ICSA Bulletin

Volume 34/2, Sep 2022
ISSN 2226-2393

Editorial Staff

From the Editor 1
From the 2022 President, ICSA 2
From the Executive Director (2020-2022) 3
The 2023 ICSA Official Election Results 4
The 2022 New Fellows of ASA and IMS 4
The 2022 ICSA Awards 6
Report from the ICSA 2022 Applied Statistics Symposium . . 7
Report from the 2022 ICSA Program Committee 11
News from the ICSA-Canada Chapter 12
Report from the 2022 ICSA China Conference 14
Report from Statistica Sinica 15
Report from Statistics in Biosciences 19
ICSA Financial Report 21
Congratulations to Song Xi Chen 22
Congratulations to Heping Zhang on Winning IMS Neyman Lecture 22
Congratulations to Xuming He on Winning Carver Award . . . 23
XL-Files: When a Statistician becomes a (COVID) Statistic . . 23
On the Convergence of Epidemiology, Biostatistics, and Data Science 26
Statistics in Biosciences CALL FOR PAPERS 35
Upcoming Events 36

From the Editor

Ming Wang

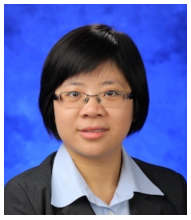
Dear ICSA Members:

Welcome to our 2022 Sep issue of the ICSA Bulletin! This year marks a step forward for our in-person meetings and social gatherings. As Dr. Liu (Executive Direction of ICSA) mentioned, our ICSA Applied Statistics Symposium was successfully held in University of Florida, Gainesville, FL in June 2022, and the committee meetings were returned in major statistical conferences of ENAR and JSM. I was fortunate and felt excited to attend these events in person. While COVID-19 brings disruptions to all, our ICSA members have accomplished much and adapted to a new norm with continuous and productive work.

In this issue, we would like to share with you several articles that may be interested to your career and research interest. In the column "XL-Files", Professor Xiao-Li Meng shared his most recent updates during the pandemic and provided insights being a statistician when getting COVID infection. This is a reprint from the author's column "When a Statistician becomes a (COVID) Statistic" in the IMS Bulletin, August 2022. Also, we have one article entitled "On the Convergence of Epidemiology, Biostatistics, and Data Science" from Neal D. Goldstein, Michael LeVasseur and Leslie A. McClure. They presented throughout review and discussion of three major disciplines, epidemiology, biostatistics, and data science, with focus on their convergence in the era of electronic health data, tackling challenges and offering suggestions for collaborative teamwork. This article is reprinted from the Harvard Data Sciences Review (HDSR) with permission (https://hdsr.mitpress.mit.edu/pub/twqhhlhr/release/3). Of note, the column of "Yi's FDA Stories" will continue in the upcoming issue, and we look forward to Dr. Yi Tsong's experience sharing in 2023. Besides, we have a memorial article of Dr. Don Ylvisaker written by Drs. Ching-Shui Cheng, Ker-Chau Li and Jane-Ling Wang. Dr. Ylvisaker (1933-2022) is truly a great friend of our ICSA and received the ICSA Outstanding Service Award during the Applied Statistics Symposium held in San Diego in 2004. He will forever in our hearts, and his immense service and contributions to our ICSA, society and science are appreciated!

Turning to ICSA business, we include the letter from the 2022 ICSA President, Dr. Zhezhen Jin, and the letter from the executive director (2020-2022), Dr. Mengling Liu; announcement of the recipients of the 2022 ICSA Awards and the other prestigious awards; new fellows of ASA and IMS in our ICSA community; reports from our ICSA treasure and program committee; reports from ICSA sponsored or co-sponsored journals; summary report from the 2022 ICSA China conference and Canada chapter; and the announcements of the upcoming meetings/conferences at the end of this issue. In particular, we have the 2022 ICSA Officers successfully selected, and congratulations to Dr. Xun Chen, Global Head of Biostatistics and programming Department at Sanofi on being elected as 2023 President Elect, and Drs. Huazhen Lin, Ming Tan, Li Wang, Yanping Wang and Min Zhang as 2023-2025 Board of Directors. With the leadership of this new team, we believe that our ICSA will keep growing and stronger!

In the end, I am extremely grateful to all the contributors, ICSA executives and committee members for their support on our bulletin. Thanks to my great assistant, Dr. Chixiang Chen, for putting these files together. Hope you' ll enjoy our issue! As always, welcome to submit any news, updates, academic opinions, or anything else you want to share with our members! Cheers! Keep moving forward!



Ming Wang, Ph.D.
Editor-in-Chief, ICSA Bulletin
Associate Professor
Department of Population and Quantitative Health Sciences
Case Western Reserve University

From the 2022 President, ICSA

Zhezhen Jin

Dear ICSA Members:

Despite the COVID-19 pandemic, ICSA has had an outstanding year thus far. For instance, ICSA was able to successfully hold the 2022 Applied Statistics Symposium in person after two years of virtual meetings. The symposium was held on June 19-22, 2022 at the University of Florida, Gainesville, FL. The organizing committee was led by Dr. Peihua Qiu, Dr. Samuel Wu, Dr. Somnath Datta, and Dr. Ji-Hyun Lee. The theme of the symposium was "Statistical Innovations in the Era of Artificial Intelligence and Data Science." It featured 3 plenary lectures, 2 special invited lectures, and banquet speeches delivered by David Siegmund, Jianqing Fan, Nilanjan Chatterjee, Susan Murphy, Xihong Lin and L.J. Wei. In addition, the 2022 ICSA China conference was very successful. It was held on July 1-4, 2022 at Xi' an University of Finance and Economics in Xi' an, China. Due to travel restriction, it was held both in-person and virtually. Dr. Jianguo (Tony) Sun at the University of Missouri and Dr. Ming Fang at Xi' an University of Finance and Economics chaired the organizing committee. Dr. Yingying Fan at the University of Southern California and Dr. Chunjie Wang at Changchun University of Technology chaired the scientific program committee. The theme of the conference was "Statistics in the Era of Big Data and Artificial Intelligence." It featured 3 keynote lectures delivered by Jane-Ling Wang, Richard Samworth, and Shurong Zheng. And finally, the Canada Chapter successfully held its 5th symposium at Banff, Canada on July 8-10, 2022. Dr. Dehan Kong and Dr. Linglong Kong led the organizing committee. I would like to thank all organizing committee chairs, members, and volunteers for their time and dedication in making these successes a reality.

In the remainder of the year 2022, ICSA will continue to have activities. The Midwest Chapter plans to have a joint symposium with NIC-ASA during October 5-6, 2022 at Northbrook, IL. Next, the Taiwan Chapter plans to have a joint conference with ISSAS; the theme will be in Memory of Academician Yuan-Shih Chow, and the conference will be held on December 15-16, 2022 at Taipei, Taiwan. Due to continuous COVID-19 pandemic, the 12th ICSA International Conference is postponed to next year. It was planned to be held in person from Sunday, December 18 to Tuesday, December 20, 2022 at the Chinese University of Hong Kong, Hong Kong. The organizing committee, led by Dr. Xin-Yuan Song, Dr. Jianguo (Tony) Sun, and Dr. Xingqiu Zhao, has been devoting a tremendous amount of time and effort to organizing this triennial conference.

Led by Dr. Pei Wang, ICSA organized 3 invited, 3 topic-contributed sessions, and 12 co-sponsored scientific sessions for 2022 JSM. In collaboration with Korean International Statistical Society (KISS), International Indian Statistical Association (IISA), and ASA, ICSA also organized a career development workshop with the theme of ASA Asian Forward. In addition, we had in person general member meeting and banquet during the 2022 JSM.

Recently, the ICSA board of directors and executives approved Dr. Jun Zhao (Director of Biostatistics, Astellas Pharma Global Development Inc.) to be the ICSA executive director for the 2023-2025 term, and Dr. Xin He (a faculty member in the Department of Epidemiology and Biostatistics, School of Public Health at University of Maryland) to be the ICSA treasurer for the 2025-2027 term. Given their impressive history of experience and service to ICSA, I have no doubt that Jun and Xin will continue to provide excellent service to ICSA.

For the journal Statistica Sinica, ICSA and ISSAS (Institute of Statistical Science Academia Sinica) have jointly formed a committee to search for new 3-year co-editors for the 2023-2025 term. For the journal Statistics in Biosciences, co-editors Dr. Joan Hu and Dr. Hongkai Ji have been actively identifying several important topics to be covered as special issues.

It is again our election time, so please cast your votes on time to exercise your rights and responsibilities as an ICSA member. The success of our society depends on the active participation of members like you.

Many of our ICSA members have received various honors and awards in 2022; congratulations to all those recipients! While I hope that we can share such news with our members through Bulletin and monthly newsletters, it has been challenging to track and give recognition where due without a notice. I hope you can inform editors on time if you have any news that you would like to share in the fu-

ture. Finally, I would like to thank the board of directors, committee members, volunteers and our friends for their time and help throughout the past several months. In particular, I would like to thank the ICSA Executive Director, Dr. Mengling Liu, for effectively managing ICSA operations during this COVID-19 pandemic.

Please let me know if you have any suggestions or innovative ideas for ICSA.

Zhezhen Jin, Ph.D.
2022 President,
Professor,
Department of Biostatistics,
Mailman School of Public Health,
Columbia University.

From the Executive Director (2020-2022)

Mengling Liu

This year, I was fortunate to make it to three major conferences, ENAR, ICSA Applied Statistics Symposium, and JSM. At many moments when I was standing among the hustle and bustle of these meetings, I felt deep appreciations for in-person meetings and hope you share the excitement as well.

After more than two years of virtual gatherings, the 2022 ICSA Applied Statistics Symposium returned in-person with over 400 attendees to University of Florida, Gainesville, FL in June 2022. The symposium was filled with idea-stimulating talks, informative short-courses, delicious banquet and fun-filled mixer, and most importantly, our long-time-no-see colleagues and friends. The 2022 ICSA China Conference overcame many hurdles and was successfully held in a hybrid mode on the campus of Xi'an University of Finance and Economics (XUFE), Xi'an, China. Then at the 2022 JSM, ICSA returned to our long-time tradition of ICSA General Member meeting and banquet during JSM, and over 100 attendees enjoyed famous Peking duck at Peking Gourmet Inn.

We also successfully held two ICSA Board meetings this year. During the first board meeting in June 2022, ICSA Board of Directors and members of the Executive committee received reports and updates from many committees, whose continuous and diligent efforts carried ICSA through many challenges during the pandemic. The Board also approved the candidates for the election of 2023 ICSA President-Elect and Board of Directors, nominations for ICSA awards, and the location of the 2024 ICSA Applied Statistics Symposium at Vanderbilt University in Nashville, Tennessee. Immediately following the Board' s approval, ICSA election was launched, and all eligible ICSA members had a chance to cast their votes to shape the ICSA future. In August 2022 during the JSM, the second board meeting was held, and the board approved the final election results with details included in this Bulletin. During the meeting, the Board discussed extensively several important items, including how to promote ICSA in the social-media era, plans to update ICSA by-laws, supports to ICSA sponsored and co-sponsored journals, and ICSA fund-raising and donations.

I would also like to call your attention to the lineup of future ICSA meetings in the Program Committee Report in this Bulletin. Your participation of ICSA activities will be the most important for continuing the success of ICSA. We welcome your ideas and suggestions. Lastly, I sincerely express my deep appreciation and gratitude to all of your supports and contributions that have made my job as the Executive Director possible. I' m very thankful to all the wonderful opportunities that this position allowed me to interact with you.

Mengling Liu, Ph.D.
ICSA Executive Director (2020-2022)
Professor of Biostatistics,
Department of Population Health,
Department of Environmental Medicine,
NYU Langone Health.

The 2023 ICSA Official Election Results

ICSA 2023 President Elect:

Dr. Xun Chen
Global Head of Biostatistics and programming Department,
Sanofi

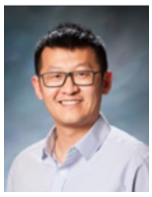
ICSA Board of Directors (2023-2025):

Dr. Huazhen Lin

Univ. of Finance and Economics, Chengdu, Sichuan, China

Dr. Ming Tan

Georgetown Univ. Medical Center



Dr. Li Wang

AbbVie

Dr. Yanping Wang

Eli Lilly and Co.

Dr. Min Zhang

University of Michigan, Ann Arbor

The 2022 New Fellows of ASA and IMS

The 2022 ASA Fellows

Congratulations to the following ICSA members who are bestowed the prestigious distinction of the 2022 American Statistical Association (ASA) Fellow for their professional contributions, leadership, and commitment to the field of statistical science:

- Kun Chen, University of Connecticut
- Yuehua Cui, Michigan State University
- Yang Feng, New York University
- Haiyan Huang, University of California, Berkeley
- Wei-Ting Hwang, University of Pennsylvania Perelman School of Medicine
- Fan Li, Duke University

- Lu Tian, Stanford University
- Peng Wei, The University of Texas MD Anderson Cancer Center
- Michael C. Wu, Fred Hutchinson Cancer Research Center
- Xian-Jin Xie, University of Iowa
- Eric Poe Xing, Carnegie Mellon University
- Wenxuan Zhong, University of Georgia
- Jianhui Zhou, University of Virginia

The 2022 Class of IMS Fellows

Congratulations to the following ICSA members on being the 2022 class of IMS fellows. The designation of IMS Fellow has been a significant honor for over 85 years. Each Fellow has demonstrated distinction in research in statistics or probability or has demonstrated leadership that has profoundly influenced the field:

- Haoda Fu, Eli Lilly and Company
- Haiyan Huang, University of California, Berkeley
- Ying Hung, Rutgers University
- Zhezhen Jin, Columbia University
- Jialiang Li, National University of Singapore
- Pengfei Li, University of Waterloo
- Shujie Ma, University of California, Riverside
- Limin Peng, Emory University
- Cheng Yong Tang, Temple University
- Ruodu Wang, University of Waterloo
- Jun Yan, University of Connecticut
- Tian Zheng, Columbia University
- Yong Zhou, East China Normal University

The 2022 ICSA Awards

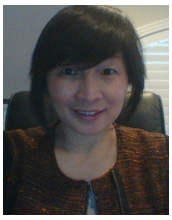
Awards Committee Chair: Chunming Zhang

2022 ICSA Distinguished Achievement Award

Xiaotong Shen, University of Minnesota
Citation:
For fundamental and influential contributions in statistical theory and methodology, especially in machine learning, and non- and semi-parametric statistics; for excellent service to and leadership in statistical communities, including the International Chinese Statistical Association; for exemplifying mentorship of students and junior faculty.

