ISSN: 2226-2393

www.icsa.org

January, 2021

Volume 33, Issue 1

International Chinese Statistical Association

泛華統計協會

Bulletin

會刊

ICSA Bulletin

Volume 33/1, February, 2021

ISSN 2226-2393

Editorial Staff

From the Editor

Ming Wang

Dear ICSA Members:

Welcome to our first issue of the 2021 ICSA Bulletin! Happy New Year!

During this special period, it is my great pleasure and honor to serve as the new editor-in-chief of the ICSA Bulletin for 2021-2023. First, I would like to thank Dr. Yi Huang's hard work on previous issues, and appreciate her generous support to help me get familiar with the contents and the generation procedure of our ICSA bulletin! Also, with the guidance of the 2021 President Dr. Colin Wu and Executive Director Dr. Mengling Liu, this transition period went smoothly. I have a depth learning about the foundation of ICSA and its revolution, remarkable leaders and many excellent members who make our association larger and more impactful to the society. I will endeavor to do my best to maintain our ICSA bulletin quality as high as I can.

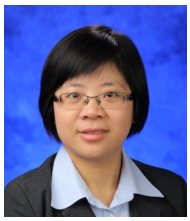
In this issue of the ICSA Bulletin, there are two articles, one memorial article and four column articles along with prior bulletins' tradition. The first invited article is an interview article from Professor Bin Yu, written by Jon Bashor (reprinted with permission from the website of University of California Berkeley Division of Computing, Data Science and Society). As the Chancellor's Professor in the Departments of Statistics and of Electrical Engineering and Computer Sciences at the University of California at Berkeley, Dr. Yu provides a novel insight on a better framework for more robust, trustworthy data science. Furthermore, in the second article, Dr. Yu presented the most exciting 10 challenges in data science to solve real-world data problems with positive impacts. This article is reprinted from the Harvard Data Sciences Review with permission (https://hdsr.mitpress.mit.edu/pub/ekrhsui8/release/1). Dr. Yu and her group at Berkeley is engaged in interdisciplinary research by employing quantitative critical thinking and developing statistical and machine learning algorithms and theory. Now they are actively involved in research for fighting the COVID-19 pandemic, and look forward to their important findings! In addition, we include one memorial article assembling memories and stories for remembering Dr. Xiangrong Yin. We feel sad to lose such a great member and friend, and thank for his joyful spirit and inspiration brought to everyone that knew him!

In the column "Hints from Hans," Professor Hans Rudolf Künsch talks about some stories during the pandemic to exemplify a dilemma of science, and emphasize the awareness of some issues in the communication of scientific evidence and uncertainty. In the column of "Yi's FDA Stories," Dr. Yi Tsong shared with us important historical events happened in 1994, and mentioned his experience in analyzing the Upper gastrointestinal bleeding, perforation, and ulcer (UPU) events from Nonsteroidal anti-inflammatory drugs in 1990s. In the column "Terence's Stuff," Professor Terry Speed discusses the concerns and the importance of assumptions for statistical analysis. How we can make a reasonable statistical assumption, apply our statistical methods correctly and then infer a valid statistical conclusion about the real world still has a long way to go. This turns out to be more urged and critical during this pandemic. This is a reprint from the author' s column "Assumptions" in the IMS Bulletin, November 2016. In the column "XL-Files," Professor Xiao-Li Meng shares his opinions on polling, and provides a general idea of representative sampling by utilizing a vivid example of "soup" for explanation. This is a reprint from the author' s column "Opinion Polling: Its Secret Sauce is also its Spoilage Source" in the IMS Bulletin, February 2021.

Regarding to ICSA business, this issue of the Bulletin includes a message from the 2021 ICSA President, Dr. Colin Wu, and also the 2020 ICSA President, Dr. Jianguo Sun; the announcement of ICSA 2021 official election results; the call for Nominations of Candidates for 2022 ICSA Officers; the call for Nominations of the 2021 ICSA Awards; report from the ICSA Springer Book Series in Statistics. This issue also contains 2020 ICSA Financial Report, and also summary report of the 2020 ICSA Applied Statistics Symposium provided by Dr. Hulin Wu. The coming ICSA sponsored or cosponsored meetings and conferences are announced at the end of this issue. Of not is that we also have a two-page introduction on Harvard Data Sciences Review (HDSR), an open access resource for leading and advanced content in data science, with the founding editor-in-chief of Professor Xiao-Li Meng. In our future issues of ICSA Bulletin, we can reprint any HDSR article of interest to our members. Appreciate Dr. Meng's generosity and strong support on our ICSA! In the meantime, our ICSA members are also encouraged to contribute your work or ideas to the HDSR in future.

In the end, I would like to thank all the con-

tributors, ICSA executives and committee members for their support on our bulletin work. In particular, I would like to thank Dr. Xiaoyu Cai for her time helping us to figure out the LaTeX barriers in the construction of ICSA bulletin for the first time. Also, I want to thank Dr. Chixiang Chen to serve my assistant and helping me on sending follow-up emails and file generation. Their help, support and enthusiasm are critical to ensure this issue come out on time. Lastly, I have to say that many of us experienced anxiety and depression in the past year, but with hopes, I truly believe that there' s a light at the end of the tunnel. Let's keep going by supporting each other and contributing ourselves to the society, and our normal life will come back soon!



Ming Wang, Ph.D.
Editor-in-Chief, ICSA Bulletin
Associate Professor
Division of Biostatistics and Bioinformatics
College of Medicine
Pennsylvania State University

From the 2021 President, ICSA

Colin Wu

Dear ICSA members and friends,

With pride and resilience, we said goodbye to 2020 and entered 2021. Although the past year was tumultuous as we endured the COVID-19 pandemic as well as significant social and economic changes, the ICSA community has been flourishing and growing strong. Thanks to the leadership of the 2020 ICSA leadership team led by 2020 President Dr. Jianguo (Tony) Sun, Executive Director Dr. Mengling Liu, and members of the Bord of Directors, 2020 was a fruitful and productive year for our society. Although we were forced to cancel our 2020 China Conferences due to restriction on international travel, we had a "virtually amazing" 2020 Applied Statistics Symposium, which attracted over six hundred participants, showcased state-of-the-art scientific presentations, and was highlighted by a brilliant virtual talent show. The brilliance exhibited in the presentations and the talent show was an eye opener for me to observe the multifaceted talents of our members and our next generation of productive citizens. It's a demonstration that ICSA is a big warm family that connects us both professionally and socially. I would like to give a big thanks to Dr. Hulin Wu and all the members of the 2020 Applied Statistics Symposium organizing committee. Their dedication and creativity ensured us a productive and enjoyable event during the COVID-19 lockdown even though we were not able to enjoy our annual gatherings as usual. Of course, the success of our society in 2020 went far beyond the Applied Statistics Symposium. Our members have been highly productive in all kinds of scientific and professional activities, which makes our society a highly reputable and respectable entity within the statistical community. Although we statisticians are often not the public faces of high profile scientific projects, in this era of "big data and artificial (actually human) intelligence," our presence and contributions are the backbone for justifying the science behind these projects. Specifically, I would like to thank those ICSA members who have been actively and directly participated in various projects related to the fight against COVID-19. Their dedication, ingenuity, and sacrifice are key components that save countless lives and provide us hope for winning the fight against the pandemic.

From the founding of ICSA in 1988, I have personally witnessed its growth from a small group of statisticians to a major force with more than 1000 active members. Nowadays you cannot miss the sight of our members at all kinds of statistical meetings, such as the JSM, ENAR, WNAR and IMS meetings. After a year of learning the job as the ICSA President-Elect, I am honored to assume the responsibility of ICSA presidency this year. At the same time, I am humbled by the support from the ICSA family, particularly those who had served or are currently serving on various ICSA committees. To ensure the continued success of our society, we need the participation and cumulative effort from all our members. To better serve our existing members and attract new members, we need to have forward visions on what the future may bestow upon our profession and our society. We need to be proactive and take quick actions in this rapidly changing world. I

will follow the time-honored tradition of our past ICSA leaders and constantly seek your comments and suggestions. We could act upon your good suggestions to further improve our society. With the leadership of our Executive Committee and Board of Directors and the hard work of our committee members, we can further enhance our reputation within the statistical community and make our society a fertile ground of professional growth for both young and "young-at-heart" members.

Looking forward to 2021, I am sure that, with the help of our members, we can be a more inclusive professional society and expand our influence into a number of new areas of data science. In particular, we can use the ICSA activities, such as our large meetings, local workshops, webinars and online conferences, to exchange promising ideas in new frontiers, facilitate collaboration among researchers in academia, industry and government, and promote young researchers in our profession. I am proud of the fact that many of our ICSA members are world-renowned experts of statistical machine learning and artificial intelligence (ML/AI). In this era of "Real World Data/Real World Evidence" (RWD/RWE), ICSA has the unique opportunity to play a leadership role in promoting the use of statistical ML/AI methods to other areas of data science and beyond. Given the expanding job market for statisticians in scientific fields utilizing ML/AI, visibility of ICSA activities will be tremendously helpful for our young researchers. My own experience as a proud researcher at the National Institutes of Health (NIH) is that many of the available resources for data, statistical methods and computational tools are still under-utilized and this issue has been well-recognized by the general scientific community. As a result, the NIH has been increasing its efforts through funding and collaborative opportunities in data science. I believe that ICSA can and should participate in these activities through our journals, conferences, workshops and other novel ideas. Our reputation as a leading statistical society will also be tremendously helpful to promote ICSA members to leadership positions in the American Statistical Association (ASA). Given that Asian Americans are still under-represented in many leadership positions of ASA, one of my goal in 2021 is to nominate our prominent members to serve on various leadership positions at ASA and its committees.

As an example of innovative ideas, in this issue of the Bulletin, you can find an interesting interview of Professor Bin Yu together with one of her articles on the new frontiers of data science. In these pieces, Professor Bin Yu shared her experience over the past 30 years pushing the boundary of statistical research to previously unknown territories. These innovative research and discoveries led us to the tools for fighting the COVID-19 pandemic, advancing science and being part of the information revolution. I am sure that you will find these and all other articles in this issue interesting reads. I hope these thought-provoking articles will lead us to actionable items for ICSA in the new year.

