Delhi Seroprevalence Study and a single number that was reported? Medium. Available at https://brhamarm.medium.com/what-can-we-learn-from-the-delhi-seroprevalence-study-and-a-single-number-that-was-reported-45e4cf185357 (accessed Mar. 03, 2022).


[36] GRAM, M. A. et al. Vaccine effectiveness against SARS-CoV-2 infection, hospitalization, and death when combining a first dose ChAdOx1 vaccine with a subsequent mRNA vaccine in Denmark: A nationwide population-based cohort study. PLOS Med. 18 e1003874. https://doi.org/10.1371/journal.pmed.1003874


Lessons Learned from the COVID-19 Pandemic: A Statistician’s Reflection

Xihong Lin

Abstract. In this article, I will discuss my experience as a statistician involved in COVID-19 research in multiple capacities in the last two years, especially in the early phase of the pandemic. I will reflect on the challenges and the lessons I have learned in pandemic research regarding data collection and access, epidemic modeling and data analysis, open science and real time dissemination of research findings, implementation science, media and public communication, and partnerships between academia, government, industry and civil society. I will also make several recommendations on navigating the next stage of the pandemic and preparing for future pandemics.

Key words and phrases: COVID-19 research, data collection and access, data analysis, epidemic modeling, implementation science, open science, media and public communication.

REFERENCES


NATURE EDITORIAL TEAM (2022). This is no time to stop tracking COVID-19. Available at https://www.nature.com/articles/d41586-022-00788-y.


Xihong Lin is Professor of Biostatistics at Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA, and Professor of Statistics at Harvard University, Cambridge, Massachusetts 02138-2901, USA (e-mail: xlin@hsph.harvard.edu).
Data, Science, and Global Disasters

John M. Chambers

Abstract. The spread and impact of COVID-19 have disrupted human activities and energized a response of scientific activity on a remarkable, nearly unprecedented scale. This has somewhat distracted attention from a broad range of less immediate but fundamentally more serious global threats resulting from human actions. These can be collectively labelled the anthropocene disasters.

Science cannot itself prevent or mitigate them. To do so requires a global policy resolve not currently existing. When and if that resolve emerges, science will be essential for guiding action. This science will be radically data-intensive, global and inclusive. Teams will be required that include the best and most motivated individuals from all relevant scientific disciplines, plus members knowledgeable about implementing likely policy recommendations. Such participants must be attracted to join and then properly supported and rewarded—not likely with current academic structures. Some insights can be gained from the recent experience with COVID-19 and the much less recent example of research at Bell Labs.

Key words and phrases: Data science, climate change, biodiversity, pandemic, Bell Labs.

REFERENCES


John M. Chambers is Adjunct Professor in Statistics and Senior Advisor for Data Science, Stanford University, Stanford, California 94305-4020, USA. He was in Bell Labs Research from 1966 to 2005 (e-mail: jmc4@stanford.edu).
INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

President: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, Washington 98195-4350, USA

President-Elect: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

Past President: Regina Y. Liu, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854-8019, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. Editors: Enno Mammen, Institute for Applied Mathematics, Heidelberg University, 69120 Heidelberg, Germany. Lan Wang, Miami Herbert Business School, University of Miami, Coral Gables, FL 33124, USA

The Annals of Applied Statistics. Editor-in-Chief: Ji Zhu, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The Annals of Probability. Editors: Alice Guionnet, Unité de Mathématiques Pures et Appliquées, ENS de Lyon, Lyon, France. Christophe Garban, Institut Camille Jordan, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France

The Annals of Applied Probability. Editors: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. Qi-Man Shao, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P.R. China

Statistical Science. Editor: Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

The IMS Bulletin. Editor: Vlada Limic, UMR 7501 de l’Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France
BRADLEY EFRON
TREVOR HASTIE

COMPUTER AGE
STATISTICAL
INFERENCE

ALGORITHMS, EVIDENCE, AND DATA SCIENCE