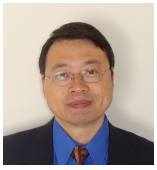
2022 Outstanding Service Award

Jianguo (Tony) Sun, University of Missouri
Citation:
In recognition with a sincere appreciation for his outstanding and dedicated service to ICSA, his leadership across ICSA board and committees as the 2020 ICSA President, as a member of the ICSA Board of Directors 2010-2012, as the Chair of the Membership Committee 2016-2017, as the Chair of the Program Committee 2018, as the Chair of the Program Committee of 2017 ICSA China Conference, as the Co-Chair of 2022 ICSA China Conference, and as the Chair of the organizing Committee of 2022 ICSA International Conference.



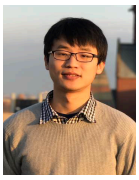
Rongwei (Rochelle) Fu, Oregon Health & Science University
Citation:
In recognition with a sincere appreciation for her outstanding and dedicated service to ICSA, her service across ICSA executives and committees as the Co-Chair of 2014 ICSA/KISS joint statistical conference, as the ICSA treasurer 2019-2021, as the member of ICSA Investment Ad Hoc Committee (2015-2018), and as the member of ICSA Financial Advisory Committee (2019-2022).

2022 President Citation Award

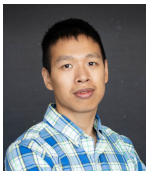


Hulin Wu, University of Texas Health Science Center at Houston
Citation:
In great appreciation of his generosity, dedication and devoted effort for ICSA as the Chair of the organizing committee of the 2020 ICSA Applied Statistics Symposium, and as the Chair of the ICSA Program Committee (2021-2022).

2022 ICSA Outstanding Young Researcher Award



Xinran Li, University of Illinois at Urbana-Champaign
Citation:
For his significant contributions in the areas of causal inference, observational studies, general statistical theory and methods, and for his novel research on rerandomization in randomized experiments.



Linbo Wang, University of Toronto

Citation:
For outstanding research in causal inference, notably in instrumental variable estimation, variable selection, and identification of causal effects; for innovative developments in categorical data analysis; and for important contributions to biomedical data analysis and personalized medicine.

Report from the ICSA 2022 Applied Statistics Symposium

Samuel S. Wu and Somnath Datta

2022 Applied Statistics Symposium

The 2022 Applied Statistics Symposium, the first in-person gathering after 2 years of virtual meetings, was successfully held on the University of Florida campus in Gainesville, Florida, USA, from June 19 through 22. Over four hundred participants from 8 countries attended the conference on the theme of "Statistical Innovations in the Era of Artificial Intelligence and Data Science." The conference program consists of 3 keynote presentations by Dr. Jianqing Fan of Princeton University, Dr. Susan Murphy of Harvard University, and Dr. David O. Siegmund of Stanford University; 2 special invited talks by Dr. Xihong Lin of Harvard University and Dr. Nilanjan Chatterjee of Johns Hopkins University; and banquet speech by Dr. Lee-Jen Wei of Harvard University. There were 70 invited sessions, 38 posters, 6 short courses and 5 oral presentations by student paper award recipients (See a separate article of this issue for more details).

The organizing committee would like to express our appreciation to all individuals and entities who contribute to the success of this event. Notably, we would like to acknowledge the strong support from ICSA Executive Committee, ICSA Program Committee, faculty and staff members as well as student volunteers from the Department of Biostatistics at the University of Florida, all conference committee members and symposium participants. In addition, we gratefully appreciate financial support from the National Science Foundation, 7 gold sponsors (Abbvie, BeiGene, Boehringer Ingelheim, Bristol Myers Squibb, ClinChoice, The Lotus Group, and Regeneron), 7 silver sponsors (Amgen, Everest, Gilead, Merck, Sanofi, Takeda, and Vertex) and 1 bronze sponsor (Servier).

2022 Student Paper Awards and Best Poster Awards Recipients

A student paper award competition and a best poster award competition were held for the 2022 ICSA Applied Statistics Symposium. The student paper competition award committee is chaired by Zhigang Li (University of Florida) and members include:

- Hsin-wen Chang (Institue of Statistical Science, Academia Sinica)
- Guanhua Chen (University of Wisconsin-Madison)
- Li Chen (Indiana University)
- Subharup Guha (University of Florida)
- Pengsheng Ji (University of Georgia)
- Fan Li (Yale University)
- Muxuan Liang (University of Florida)
- Xiangyang Lou (University of Florida)
- Qing Lu (University of Florida)
- Jing Ma (Fred Hutchinson Cancer Research Center)
- Guogen Shan (University of Florida)
- Zhihua Su (University of Florida)
- Zhengzheng Tang (University of Wisconsin-Madison)
- Xuefeng Wang (Moffitt Cancer Center)
- Feifei Xiao (University of Florida)
- Kai Yang (Medical College of Wisconsin)
- Yang Yang (University of Florida)
- Lu You (University of South Florida)
- Lihui Zhao (Northwestern University)

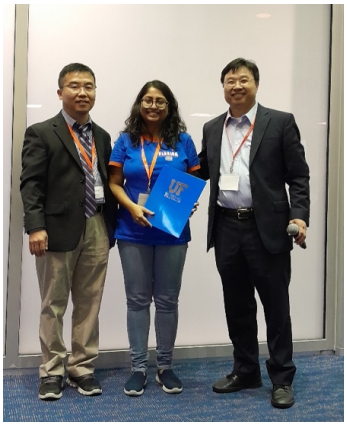
The best poster competition award committee is chaired by Guogen Shan (University of Florida) and members include:

- Rhonda Bacher (University of Florida)
- Li Chen (Indiana University)
- Jonathan Fischer (University of Florida)
- Steven Foti (University of Florida)
- Matt Hitchings (University of Florida)
- Zhigang Li (University of Florida)
- Muxuan Liang (University of Florida)

2022 ICSA Applied Statistics Symposium Student Volunteers

Dr. Ji-Hyun Lee (local committee chair) with the Graceful Orchid Dance Studio members at welcome mixer

2022 ICSA Applied Statistics Symposium student paper awardees Xiulin Xie, Shuting Shen, Ying Jin, Margaret Banker



Best poster award: Srijata Samanta

Keynote speakers (from left to right): Jianqing Fan, Susan Murphy and David O. Siegmund

- Ruitao Lin (MD Anderson Cancer Center)
- Qing Lu (University of Florida)
- Arlene Naranjo (University of Florida)
- Robert Parker (University of Florida)
- Qinglin Pei (University of Florida)
- Arkaprava Roy (University of Florida)
- Wei Xue (University of Florida)

ICSA Applied Statistical Symposium Student Paper Awards

Margaret Banker, University of Michigan

Title: Supervised Learning of Physical Activity Features from Functional Accelerometer Data.

Abstract: Accelerometry data enables scientists to extract personal digital features that can benefit precision health decision making. Existing methods in accelerometry data analysis typically begin with discretizing summary single-axis counts by certain fixed cutoffs into several activity categories, such as Vigorous, Moderate, Light, and Sedentary. One well-known limitation is that the chosen cutoffs have often been validated with restricted settings, and thus they cannot be generalizable across populations, devices, or studies. In this paper, we develop a data-driven approach to overcome this bottleneck in the analysis of activity data, in which we holistically summarize a subject' s activity profile using Occupation-Time curves (OTCs). Being a functional predictor, OTC describes the percentage of time spent at or above a continuum of activity count levels. We develop multi-step adaptive learning algorithms to perform a supervised learning via a scale-functional regression model that contains OTC as the functional predictor of interest as well as other covariates. Our learning algorithm first incorporates a hybrid approach of fused lasso for grouping and Hidden Markov Model for change-point detection, and then executes a few refinement learning steps to yield activity windows of interest. We demonstrate good performances of this learning algorithm using simulations as well as real world data analysis to assess the influence of physical activity on biological aging.

Ying Jin, Stanford University

Title: Sensitivity Analysis under the f-Sensitivity Models: A Distributionally Robust Optimization Viewpoint.

Abstract: The growing availability of observational data in various application regimes provides exciting opportunities for causal inference, especially in situations where randomization can be costly or unethical. However, observational data face the key challenge of unmeasured confounding: there might be unobserved variables that influences both the treatment mechanism and the outcomes, invalidating the causal conclusions inferred from the data. While such violation is not verifiable, sensitivity analysis could help gauge the robustness and credibility of the causal conclusions. In this paper, we propose a new class of sensitivity modes where the selection bias caused by unmeasured confounding factors is bounded on average. It generalizes commonly adopted uniform bound on selection bias in the literature, and might be more suitable for situations where the selection bias grows unbounded locally but, due to the impact of such region being small overall, is controlled at the population level. The new class of models motivates an illuminating perspective from distributionally robust optimization for sensitivity analysis: we show that the counterfactual distribution, the essential for inferring treatment effects, is within certain distance from the observed distribution. Based on this perspective, we represent the bounds on counterfactual means via distributionally robust optimization programs. We then design procedures to estimate these bounds, and show that our procedure is doubly robust to the estimation error of nuisance components and remains valid even when the optimization step is off. In addition, we establish Wald-type inference guarantee that is again robust to the optimization step. We demonstrate our method and verify its validity with numerical experiments.

Shuting Shen, Harvard University

Title: Fast Distributed Principal Component Analysis for Large-Scale Federated Data.

Abstract: Principal component analysis (PCA) is one of the most popular methods for dimension reduction. In light of the rapidly increasing large-scale data in federated ecosystems, the traditional PCA method is often not applicable due to privacy protection considerations and large computational burden. Fast PCA algorithms have been proposed to lower the computational cost but cannot handle federated data. Distributed PCA algorithms have been developed to handle federated data but are not computationally efficient when data at each site are very large. In this paper, we propose the FAst DIstributed (FADI) PCA method which applies fast PCA to site specific data using multiple random sketches and aggregates the results across sites. We perform a non-asymptotic theoretical study to show that FADI enjoys the same error rate as the traditional full sample PCA and a much smaller order of computational burden compared to existing methods. We perform extensive simulation studies and show that FADI substantially outperforms the other methods in computational efficiency without sacrificing statistical accuracy. We apply FADI to the analysis of the 1000 Genomes data to study the population structure.

Xiulin Xie, University of Florida

Title: High-Dimensional Dynamic Process Monitoring By PCA-Based Sequential Learning.

Abstract: Sequential process monitoring has broad applications. In practice, process characteristics to monitor often have a high dimensionality, partly due to the fast progress in data acquisition techniques. Thus, statistical process control (SPC) research for monitoring high dimensional processes is in rapid development in recent years. Most existing SPC charts for monitoring high-dimensional processes are designed for conventional cases in which the in-control (IC) process observations at different time points are assumed to be independent and identically distributed. In practice, however, serial correlation almost always exists in the observed sequential data, and the longitudinal pattern of the process to monitor could be dynamic in the sense that its IC distribution would vary over time (e.g., seasonality). In this paper, we develop a novel SPC chart for monitoring high-dimensional dynamic processes. The new method is based on nonparametric longitudinal modeling for describing the longitudinal pattern of the process under monitoring, principal component analysis for dimension reduction, and a sequential learning algorithm for developing an effective decision rule. It can well accommodate time-varying IC process distribution, serial data correlation, and nonparametric data distribution. The proposed method has been shown effective for air pollution surveillance.

Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award

Kan Chen, University of Pennsylvania

Title: Covariate-Balancing-Aware Interpretable Deep Learning models for Treatment Effect Estimation.

Abstract: Estimating treatment effects is of great importance for many biomedical applications with observational data. Particularly, interpretability of the treatment effects is preferable for many biomedical researchers. In this paper, we first give a theoretical analysis and propose an upper bound for the bias of average treatment effect estimation under the strong ignorability assumption. The proposed upper bound consists of two parts: training error for factual outcomes, and the distance between treated and control distributions. We use the Weighted Energy Distance (WED) to measure the distance between two distributions. Motivated by the theoretical analysis, we implement this upper bound as an objective function is minimized by leveraging a novel additive neural network architecture, which combines the expressivity of deep neural network, the interpretability of generalized additive model, the sufficiency of the balancing score for estimation adjustment, and covariate balancing properties of treated and control distributions, for estimating average treatment effects from observational data. Furthermore, we impose a so-called weighted regularization procedure based on nonparametric theory, to obtain some desirable asymptotic properties. The proposed method is illustrated by re-examinning the benchmark datasets for causal inference, and it outperforms the state-of-art.

The Best Poster Awards

- 1st place: Srijata Samanta, University of Florida: Title: "A Generalized Likelihood Based Bayesian Approach for Scalable Joint Regression and Covariance Selection in High Dimensions"
- 2nd place: Mengxin Yu,Princeton University: Title: "Are Latent Factor Regression and Sparse Regression Adequate?"
- 3rd place: Doudou Zhou, University of California Davis: Title: "RISE: Rank in Similarity Graph Edge-Count Two-Sample Test"

Samuel S. Wu, PhD,
Co-chair of Organizing Committee,
Professor of Biostatistics, University of Florida.
Somnath Datta , PhD,
Co-chair of Organizing Committee,
Professor of Biostatistics, University of Florida.