Finally, I would like to express my deepest sympathies to Dr. Xiangrong Yin's family. Dr. Xiangrong Yin, professor of statistics at the University of Kentucky, passed away suddenly on August 11, 2020, at the age of 54. He was a prolific and well-recognized researcher and teacher. We miss him dearly. You can find in this issue of the Bulletin a memorial article about Dr. Yin written by Dr. John Stufken, Bank of America Excellence Professor at UNC Greensboro. There will also be an invited memorial panel for Dr. Xiangrong Yin in the 2021 JSM.

Colin Wu Ph.D.
2021 President, ICSA
Mathematical Statistician National Heart, Lung and Blood Institute
National Institutes of Health

From the 2020 President, ICSA

Jianguo (Tony) Sun

Dear ICSA Members and Friends,

Happy New Year! It has been my great pleasure and honor to serve as the president of the ICSA in 2020 and what a year we had. Although COVID-19 made many things difficult and changed many things, I am happy that our association has been doing well in the past year largely due to your involvement and great support and also I feel so grateful to so many people and many things. Among them, I appreciate so much the opportunity of working closely with other ICSA EC members, Drs. Rochelle Fu, Mengling Liu, Colin Wu and Heping Zhang, and the chairs of various ICSA Committees and learned a great deal from each of them. In particular, I learned so much from Dr. Heping Zhang, who served EC for last three years and made tremendous contributions to our ICSA.

I am sure that we all agree that 2020 was a difficult and challenging year but also good in the sense that it gave us the opportunity for ICSA members to show their care and love about our profession and society. In particular, I am very glad to see that many of our members in all fields, including academia, governments and industry, have been doing cutting-edge and timely work to help to fight COVID-19. To pick a couple of examples, Dr. Xihong Lin from Harvard University along with her collaborators has developed an app that uses the community' s help to track COVID-19 and her work on COVID-19 has made impact on people from many fields, including government agencies, and been covered by various news media. Dr. Jun Yan from the University of Connecticut along with other ICSA members organized a Webinar Series: Data Science in Action in Response to the Outbreak of COVID-19 and a special issue on the same topic in Journal of Data Science. Also ICSA co-sponsored journal, Statistics and Its Interface, will have a special issue on emerging issues of COVD-19 too. I applaud all of your significant contributions on this and hope that every ICSA member will continue to make contributions in their own ways.

Each year ICSA holds two major events, ICSA China Conference and ICSA Applied Statistics Symposium. As you have already known and I mentioned at the second issue of 2020 ICSA Bulletin, unfortunately, we had to cancel 2020 ICSA China Conference due to COVID-19 that was originally scheduled to be held at Zhongnan University of Economics and Law, Wuhan, China during June 26-29, 2020. Before the cancelation, a great deal of effort and work from both ICSA side and local side had been put into it and among others, again I wish to thank the Program Co-Chairs Drs. Ying Zhang and Hui Zhao, who had organized an excellent program, and the Local Organizing Committee Chairs, Drs. Hu Zhang and Qinglong Yang, who had planned and booked almost everything needed for a great conference. Although we had to change the 2020 Applied Statistics Symposium to a virtual meeting, I believe that many of you agree with me that we did have a very successful one, held during December 13-16, 2020, largely due to the leadership and dedications of Dr. Hulin Wu, the Executive Committee Chair, and his amazing team. Because of the changes both in timing and format, they essentially prepared two conferences. In addition to three well attended, excellent keynote sessions, the symposium had 92 invited sessions, 50 posters and 2 panel discussion sessions. Also it received over 70 submissions for student paper awards, offered 9 short courses, and attracted over 100 attendees for the ICSA general member meeting, award ceremony and talent show at the evening of December 15, 2020. What an accomplishment!

Among the things that happened in ICSA in 2020, you had elected Dr. Zhezhen Jin from Columbia University as 2021 ICSA President-Elect, an excellent choice. Dr. Jin has served ICSA for many years in various other roles and I am so happy and appreciative that he will continue to serve ICSA. With the joining of Dr. Zhezhen Jin to the EC in 2021, I am sure that ICSA will continue to do well and provide better services to our members. During the year, three new Co-Editors, Dr. Rong Chen, Su-Yun Huang and Xiaotong Shen, for ICSA journal Statistica Sinica started their work, replacing the three retired Co-Editors Drs. Yuan-Chin Ivan Chang, Hans-Georg Muller and Yazhen Wang. Also during the year, Drs. Joan Hu and Ming Wang were selected to replace Drs. Mei-Cheng Wang and Yi Huang as the new Editor-in-Chief for another ICSA journal Statistics in Biosciences and the new ICSA Bulletin editor, respectively. I truly believe that all new editors will continue outstanding work that the former editors had done and make our journals and

bulletin better. Congratulations to all new editors and a big thank to all retired editors for your excellent and hard for ICSA. As usual, we also handled out a number of ICSA awards, including 2020 ICSA Distinguished Achievement Award to Dr. Ming-Hui Chen and 2020 ICSA Outstanding Service Awards to Drs. Hongzhe Lee, Gang Li and Aiyi Liu. Congratulations to Drs. Chen, Lee, Li and Liu and other awardees for your outstanding work.

As mentioned above, I had a very pleasant year to serve ICSA in 2020 mainly because I had so many great people to learn and to rely on and work with for all ICSA activities. In particular, it has been my privilege and great honor to closely work with Dr. Mengling Liu, ICSA Executive Director who leads the ICSA daily operations and makes sure everything on track, and all ICSA committee chairs. Among them, Dr. Joan Hu chaired the Nomination Committee that identified and recommended outstanding and well-qualified candidates for the future leadership of ICSA, and Dr. Zhezhen Jin led the Program Committee that coordinated all ICSA meetings in 2020 and worked really hard to make excellent plans for several future ICSA meetings. Also, Dr. Xiangrong Yin, who unfortunately passed away in August, and Dr. Yichuan Zhao led the Award Committee and the Publication Committee that came up excellent candidates for the ICSA awards and the new editors for Statistics in Biosciences and the ICSA Bulletin, respectively. I am deeply grateful to you and other committee chairs, Drs. Ivan Chan (Lingzi Lu Award), Joyce Chang (Archive), Jianqing Fan (Special Lectures), Rui Feng (Financial Advisory), Bo Fu (Membership), Rochelle Fu (Finance) and Chengsheng Jiang (IT). Without your support and involvement, the success of ICSA would not be possible.

To conclude, I wish to thank all of you for being an ICSA member or friend, for your outstanding and dedicated contributions to ICSA, and for your continuous support, participation of various ICSA activities and volunteer work. With your support and involvement and the leadership of Dr. Colin Wu, 2021 ICSA President, our association will be stronger and succeed in its missions. I hope that we all will stay connected in our big ICSA family and think of others as both ICSA and each of you get better and stronger. I wish all of you and your loved ones staying safe and healthy and having a great, productive and successful 2021!

Jianguo (Tony) Sun, PhD
2020 ICSA President
Professor, Department of Statistics
University of Missouri

From the 2021 President-Elect, ICSA

Zhezhen Jin

Dear ICSA Members,

I am honored and privileged to be elected as 2022 ICSA President. The growth and success of ICSA would not have been possible without the vision, enthusiasm and hard work of our members, and strong leadership of our past and current presidents, executive directors, board of directors and various committee members. As we move into a new era of big data with data science, artificial intelligence, and data analytics, it becomes increasingly critical to have our association provide better platforms for our members to adapt to new challenges and capitalize opportunities for our profession while continuing established successes. The development of online learning and digital media is essential for virtual activities as shown through the COVID-19 pandemic. I will closely work with ICSA executives, board of directors and members to continue to pursue ICSA educational, charitable, and scientific objectives outlined in the ICSA constitution and by-laws. Please feel free to email me (zj7@cumc.columbia.edu) with your ideas and suggestions for how our association can better meet your needs and interests.

Zhezhen Jin, Ph.D.
2021 President-elect
Professor, Department of Biostatistics
Mailman School of Public Health
Columbia University

From the Executive Director 2020-2022

Mengling Liu

Dear ICSA members,

Happy New Year 2021!

In my first year as the Executive Director of ICSA, it has been exhilarating experience for me to work with the ICSA executive members, board of directors, all standing and ad-hoc committees' chairs and members, and office manager, and to witness their tireless efforts and dedication given to ICSA. During 2020, everyone was stretched very thin in every dimension yet still plowed through many challenges and made the ICSA a stronger community. The executive committee held monthly meetings to ensure the ICSA normal business and functionality, and the committees carried out all their tasks in efficient and professional manners. We held two board meetings and one general member meeting virtually. In December 2020, after two rounds of changing schedules and intensive re-planning, the 2020 Applied Symposium Organizing Committee and their members, led by Dr. Hulin Wu, successfully held the virtual ICSA 2020 Applied Statistics Symposium, which certainly brought highlights to the entire ICSA community.

Looking into the new year of 2021, the ICSA 2021 Applied Statistics Symposium will be held virtually in September 12-15, 2021. We would like to call out to our ICSA community to actively participate the symposium, by organizing sessions, giving presentations, serving on committees, chairing a session, and attending the symposium. Even though 2021 is still full of uncertainties, I look forward to it with hopes. In particularly, I look forward to providing service to the ICSA and working with the ICSA community to make another prosperous year.