Report from the 2022 ICSA Program Committee

ICSA Program Committee

- Hulin Wu, Chair, ICSA 2022 Program Committee (Hulin.Wu@uth.tmc.edu)
- Aiyi Liu (2020-2022, JSM Representative 2021, liua@mail.nih.gov)
- Pei Wang (2021-2023, JSM Representative 2022, pei.wang@mssm.edu)
- Jianguo (Tony) Sun (2022-2024, JSM Representative 2023, sunj@missouri.edu)
- Guoqing Diao (2020-2022, ICSA Symposium 2021, gdiao@email.gwu.edu)
- Samuel Wu (2021-2023, ICSA Symposium 2022, samwu@biostat.ufl.edu)
- Jian Kang (2022-2024, ICSA Symposium 2023, jiankang@umich.edu)
- Gongjun Xu (2022-2024, ICSA Symposium 2023, gongjun@umich.edu)
- Hongzhe Lee (2021-2022, ICSA International Conference 2019, hongzhe@pennmedicine.upenn.edu)
- Xin-Yuan Song (2021-2022, ICSA International Conference 2022, xysong@sta.cuhk.edu.hk)
- Qingning Zhou (2020-2022, qzhou8@uncc.edu)
- Liang Zhu (2020-2022, Liang_Zhu@eisai.com)
- Jie Chen (2020-2022, jiechen0713@gmail.com)
- Lihui Zhao (2022-2024, Lihui.zhao@northwestern.edu)

Conferences and Meetings in 2022

- The 2022 ICSA Applied Statistics Symposium, the first in-person gathering after 2 years of virtual meetings, was successfully held on the University of Florida campus in Gainesville, Florida, USA, from June 19 through 22. Over four hundred participants from 8 countries attended the conference. The conference program consists of three keynote lectures by Drs. Jianqing Fan (Princeton), Susan Murphy (Harvard), and David Siegmund (Stanford), three special invited talks, and 70 invited sessions on the theme of "Statistical Innovations in the Era of Artificial Intelligence and Data Science." In addition, there were 38 posters, 6 short courses and 5 oral presentations by student paper award recipients. The executive committee is chaired by Dr. Peihua Qiu and the Scientific Program Committee is co-chaired by Dr. Samuel Wu and Dr. Somnath Datta from the Department of Biostatistics, University of Florida.

- The 2022 International Chinese Statistical Association (ICSA) China Conference was successfully held in hybrid format on the campus of Xi' an University of Finance and Economics (XUFE), Xi' an, China, from Friday July 1 to Monday July 4, 2022. The conference was originally scheduled to be held in July, 2021, but was postponed due to COVID-19. The conference attracted more than 300 participants and offered three plenary keynote lectures and 74 invited scientific sessions. In addition, the conference awarded six Junior Researcher Awards. The Executive Committee was chaired by Dr. (Tony) Jianguo Sun, Curator' s Distinguished Professor of the University of Missouri, and Dr. Ming Fang, President and Professor of XUFE, and the Program Committee was chaired by Dr. Yingying Fan, Professor of University of Southern California, and Dr. Chunjie Wang, Dean and Professor of Changchun University of Technology, China. The three keynote lectures were provided by three distinguished statisticians: Professor Jane-Ling Wang from the University of California, Davis, Professor Richard Samworth from the University of Cambridge of UK, and Professor Shurong Zheng from Northeast Normal University of China.

- 2022 12th ICSA International Conference at the Chinese University of Hong Kong (Program Committee Chair: Dr. Tony Sun and Local Committee Chair: Dr. Xin-Yuan Song): Postponed to 2023.

Future Conferences and Meetings

The program committee has reviewed and approved one proposal to host the 2024 ICSA Applied Statistics Symposium from Vanderbilt University (Co-Chairs: Drs. Dandan Liu and Qingxia Chen). Two universities in USA expressed interest in hosting 2025 ICSA Applied Statistics Symposium and three universities in China expressed interest in hosting the ICSA China Conference in future. Below is a list of planned upcoming events sponsored or co-sponsored by ICSA. We will keep you updated as the COVID-19 situation evolves.

- ICSA co-sponsored the 8th workshop on Biostatistics and Bioinformatics, Atlanta, GA: Postponed to Spring 2023.
- The ICSA 2023 China Conference will be held at 金牛宾馆 (The information about the hotel can be found at http://www.jnhotel.com/), Chengdu, Sichuan, China from June 30 - July 3. It will be co-sponsored by Southwest Jiaotong University (SWJTU). Please contact Dr. Yichuan Zhao (yichuan@gsu.edu) for more information.
- 2023 ICSA Applied Statistics Symposium: June 11-14, 2023, hosted by the University of Michigan in Ann Arbor, Michigan (Co-Chairs: Drs. Jian Kang and Gongjun Xu)
- ICSA co-sponsored the IMS Asia Pacific Rim Meeting: January 4-7, 2024, Melbourne, Australia
- 2024 ICSA Applied Statistics Symposium: Nashville, TN (date to be determined), to be hosted by Vanderbilt University (Co-Chairs: Drs. Dandan Liu and Qingxia Chen).

The program committee also developed a guideline on how to prepare and evaluate a proposal to host the ICSA Symposium and the work procedure for the ICSA program committee is also developed. If you have any comments or suggestions on ICSA programs, please contact Professor Hulin Wu at the University of Texas Health Science Center at Houston (`Hulin.Wu@uth.tmc.edu`).

Hulin Wu, PhD,
ICSA Program Committee Chair,
Professor,
University of Texas Health Science Center at Houston.

News from the ICSA-Canada Chapter

Yingwei Peng and Leilei Zeng

The year of 2022 is a special year for the ICSA Canada Chapter. It is the 10th anniversary of the Chapter. We also successfully held its Fifth Chapter Symposium in Banff, Alberta, Canada after a one-year delay due to COVID pandemic. Following are some details about the Chapter and the major events that happened in the Chapter in the past 12 months.

1. ICSA-Canada Chapter is governed by its Executive Committee. The 2021-2022 Executive Committee includes

- Chair: Yingwei (Paul) Peng from Queen' s University
- Past Chair: Liqun Wang from the University of Manitoba
- Chair-Elect: Joan Hu from Simon Fraser University
- Secretary/Treasurer: Leilei Zeng from the University of Waterloo

The Chapter also has three regional representatives appointed by the Chapter Chair. They are

- Cindy Feng, Dalhousie University (East Canada),
- Sunny Wang, Wilfrid Laurier University (Central Canada),
- Juxin Liu, University of Saskatchewan (West Canada)

2. The Fifth ICSA-Canada Symposium was successfully held in person at Banff Centre, Alberta, Canada on July 8-10. The Organizing Committee of the Symposium includes

- Dehan Kong (Program Chair, University of Toronto)
- Linglong Kong (Local Chair, University of Alberta)
- Joan Hu (Chapter Chair-Elect, Simon Fraser University)
- Yingwei Peng (Chapter Chair, Queen' s University)
- Liqun Wang (Chapter Past Chair, University of Manitoba)
- Leilei Zeng (Chapter Secretary/Treasurer, University of Waterloo)

The keynote speakers of the Symposium are

- Jianqing Fan, Princeton University
- Xiaotong Shen, University of Minnesota
- Heping Zhang, Yale University

The Symposium was co-sponsored by CANSSI (Canadian Statistical Sciences Institute) and PIMS (The Pacific Institute for Mathematical Sciences). It attracted nearly 200 participants, and among them, nearly 50 are new members of ICSA. The event was well attended by graduate students, five best student' s posters were chosen by the award committee and a ceremony was held at the Banquet. The President of ICSA, Professor Zhezhen Jin, also attended the Symposium.

3. The Chapter 2022 Annual General Meeting was held at Banff Centre, Alberta, Canada on July 9th, 2022 during the Symposium

The members of the executive committee reported to the members on the Chapter' s activities and financial status in 2021, as well as the plans for the upcoming events. Potential amendments to improve the clarity of the Chapter' s constitution and By-laws were also discussed. The ICSA President Zhezhen Jin attended the AGM meeting.

4. The Chapter Chair-Elect, Joan Hu, organized a Chapter-sponsored invited session for the Statistical Society of Canada (SSC) Annual Meeting held virtually on May 30 to June 3, 2022.

This has been organized at the past few SSC annual meetings.

5. The Chapter Chair, Yingwei Peng, was invited to attend the ICSA board meeting held at the 2022 ICSA Applied Statistics Symposium in Florida, USA on June 19th, 2022, and reported the Chapter's activities.

6. The Chapter Chair, Yingwei Peng, had an online meeting with some of the ICSA Executive Committee members including ICSA President Zhezhen Jin and Executive Director Mengling Liu in April 2022.

The meeting discussed ways to strengthen the ties between the Chapter and the ICSA and attain support from ICSA to the Chapter.

7. The Chapter organized an election in late 2021 to elect Joan Hu as the Chapter' s Chair-Elect. Her Chair-Elect term is 2021-2022.

8. In lieu of the postponed Symposium, the Chapter organized a half-day online meeting in August 2021 for the Chapter members. The meeting included three keynote speeches by

- Jiguo Cao, Simon Fraser University
- Peng Ding, University of California, Berkeley,
- Hongtu Zhu, University of North Carolina at Chapel Hill

The meeting attracted about 40 participants. An annual general meeting of the Chapter members was held after the keynote speeches.

Yingwei Peng, PhD,
Chair, ICSA Canada Chapter,
Department of Public Health Sciences,
Queen' s University, Canada.

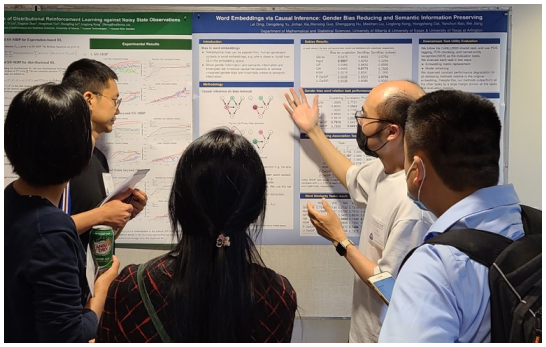
Leilei Zeng , PhD,
Secretary/Treasurer, ICSA Canada Chapter,
Department of Statistics and Actuarial Science,
University of Waterloo, Canada.

Heping Zhang, Yale University


Jianqing Fan, Princeton University


Xiaotong Shen, University of Minnesota


Poster Awards


Poster Session


The Fifth ICSA-Canada Chapter Symposium - Statistics: From Data to Knowledge
July 8-10, 2022 | Banff Center, Banff, Alberta, Canada

Report from the 2022 ICSA China Conference

The 2022 International Chinese Statistical Association (ICSA) China Conference was successfully held in hybrid format on the campus of Xi'an University of Finance and Economics (XUFE), Xi'an, China, from Friday July 1 to Monday July 4, 2022. The conference was originally scheduled to be held in July, 2021 but was postponed due to COVID-19 with the hope of being held in person in 2022. The conference attracted more than 300 participants and offered three plenary keynote lectures and 74 invited scientific sessions. In addition, the conference awarded six Junior Researcher Awards. The Executive Committee was chaired by Dr. (Tony) Jianguo Sun, Curator's Distinguished Professor of the University of Missouri, and Dr. Ming Fang, President and Professor of XUFE, and the Program Commit-

tee was chaired by Dr. Yingying Fan, Professor of University of Southern California, and Dr. Chunjie Wang, Dean and Professor of Changchun University of Technology, China.

The opening ceremony was held in the auditorium of the main building of XUFE on July 1, 2022. It was chaired by Dr. (Tony) Jianguo Sun and three opening remarks were delivered by Professor Ming Fang, Dr. Zhezhen Jin, ICSA President and Professor of Columbia University, and Professor Yingying Fan. It was followed by the Junior Research Award ceremony chaired by Dr. Hongyuan Cao, Chair of the Committee and Professor of Florida State University. The three keynote lectures were provided by three distinguished statisticians: Professor Jane-Ling Wang from the University of California, Davis, Professor Richard Samworth from the University of Cambridge of UK, and Professor Shurong Zheng from Northeast Normal University of China.

Professor Wang presented a lecture on "The Three Faces of Functional Data," Professor Samworh's presentation was entitled "Optimal Nonparametric Testing of Missing Completely at Random, and its Connections to Compatibility," and the title of Professor Zheng's lecture is "Spectral Properties of High-Dimensional Sample Correlation Matrix and its Applications." The 74 parallel scientific sessions covered a wide range of topics including biostatistics, bioinformatics, statistics, engineering, finance, economics, genetics and genomics, big data computing, clinical trials, health policy and data science. The six Junior Researcher awardees are Edgar Dobriban from University of Pennsylvania, Mingyue Du from Hong Kong Polytechnic University, Ping Hu from University of Western Ontario of Canada, Luhua Lei from Stanford University, Yan Li from Jilin University of China, and Shuting Shen from Harvard University. We would like to extend our genuine gratitude to all the committee members, volunteers, session organizers, speakers, and conference participants whose effort and time made the conference such a success. In particular, we want to thank Drs. Weiqun Zhang and Shuang Li, Deans and Professors of the School of Statistics of XUFE, many leaders of XUFE and all faculty and students of the School of Statistics of XUFE for their many meetings, countless hours and enormous efforts for preparing this conference in this special period.

(Tony) Jianguo Sun, PhD,
Co-Chair of 2022 ICSA China Executive Committee,
Curator' s Distinguished Professor of Statistics, University of Missouri.

Yingying Fan, PhD,
Co-Chair of 2022 ICSA China Executive Committee,
Professor of Data Sciences and Operations, University of Southern California.

Report from Statistica Sinica

Rong Chen, Su-Yun Huang, and Xiaotong Shen

Executive Summary: In the past 12 months, from July 1, 2021, to June 30, 2022, Statistica Sinica receives 416 submissions. The total number of submissions remains relatively stable in the recent years. Of the 416 submission papers, we have accepted 32 and rejected 246, with other 138 under various revision stage. Overall, the eventual acceptance rate in the past 12 months is lower than that in the past. It is very possible that we have raised our acceptance standard slightly higher, as we have a strict limit on the number of pages for publication each year.

Up to 2022.6.30, there are 191 accepted papers (Vol.32 No.3 Vol.34 No.3) in the backlog, including 9 papers for special issue (In honor of Prof. K.C. Li) and one short note paper. Vol.32 No.3 is going to be published in July this year, and Vol.34 No.3 will be published in July, 2024. So, currently we have enough backlog of accepted papers to be published formally, though we post

Table 1. Submissions and acceptance rate for the past 6 years.

	Aug 1, 2016 – Jul 31, 2017	Aug 1, 2017 – Jul 31, 2018	Aug 1, 2018 – Jul 31, 2019	Aug 1, 2019 – Jul 31, 2020	Aug 1, 2020 – Jul 31, 2021	Jul 1, 2021 – June 30, 2022
Submission	531	486	448	458	447	416
Accepted papers	150	146	76	85	139	32
Acceptance rate	28%	30%	16.9%	18.5%	31%	7.6%

Table 2. Paper status for the past 3 years.

	Aug 1, 2019– Jul 31, 2020		Aug 1, 2020– Jul 31, 2021		Jul 1, 2021– June 30, 2022	
Rejected w/o external review	172	37.5%	187	41.8%	154	37%
Rejected with external review	144	31.4%	98	21.9%	82	19.7%
Rejected with revision allowed	53	11.5%	7	1.6%	5	1.2%
Major/Minor revision	2	0.4%	6	1.3%	48	11.5%
First submission under review	0	0.0%	1	0.2%	61	14.6%
Revision under review	0	0.0%	5	1.1%	15	3.6%
Withdrawn	2	0.4%	0	0.0%	1	0.2%
Accepted	85	18.5%	139	31%	32	7.6%
Total	458		443		398	

all accepted papers online immediately after acceptance. In order to shorten the backlog without reducing the number of accepted papers, we plan to publish more additional online-only special issues that are not restricted to the publication page limit.

Thanks to the efforts of the former editors, associate editors and contributing authors, the journal' s two-year impact factor in 2021 reaches up to 1.330, which is higher than that in 2020 (1.26). But its five-year IF in 2021 is 1.481, which is lower than that in 2020 (1.64). In the future, we'll keep organizing more special issues on trendy topics; hopefully, our impact factor can grow steadily.

1. Submissions and Acceptance Statistics Table 1 shows a comparison of the number of submissions and the number of accepted papers for the past 6 years. Note that the accepted papers may have been submitted in the years before. Table 2 shows the current status of the papers for the past 3 years. Table 3 shows the number of submissions by country in the past three years.

2. Manuscript Processing Time Table 4 shows the turnaround statistics of initial decisions for the past three years, with the decision times censored on July 3, 2022. About 75% of the editorial decisions during 2021-2022 take less than 67 days, but 5% take over 127 days. For the past 3 years, our average reviewing time is shorter and shorter. From table 5, it can be seen that a large percentage of papers get the reviewing decisions within 30 days.

3. Backlog for Publication There remain 191 accepted manuscripts waiting to be published. Among them, 24 will appear in the general issue in July 2022 (32-3) and 23 will show in the general issue in Oct 2022 (32-4). Besides, 9 will be published as online-only issue around September ("Sliced Inverse Regression after 30 Years" in honor of Prof. Ker-Chau Li). For the past year, the number of days from acceptance to publication averages 389 days.

4. Rankings and Impact Factors Table 6 shows the ranks of Statistica Sinica based on the 2-Year Impact Factor and the 5-Year Impact Factor provided by the Journal Citation Reports (JCR) in the area of Statistics and Probability from 2012 to 2021. Table 7 shows the ranks of Statistica Sinica in Scimago Journal Rankings among all journals of Statistics and Probability in the Scopus database from 2012-2021. The ranking is performed using the algorithm Google PageRank.

5. Special Issue In the past year, we have published four regular issues and two online special issues (In Honor of Professor Tze Leung Lai & Causal Inference and Short Notes) containing 113 articles.

Table 3. Top ten countries with the highest submissions for the past 3 years.

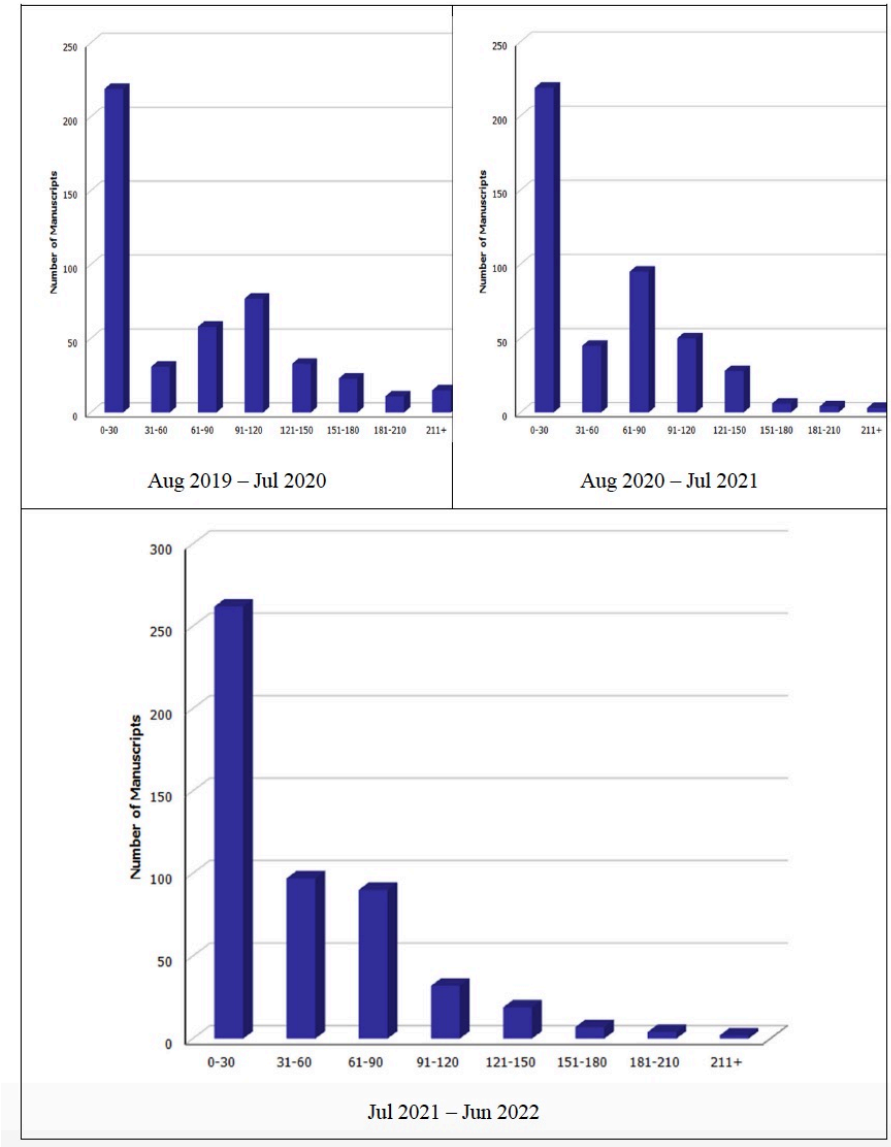
Rank	Aug 1, 2019 – Jul 31, 2020		Aug 1, 2020 – Jul 31, 2021		Jul 1, 2021– June 30, 2022	
1	USA	173 (30.1%)	USA	218 (35.2%)	USA	212 (35.3%)
2	China	170 (29.6%)	China	162 (26.2%)	China	180 (30%)
3	Canada	22 (3.8%)	Hong Kong	31 (5%)	Canada	29 (4.8%)
4	Hong Kong	20 (3.5%)	Canada	25 (4%)	Taiwan	26 (4.3%)
5	United Kingdom of Great Britain and Northern Ireland	18 (3.1%)	Taiwan	18 (2.9%)	Hong Kong	20 (3.3%)
6	Iran	15 (2.6%)	Italy	16 (2.6%)	Italy	15 (2.5%)
7	Singapore	14 (2.4%)	Japan	12 (1.9%)	Japan	12 (2%)
8	Taiwan	13 (2.3%)	Brazil	11 (1.8%)	United Kingdom of Great Britain and Northern Ireland/ Germany	9 (1.5%)
9	Japan	12 (2.1%)	Australia	10 (1.6%)	Singapore/ Kenya	8 (1.3%)
10	Germany/ Australia	10 (1.7%)	United Kingdom of Great Britain and Northern Ireland/ Singapore/ Pakistan/ India/ Chile	9 (1.5%)	India/ Australia	6 (1%)

Table 4. Percentiles of review time in days for the past 3 years.

Period	5th	25th	50th	75th	95th	Sample Size	**Average Review Days**
Aug 2019 – Jul 2020	5	13	37	105	185	467	67
Aug 2020 – Jul 2021	3	9	33	82	135	450	51
Jul 2021 – Jun 2022	3	9	29	67	127	513	43

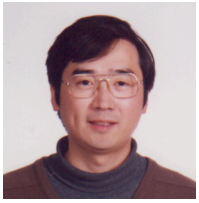
Currently, we are organizing three special issues: "Sliced Inverse Regression after 30 Years" in honor of Prof. Ker- Chau Li, "Data Privacy," and "sequential Monte Carlo". Up to now, 9 papers has been included in "Sliced Inverse Regression after 30 Years" in honor of Prof. Ker-Chau Li and all will be

Table 5. Comparison of review time in days for the past 3 years (with 30 days group).



published online around September, 2022. All these three special issues will be published as online-only issues.

Su-Yun Huang, PhD,
Researcher,
Academia Sinica, Taipei.



Rong Chen , PhD,
Distinguished Professor of Statistics,
Rutgers University.

Xiaotong T. Shen, PhD,
Distinguished Professor,
College of Liberal Arts,
University of Minnesota.

Table 6. JCR rankings for the recent 10 years.

Year	Number of Journals	Ranking (2-Year Impact Factor)		Ranking (5-Year Impact Factor)	
2021	125	78	(1.330)	70	(1.481)
2020	125	76	(1.261)	64	(1.647)
2019	124	72	(0.968)	67	(1.230)
2018	123	71	(0.947)	66	(1.256)
2017	123	71	(0.886)	51	(1.399)
2016	124	70	(0.899)	46	(1.632)
2015	123	66	(0.838)	42	(1.611)
2014	122	44	(1.158)	36	(1.591)
2013	119	37	(1.226)	44	(1.365)
2012	117	25	(1.440)	41	(1.418)

Table 7. SCImago journal rankings for the recent 10 years.

Year	Total Number of Journal	Journal Rank	Quartile
2021	250	45	Q1
2020	257	50	Q1
2019	246	41	Q1
2018	219	41	Q1
2017	196	23	Q1
2016	183	26	Q1
2015	179	20	Q1
2014	179	14	Q1
2013	179	12	Q1
2012	176	19	Q1

Report from Statistics in Biosciences

Joan Hu and Hongkai Ji

Statistics in Biosciences (SIBS), one of the two statistical journals established by ICSA, has currently three issues a year in print and electronic form. It publishes articles on development and application of statistical methods and their interface with other quantitative methods, such as computational and mathematical methods, in biological and life science, health science, and biopharmaceutical and biotechnological science. The journal has published both regular articles and topic-oriented papers in its special issues. See more information on the journal' s website, `https://www.springer.com/journal/12561`. The journal has published 12 articles in a regular issue and 10 articles in a special issue this year. The special issue is on "Leveraging External Data to Improve Trial Efficiency", guest-edited by Lanju Zhang at Vertex Pharmaceuticals and Naitee Ting at Boehringer Ingelheim.

SIBS currently has four special issues in preparation for publications (`https://www.springer.com/journal/12561/updates/23349828`):

- "Machine Learning Algorithms in Genomics and Genetics" (Guest-editor: Yingying Wei, Chinese University of Hong Kong)
- "Novel Statistical Approaches for Modeling Exposure Mixtures and Health Outcomes" (Guest-editors: Zhen Chen and Paul Albert with NIH)
- "Machine Learning in Biomedical Sciences" (Guest-editors: Dehan Kong, University of Toronto and Bingxin Zha, University of Pennsylvania
- "Statistical Methods, Algorithms and Applications in Biomedical Data Integration" (Guest-editors: Peter X.-K. Song, University of Michigan and Lu Tang, University of Pittsburgh).

We welcome proposals of new special issues for SIBS from ICSA members.

Initiating its annual Best Paper Award in 2021, Statistics in Biosciences (SIBS) selects from SIBS

papers, including review articles, published in the previous year for one or two Best Paper Awards each year. The award committee consists of the most recent Past Editor (chair) and the Present Co-Editors. The SIBS's Best Paper Awards are announced at the ICSA Applied Statistics Symposium of the year. Each of the awards includes a certificate from Springer/SIBS and a check up to $1,000 (USD) from ICSA/SIBS. The authors of the papers are invited to talk in the SIBS sponsored session at the ICSA Applied Statistics Symposium of the year. The 2022 SIBS Best Paper Awards are given to

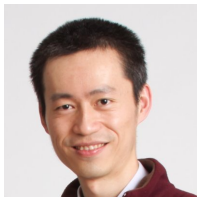
- Yifan Zhu and Ying Qing Chen, "On a Statistical Transmission Model in Analysis of the Early Phase of COVID-19 Outbrea", Statistics in Biosciences, Vol. 13 Issue 1, p56-72.
- Jonggyu Baek, Magaret Banker, Erica C. Jansen, Xichen She, Karen E. Peterson, E. Andrew Pitchford, and Peter X. K. Song, "An Efficient Segmentation Algorithm to Estimate Sleep Duration from Actigraphy Data", Statistics in Biosciences, Vol. 13 Issue 3, p563-583.