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

2021 ICSA Core Members

EXECUTIVES:

- President: Colin Wu (wuc@nhlbi.nih.gov)
- Past-President: Jianguo (Tony) Sun (sunj@missouri.edu)
- President-Elect: Zhezhen Jin (zj7@columbia.edu)
- Executive Director: Mengling Liu (2020-2022, executive.director@icsa.org)
- ICSA Treasurer: Rochelle Fu (2019-2021, fur@ohsu.edu)
- The ICSA Office Manager: Grace Ying Li (oicsa@icsa.org, Phone: 317-287-4261)

BOARD of DIRECTORS:

- Yinglei Lai (2019-2021, ylai@gwu.edu)
- Lei Shen (2019-2021, shen_lei@lilly.com)
- Yifei Sun (2019-2021, ys3072@cumc.columbia.edu)
- Xin Tian (2019-2021, tianx@nhlbi.nih.gov)
- Kelly Zou (2019-2021, KelZouDS@gmail.com)
- Jason Liao (2020-2022, jason_liao@merck.com)
- Bin Nan (2020-2022, nanb@uci.edu)

- Peihua Qiu (2020-2022, pqiu@phhp.ufl.edu)
- Jane Zhang (2020-2022, Zhang_Jane@Allergan.com)
- Yichuan Zhao (2020-2022, yichuan@gsu.edu)
- Shu-Hui Chang (2021-2023, shuhui@ntu.edu.tw)
- Yong Chen (2021-2023, ychen123@mail.med.upenn.edu)
- Chenlei Leng (2021-2023, c.leng@warwick.ac.uk)
- Xiaodong Luo (2021-2023, xiaodong.luo@sanofi.com)
- Yang Song (2021-2023, Yang_Song@vrtx.com)

COMMITTEES:

Program Committee:

- Chair: Hulin Wu (Hulin.Wu@uth.tmc.edu)
- Members:
  - Xuming He (2019-2021, JSM Representative 2020, xmhe@umich.edu)
  - Aiyi Liu (2020-2022, JSM Representative 2021, liua@mail.nih.gov)
  - Pei Wang (2021-2023, JSM Representative 2022, pei.wang@mssm.edu)
  - Guoqing Diao (2020-2022, ICSA Symposium 2021, gdiao@email.gwu.edu)
  - Samuel Wu (2021-2023, ICSA Symposium 2022, samwu@biostat.ufl.edu)
  - Hongzhe Li (2021-2022, ICSA International Conference 2019, hongzhe@mail.med.upenn.edu)
  - Xin-Yuan Song (2021-2022, ICSA International Conference 2022, xysong@sta.cuhk.edu.hk)
  - Alan Y Chiang (2019-2021, achiang@celgene.com)
  - Bin Nan (2019-2021, nanb@uci.edu)
  - Ji Zhu (2019-2021, jizhu@umich.edu)
  - Qingning Zhou (2020-2022, qzhou8@uncc.edu)
  - Liang Zhu (2020-2022, Liang.Zhu@uth.tmc.edu)
  - Jie Chen (2020-2022, jiechen0713@gmail.com)

Awards Committee:

- Chair: Judy Wang (judywang@gwu.edu)
- Members:
  - Jie Chen (2019-2021, jiechen0713@gmail.com)
  - Ning Hao (2019-2021, nhao@math.arizona.edu)
  - Hongyuan Cao (2020-2022, hcao@fsu.edu)
  - Xiaofeng Shao (2020-2022, xshao@illinois.edu)
  - Xiaogang Su (2020-2022, xsu@utep.edu)
  - Yichao Wu (2020-2022, yichaowu@uic.edu)
  - Mingxiu Hu (2021-2023, mhu@nektar.com)
  - Jianxin Shi (2021-2023, jianxin.shi@nih.gov)
  - Ying Wei (2021-2023, yw2148@cumc.columbia.edu)
  - Ji Zhu (2021-2023, jizhu@umich.edu)

Nominating and Election Committee:

- Chair: Ming Tan (Ming.Tan@georgetown.edu)
- Members:
  - Bo Huang (2019-2021, Bo.Huang@pfizer.com)
  - Chunling Liu (2019-2021, catherine.chunling.liu@polyu.edu.hk)
  - Yiyuan She (2019-2021, yshe@stat.fsu.edu)
  - Jiayang Sun (2019-2021, jsun@case.edu)
  - Hailong Cheng (2020-2022, hailong.cheng@sunovion.com)
  - Bin Zhang (2020-2022, Bin.Zhang@cchmc.org)

Special Lecture Committee:

- Chair: Gang Li (2021, vli@ucla.edu)
- Members:
  - Haiqun Lin (2019-2021, haiqun.lin@yale.edu)
  - Huazhen Lin (2019-2021, linhz@swufe.edu.cn)
  - Aiyi Liu (2021-2023, liua@mail.nih.gov)
  - Gang Li (2021-2023, GLi@its.jnj.com)

Publication Committee:

- Chair: Ming-Hui Chen (2021, ming-hui.chen@uconn.edu)
- Members:
  - Hongzhe Li (Co-Editors of SIB, hongzhe@mail.med.upenn.edu)
  - Joan Hu (Co-Editors of SIB, joan_hu@sfu.ca)
  - Rong Chen (Co-Editors of Statistica Sinica, rongchen@stat.rutgers.edu)
  - Su-Yun Huang (Co-Editors of Statistica Sinica, syhuang@stat.sinica.edu.tw)
  - Xiaotong Shen (Co-Editors of Statistica Sinica, xshen@umn.edu)
  - Ding-Geng (Din) Chen (Editor of ICSA book series, dinchen@email.unc.edu)
  - Ming Wang (Editor for ICSA Bulletin, mwang@phs.psu.edu)
  - Mengling Liu (Executive Director of ICSA, executive.director@icsa.org)

Membership Committee:

- Chair: Bo Fu (2021, bo.fu@abbvie.com)
- Co-Chair: Liuquan Sun (2021, slq@amt.ac.cn)
- Members:
  - Niansheng Tang (2019-2021, nstang@ynu.edu.cn)
  - Lei Shen (2019-2021, shen_lei@lilly.com)
  - Shuwei Li (2020-2022, lishuwstat@163.com)
  - Yifei Sun (2021-2023, ys3072@cumc.columbia.edu)

IT Committee:

- Chair: Chengsheng Jiang (2021, website@icsa.org)

Archive Committee':

- Chair: Xin Tian (2019-2021, tianx@nhlbi.nih.gov)
- Members:
  - Xin (Henry) Zhang (2021-2023, henry@stat.fsu.edu)
  - Naitee Ting (2021-2023, naitee.ting@boehringer-ingelheim.com)

Finance Committee:

- Chair: Rochelle Fu (2019-2021, fur@ohsu.edu)
- Members:
  - Hongliang Shi (hongliangshi15@gmail.com)
  - Rui Feng (ruifeng@pennmedicine.upenn.edu)
  - Xin He (2021 ICSA Applied Statistics Symposium treasurer, xinhe@umd.edu)

Financial Advisory Committee:

- Chair: Rui Feng (ruifeng@upenn.edu)
- Members:
  - Hongliang Shi (hongliangshi15@gmail.com)
  - Nianjun Liu, (liunian@indiana.edu)
  - Xiangqin Cui (xiangqin.cui@emory.edu)
  - Fang Chen (FangK.Chen@sas.com)
  - Rochelle Fu (fur@ohsu.edu)

Lingzi Lu Award Committee (ASA/ICSA):

- Chair: Chan, Ivan (2019-2021, ivan.chan@abbvie.com)
- Members:
  - Jichun Xie (2018-2020, jichun.xie@duke.edu)
  - Shelly Hurwitz (2017-2022, hurwitz@hms.harvard.edu)
  - Laura J Meyerson (2020-2022, laurameyerson@msn.com)

ICSA Representative to JSM Program Committee:

- Aiyi Liu (2021, liua@mail.nih.gov)
- Pei Wang (2022, pei.wang@mssm.edu)

AD HOC COMMITTEES:

- 2021 ICSA Applied Statistics Symposium: Guoqing Diao (Chair, gdiao@email.gwu.edu)
- 2021 ICSA China Conference: Yingying Fan (Chair, fanyingy@usc.edu); Chunjie Wang (Co-Chair, wangchunjie@ccut.edu.cn)
- 2022 ICSA Applied Statistics Symposium: Samuel Wu (Chair, samwu@biostat.ufl.edu)
- CHAPTERS
  - ICSA-Canada Chapter: Liqun Wang (Chair, Liqun.Wang@umanitoba.ca)
  - ICSA-Midwest Chapter: Li Wang (Chair, li.wang1@abbvie.com)
  - ICSA-Taiwan Chapter: Chao A. Hsiung (Chair, hsiung@nhri.org.tw)

ICSA Financial Report

Profit and Loss: July 1, 2020 through December 31, 2020

Beginning Cash Balance (Bank/Paypal accounts)	**7/1/2020**	**$**	**627,071.28**
Income:		**$**	**9,190.87**
Membership from Paypal Account		$	4,360.00
Membership from Institute of Mathematical Statistics		$	180.00
Springer		$	2,500.00
Job Posting		$	1,032.00
Interest		$	118.87
Donation to Peter Hall Lecture		$	1,000.00
Total Income		**$**	**9,190.87**
Expense:		**$**	**(8,894.38)**
ICSA Old Website Hosting Fee from 08/2016 to 07/2020		$	(1,250.34)
IT Cost		$	(4,556.67)
ICSA PO Box Reservation Fee		$	(226.00)
Statistica Sinica Mailing Fee		$	(1,073.76)
Tax Filing Fee		$	(1,294.00)
Sympathy Spray		$	(296.79)
Mailing Postage		$	(28.35)
Paypal Fee		$	(168.47)
Total Expense		**$**	**(8,894.38)**
Net Total Income		**$**	**296.49**
Transfer			
To Vanguard Investment Account		**$**	**(120,000.00)**
Ending Cash Balance (Bank/Paypal accounts)	**12/31/2020**		**$507,367.77**

ASSETS		
Main Checking/Savings/PayPal	$	507,367.77
Vanguard Investment Balance	$	524,890.52
TOTAL ASSETS	**$**	**1,032,258.29**
LIABILITIES & EQUITY		
Equity		
Main Accounts Opening Balance July 1, 2020	$	627,071.28
July 1 to December 31, 2020 Net Income(+)/Expense(-)	*$*	*296.49*
Transfer to Vanguard Investiment Account	*$*	*(120,000.00)*
2020 Symposium Bank Accounts Opening Balance July 1, 2020	$	90,550.01
July 1 to December 31, 2020 Net Income(+)/Expense(-)	*$*	*20,742.86*
Vanguard investment account opening balance on July 1, 2020	$	337,005.92
Transfer from Main Bank Account	*$*	*120,000.00*
July 1 to December 31, 2020 Investment Profit(+)/Loss(-)	*$*	*67,884.60*
Total Equity	**$**	**1,143,551.16**
TOTAL LIABILITIES & EQUITY	**$**	**1,143,551.16**

Rongwei (Rochelle) Fu, PhD
Treasurer (2021-2023), ICSA
Professor
OHSU-PSU School of Public Health

2020 ICSA Applied Statistics Symposium: Summary Report

Executive Organizing Committee

Time and location: The 2020 ICSA Applied Statistics Symposium was originally planned at the Westin Galleria Houston Hotel at Houston, Texas on May 17-20, 2020. It was postponed to December 13-16, 2020 online using the Cvent virtual conference platform.