Congratulations to the winners on their excellent contributions to SIBS!

SIBS is working on enriching its publication portfolio by attracting more software and resource manuscripts, review articles, and commentaries. We have reached out to a number of potential contributors and will continue to solicit submissions in these categories from the community.

Professor Hongzhe Li at University of Pennsylvania Perelman School of Medicine finished his Co-Editor term at the end of 2021. We thank Hongzhe for his great dedication to the journal and his outstanding editorship.

Joan Hu, PhD,
Professor of Statistics,
Department of Statistics and Actuarial Science,
Simon Fraser University,
Burnaby, BC V5A1S6, Canada.



Hongkai Ji , PhD,
Professor,
Department of Biostatistics,
Johns Hopkins Bloomberg School of Public Health,
Baltimore, MD 21205, USA.

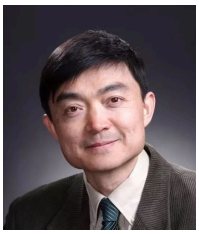
ICSA Financial Report

Profit and Loss: January 1, 2022 through June 30, 2022.

Beginning Cash Balance (Bank/Paypal accounts)	**1/1/2022**	**$**	**248,697.77**
Income:		**$**	**161,544.10**
Membership from Paypal Account		$	23,034.00
Membership from Institute of Mathematical Statistics		$	1,700.00
2022 ICSA Symposium		$	89,175.66
Registration from 2022 China Conference		$	31,892.60
Donation to JP Hsu Account		$	10,000.00
General donations		$	2,000.00
Springer Science & World Scientific Publishing Co.		$	3,099.11
Job Posting		$	600.00
Interest		$	42.73
Total Income		**$**	**161,544.10**
Expense:		**$**	**(14,978.50)**
Cevent		$	(8,785.00)
IT Cost		$	(4,060.03)
ICSA Office Cost		$	(1,421.48)
Paypal Fee		$	(711.99)
Total Expense		**$**	**(14,978.50)**
Net Total Income		**$**	**146,565.60**
Transfer			
From 2021 Applied Symposium Account		$	63,680.31
For JP Hsu Account		$	2,951.96
Ending Cash Balance (Bank/Paypal accounts)	**6/30/2022**		**$461,895.64**
ASSETS			
Main Checking/Savings/PayPal		$	461,895.64
Vanguard Investment Balance		$	797,929.25
TOTAL ASSETS		**$**	**1,259,824.89**
LIABILITIES & EQUITY			
Equity			
Main Accounts Opening Balance January 1, 2022		$	248,697.77
January 1 to June 30, 2022 Net Income(+)/Expense(-)		*$*	*146,565.60*
From 2021 Applied Symposium Account		*$*	*63,680.31*
For JP Hsu Account		*$*	*2,951.96*
2022 Symposium Bank Accounts Opening Balance January 1, 2022		$	-
January 1 to June 30, 2022 Net Income(+)/Expense(-)		*$*	*89,175.66*
Vanguard investment account opening balance on January 1, 2022		$	935,986.18
January 1 to June 30, 2022 Investment Profit(+)/Loss(-)		*$*	*(138,056.90)*
Total Equity		**$**	**1,349,000.58**
TOTAL LIABILITIES & EQUITY		**$**	**1,349,000.58**

Rui Feng, PhD,
Treasurer, ICSA
Associate Professor of Biostatistics,
University of Pennsylvania

Congratulations to Song Xi Chen



ICSA member Song Xi Chen of Peking University has been elected as a member of the Chinese Academy of Science (CAS). He and other 64 scholars were elected to the academy during the 2021 Assembly of Academicians of CAS held in early November 2021. Chen is elected to the Mathematics and Physics Division. The previous statistics members of CAS were late Professors Pau Lu Hsu (1948-1970) and Xiru Chen (1997-2005).

In addition, the 2022 Peter Hall Lecture will be delivered by Professor Song Xi Chen of Peking University, China, at the 2022 ICSA International Conference in Hong Kong with the specific dates to be determined. In honor of Professor Peter Hall's life and achievements and his contributions to the ICSA, the ICSA has established the Peter Hall Lecture. Professor Peter Gavin Hall was an Australian statistician who made profound and creative contributions in mathematical statistics and probability theory. He was a strong supporter and friend of the ICSA and Chinese statisticians, especially young researchers. He was a constant source of inspiration for the ICSA members.

Professor Song Xi Chen is a University Chair Professor affiliated with School of Mathematical Sciences, Guanghua School of Management and Center for Statistical Science in Peking University in China. Prior to that, he was on the faculty in the Department of Statistics at Iowa State University and was Associate Professor of Statistics in the National University of Singapore. His research interests are in the general area of non-parametric statistics, and more specifically in inference for high dimensional data; environmental modeling; empirical likelihood; inference for stochastic processes; econometric theory and financial econometrics; missing data problems; multiple system surveys for the US census. He has published over 110 manuscripts in the statistical journals and has mentored 21 doctoral students. His recent work on air quality in China was widely cited and reported in the New York Times.

Professor Chen is a Fellow of the American Association for the Advancement of Science (AAAS), a Fellow of the Institute of Mathematical Statistics and a Fellow of the American Statistical Association and is an elected member of the International Statistical Institute. He has served in various editorial roles for the top-tier journals in statistics and served as a board of director for the International Chinese Statistical Association. He received the PhD in Statistics from the Australian National University (1993), an MSc in Statistics and Operations Research from Victoria University in New Zealand (1990), and an MSc in Mathematical Statistics (1988) and a BSc in Mathematics from Beijing Normal University (1983).

Congratulations to Heping Zhang on Winning IMS Neyman Lecture

Heping Zhang is Susan Dwight Bliss Professor of Biostatistics, Professor of Child Study, and Professor of Statistics and Data Science at Yale University. He has published over 350 research articles and monographs in theory, methodology, and applications of statistics. He is particularly interested in biomedical research including epidemiology, genetics, child and women health, mental health, and substance use. He directs the Collaborative Center for Statistics in Science that coordinates major national research networks to understand the etiology of pregnancy outcomes and to evaluate treatment effectiveness for infertility.

Zhang is a fellow of the American Statistical Association, a fellow of the Institute of Mathematical Statistics, and ICSA President of 2019. He was named the 2008 Myrto Lefkopoulou distinguished lecturer by Harvard School of Public Health and a 2011 IMS Medallion Lecturer. Dr. Zhang was the founding Editor-in-Chief of Statistics and Its Inter-

face and is the past coordinating Editor of the Journal of the American Statistical Association.

Heping Zhang's Neyman Lecture was given at the IMS Annual Meeting in London, June 27–30, 2022. in Yu by journalist Nathalie Randin.

Congratulations to Xuming He on Winning Carver Award

Dr. Xuming He was selected for the 2022 IMS Carver Medal, an award for service to the IMS for "his decades-long contributions to the IMS in multiple capacities including Editor of the IMS Bulletin, IMS Council member, Committee Chairs, and conference program co-chairs; and for his strong and conscientious leadership in a wide range of other professional services."

XL-Files: When a Statistician becomes a (COVID) Statistic

Xiao-Li Meng

Editorial: This is a reprint from a column article published in the *IMS* ; `https://imstat.org/2022/07/18/xl-files-when-a-statistician-becomes-a-covid-statistic/`) with IMS permission.

AWhat happens when a statistician becomes a COVID statistic? Well, first of all, a COVID fever reignited my XL fervor. No, the missing "XL-Files" have not been due to lack of shareable stories in my life (if I have one). Launching and editing Harvard Data Science Review (HDSR) alone has given me abundant excitements to regale and frustrations to vent, including not having time to vent. The general data science community apparently lifts up more those who have developed the skills to benefit from NeurIPS or ICML kinds of pressing deadlines, instead of Annals-esque requests for deeper probing. However, illness tends to reset priorities, rearrange calendars, and remind us of our roots. Usually temporary, unfortunately, or maybe fortunately, depending how you look at it.

Regardless of how long the fever lasted, here I am, reopening the "XL-Files," while appreciating the luxury of being able to repeat this mantra for the $(n+1)$st time: Every cloud has a silver lining.

The unexpected motivation, from a participant of a conference – the one where COVID finally got hold of me after chasing me unsuccessfully for over two years – was also responsible for this reopening. Her message that she had used some "XL-Files" for teaching was as encouraging to me as – I imagine – a novice wine maker who finds its product is being sampled by a WSET class.

Secondly, my COVID episode personalized several research areas that have been competing for much of my (non-feverish and non-HDSR) time: individualized risk and prediction, imprecise probability, data quality, and data privacy. I submit these stories of personalization for your judgment as to whether they are results of an overfitting neural network attempting to bootstrap itself out of a natural annealing process. But regardless how your non-artificial neural network differs from mine, I hope we share a time-honored lesson: preaching is far easier than practicing.

The onset of the COVID was signaled by a rather sudden sense of chill, much like entering a wine cellar without being prepared for the immediate temperature drop. There could be a variety of reasons for feeling sick after a week-long travel but know-

ing someone at conference had just tested positive for COVID obviously should increase my chance of being infected. However, what does "my chance" actually mean here, and in what ways it is affected by my other data? As my body was getting busy with a rising temperature, my brain had its own fervent self-dialogue. "I just got tested negative, and I am fully vaccinated and boosted." "But I have symptoms." "But I had symptoms before, and most COVID cases have no symptoms." "This feels more like a bad flu." "But most infected people reported that it is like a bad flu." "I wore masks." "But I went out for lunch with others, and at the banquet few had masks on." "But ···""

Wait. Where was the Bayes theorem? What events were being conditioned upon? Did my brain just commit a prosecutor's fallacy? Wait, wait. How could anyone apply the Bayes theorem here? What numbers can be plugged in? Where could those number—any number—come from? Wait, wait, wait. Where was the Dempster's rule of combination when it's most needed? And what were pieces to be combined? Where were p(I have COVID), q(I don't have COVID), and r(I have no idea)? Did my brain just convince itself that r increases with the duration of the dialog? Is that a form of dilation or more a hallucination?

Fortunately, there is a rapid test that can rapidly stop the hallucination, or at least give me an instrument to greatly reduce r. Having reduced r, I could concentrate on reducing the fever. Another instrument came to help: the fever was 99.9^{o}F (37.7^{o}C) the first night I returned from the conference and got a digital thermometer from a local pharmacy store. The measuring process took longer than I expected, but everything I preached about using n>1 was completely suppressed by my annealed brain. I surmised that it was afraid of engaging in another r-increasing exercise.

Waking up soaked in Tylenol-enhanced sweat the next morning, I took another measure. The thermometer quickly "peeped," reporting 100^{o}F (37.8^{o}C). Wait. That couldn't be right. I felt less feverish, and the peep came way too fast, compared to that of the night before. I had to measure it again. It took a bit longer, but it gave a number that stopped me from employing n=3: 99.4 (37.4^{o}C). It just felt right. And when my concerned family members call, I could honestly tell them not to worry as my fever had gone down.

Honestly? Well, I don't think I need to insult any IMS Bulletin reader's intelligence by explaining the logical equivalence between, "What's wrong with choosing numbers that can make me feel better and comfort others?" and, "What's wrong with choosing data to support my values and ideology and to unite all people who support the same?" But actively reflecting upon how we behave differently, consciously, or subconsciously as private individuals as opposed to professional members can remind us, minimally, of paying more attention to data minding before data mining or analysis. For example, modeling measurement errors for self-reported measures such as blood pressures, weight, food intake, amount of exercise, etc., should never be done by only considering adding a convenient Gaussian error or any symmetric error.

Yes, our individual behaviors are sufficient to cast strong doubts about such convenient assumptions when we have reliable priors on the similarities of human behaviors. I have never witnessed any of you ignoring or discarding any survey and hence making yourself a contributor to the big headache of non-response bias, in which I invested a considerable amount of my professional time to ease. Yet, I am willing to put my professional reputation (if I have one) on the line to state that statistically speaking, we all have contributed to this problem multiple times in our private lives, drawing from my experience of not being able to answer over 95% of the surveys I receive every year, no matter how hard I compel myself on a professional and moral ground. (If I have just insulted you by implying your moral standard is as low as mine, please be in touch so I can send you an HDSR readership survey as a token of my apology.)

Data privacy is another area where reflecting on our private behaviors may have professional benefits (and vice versa). Bluntly, data privacy is an oxymoronic term, because data are born to reveal, yet privacy requires us to conceal. Periodically reflecting upon our private behaviors should help us better appreciate how complex the issue is, be more sensible in making professional demands, and give others the benefit of doubt when they seem to make the data less private or useful than our preferred level of trade-off (or lack thereof).

My COVID encounter reminded me of this possibility because of the actual instance of trading between providing timely contact tracing information and protecting the privacy of the infected individuals. Identities are critical for contact tracing, and I'm deeply grateful for such information volunteered by an Individual, who also shared the experience of the rapid onset without warning signs. This timely information, and the knowledge that it could strike rather suddenly, gave me just enough time, and reason, to make arrangements with my family for a

minimax quarantine strategy before I got home—an arrangement that, retrospectively, we are all glad that we made.

If physical health is the only metric for optimization, one may argue that any personal information that can help others to reduce the risk of delayed treatments or the spreading of the virus should be shared as quickly and as widely as possible among (in this case) the conference attendees, regardless of how the information is obtained. Again, I don't need to insult anyone's intelligence by explaining why a single metric, however well-intended and well-designed, would almost always fall short in addressing problems in the human ecosystem. But my intelligence is seriously auto-insulated by my failure to find a privacy-preserving narrative that would reveal an additional privacy dilemma this conference faced, but without increasing the privacy-loss budget for any meeting attendees, especially those who do not wish to disclose their COVID status. Protecting privacy is extremely hard, because information travels like a virus. (It also mutates as it spreads.)

I may as well take the cue and stop here before this self-invited feverish reopening remark becomes an editor-invited closing remark for the "XL-Files." But I still need to credit where credit is due. Can anyone help to locate the original source of this inspiration for the title of this column? *"Don't become a statistic, drive safely. Go to graduate school-become a statistician"?*

Xiao-Li Meng, Ph.D.
Whipple V. N. Jones Professor of Statistics
Department of Statistics
Harvard University

On the Convergence of Epidemiology, Biostatistics, and Data Science

Neal D. Goldstein, Michael LeVasseur and Leslie A. McClure

Editorial: This artical is originally published on Harvard Data Science Review, Issue 2.2, Spring 2020. DOI: https://doi.org/10.1162/99608f92.9f0215e6

ABSTRACT

Epidemiology, biostatistics, and data science are broad disciplines that incorporate a variety of substantive areas. Common among them is a focus on quantitative approaches for solving intricate problems. When the substantive area is health and health care, the overlap is further cemented. Researchers in these disciplines are fluent in statistics, data management and analysis, and health and medicine, to name but a few competencies. Yet there are important and perhaps mutually exclusive attributes of these fields that warrant a tighter integration. For example, epidemiologists receive substantial training in the science of study design, measurement, and the art of causal inference. Biostatisticians are well versed in the theory and application of methodological techniques, as well as the design and conduct of public health research. Data scientists receive equivalently rigorous training in computational and visualization approaches for high-dimensional data. Compared to data scientists, epidemiologists and biostatisticians may have less expertise in computer science and informatics, while data scientists may benefit from a working knowledge of study design and causal inference. Collaboration and cross-training offer the opportunity to share and learn of the constructs, frameworks, theories, and methods of these fields with the goal of offering fresh and innovate perspectives for tackling challenging problems in health and health care. In this article, we first describe the evolution of these fields focusing on their convergence in the era of electronic health data, notably electronic medical records (EMRs). Next we present how a collaborative team may design, analyze, and implement an EMR-based study. Finally, we review the curricula at leading epidemiology, biostatistics, and data science training programs, identifying gaps and offering suggestions for the fields moving forward.

Keywords: epidemiology, biostatistics, data science, training and education, causal inference, study design, electronic medical records.

1. Introduction: A Confluence of Concepts

The fields of epidemiology, biostatistics, and data science, while very distinct in their focus on training, share much in common in that they all rely upon an intersection of various and overlapping concepts. These concepts include statistical methods, research design, and substantive expertise. Rigorous analysis of quantitative data is the common thread among them. When data science is applied to health and medicine for understanding disease etiology, the distinction between the fields becomes blurred. For the data scientist engaging in healthrelated research, epidemiology and biostatistics provide appropriate complementary knowledge and skillsets through the application of causal inference theory, meticulous study design and measurement, and the development of new statistical methods. Likewise, for the epidemiologist working with massive amounts of health care data, data science provides innovative and robust computational and visualization approaches for high dimensional data that may not be traditionally taught in epidemiology training programs, while biostatistics brings novel statistical methods that could improve inference about the data. For a biostatistician concerned with developing new methods that lessen bias or reduce variance in a particular field, the epidemiologist can bring topic-matter expertise and data, while the data scientist can play a key role in improving the computational aspects of the approach. In short, there is much to be shared across fields, as well as much contributed from each expert, as exemplified in Figure 1. The epidemiologist and biostatistician may lack computer science skills (labeled as hacking skills and including database design, data management, and informatics), the data scientist may lack research expertise in terms of causal inference, study design, and measurement (envisioned as a third dimension in this figure, intersecting the center), and both the data scientist and epidemiologist may lack the statistical background to sufficiently improve on the current methodology.

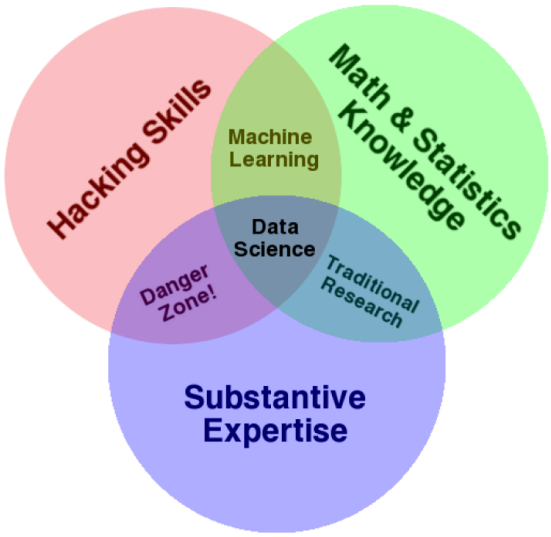


Figure 1. The data science Venn diagram. Reprinted under the Creative Commons license (Conway, 2013).

A brief examination of the history of these fields reveals a natural convergence over time centered around the increasing amount of data available for analyses. Biostatistics surfaced around the mid-1800s for measuring human traits, as well as quantifying morbidity and mortality, but statistical methods applied to health data really took off during the late 1800s with the availability of genetics data (Salsburg, 2001). Meanwhile, epidemiology as a distinct discipline evolved from medicine in response to public health infectious disease crises of the 19th century (Rosen, 1993). As public health research diverged from an infectious disease perspective to a chronic disease perspective, methods were developed to specifically mitigate the effects of bias and confounding resulting from nonexperimental study designs (Greenland, 1987; Susser & Stein, 2009). Figure 1. The data science Venn diagram. Reprinted under the Creative Commons license (Conway, 2013). Biostatistics has provided much of this methodology, and training in biostatistics has emphasized basic programming and data management skills, with this emphasis growing as statistical software has become more readily available. With the rise of electronic medical records (EMRs) in the later part of the 20th century (Henry et al., 2016), as well as the ever-increasing amount of health-related data in other disciplines, the modern-day epidemiologist and biostatistician continue to evolve to better understand disparate data sources.

Data science became formalized during the mid-20th century and is a comparatively new field. Recognizing the need for computer scientists to not only define and develop software and hardware platforms, but to analyze the data captured electronically therein, a cross-disciplinary approach was proposed that incorporated the rigor of various computational approaches with statistics (Cleveland, 2014). Yet it is not solely an applied discipline with a focus on algorithmic development, such as machine learning, or statistics (Meng, 2019). As with epidemiology, it extends beyond methods: formalized theory and frameworks help to define the training and skillset necessary for the data scientist (Borgman, 2019; Floridi & Cowls, 2019). Rooted in common among epidemiology, biostatistics, and data science has been how quantitative—and more recently, qualitative – data can be used to answer research and programmatic questions, including important questions that can be answered with electronic health data. We believe these disciplines have much to learn from and share with each other, and thus we discuss the education, skills, and competencies that a modern-day researcher who works with electronic health data must possess.

We begin with a motivational research question using data derived from the electronic medical record (EMR), which has become frequent with the near ubiquity of EMRs in medical practice (Henry et al., 2016). We then proceed with a broad overview of health-related research as it applies to etiological questions: did some exposure cause disease? We return to our example of an EMR-based study to demonstrate the complementary roles of epidemiology, biostatistics, and data science in addressing our research question, and then conclude by discussing the state of formalized educational programs in the United States and provide recommendations

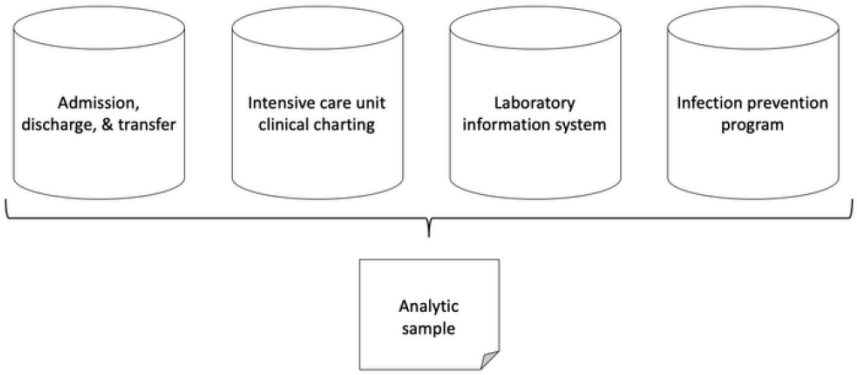


Figure 2. Simplified architecture of an electronic medical record system as it relates to our research question: Does the number of occupied beds in an intensive care unit increase risk for infection?

Figure 3. The traditional systems development lifecycle. Adapted from Information Management and Security Staff, 2003, Chapter 1.

for cross-training moving forward.

2. A Motivational Example of an EMR-Based Research Question

Suppose a research group is interested in conducting a study on whether the number of occupied beds in an intensive care unit (ICU) is related to risk for infection (Goldstein et al., 2017). The researchers hypothesize that the higher the ICU' s occupancy rate, the more likely it is for basic hygiene practices to break down, thus leading to increased exposure to pathogenic organisms such as methicillin-resistant Staphylococcus aureus. The researchers expect the patient' s admitting diagnosis, comorbidities, and length of ICU stay may also be related to the hypothesis. These data can all be ascertained from the EMR.

Answering research questions that involve EMRs is inherently cross-disciplinary. EMRs are complex data systems, and require expertise in databases, data linkage, and data abstraction to compile the analytic sample (Figure 2). Meanwhile, understanding risk of infection in a health care environment represents a web of causality: there are many potential factors that could explain the outcome, requiring sophisticated methodologies to unpack. Fur-

ther, approaches to assessing and mitigating potential biases arising from the data, including from incomplete data, are important to providing the best answer to the researchers' question. In our view, research that utilizes EMR data is beyond the bounds of any single field in isolation, with the best answer to questions such as this arising as a result of team science, including our clinical colleagues. Indeed, a 2018 article exemplifies the potential of data science as part of the team when conducting EMR research: the authors had to mine free text clinical notes in an EMR to derive social risk factors that may otherwise be discrete variables in a prospectively designed epidemiological study (Navathe et al., 2018).

Continuing our hypothetical example, the epidemiologist suggests assembling a retrospective cohort from the EMR records for a one-year period and the data scientist is able to interface with the EMR, retrieve a patient list, and abstract all of the variables necessary for analysis. The biostatistician conducts a rigorous analysis, including assessing completeness of the data and identifying potential biases in the analysis. The team observes a strong relationship between an increased number of occupied beds and increased risk for infection in the data. Does this reflect some underlying causal relation?

Public Health Methodology for the Data Scientist: When Does Correlation Equal Causation?

Public health researchers are trained in the art and science of causal inference—the process of evaluating whether a health-related outcome would have been affected given a change in an exposure. Epidemiologists evaluate causal inference using two separate – but equally important – factors: internal and external validity. Internal validity refers to the ability of a study to correctly ascribe the true underlying relationship within the confines of the study. External validity refers to the ability of a study to correctly ascribe the true causal relationship outside of the confines of the study – that the results are generalizable and transportable. Biostatisticians help ensure studies are designed to maximize both internal and external validity, while also developing statistical methods that better answer the scientific questions posed by public health and clinical researchers. Together, biostatisticians and epidemiologists have developed and adapted numerous methods and study designs to reduce the threats to both internal and external validity and have employed methods for assessing the effect of these threats (Morabia, 2004).

Threats to internal validity include random error, bias (aka systematic error), and confounding. These threats are most often, but not exclusively, found in observational studies, such as our EMR-based example. Random error can be minimized through appropriate sample size and power calculations, although any given study may have the possibility of arriving at an erroneous conclusion on the basis of chance alone. Multiple studies conducted using different study designs in similar settings can increase confidence that the results are not due to chance alone, although researchers need to be aware of effect heterogeneity whereby the same analysis conducted in different samples may produce striking, albeit real, differences (Madigan et al., 2013). Broadly speaking, bias can be classified as selection bias or information bias. Selection bias occurs when individuals from an eligible population have a differential probability of being included in a study based on both their exposure and outcome status. Information bias occurs when there is a systematic tendency to erroneously measure the effect, its antecedent cause, or any other covariates that are involved in the exposure to outcome relation. It is important to note that while increasing sample sizes (as the 'big data' movement is witness to) can increase the precision of an estimate, it does nothing to mitigate the effects of bias. That is to say, if there is bias present in one's data, having a larger sample size only means that one has more precisely measured a biased effect.

Confounding occurs when some factor is causally associated with the outcome and noncausally associated with the exposure. This results in a 'mixing' of effects that distorts the true effect between exposure and outcome. While some consider confounding to be a form of bias, the key difference between bias and confounding is that bias is artificially introduced by the researcher whereas confounding exists in nature. Bias and confounding may never be completely removed from a study: the goal is to understand its presence and potential impact on the observed association. Interested readers are referred to the field of quantitative bias analysis (Lash et al., 2009).

Mitigating the effects of bias and confounding occurs at every stage of public health research, including study design (understanding the influence of study design on bias, reducing selection bias in sample selection), data collection (properly measuring variables of interest, reducing the amount of missing data, collecting all data that may be relevant to the evaluation of confounding), data management (properly coding variables, formatting data sets to best answer the research question, sum-

marizing variables in meaningful ways, managing missing data), analysis (using the correct statistical techniques, model building, confounder assessment), and interpretation (appropriate interpretation of the results, sensitivity analyses to assess the influence of bias, error, and assumptions). Even the best designed and executed studies have flaws and may not be externally valid. In fact, randomized controlled trials, which aim to eliminate bias and confounding by randomizing people to treatment, often have strict inclusion criteria that make their inferences nongeneralizable (Rothwell, 2005). Thus, solving one problem often leads to others.