Executive Organizing Committee:

- Chair: Hulin Wu, University of Texas Health Science Center at Houston (UTHealth)
- Scientific Program Co-Chairs:
  Momiao Xiong , UTHealth
  Jianhua Huang , Texas A&M University
- Poster Session Committee Chair: Xi Luo, UTHealth
- Program Book and Website Committee Co-Chairs: Yunxin Fu and Ashraf Yaseen, UTHealth
- Local Committee Chair: Hongyu Miao, UTHealth
- Treasurer: Dejian Lai, UTHealth
- Student Paper Competition Committee Co-Chairs:
  Ruosha Li, UTHealth
  Jing Ning, University of Texas MD Anderson Cancer Center (MDA)
- Short Course Committee Chair: Wenyi Wang, MDA
- Fund Raising Committee Chair: Rui (Sammi) Tang, Servier Pharmaceuticals
- Talent Show Committee Chair: Kelly H. Zou, Viatris
- Strategic Advisors:
  Zhezhen Jin, Columbia University
  Wenbin Lu, North Carolina State University
  Lanju Zhang, Abbvie Inc
- Conference Secretary: Gen Zhu, UTHealth
- ICSA Leaders:
  ICSA President: (Tony) Jianguo Sun
  ICSA Board Executive Director: Mengling Liu
  ICSA IT Support: Chengsheng Jiang

Summary statistics

- 635 people registered: 572 attended the symposium
- 3 keynote sessions
- 2 panel discussion sessions during the lunch break
- 355 invited talks in 92 invited sessions
- 50 posters 50 in 7 poster sessions: 3 poster awards
- Received 73 student papers for student paper awards: 6 student paper awards
- 9 short courses (4 half-day short courses and 5 full-day courses): 164 people registered
- 9 sponsors (4 gold sponsors, 3 silver sponsors and 2 special sponsors): 31.5k
- 103 people attended Tuesday night "ICSA General Member Meeting, Award Ceremony and Talent Show."

Student paper awards: 6 awards

- Xinyue Qi, Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston. "*Bayesian Meta-analysis of Censored Rare Events with Stochastic Coarsening.*"

- Xinjun Wang, Department of Biostatistics, University of Pittsburgh. "*BREM-SC: A Bayesian Random Effects Mixture Model for Joint Clustering Single Cell Multi-omics Data.*"
- Zhengjia Wang, Department of Statistics, Rice University. "*Functional Group Bridge Regression with Application to iEEG Data.*"
- Yizhen Xu, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health. "*Inference for BART with Multinomial Outcomes.*"
- Huijuan Zhou, Institute of Statistics and Big Data, Renmin University of China & Department of Statistics, Texas A&M University. "*Covariate Adaptive Family-wise Error Rate Control for Genome-Wide Association Studies.*"
- Peng Jin, Division of Biostatistics, New York University Grossman School of Medicine. "*Generalized Mean Residual Life Models for Case-Cohort and Nested Case-Control Studies.*"

Poster Awards: 3 awards

- Yichen Jia, Department of Biostatistics, University of Pittsburgh. "*Deep Learning for Quantile Regression: DeepQuantreg.*"
- Dayu Sun, Department of Biostatistics, Emory University. "*Semiparametric Maximum Likelihood Estimation of Panel Count Data with Time-dependent Covariates.*"
- Jing Zhang, Dept of Biostatistics and Data Science, The University of Texas Health Science Center at Houston. "*On the Time-varying Predictive Performance of Longitudinal Biomarkers: Measure and Estimation.*"

Short Courses	Instructors	Email
A Short Course on Absolute Risk Prediction (Full Day)	Mitchell H. Gail	gailm@mail.nih.gov
	Ruth Pfeiffer	pfeiffer@mail.nih.gov
Empower Statistician with Spark, Machine Learning and Deep Learning (Full Day)	Ming Li	mli@alumni.iastate.edu
	Hui Lin	hui@linhui.org
Statistics and Machine Learning Methods for EHR Data: From Data Extraction to Data Analytics/Predictions (Full Day)	Hulin Wu	Hulin.Wu@uth.tmc.edu
	Vahed Maroufy	Vahed.Maroufy@uth.tmc.edu
	Ashraf Yaseen	ashraf.yaseen@uth.tmc.edu
Statistical Analysis of Microbiome Data with R (Full Day)	Yinglin Xia	yxia@uic.edu
	Ding-Geng Chen	dinchen@email.unc.edu
Utilizing Real-World-Data and Real-World-Evidence in Drug Development and Evaluation (Full Day)	Binbing Yu	binbing.yu@astrazeneca.com
	Bo Lu	lu.232@osu.edu
	Qing Li	qing.li2@takeda.com
Multivariate meta-analysis methods (Half Day – morning)	Haitao Chu	chux0051@umn.edu
	Yong Chen	ychen123@upenn.edu
Including historical data in clinical trial design and analysis (Half Day – morning)	Frank Fleischer	frank.fleischer@boehringer-ingelheim.com
	Martin Oliver Sailer	martin_oliver.sailer@boehringer-ingelheim.com
Estimands and Statistical Methods for Missing data in Clinical Trials (Half Day – afternoon)	Guanghan (Frank) Liu	Guanghan_frank_liu@merck.com
	Man (Mandy) Jin	manmandy.jin@abbvie.com
Statistical Remedies for Flawed Conventions in Medical Research (Half Day – afternoon)	Peter F. Thall	rex@mdanderson.org

Nine Short courses

Financial summary: Dejian Lai

- Incomes: Total $138,000
  i. ICSA initial deposit: 5k
  ii. Short course registration fee: 41.5k
  iii. Conference registration fee: 60k
  iv. Sponsors: 31.5k
- Estimated Expenses: Total $52,500
  i. Virtual conference platform (Cvent) Cvent: 15k
  ii. Honorarium to short course instructors: 11.5k
  iii. Honorarium to three keynote speakers: 3k
  iv. Payment to student paper awards and poster awards: 5k?
  v. Conference software purchase: 1k
  vi. Conference bag, souvenir etc: 11k
  vii. Committee and volunteer appreciation diner (coupons): 6k
  viii. Other expenses?
- Estimated Balance: $85,500 (current balance 115k + 10k from hotel)

Call for Nominations for 2022 ICSA Officers

Deadline: March 31, 2021

The ICSA 2021 Nomination for Election Committee is seeking nominations for ICSA 2022 officers: two (2) candidates for ICSA President-Elect 2022 and twelve (12) for ICSA Board of Directors. Candidates for all positions need to be active ICSA members in 2021 and have strong interests in serving ICSA. According to the ICSA Bylaws, President-Elect should be from academia, non-academia, or no restriction, on a three-year rotational basis-one year from academia, another from non-academia, and the third year open. So the candidates will be from academia this year. We hope that the candidates for Board of Directors are balanced in gender, region, professional interests, and sector of primary employment (academia, industry/business, or government). Please email your nominations to Dr. Ming Tan (ming.tan@georgetown.edu) with the subject: ICSA Nomination by March 31, 2021.

Call for Nominations for 2021 ICSA Awards

Distinguished Achievement Award

This award is in recognition of the distinguished achievement in statistical research and unselfish support of the association. *More details will be announced by ICSA at https://www.icsa.org/distinguished-achievement-award/*

Most recent winner: Ming-Hui Chen, University of Connecticut.

Outstanding Young Researcher Award

This award is in recognition of the outstanding research in statistical theory, methodology, and/or applications. *More details will be announced by ICSA at https://www.icsa.org/awards/outstanding-young-research-award/*

Most recent winners:

- Guan Yu, University of Buffalo.
- Xin Zhang, Florida State University.

Pao-Lu Hsu Award

The Pao-Lu Hsu Award is presented every three years by the International Chinese Statistical Association (ICSA), usually at an ICSA conference, to an individual under age 50, who makes influential and fundamental contributions to any field of statistics and probability, and exemplifies Hsu' s deep involvement in developing statistics and probability research with significant impact on education.

Professor Hsu, born in 1910, was a pioneer and founder of the newly formed discipline of statistics and probability in China. He was best known for his rigorous research with depth and breadth, and for his profound impact on younger generations. See "Pao-Lu Hsu Memorial Collection" published by Peking University Press for more details.

The prize is open to all nationalities. Priorities are given to the candidates whose work contributes greatly to the research and education of Chinese statisticians. The award recipient will speak at an ICSA International Conference. The award includes $3000 in cash prize. *Deadlines will be announced by the ICSA and more details could be found at https://www.icsa.org/about-pao-lu-hsu-award/*

Most recent winner: Hongyu Zhao, Yale School of Public Health.

ICSA Student Travel Awards and Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award

To encourage student members of ICSA to participate and to share their research at the annual ICSA applied Symposium, ICSA offers Student Travel Awards and one Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award for outstanding student papers.

The Student travel award is sponsored by the Annual ICSA Applied Statistics symposium. The main purpose of the award is to encourage student members of ICSA to participate and present their research work at the annual symposium. Travel award winners will be selected by the Awards Committee annually.

The recipients of the "ICSA Student Travel Award" will present the award papers at an invited paper session entitled "ICSA Student Award Session" during the symposium. Each recipient will also receive a plaque commemorating the event and the same monetary sum as the Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award. In normal situations, up to six student award winners (five Student Travel Awards, one Jiann-Ping Hsu Pharmaceutical and Regulatory Sciences Student Paper Award) will be selected. Each winner will receive a plaque, an award for travel and registration reimbursement up to $1000 or a cash award of $550, whichever is bigger, as well as a free registration for a short course. *More details will be announced at* *https://www.icsa.org/awards/icsa-awards-and-honors/*

Most recent winner: Announced at the summary report of 2020 ICSA Applied Statistics Symposium.

Report from the ICSA Springer Book Series in Statistics

Ding-Geng Chen

ICSA Book Series in Statistics (ISSN:2199-0980) was established in the year 2012 between ICSA and Springer. With this initiative, Professor Jiahua Chen was appointed as the editor of this book series. In 2014, Professor Ding-Geng Chen joined with Professor Jiahua Chen as co-editorship for this book series.

From 2012 to 2020, this book series has published twenty-two (22) books in statistics, biostatistics, bioinformatics, biopharmaceutical biostatistics, data sciences, and public health, as listed online at https://www.springer.com/series/13402. In 2020, Professor Jiahua Chen stepped down as the editor after his successful five-year editorship, and Professor Ding-Geng Chen is the current editor of the ICSA Book Series in Statistics.

The ICSA Book Series in Statistics is aimed to showcase research from the International Chinese Statistical Association that has an international reach. It publishes books on statistical theory, applications, and statistical education. All books are associated with the ICSA or are authored by invited contributors. Books may be monographs, edited volumes, textbooks, and proceedings.

To all ICSA members, you and your colleagues are professionally welcome to contribute to this book series to make it successful to our International Chinese Statistical Association. Please contact Professor Ding-Geng Chen at dinchen@email.unc.edu for your interest.

Ding-Geng Chen , PhD
ASA Fellow, Wallace H. Kuralt Distinguished Professor,
University of North Carolina, Chapel Hill, NC 27517, USA.

A Memorial Article for Dr. Xiangrong Yin

John Stufken,
Director of Informatics and Analytics,
Bank of America Excellence Professor,
UNC Greensboro.

While about 4 months have passed since the sad news, it is still hard to believe that Xiangrong is no longer with us. The only thing that helps a little is that I have many fond memories of Xiangrong.

1. We met in 2003 when I joined the University of Georgia as Head of the Department of Statistics and were colleagues for 11 years until we both left UGA in 2014.

2. Xiangrong liked to call me his boss, but in some ways, through the ideas that he planted in me, he was maybe my boss!

3. With Xiangrong you didn' t have to guess what he thought about an issue; he would let you know. He wore his emotions on his sleeve. He loved to engage in passionate discussions about statistics, research, and professional issues. He had strong opinions about many things, and we didn't always agree, but that made the discussions interesting.

4. I had many opportunities to have such discussions with him, such as at the many parties that Xiangrong and his wife, Xiaofang, held at their home, or while playing racquetball, or during our joint trip to China in 2007 to Shanghai, Hangzhou, and his hometown of Changxing-an unforgettable trip.

5. Professionally, he was very passionate about his research, working incredibly hard, seven days per week. And he built a great research record, which is why he was elected to ASA and IMS Fellowship last year. I prepared the nominations and the letters that I collected from prominent researchers spoke to the groundbreaking impact of Xiangrong's research.

6. He was especially passionate about his students and built an incredibly strong record of supervising students. He could be tough on them but was also very supportive and proud of them.

7. His passion was also visible and audible in the classroom. It was hard to have a meeting next to a classroom where Xiangrong was teaching!

My last correspondence with Xiangrong was about a week before he died. He had just accepted to give an online seminar for us at UNCG, and he thanked me for the invitation. I wrote to him that he and Xiaofang were welcome to visit Lili and me, but he wrote back that traveling was no longer fun because he was becoming old. We have lost a very special and loyal friend, far too early. We must remember Xiangrong for the infectious passion that he brought to everything he did in life, and that is the best way to honor him. He was truly an inspiration for me and, I know, for many others.

Lastly, if you go to JSM in 2021, there is an invited memorial panel for Xiangrong, which is organized by his former University of Kentucy student, Jiaying Weng. The scheduled panelists include: Cook, Dennis (University of Minnesota); Stufken, John (UNC Greensboro); Li, Bing (Penn State University); Stromberg, Arnold (University of Kentucky); Harrar, Solomon (University of Kentucky).

There is a poem that Dr. Xiaogang Su, Professor in the Department of Mathematical Sciences from University of Texas at El Paso, wrote for memorizing Dr. Yin.

【吊殷向荣教授】- 悼亡诗

Xiaogang Su 12/16/2020 于 ICSA 年会

万湖之州双子城，月冷窗寒雪色浑。
精诚深得名师器，如切如磋立程门。
博观约取积已厚，一朝引吭众人惊。
声名鹊飞东南起，文章喷薄气如奔。
大浪淘淘中流砥，高山仰止雪分明。
后来形迹亦汗漫，接物始终善也温。
灯前伏案年共岁，秋往冬复石与金。
桃李依依春雨细，世事浮沉长者心。
白驹骎骎何太遽！况似使君跨鹤身？
宛然音容犹历历，念之使人泪倾盆。

Bin Yu's Interview

Jon Bashor

Editorial: This interview written by Jon Bashor is reprinted with permission from the website of UC Berkeley Division of Computing, Data Science and Society (CDSS): https://data.berkeley.edu/people/bin-yu. More information about the PCS framework mentioned in this article can be found from

- Veridical data science (PCS framework, https://www.pnas.org/content/117/8/3920), PNAS, 2020. (QnAs with Bin Yu, https://www.pnas.org/content/117/8/3893).
- Breiman Lecture (video) at NeurIPS "Veridical data Science" (PCS framework and iRF, https://slideslive.com/38922599/veridical-data-science), 2019; updated slides (https://www.stat.berkeley.edu/~binyu/ps/papers2020/Breiman19-NeurIPS-yu.pdf), 2020.

Over the past 30 years, UC Berkeley statistics Professor Bin Yu has covered a lot of territory, both in her research field and in sharing her knowledge with others. And now she has used it to create a framework that she believes will lead to a more rigorous and trustworthy data science process, including the use of methods such as machine learning.

Yu and her team at Berkeley have developed novel statistical machine learning approaches and are combining their work with the domain expertise of collaborators to solve important problems in the fields of neuroscience, genomics and precision medicine.

"Artificial intelligence has huge potential to help us solve critical problems," said Yu, who is also a professor in UC Berkeley's Department of Electrical Engineering and Computer Science and the Division of Computing, Data Science, and Society (CDSS). "But there is a lot of misunderstanding as well. We need to have realistic optimism."

Yu points to applications ranging from self-driving vehicles to precision medicine to medical imaging where AI is seen by many as the answer to the problems. The challenge of creating safe self-driving cars can be seen in the number of startups in the field that have folded, she added.

Behind all of these efforts is the discipline of data science, which Yu describes as a "field of evidence seeking that combines data with information from a research domain information to generate new knowledge," Yu wrote. It' s this process that she wants to make more consistent and trustworthy.

Yu laid out her framework for integrating predictability, computability and stability, which she calls PCS, in her paper "Veridical data science" (link is external) co-authored with her former student Karl Kumbier (now a postdoc at UCSF) and published in the Proceedings of the National Academy of Sciences in February 2020.

She and Rebecca Barter, Yu' s former student and current postdoc, are now adapting the material into a textbook to be published by the MIT Press in 2021. They also plan to make the material available online at no cost. She adds that "veridical" refers to something that is truthful or coincides with reality.

Yu was inspired to develop the PCS framework as a result of her interdisciplinary research projects with the Gallant Lab in neuroscience on the Berkeley campus and the Celniker and Brown Labs in genomics at Lawrence Berkeley National Laboratory. Recently, the PCS framework successfully guided the development of novel statistical pipelines epiTree and staDISC by the Yu Group and collaborators to recommend possible genetic drivers of disease and subgroups of people for which a particular treatment is effective, respectively. The projects are part of a research program to advance precision medicine.

The key, she says, is to look at the use of data science as a cycle, not a set of linear steps. In this scenario, the cycle of steps begins with the posing of a science question in a particular domain, and proceeds through collecting, managing, processing (or cleaning), exploring, modeling, and interpreting data results to guide new actions.

"We need to look at the whole data science life cycle and make sure it is trustworthy," Yu said. "Every step needs to be vetted."

According to Yu, since data science typically crosses over multiple research disciplines, it requires human involvement from experts who understand both the domain and the tools used to collect, process, and model data.

"These individuals make implicit and explicit judgment calls throughout the data science life cycle," she said. "Since there is limited transparency in reporting these judgment calls, the evidence behind many analyses is blurred and we are seeing more false discoveries than might otherwise occur."

She describes PCS as a conceptual framework for asking critical questions and documenting them at

every step of the data science life cycle. In fact, she sees an important role in data science for the role of critical thinking as it is taught in the liberal arts.

"The first step is to make an argument to yourself, then make an argument to the reader as to why your thinking is sound, making the process transparent," Yu said. "Documenting these steps is integral to the process. You need to make a concise summary of why the work is responsible, reliable, reproducible and transparent."

The core principles of predictability, computability, and stability are the basis for such a unified data analysis framework, which builds and expands on principles of statistics, machine learning, and scientific inquiry. Yu points out that many of the ideas embedded in PCS have been widely used across various areas of data science and sees them as the minimum requirements for achieving her goal of veridical data science.

In other words, PCS synthesizes, streamlines and expands on these ideas as an accessible protocol or pipeline to share best practices for a quality-controlled data science life cycle. At the same time, it emphasizes the importance of domain knowledge and critical thinking and communication skills of a data scientist.

Even though many research fields have successfully embraced artificial intelligence (AI), Yu notes that there is still a lot about AI and machine learning that is not understood. She is a co-principal investigator on a Berkeley-led 10 million project funded by the National Science Foundation and Simons Foundation to gain a theoretical understanding of deep learning–how it works and why it works.

Why now? Yu said she made the decision to push forward her PCS framework for both personal and professional reasons. About the same time she hit a milestone age of 50 years, her mother became seriously ill. So, she wanted to make a statement about an issue she feels strongly about. Although she was eligible to submit a paper to the Proceedings of the National Academy of Sciences upon her election in 2014, she deliberately took her time to refine and polish her ideas before publishing them.

Her work draws on a wide range of perspectives she has gained. Since earning her bachelor's degree in mathematics at Peking University, Yu went on to earn her M.S. and Ph.D. degrees in statistics from UC Berkeley. In addition to working for two years at the Bell Labs based in New Jersey, she has been a professor at the University of Wisconsin-Madison, visiting professor at Yale University and a visiting faculty member at MIT, ETH (the Swiss Federal Institute of Technology in Zurich), Poincare Institute in Paris, Peking University, Inria-Paris (the French Institute for Research in Computer Science and Automation), Fields Institute at University of Toronto, Newton Institute at Cambridge University, and Flatiron Institute in New York. Yu is an investigator with the Chan-Zuckerberg Biohub and the Weill Neurohub. She is also a member of the American Academy of Arts and Sciences.

Stability Expanded, in Reality

Bin Yu

It is thought-provoking to read the pair of articles on 10 challenges in data science by Xuming He and Xihong Lin from a statistics perspective and Jeannette Wing from a computer science perspective. Unsurprisingly, there is a good overlap of important topics including multimodal and heterogenous data, data privacy, fairness and interpretability, and causal inference or reasoning. This overlap reflects and confirms the foundational and shared roles of statistics and computer science in data science, which is the merging of statistical and computing thinking in the context of solving domain problems. The challenges in both articles are presented as separate, not integrated, topics, and mostly decoupled from domain problems, possibly because of the mandate of "10 challenges."