Epidemiologists reflect upon philosophical as much as practical matters when it comes to their approach to science. Before engaging in designing a research study, the epidemiologist, often in collaboration with a biostatistician, will formulate a research question ensuring that it is answerable methodologically. This question ultimately influences the type of study to undertake. Study design has important implications for application of correct analytical procedures and causal inference. Practically, health researchers consider two main categories of study design—observational and experimental—the key difference being the manipulation of the exposure. Experimental studies, such as randomized clinical trials, allow the researcher to manipulate the exposure, whereas observational studies do not. For example, had our motivational research question been, 'Will altering the ICU admission process and bed location reduce the risk for infection?' the study design would have been experimental, as the investigators would be directly manipulating the treatment, in this case, the patient admission process. Randomized trials are often considered the gold standard for assessing causality but are infeasible in many research projects. Contrast this to our stated research question, "Does the number of occupied beds in an ICU relate to risk of infection." This question does not manipulate any exposure effect (we are not moving around patients after all); rather, we simply observe what happens naturally over time in the ICU. This is considered an observational study, the mainstay of epidemiology. Observational studies can further be subdivided into other study types: cross-sectional studies (sometimes termed a prevalence study), case-control studies, and cohort studies, with a variety of hybrid designs possible (Celentano & Szklo, 2018; Rothman et al., 2008; Szklo & Nieto, 2018). A defining factor among different observational study designs is the timing of the exposure and outcome. Cross-sectional studies evaluate both exposure and outcome at a single time point, case-control studies can assess retrospective exposures against an observed health outcome, while cohort studies can either prospectively or retrospectively assess new cases of some outcome given an exposure. Cohort studies are considered the most flexible, albeit the most time-consuming and can be expensive. Studies performed from a health care system' s EMRs, in which a group of patients are followed over time in either the inpatient or outpatient settings, are typically of a (retrospective) cohort nature, as in our example.

Before data are extracted and analyses commence, proper study design can help minimize selection bias and random error. Having a sound theoretical model can help to identify all relevant confounding variables included in the analysis, and equally important, exclude nonrelevant variables. Causal diagrams such as directed acyclic graphs, are conceptual tools that help with variable selection and understanding variable interplay (Rothman et al., 2008). This is especially important given the vast amount of data available in the EMR. During data abstraction and variable operationalization, the research team needs to ensure that all variables have been recorded properly and, to the best of their knowledge, represent the truth, by working with clinical and informatics colleagues. This will hopefully mitigate information bias.

4. Bringing Data Science to Health Research: More Than Just

Machine Learning Data scientists employ a variety of sophisticated methods that noncomputational researchers may not be aware of. Machine learning and artificial intelligence algorithms, one of the many methodological tools of the data scientist, are becoming increasingly utilized in a variety of fields and have advanced causal inference approaches used by epidemiologists and biostatisticians. Various algorithms exist that represent a data-adaptive approach in estimation of causal inference parameters, including targeted minimum loss-based estimation, double/debiased machine learning, and improved construction of propensity scores and their inverse probability weights for predicting exposures (Blakely et al., 2019; Diaz, 2020). So-called Super Learner algorithms may prove useful in the search for candidate risk or protective factors for a given health outcome (Naimi & Balzer, 2018).

At this point, we wish to draw a clear distinction between predictive and causal modeling, and caution against using machine learning and artificial intelligence for the latter, as others have noted (Lin & Ikram, 2019). In a predictive model, one

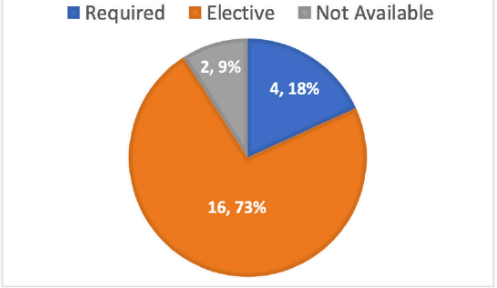


Figure 4a

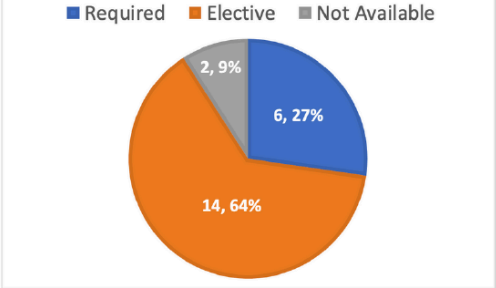


Figure 4b

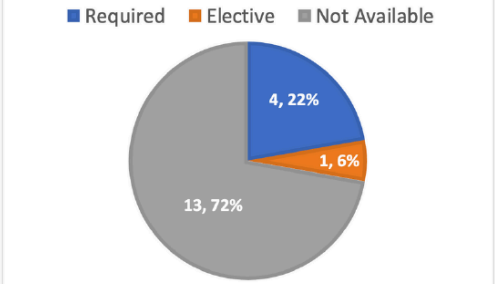


Figure 4c

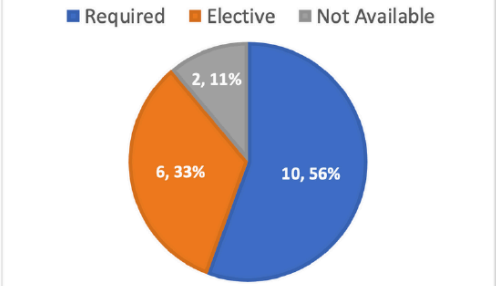


Figure 4d

seeks candidate factors from among a larger set that are statistically associated with an outcome. This is useful for hypothesis generation but may lead to the identification of spurious associations. Occasionally a correlation between covariates in the data may be present but not intervenable and possibly irrelevant from a clinical perspective. This can be especially problematic with high-dimensional data where statistical associations may arise but have little meaning (Lin et al., 2013). In causal modeling, a specific exposure is examined to test its causal relation with the outcome, potentially to intervene upon (if harmful) or promote (if protective). Indeed, machine learning and artificial intelligence may not be the panacea to health and health care problems that some have anticipated; without careful scrutiny and regulation, there is the potential for harm (Kaiser Health News, 2019).

Importantly, the data scientist' s repertoire extends beyond the more recent innovations of machine learning and artificial intelligence (Meng, 2019). For example, data scientists may be versed in sophisticated approaches to data collection, database system management, novel visualization techniques, complex system modeling, the software development lifecycle, data security, data privacy, and algorithm ethics, among others areas. Depending upon the data scientists formalized training, even more specialized expertise may be available. For example, data scientists with backgrounds in computer science or software engineering can develop algorithms, program statistical simulations, and optimize existing analyses. Data scientists with expertise in privacy and security can help unpack the complex requirements of sharing and releasing health data inherent in many types of epidemiological research and implement innovative solutions (Goldstein & Sarwate, 2016). Data scientists who are knowledgeable in linguistics can help create discrete variables from free text in the EMR, such as progress notes, through natural language processing, and data scientists who work with highdimensional data can assist with automated extraction of data from the EMR.

Returning to our hypothetical example, the methods and tools of the data scientist can aid in the investigative process. Suppose the researchers are confident that the observed relation—the number of occupied beds and risk for infection—is not due to chance, bias, or confounding. Attention turns to understanding the mechanism of risk, as well as possible interventions. The data scientist may employ novel visualization techniques to reveal time- and place-based depictions of patients in the ICU, as well as health care workers serving as the pathogen vectors. There may be algorithmic approaches to identifying other salient risk factors in the environment that are intervenable. For example, if the re-

searchers were considering several candidate factors and how they might relate to the infection, one may decide to employ a predictive Super Learner model to generate hypotheses. The data scientist may further be able to collaborate on the development of a complex system simulation of the ICU environment and introduce infection prevention practices to evaluate the potential for staving off pathogen transmission. Theoretically, this simulation model may even reveal the opportune time for an infection prevention practice, such as hand hygiene, to occur.

It is also important to note that the sophisticated techniques we describe are not employed haphazardly. Rather, there is a focus on sound engineering principles, such as testability, maintainability, integrity, reproducibility, and so on. The systems development lifecycle taught in many engineering programs creates a formalized process for planning, creating, testing, and deploying a data system (Figure 3) (Information Management and Security Staff, 2003, Chapter 1). This process can further be decomposed; for example, deploying includes implementation, operations, maintenance, and obsolescence. Continuing with the hypothetical example, the researchers have observed an empiric association in the data between ICU census and risk for infection. The collaboration to develop a complex systems model of the ICU has identified a point of intervention: namely, a hand hygiene reminder at the opportune time. Now, the data scientist, armed with the empiric data obtained from the simulation, can begin the process of deploying such an intervention into the ICU in collaboration with the research team, with careful consideration of testing the algorithm, the appropriate type of implementation (e.g., integrated within the EMR versus a standalone application), evaluating the ongoing operation of this algorithm, including any corresponding maintenance, and planned obsolescence. The epidemiologist and biostatistician can provide expertise in implementation of the intervention and can design an implement an evaluation of its effectiveness, which could in turn result in further refinement by the research team. Truly, this is a cross-collaborative iterative process.

5. Opportunities for Training: Brick and Mortar Barriers to Collaboration

Given the importance of a collaborative model in health research, the question as to whether students are afforded an opportunity to cross-train arises. To assess the current state of formalized training in epidemiology, biostatistics, and data science, we undertook a review of curricula as of the Fall 2019 academic year at the top 20–ranked U.S. News and World Report public health programs (U.S. News and World Report, 2019). For each program, we evaluated the curriculum for each master' s level epidemiology, biostatistics, and data science degree-granting program to assess three factors: 1) the program offering the degree, 2) whether an epidemiology or statistics course (for data science) or data science course (for epidemiology or biostatistics) was required, and 3) if not required, whether these courses were available as an elective. We chose to use as the basis of our review public health program rankings in the United States as opposed to data science program rankings for several reasons. First, to our knowledge, no equivalent list exists ranking the top data science programs. Second, data science is inherently a cross-discipline field and can be housed in schools of engineering, computer science, business, and others. Thus, any ranking system specific to these broader disciplines would be incomplete or include unrelated degrees. Third, as one of our aims was to assess whether an epidemiology component was included in the data science programs, having access to the appropriate faculty would likely necessitate formalized public health degrees at the institution. Therefore, this review can be viewed as the top 20 public health programs in the United States, and whether these universities also offer master' s-level degrees in data science.

There are several other qualifiers to our review we wish to highlight at the outset. Our interest in master' s programs is because they represent a degree that is most likely to be sought by those doing applied work, as opposed to the more academic-focused goals of a doctorate. When assessing whether an epidemiology or biostatistics training program included coursework in data science, we considered courses with a primary focus in computational science (aside from statistical computing), informatics, or data management to be sufficient to label as data science. Likewise, when assessing whether a data science training program included coursework in epidemiology, we considered courses with a primary focus in study and experimental design or causal inference to be sufficient to be labeled as epidemiology, even if not taught by faculty in public health. Occasionally, a university offered competing or similar programs of study out of different schools. In this case, we focused on the program labeled as or most directly aligned with data science (e.g., as opposed to a degree in health informatics or business analytics). Additionally, if a program offered multiple master' s degrees, we evaluated the research-focused degree (e.g., a master' s in science

superseded a master' s in public health). When evaluating electives, unless explicitly indicated in the curriculum, we included only courses within the school/college offering the degree.

Among the 22 reviewed programs (top 20 plus ties; Figure 4 and Supplement 1), all conferred epidemiologyand biostatistics-related degrees as part of a master' s in public health or a master' s in science, and 18 (82%) conferred a data science–related degree most often as a master' s of science, although several nonthesis and engineering degrees were available. Data science degrees were offered out of a variety of schools and colleges, indicating the cross-disciplinary nature of the field (Supplement 1). Most commonly, these data science programs were found in Schools and Colleges of Arts and Science ($n = 5$, 28%) and Engineering ($n = 6$, 33%). In some cases, the program was housed in an interdisciplinary institute such as Brown University' s Data Science Initiative and University of Washington's eScience Institute.

A data science component was more often required in a biostatistics training program ($n = 6$, 27%) than in an epidemiology training program ($n = 4$, 18%), and it was frequently available as an elective for both (n = 14, 64% for biostatistics; $n = 16$, 73% for epidemiology). Most often, data science coursework was available through a biostatistics course designation, suggesting an alignment of data science with biostatistics. Contrast this with an epidemiology component in a data science program, where it was required in a similar proportion ($n = 4$, 21%) but less frequently available as an elective ($n = 2$, 11%). Statistical coursework in a data science program, being a core component of the discipline, was offered more frequently ($n = 16$, 89%), though it was not ubiquitous.

Several programs warrant specific comments or highlighting. Harvard University offers both a master' s of science in data science through the School of Engineering and Applied Sciences and a master' s of science in health data science through the School of Public Health. This latter degree explicitly included an epidemiology requirement, whereas the former did not. The University of North Carolina (epidemiology and biostatistics), University of Washington (biostatistics), and University of Pittsburgh (biostatistics) offered a data science–specific track in their public health programs. The University of Michigan at Ann Arbor School of Information offered a health data science concentration in their master' s of applied data science program, which included coursework in experimental design and analysis. The University of Pittsburgh offered data science tracks within two programs: a master' s in health informatics through their School of Health and Rehabilitation Sciences and a master' s in information science through their School of Computing and Information, with both emphasizing the requirements of analyzing large data sets regardless of discipline, common in data science. Yale University' s Department of Statistics and Data Science offered a postgraduate certificate in data science without a formal degree program—a nondegree option we suspect is available elsewhere—while the University of Iowa College of Liberal Arts and Sciences offered an undergraduate degree in data science. The only program to offer a specific course in causal inference in a data science program was the University of California at Berkeley School of Information, albeit as an elective. More often than not, the epidemiology component was either satisfied by taking a course directly through the public health program or an independent course in study and experimental design.