In my mind, the most exciting 10 challenges in data science are to solve 10 pressing real-world data problems with positive impacts. For example, how is data science going to help control covid-19 spread while allowing a healthy economy? To mitigate climate change so that its negative impact on human and economics can be minimized and in time? To bring precision medicine to every patient safely and

timely? To unlock the mysteries of the unconscious brain? To design genomic therapies for Alzheimer's? To design wearables that interact with multiple sclerosis patients to keep them safe? To help discover chip materials for the next generation of computers? To understand the origins of universe? To prevent cyberattacks on democracies all over the world? To self-regulate interactions of digital media with kids? To help people retool skills needed by the rapidly changing economy while allowing them to stay in familiar physical environments of friends, families, mountains, and rivers? Such real-world problems have to be the mission, the anchor, and the goal of data science, while methodologies/algorithms, approaches, and theories have to be at their service and appraised relative to how well they help solve them.

To solve any of these 10 real-world challenges and more, an integrated- and system-framing of data science needs to be embraced. Real-world data science problems are multidisciplinary, multidimensional, and multiphased. Each data science life cycle (DSLC) consists of domain problem formulation, data collection, data cleaning/preprocessing, visualization, analytical problem formulation/modeling, interpretation, evaluation/validation, data conclusions and decisions, and communication of decisions and conclusions. The steps are not at all linear but nonlinear and iterative. The challenges in He and Lin (this issue) fall mostly in the analytical problem-formulation or modeling stage and some on data preprocessing and one on issues in decision making. They do not touch other important steps such as data cleaning, problem formulation, and communication of decisions. Wing (this issue) covers emerging conceptual topics such as trustworthy AI and automating data preparation/preprocessing. Even though I believe some automation in the data cleaning step is necessary, I believe humans have to be in the loop to monitor, check, and make judgment calls in ambiguous situations flagged by machines. That is, I see a human–machine collaboration future, not automation, for "front-end stages of the data life cycle" (Wing 2020).

The challenges in both articles are important, yet incomplete, components of a data science life cycle or system. Unless the entire system or all the components are integrated and connected together and owned as the traditional topics, there is no insurance that real-world problems such as the 10 challenges above will be solved with positive impacts. In particular, neither article recognizes the many human judgment calls in DSLC or discusses the stability or robustness or reproducibility issues in, say, the choices of data leaning and algorithm in solving a data problem. Data cleaning/preprocessing and coding irreproducibility has led to grave consequences in the past. An article called "Growth in a Time of Debt" was published by economists Carmen Reinhart and Kenneth Rogoff (2010). They concluded that public debt is not good for growth. Such a conclusion was widely used as evidence to argue for austerity policies in Europe and the United States after the 2008 financial crisis. Four years later, Thomas Herndon, Michael Ash, and Robert Pollin (2014) invalidated this conclusion when they included the few data points from New Zealand and corrected the coding errors. (It is not clear why these data points were omitted in the first place.)

When we embrace the data science life cycle as a system, it is clear that the elephant in the room is the human judgment calls made in every step. That is, stability (or robustness) relative to reasonable or appropriate perturbations to the system, including human judgment calls on data-cleaning choices, data perturbation, and model choices, has to be among the core considerations and a key metric for success. This is to makes sure that these perturbations and judgment calls are not driving the data conclusions and decisions, unless justified with well-explained documents. Equally important is to ensure a reality check through prediction into the future (or its good surrogate). Stability is a fundamental and commonsense principle in knowledge seeking and decision making. In fact, when I asked philosopher colleague Branden Fitelson at Northeastern whether considerations of stability of belief/judgment go back to the Greeks, his answer was an affirmative yes and he pointed me to Plato's quotes, here.

In the Meno, Plato writes:

For true opinions, as long as they remain, are a fine thing and all they do is good, but they are not willing to remain long, and they escape from a man's mind, so that they are not worth much until one ties them down ... That is why knowledge is prized higher than correct opinion, and knowledge differs from correct opinion in being tied down.

And, in Protagoras, Plato writes:

[K]nowledge is something noble and able to govern man, and that whoever learns what is good and what is bad will never be swayed by anything to act otherwise than as knowledge bids, and that intelligence is a sufficient succor for mankind.

Fitelson also told me, "Hume was one of the first

to emphasize that even (mere) belief needs to be stable (if it is to guide action in the right ways, etc.). Much of the contemporary work has shifted to arguing that even (mere) belief must also be stable in various ways, in order to perform its functions." (For more information on stability of belief, please see Leitgeb, 2017).

In order for a data science life cycle to "perform its functions" and "guide action in the right ways," say, to find a gene therapy for Alzheimer' s, the DSLC process has to be stable and capture reality in the data and neuroscience. Predictability (reality check), stability, and computability were argued as the three pillars to support the PCS (predictability, computability, stability) framework for veridical data science (Yu, 2013; Yu & Kumbier, 2020). The PCS framework bridges Breiman' s two cultures. It unifies and expands on ideas from machine learning (P and C) and statistics (P and S). Stability in PCS is a significant expansion on the concept of sample-to-sample variability in statistical uncertainty assessment and robust statistics to the entire DSLC including linguistic stability of the same word meaning the same thing for a multidisciplinary team. The PCS framework contains PCSF workflow and PCS documentation on GitHub in R Markdown or Jupyter Notebook to record human judgment calls and choices in the DSLC.

PCS was motivated and developed in the context of multidisciplinary projects in neuroscience and genomics. It has led to the developments of cutting-edge statistical machine learning algorithms ESCV (estimation stability with cross-validation) for Lasso model selection (Lim & Yu, 2015), staNMF for statbility-driven NMF (nonnegative matrix factorization) (Wu et al., 2016), iterative random forests (iRF) for predictive and stable discovery of high-order Boolean interactions (Basu et al., 2018), and DeepTune for visually characterizing V4 neurons (Abbasi-Asl et al., 2018) (corresponding codes can be found at `https://www.stat.berkeley.edu/~yugroup/code.html`). A recent article of ours (Dwivedi et al., 2020) articulated PCS in the context of causal inference to propose staDISC (stable discovery of subgroups via calibration). It is the first to propose a general model-checking device in causal studies, or calibration as reality checking, as an implementation of P from PCS. Simultaneous to the development of statDISC, we reanalyzed the 1999–2000 VIGOR study, which is an 8,076-patient randomized controlled trial that compared the risk of adverse GI and TC events from a then newly approved drug, rofecoxib (Vioxx), to that from an older drug, naproxen. StaDISC found a subgroup of patients with a prior history of GI events not only has a disproportionately reduced risk of GI events but also does not experience an increased risk of TC events. Building and employing the PCS framework, my group has had very fruitful outcomes in solving multidisciplinary data problems and developing new general machine learning methodologies. I hope other teams join us in using it in their data science projects and developing if further together.

Finally, I believe a healthy and imperative criterion for designing a new data science algorithm, concept, or framework is to make a serious attempt at solving at least one new data problem as we did in developing algorithms such as iRF. It is a disturbing problem and wasteful of human and computing resources that in statistics, machine learning, or data science, we have way too many new algorithms (and way too many articles) relative to the new data problems that we solve. To solve real-world problems most efficiently from the point of view of society, the reward system in academia needs revamping so that research quality and positive impact are more valued and better incentivized. I believe that, if we willing to improve our reward system, and if we take on the real-world data challenges, embrace reality-check and stability considerations in the entire DSLC, we stand a much higher chance to meet the challenges outlined in the pair of articles by He and Lin, and Wing, respectively.

Disclosure Statement

Partial supports are gratefully acknowledged from ONR grant N00014-17-1-2176, NSF grants DMS-1613002, and IIS 1741340, and the Center for Science of Information (CSoI), a US NSF Science and Technology Center, under grant agreement CCF-0939370.

Acknowledgments

The author would like to thank Dr. Xiao-Li Meng for helpful comments on a draft of this discussion and Branden Fitelson for the permission to quote him.

References

1. Abbasi-Asl, R., Chen, Y., Bloniarz, A., Oliver, M., Willmore, B. D. B., Gallant, J. L., & Yu, B. (2018). The DeepTune framework for modeling and characterizing neurons in visual cortex area V4. bioRxiv. `https://doi.org/10.1101/465534`

2. Basu, S., Kumbier, K., Brown, J. B., & Yu, B. (2018). Iterative random forests to discover

predictive and stable high-order interactions. Proceedings of the National Academy of Sciences, 115(8), 1943–1948.

3. Dwivedi, R., Tan, Y., Park, B., Wei, M., Hogan, K., Madigan, D., & Yu, B. (2020). Stable discovery of interpretable subgroups via calibration in causal studies (staDISC). arXiv. https://arxiv.org/abs/2008.10109

4. Herndon, M., Ash, T., & Pollin, R. (2014). Does high public debt consistently stifle economic growth? A critique of Reinhart and Rogoff. Cambridge Journal of Economics, 38(2), 257–279.

5. Leitgeb, H. (2017). The stability of belief: How rational belief coheres with probability. Oxford University Press.

6. Lim, C., & Yu, B. (2016). Estimation stability with cross-validation (ESCV). Journal of Computational and Graphical Statistics, 25(2), 464–492.

7. Plato. Meno. Translated by Benjamin Jowett. Retrieved from http://classics.mit.edu/Plato/meno.html

8. Plato. Protagoras. Translated by Benjamin Jowett. Retrieved from http://classics.mit.edu/Plato/protagoras.html

9. Reinhart, C., & Rogoff, K. (2010). Growth in a time of debt. American Economic Review: Papers and Proceedings, 100(2), 573–578.

10. Wu, S., Joseph, A., Hammonds, A., Celniker, S., Yu, B., & Fris, E. (2016). Stability driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. Proceedings of the National Academy of Sciences, 113(16), 4290–4295.

11. Yu, B. (2013). Stability. Bernoulli, 19(4), 1484–1500.

12. Yu, B., & Kumbier, K. (2020). Veridical data science. Proceedings of the National Academy of Sciences, 117(8), 3920–3929.

Bin Yu, PhD
Chancellor's Distinguished Professor,
Department of Statistics,
Department of Electrical Engineering & Computer Science,
University of California Berkeley.

Terence's Stuff: Assumptions

Terry Speed

Editorial: This is a reprint from a Terry Speed's column published in the *IMS website* https://imstat.org/2016/11/17/terences-stuff-assumptions/) with IMS's permission.

Many of us were brought up to think that we cannot do a statistical analysis without making assumptions. We were taught to pay attention to assumptions, to consider whether or not they are approximately satisfied with our data, and to reflect on the likely impact of a violation of our assumptions. Our assumptions might concern independence, identical distribution, normality, additivity, linearity or equality of variances. (Needless to say, we have some tests for all these assumptions.)

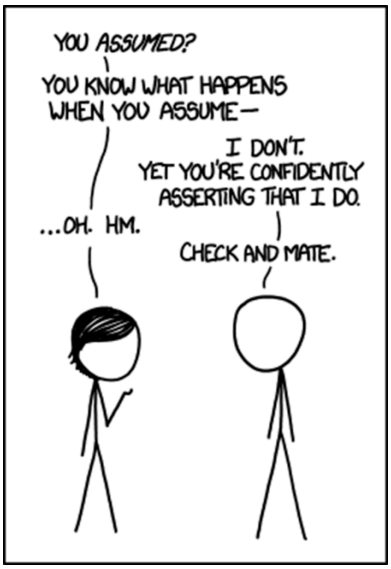
I see much of this concern with assumptions as coming from the desire to reduce statistics to something close to mathematics. A typical mathematical theorem starts with hypotheses (e.g. axioms), and proceeds through an accepted process of reasoning, to a conclusion. If your hypotheses are true, and your reasoning is correct, then your conclusions will be true. The statistical version of this is that we have data and a question about the real world. We take a statistical model involving statistical assumptions, we make use of it and some statistical methods with our data, and we draw statistical conclusions about the real world. Our hope is that if our statistical assumptions are true, and our statistical methods are correctly applied, then our statistical conclusions about the real world will be true.

This analogy between statistical and mathemat-

ical reasoning is flawed, and not just because statistics involves uncertainty. Mathematical truth and real world truth are very different things. Data cannot always be understood by making statistical assumptions; sometime we need to look at it and think hard about it in its context. Only rarely are all our assumptions testable, and often some will be implicit-we don't even notice we are making them. J.W. Tukey has highlighted one: a belief in omnicompetence, the ability of a methodology to handle any situation. Also, seeking real world truth with (near) certainty is just one goal; there are many other performance metrics for statistical analyses. Most of us will be familiar with the effective use of a statistical model or method, even though the standard assumptions that go with it are transparently violated. This would just be a paraphrase of George E.P. Box' s aphorism on models and usefulness. Assumptions don't always matter. In this context, Tukey strikingly observed that, "In practice, methodologies have no assumptions and deliver no certainties." He goes on to say, "One has to take this statement as a whole. So long as one does not ask for certainties one can be carefully imprecise about assumptions."

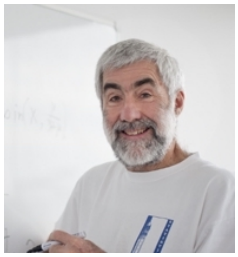
But sometimes assumptions really do matter. Many of you will know Nate Silver' s The Signal and the Noise, in which he discusses the global financial crisis, among other topics. He explains how the financial services company Standard and Poor' s calculated the chance that a collateralized debt obligation would fail over a five-year period to be 0.12%, whereas around 28% of them failed. As a result, Silver writes, "trillions of dollars in investments that were rated as being almost completely safe instead turned out to be almost completely unsafe." The problem was an inappropriate assumption of independence, one that could have been foreseen, had the matter been carefully considered.

Tukey is critical of talk of "checking the assumptions," saying that he suspects much of it is being used as an excuse for "looking to see what is happening," something that he applauds. He also gives four routes to understanding when a methodology functions well, and how well it functions, saying, "We dare not neglect any of them." On the other hand, Silver gives examples from weather forecasting, earthquake prediction, epidemic and climate modelling where assumptions have a big impact on the predictions. Where does that leave us?



I think we all know that (with Silver) sometimes particular assumptions matter a lot, and that (with Tukey) sometimes they don' t matter much at all. Learning how to distinguish these times can be challenging, but Tukey' s four routes certainly help. I get irritated when I explain what I did in some statistical analysis, and a listener says, "I see, you are assuming blah blah..." I reply tartly "No, I am not assuming blah blah... I am just doing what I did. Here' s why I thought it might help, and here' s why it did help."

I think behind Tukey's observations was an impatience with the mindless drive to automate what he called "routes to sanctification and certain truth." An emphasis on enumerating assumptions and their checking certainly looks like that. He gives the example of spectrum analysis, something usually associated only with stationary processes, as having its "greatest triumphs" with a non-stationary series. He would have welcomed Silver' s book, and wholeheartedly agreed with his focus on assumptions. To channel Box, "All assumptions are wrong but some are critical."



Terry Speed, Ph.D.
Professor and Lab Head
Bioinformatics Division
Walter & Eliza Hall Institute of Medical Research
Parkville, Victoria
Australia

Timely or Trustworthy?

Hans Rudolf Künsch

The pandemic has dominated our private and professional life during the last year, and I begin this column with a related story. On January 27, 2020, Camilla Rothe examined the first patient in Germany who was tested positive for Covid-19. The only way he could have been infected was through a colleague from China who had come to Munich for a business meeting. During her stay in Germany this colleague didn't feel sick except for a little fatigue that could be attributed to jet lag, but she was tested positive after her return to China. At that time it was believed that only patients with strong symptoms could spread Covid-19, as in the case of the Sars virus in 2002. Camilla Rothe and her boss realized the importance of their observation. Three days later they sent a short report to The New England Journal of Medicine that was published online immediately. Since the report contradicted the generally recognized opinion of most experts, it became a major political issue. Her observation was declared to be flawed and health officials in many countries continued to state that spreading without symptoms was extremely rare and thus irrelevant. It took at least two months until the danger of symptomless transmission was widely accepted and measures like recommending to wear masks and avoiding crowds were adopted. It is not clear how much the spread of the epidemic would have changed if the message of Camilla Rothe had been heard earlier.

However scrutiny and scepticism are essential parts of the scientific process. This is exemplified in following story. In September 2011, a group of physicists announced in a preprint that they had observed neutrinos traveling faster than the speed of light, thus violating special relativity theory. The reported p-value was $2\ 10^{-9}$. The experiment had taken place six months earlier and during this period, the group had checked the details of the experiment and had found no instrumental error. Once published, the result caused a stir not only in the physics community, but also in the mass media. Among physicists scepticism was prevalent as the theory of special relativity had been confirmed in several experiments before and had thus a much stronger basis than the belief that only patients with clear symptoms could spread the Covid-19 virus. The story ended in spring 2012 with the announcement that two equipment errors had been found, and after correcting for these the speed of neutrinos was no longer higher than the speed of light.

These two stories exemplify a dilemma of science: On the one hand, if an observation or an experiment has potentially serious consequences for human beings or shakes a widely believed dogma, it is important that this result is shared in a timely fashion at least with the scientific community, and possibly also with the general public. On the other hand, if scientists contradict each other continuously, their explanations change frequently and their predicted catastrophic outcomes repeatedly do not happen, the trust in science is undermined. In the current pandemic this is a problem when politicians ask for advice which measures are most efficient to slow down the spread of the virus and least harmful for the economy.

As it is too early to draw conclusions about the value of the many scientific contributions for the handlling of the pandemic (except the development of tests and vaccines), let me finish with the story of the discovery of the ozone hole as it is maybe less familar to younger readers. The ozone layer in the stratosphere protects the surface of the earth from ultraviolet radiation which is the main cause of skin cancer in humans. In the seventies a decrease in the ozone layer was observed, and as the main cause the release of chlorofluorocarbons (CFC) was identified. CFC's are used in refrigerators, air conditioners, foams and as aerosol propellants. This led to a ban of CFC in aerosol sprays in a few countries, but observations towards the end of the seventies showed the decrease to be much smaller than what had been predicted. Thus the early warnings of scientists were considered as exaggerated and the decisions to reduce CFC emissions as premature. But then in 1985 appeared an article in Nature showing a "large seasonal disappearance of ozone ... over the antarctic," the ozone hole, that had started already in 1976. This hole had been observed by a single ground station, and the long time between discovery and publication was due to delayed data processing and to an extremely cautious leader of the project who wanted to exclude instrumental and other errors. He was also concerned that NASA who had a satellite to monitor the ozone layer in orbit had not discovered the hole. A popular story says that this was due to an automatic rejection of these values as outliers, but it seems that the values were only flagged, but not rejected and nobody took the effort to look at the huge amount of data stored on magnetic tapes. The ozone hole was confirmed by

NASA in 1986, and the ensuing political pressure led in 1987 to the signing of the Montreal protocol banning the use of CFC by 43 countries. In this example, the early scientific warnings about an environmental problem turned out to be justified although the early models made poor predictions and the true extent of the problem was discovered late because of insufficient measuring efforts and an extremely cautious scientist.

I hope these stories help to make you aware of some of the issues in the communication of scientific evidence and uncertainty. For more thoughts and advice I recommend the article "Five rules for evidence communication" by Michael Blastland et al., published in Nature 587 (2020), 362–36.4.

This column is based on the following sources:

1. https://en.wikipedia.org/wiki/Faster-than-light_neutrino_anomaly
2. https://www.nytimes.com/2020/06/27/world/europe/coronavirus-spread-asymptomatic.html
3. Maureen Christie, Data Collection and the Ozone Hole: Too much of a good thing?. Proceedings of the International Commission on History of Meteorology (2004), 99-105.

Hans Rudolf Künsch, Ph.D.
Professor Emeritus of Mathematics
Seminar für Statistik
ETH Zürich
Switzerland

Opinion Polling: Its Secret Sauce is also its Spoilage Source

Xiao-Li Meng

Editorial: This is a reprint from a column article published in the *IMS Bulletin* (Volume 50, Issue 1: January/February 2021; https://imstat.org/wp-content/uploads/2020/11/Bulletin50_1.pdf) with IMS' permission.

On November 6, 2020, I woke up to a flood (for a statistician) of tweets about my 2018 article, "Statistical Paradises and Paradoxes in Big Data (I): Law of Large Populations, Big Data Paradox, and the 2016 US Presidential Election" (https://statistics.fas.harvard.edu/files/statistics-2/files/statistical_paradises_and_paradoxes.pdf). A kind soul had offered it as an explanation to the question: "What' s wrong with polls?" , which led to the article going viral.

As much as I was flattered by the attention, I was disappointed that no one had asked "Why would anyone expect polls to be right in the first place?" A poll typically samples a few hundred or thousand people, but it aims to learn about a population many times larger. For predicting the US presidential election, conducting a poll of size $n = 5,000$ to learn about the opinions of $N = 230$ million (eligible) voters is the same as asking just about 2 people out of every $100,000$ voters on average. Isn' t it absurd to expect to learn anything reliably about so many from opinions of so few?