Despite the importance of cross-training students to prepare them for collaboration, there are physical 'brick and mortar' barriers to doing so. Our review uncovered that epidemiology and biostatistics are traditionally taught in Schools and Colleges of Public Health (clinical epidemiology is also offered in many medical schools, although we did not explicitly evaluate this subdiscipline), whereas data science was more likely to fall under Schools and Colleges of Engineering, Arts and Science, or Information. In cases where data science was housed in Schools or Colleges of Engineering or Computing and Information, we observed statistics was less often a required course, compared to Schools or Colleges of Arts and Science or Public Health offering data science degrees. Data science appeared to be a hybrid program more often than epidemiology, meaning the training drew on expertise across departments and programs more often. Yet the fields are still siloed: aside from the programs specific to health data science, we did not observe any data science program that included coursework from an epidemiology training program, whereas most programs do include a statistics or biostatistics course. Unfortunately, being physically in separate spaces may hamper collaboration, as some training programs stipulated that coursework may not be derived from outside of the school or college. This separation may also translate into research barriers among established faculty, not just students. We believe a primary reason for the close relationship between epidemiology and biostatistics is due to the fact that they are located in the same school

or college, if not the same department.

6. Discussion: What Does the Future Hold?

Data literacy underscores our themes in this article. Data are inextricably embedded in everything we do as researchers; we all struggle with issues of data quality, measurement error, bias, and missing data. Training students to understand the possibilities, and more importantly, the limitations of data is paramount. As was argued in the first issue of HDSR, the approach to training data scientists can be tiered, with different levels of theoretical and methodological expertise depending on the type of student (Garber, 2019). This is also true for epidemiologists and biostatisticians: while they do not necessarily need to become experts in machine learning, artificial intelligence, database systems, and other data science approaches, they do need a foundational knowledge to enable them to communicate across members of the interdisciplinary teams required to answer important scientific questions today. This has been recognized by recent efforts to bring a related data-driven discipline into the public health training program: informatics (Dixon et al., 2015).

In our view, specialized health data science degrees and data science concentrations or tracks in epidemiology and biostatistics programs represent an appropriate paradigm for training the future generation of public health researchers, given access to the expertise in these areas in a School or College of Public Health. Additionally, training in the art and science of causal inference, study design, statistical methods, and measurement should be brought into pure data science programs, to guard against resources being invested in spurious or confounded associations. These methods need not always be taught in the context of public health, but may come from fields such as economics, psychology, and others. Economics, for example, has also developed rigorous causal inference techniques that, while often similar to those employed in public health research, represent a convergent evolution of methods (Angrist & Pischke, 2008, Chapter 5.2). The more we cross-train in other disciplines, the more we appropriately blur the distinction in the fields (Figure 1). The relative weightings of the constructs in Figure 1 can be aligned with a researcher' s—or student' s—interests and skills, and while an equal balance is likely unachievable, and perhaps even unnecessary, within an individual the overlap or lack thereof can emphasize a researcher' s specific skillset and draw attention to the collaborators needed for a given project. The selection of potential collaborators and mentors can play to the researcher' s respective strengths: epidemiologists are well versed in study design and causal inference, biostatisticians have an arsenal of analytic methods and the theoretical knowledge to develop new methods as needed, and data scientists understand data provenance and visualization. Further, while the availability of online computational resources is nearly endless, the appropriate use of these tools demands content-area knowledge that can only come from experience, training, and collaboration.

Lastly, we recommend that regardless of the discipline or field of study, epidemiologists, biostatisticians, and data scientists embrace transparent and open science (Hamra et al., 2019). There has been a recent push in public health and medicine toward releasing both data and code, as the description of methods alone may be insufficient for reproducibility, and we call upon our data science colleagues to do the same (Goldstein, 2018; Goldstein et al., 2019). Even though the programming languages of choice may differ, having data and code publicly available may help guard against erroneous findings and promote insight into complex methodologies as scientists adapt each others' code (Piwowar et al., 2007; Stodden et al., 2013). An example of this comes from the COVID-19 pandemic of 2019–2020, where a flurry of mathematical models were used to inform difficult policy decisions and the analytic codes to many of these models were released in the public domain (Wynants et al., 2020). This is a positive first step, but publishing a model is not an end in and of itself, especially if it has not been peer reviewed. Rather, this should be the beginning of a dialogue between the modelers, epidemiologists, and public health policymakers to ensure that assumptions entered into the model are valid and policy recommendations are in line with other considerations.

In summary, we are excited about the evolution of these fields, as we seek to answer more difficult public health questions with increasingly more complicated data sources. They are converging at an opportune time around the use of electronic health data. Studies of health phenomena are complex endeavors requiring large teams with expertise along all the steps of the research continuum. Having collaborators with complementary skills as part of the research team can provide insight and direction in the ongoing quest for better ways to prevent and treat disease, and we can only accomplish this through synergies in training epidemiologists, biostatisticians, and data scientists.

Disclosure Statement

Research reported in this publication was supported by the National Institute Of Allergy And Infectious Diseases of the National Institutes of Health under Award Number K01AI143356 (to NDG). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Statistics in Biosciences CALL FOR PAPERS

Special issue on "Statistical Methods, Algorithms and Applications in Biomedical Data Integration"

We invite submissions to the special issue of Statistics in Biosciences dedicated to statistical methods, algorithms and applications in biomedical data integration. Integrating multiple data sources has attracted wide interests as part of transition from data to knowledge with the potential to change the analytical scheme of modern biomedical research. In recent years, research initiatives have been created to leverage large-scale observational databases from multiple scientific disciplines and technologies, which present many significant methodological and computational challenges to traditional statistical methods and algorithms. Most notably, data collected from observational studies are leveraged for modern biomedical research to enrich study populations and improve controlling confounding factors. Despite infrastructure advancements, methodological and algorithmic challenges remain the barriers for data integration. Common analytical concerns in data integration include data storage and communication restrictions, statistical efficiency, protection of data privacy and against adversarial attacks, data harmonization over different formats, handling of missing data, heterogeneity across data sources, and external validity, among many others. Some new methods, algorithms and applications for data integration are being developed, but much remains unknown in terms of how well they perform or how they compare with conventional approaches. Moreover, with the ubiquitous availability of multi-source data and the increased desire to conduct research with massive data, new methodological and algorithmic developments are needed on many fronts of data integration, including uncertainty quantification, causal inference, and sparse and scalable analytical procedures.

The special issue welcomes new methodological developments as well as interesting applications in data integration. It also welcomes up-to-date reviews of current data analytic and computational tools in this area. All submissions must contain original unpublished work not being considered for publication elsewhere. Submissions will be refereed according to the standard procedures for Statistics in Biosciences. The deadline for submissions is August 31, 2023, with a rolling review.

Papers for the special issue should be submitted using the journal' s submission system at `https://www.editorialmanager.com/sibs/default1.aspx`. In the system, please choose the special issue on Statistical Methods, Algorithms and Applications in Biomedical Data Integration. The review and decision will be on a rolling basis.

Co-Editors for the special issue:

- Peter X.-K. Song, Ph.D., University of Michigan. Email: `pxsong@umich.edu`
- Lu Tang, Ph.D., University of Pittsburgh. Email: `lutang@pitt.edu`

Special issue on "Machine Learning in Biomedical Sciences"

We invite submissions to the special issue of Statistics in Biosciences dedicated to machine learning approaches in biomedical sciences. The last few years have seen a huge data collection increase in biological and biomedical research fields, such as biobanks, neuroimaging, spatial transcriptomics, single cell genomics, cancer genomics, and microbiomes. These biomedical data present numerous challenges and pitfalls for machine learning methods due to the complexity of data structures, the high dimensionality of biomarkers, and the heterogeneity of data resources. Novel statistical machine learning methods are demanded to address those problems in the context of translating clinical data into knowledge and practice. To join the force to improve data

collection, quality control, and analysis in a wide range of biomedical fields, we plan to publish in this special issue manuscripts on new statistical machine learning methods that target on emerging questions in biomedical data analyses, such as sampling bias, biomarker discovery, high-dimension data analysis, streaming and online data analysis, domain adaptation, dimensional reduction methods, data integration techniques, data privacy, missing data methods, predictive models, transfer learning, causal and mediation inference, and statistical diversity and generalizability.

The special issue welcomes new methodological developments as well as interesting applications in machine learning in biomedical sciences. It also welcomes up-to-date reviews of current tools in this area. All submissions must contain original unpublished work not being considered for publication elsewhere. Submissions will be refereed according to the standard procedures for Statistics in Biosciences. The deadline for submissions is January 31, 2023.

Papers for the special issue should be submitted using the journal' s submission system at https://www.editorialmanager.com/sibs/default1.aspx. In the system, please choose the special issue on Machine Learning in Biomedical Sciences.

Co-Editors for the special issue:

- Dehan Kong, Ph.D., University of Toronto. Email: dehan.kong@utoronto.ca
- Bingxin Zhao, Ph.D., University of Pennsylvania. Email: bxzhao@upenn.edu

Upcoming Events

Please find below a list of upcoming ICSA meetings and co-sponsored meetings. This list also appears on the ICSA website. If you have any questions, please contact Dr. Mengling Liu, the ICSA Executive Director (executive.director@icsa.org).

ICSA Sponsored Meetings:

ICSA 2023 Applied Statistics Symposium

June 11- June 14, 2023
The ICSA 2023 Applied Statistics Symposium will be held on June 11-14, 2023 at the University of Michigan, Ann Arbor, MI. For session proposals, interests, and questions, please contact the co-Chairs of the Organizing Committee, Dr. Jian Kang (jiankang@umich.edu) and Dr. Gongjun Xu (gongjun@umich.edu). For the details of symposium, please visit its website: https://symposium2023.icsa.org/

The 2023 ICSA China Conference

June 30-July 3, 2023
The 2023 ICSA China Conference will be held at Chengdu, Sichuan, China from June 30-July 3, 2023. It will be co-sponsored by the Southwest Jiaotong University (SWJTU). The conference venue is Jinniu Hotel. The hotel information can be found at http://www.jnhotel.com/. For more information, please contact the Scientific Program Committee Chair, Professor Yichuan Zhao at yichuan@gsu.edu. The scientific program committee welcomes invited session proposals. Some related information about submission link and timeline of the invited session proposal is as follows.

One-talk rule will be applied. Each speaker in our conference can only give one invited talk. Each invited session consists of either 4 presenters or 3 presenters and 1 discussant.

First check the link for the proposal submission https://china2023.icsa.org/session-submission-form/. Then, submit the session information including session title, brief description, the speakers and the talk titles, etc.

Please submit your proposed session before December 15, 2022.

ICSA Co-sponsored Meetings:

The 8th Workshop on Biostatistics and Bioinformatics

Postponed to Spring, 2023
Biostatistics and Bioinformatics have been playing key and important roles in statistics and other scientific research fields in recent years. The goal of the 8th workshop is to stimulate research and to foster the interaction of researchers in Biostatistics and Bioinformatics research areas. The workshop will provide the opportunity for faculty and grad-

uate students to meet the top researchers, identify important directions for future research, facilitate research collaborations. The workshop will be held at Atlanta, GA.

A keynote speaker is Dr. Nilanjan Chatterjee, Bloomberg Distinguished Professor of Biostatistics and Medicine at the Johns Hopkins University. For detailed information including registration, please refer to `https://math.gsu.edu/yichuan/2022Workshop/`

IIMS Asia Pacific Rim Meeting

Postponed to January, 2024
The sixth meeting of the Institute of Mathematical Statistics Asia Pacific Rim Meeting (IMS-APRM) will provide an excellent worldwide forum for scientific communications and collaborations for researchers in Asia and the Pacific Rim, and promote collaborations between researchers in this area and other parts of the world. The meeting will be held in Melbourne, Australia and please see `http://ims-aprm2021.com/fordetails`. Firm dates will be announced later.

Online Training and Seminars:

Healthcare Innovation Technology: The Pod of Asclepius

Looking to stay up to date on developments in health care technology around the world? The American Statistical Association is sponsoring "The Pod of Asclepius" , a new podcast where data scientists, statisticians, engineers, and regulatory experts discuss the technical challenges in their healthcare domain.

We have over 20 episodes published and available on YouTube, Podbean, iTunes, Stitcher, Podchaser, Tune In Radio, and Google Play. Looking for a good place to start? Check out the following episode links:

- Risks and Opportunities of AI in Clinical Drug Development with David Madigan and Demissie Alemayehu
- Kidney Injury - Biomarkers for Prediction and Prognosis with Allison Meisner
- NHS Digital Health Initiatives with Emma Hughes
- Data Platforms to Monitor Animal Health with Shane Burns
- Bayesian Approaches in Medical Devices: Part 1, Part 2, Part 3 with Martin Ho and Greg Maislin

You can catch up on all episodes on our YouTube playlists for Season 0 and Season 1. The easiest way to catch new episodes is to subscribe via our channels:

- Youtube: `https://www.youtube.com/channel/UCkEz2tDR5K6AjlKw-JrV57w`
- Podbean: `https://podofasclepius.podbean.com`
- You can see our full schedule on the website: `www.podofasclepius.com`

For detailed information, please visit: `https://www.podofasclepius.com/philosophy-of-data-science`.

2022 ICSA Awards (Jianguo (Tony) Sun, Rongwei (Rochelle) Fu, Linbo Wang, Xinran Li)

2022 ICSA Board Meeting

2022 ICSA Banquet