Indeed when Anders Kiaer, the founder of Statistics Norway, proposed the idea to replace a national census by "representative samples" during the 1895 World Congress of the International Statistical Institute (ISI), the reactions "were violent and Kiaer's proposals were refused almost unanimously!" as noted by former ISI President Jean-Louis Bodin (https://www.isi-web.org/news-featured/20200-celebratingisi-s-135th-anniversary). It took nearly half a century for the idea to gain general acceptance.

The statistical theory for polling might be hard to digest for many, but the general idea of representative sampling is much more palatable. In a newspaper story about Gallup Poll going to Canada (Ottawa Citizen, Nov 27, 1941: https://news.google.com/newspapers?nid=2194&dat=19411127&id=l-0uAAAAIBAJ&sjid=VtsFAAAAIBAJ&pg=4887,5489739&hl=en), Gregory Clark wrote,

"When a cook wants to taste the soup to see how it is coming, he doesn' t have to drink the whole boilerful. Nor does he take a spoonful off the top,

then a bit from the middle, and some from the bottom. He stirs the whole cauldron thoroughly. Then stirs it some more. And then he tastes it. That is how the Gallup Poll works."

The "secret sauce" for polling, therefore, is thorough stirring. Once a soup is stirred thoroughly, any part of it becomes *representative* of the entire soup. And *that* makes it possible to sample a spoonful or two to assess reliably the flavor and texture of the soup, regardless of the size of its container. Polling achieves this "thorough stirring" via random sampling, which creates, statistically speaking, a miniature that mimics the population.

But this secret sauce is also the source of spoilage. My 2018 article shows how to mathematically quantify the lack of thorough stirring, and demonstrates how a seemingly minor violation of thorough stirring can cause astonishingly large damage due to the "Law of Large Populations" (LLP: https://www.jbs.cam.ac.uk/insight/events/the-law-of-large-populations/). It also reveals that the polling error is the product of three indexes: *data quality, data quantity, and problem difficulty.*

To understand these terms intuitively, let's continue to enjoy soup. The flavoring of a soup containing only salt would be much easier to discern than a Chinese soup with five spices. *Problem difficulty* measures the complexity of the soup, regardless of how we stir it or the spoon size. Data quantity index captures the spoon size, relative to the size of the cooking container. This shift of emphasis from only the sample size n to the sample fraction n/N, which depends critically on the population size N is the key to LLP.

The most critical index and also the hardest one to assess is the *data quality*, a measure of the lack of thorough stirring. Imagine some spice clumps did not dissolve completely in the cooking, and if they have more chance of getting caught by the cook's spoon, then what the cook tastes is likely to be spicier than the soup actually is. For polling, if people who prefer candidate B over A are more (or less) likely to provide their opinions, than the polling will over- (or under-) predict the vote shares for B. This tendency can be measured by the Pearson correlation-let's denote it by r-between preferring B and responding (honestly) to the poll. The higher the value of $|r|$ (the magnitude of r), the larger the polling error. A positive r indicates overestimation, and a negative r underestimation.

The whole idea of stirring thoroughly or random sampling is to ensure r is negligible, or technically, to ensure it is on the order of the reciprocal of the square-root of N. Statistically, this is as small as it can be since we have to allow some sampling randomness. For example, for $N = 230$ million, $|r|$ should be less than 1 out of $15,000$. However, for the 2016 election polling, r was -0.005, or about 1 out of 200 in magnitude for predicting Trump's vote shares, as estimated in my article (based on polls carried out by YouGov). Whereas a half a percent correlation seems tiny, its impact is magnified greatly when multiplied by the square-root of N.

As an illustration of this impact, my article calculated how much statistical accuracy was reduced by $|r| = 0.005$. Opinions from 2.3 million responses (about 1% of the eligible voting population in 2016) with $|r| = 0.005$ has the same expected polling error as that resulting from 400 responses in a genuinely random sample. This is a 99.98% reduction of the actual sample sizes, an astonishing loss by any standard. A quality poll of size 400 still can deliver reliable predictions, but no (qualified) campaign manager would stop campaigning because a poll of size 400 predicts winning. But they may (and indeed some did) stop when the winning prediction is from 2.3 million responses, which amount to $2,300$ polls and each with $1,000$ responses.

What was generally overlooked in 2016, and unfortunately again in 2020 (but see this Harvard Data Science Review article https://hdsr.mitpress.mit.edu/pub/cnxbwum6/release/2), is the devastating impact of LLP. Statistical sampling errors tend to balance out when we increase the sample size, but systematic selection bias only solidifies when sample size increases. Worse, the selection bias is magnified by the population size: the larger the population, the larger the magnification. That is the essence of LLP.

When a particular bit of soup finds itself on the cook's spoon, it cannot say, "Well, I'm a bit too salty for the cook, so let me jump off this spoon!" But in an opinion poll, there is nothing to stop someone from opting out because of the fear of the (perceived) consequences of revealing a particular answer. Until our society knows how to remove such fear, or the pollsters can routinely and reliably adjust for such selective responses, we can all be wiser citizens of the digital age by always taking polling results with a healthy grain of salt.

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

The open access resource for leading and advanced content in data science, *Harvard Data Science Review* (HDSR) provides a crossroads at which fundamental data science research and education intersect directly with societally important applications from industry, government, NGOs, and more. HDSR publishes high-quality articles that help to define and shape data science as a rigorous field of study with global impact. Each issue informs, intrigues, and inspires readers with a wide breadth of content.

Selected Recent Content

Perspectives

- *Coming to Our Census: How Social Statistics Underpin Our Democracy (And Republic)* by Teresa A. Sullivan (Michigan State University)
- *Should We Trust Algorithms?* by David Spiegelhalter (Winton Centre for Risk and Evidence Communication, UK)
- *10 Research Challenge Areas in Data Science* by Jeannette M. Wing (Columbia University)
- *How to Define and Execute your Data and AI Strategy* by Ulla Kruhse-Lehtonen and Dirk Hofman (DAIN Studios, Finland)
- *Tackling COVID-19 through Responsible AI Innovation: Five Steps in the Right Direction* by David Leslie (The Alan Turing Institute, UK)
- *Doing Data Science on the Shoulders of Giants: The Value of Open Source Software for the Data Science Community* by Katie Malone (Tempus Labs) and Rich Wolski (UC Santa Barbara)
- *Differential Privacy and Social Science: An Urgent Puzzle* by Daniel L. Oberski (Utrecht University) and Frauke Kreuter (University of Maryland)
- *A Conversation with L. Rafael Reif on College of Computing, COVID-19 and the Future Workforce* by L. Rafael Reif (MIT), Liberty VIttert (Washington University), and Xiao-Li Meng (Harvard University)
- *On the Convergence of Epidemiology, Biostatistics, and Data Science* by Neil D. Goldstein, Michael LeVasseur, and Leslie A. McClure (Drexel University)

Impact

- *Geo-mapping of COVID-19 Risk Correlates Across Districts and Parliamentary Constituencies in India* by S. V. Subramanian, Omar Karlsson, Weixing Zhang, and Rockli Kim
- *Lost or Found? Discovering Data Needed for Research* by Kathleen Gregory (Royal Netherlands Academy), Paul Groth (University of Amsterdam), Andrea Scharnhorst (Royal Netherlands Academy), and Sally Wyatt (Maastricht University)
- *The Importance of Being Causal* by Iavor Bojinov, Albert Chen, and Min Liu (LinkedIn)
- *SCAM: A Platform for Securely Measuring Cyber Risk* by Leo de Castro, Andrew W. Lo, Taylor Reynolds, Fransisca Susan, Vinod Vaikuntanathan, Daniel Weitzner, and Nicholas Zhang (MIT)

Education

- *Data Science for Everyone Starts in Kindergarten: Strategies and Initiatives from the American Statistical Association* by Wendy Martinez and Donna LaLonde (American Statistical Association)
- *Beyond Unicorns: Educating, Classifying, and Certifying Business Data Scientists* by Tom Davenport (Babson College)
- *The Role of Academia in Data Science Education* by Rafael A. Irizarry (Dana-Farber Cancer Institute)

Research

- *Deep Learning with Gaussian Differential Privacy* by Zhiqi Bu, Jinshuo Dong, Qi Long and Su Weijie (The University of Pennsylvania)
- *With Malice Toward None: Assessing Uncertainty via Equalized Coverage* by Yaniv Romano, Rina Foygel Barber, Chiara Sabatti, and Emmanuel Candes (Stanford University and University of Chicago)

Columns

- **Diving into Data**
  Stop Flaunting Those Curves! Time for Stats to Get Down and Dirty with the Public by Timandra Harkness (BBC Radio)
- **Minding the Future**
  High School Data Science Review: Why Data Science Education Should be Reformed by Angelina Chen (Princeton High School)
- **Mining the Past**
  Error, Uncertainty, and the Shifting Ground of Census Data by Dan Bouk (Colgate University)
- **Recreations in Randomness**
  Recipes for Success: Data Science in the Home Kitchen by Shuyang Li and Julian McAuley (UC San Diego)

Special Offer: Inaugural Volume in Print

- Just published! Our Inaugural Volume (over 40 articles) is available for $25.
- This special print edition contains all the content published in Volume 1 PLUS an exclusive interview with Steve Ballmer, former Microsoft CEO, about his nonprofit organization, USAFacts.
- **Added bonus:** For a limited time, your purchase will include the special HDSR Commemorative issue containing a sampling of 8 articles.
- Order today at **mitpressjournals.org/HDSR**

Stay in touch with new article and issue alerts: hdsr.mitpress.mit.edu/subscribe

HDSI Harvard Data Science Initiative

The *Harvard Data Science Review* is an open access platform of the Harvard Data Science Initiative, and it is published by the MIT Press.

Yi's FDA Story: When Statistics Met Regulation 1994

Yi Tsong

**This article reflects the review of the author and should not be construed to represent FDA's views or policies.*

Remember the important historical events happened in 1994? Here are a few to bring you back to the year:

- January 1: The North American Free Trade Agreement (NAFTA) is established.
- January 11: The Superhighway Summit is held at UCLA's Royce Hall. It is the first conference to discuss the growing information superhighway and is presided over by U.S. Vice President Al Gore.
- January 14: U.S. President Bill Clinton and Russian President Boris Yeltsin sign the Kremlin accords, which stop the preprogrammed aiming of nuclear missiles toward each country's targets, and also provide for the dismantling of the nuclear arsenal in Ukraine.
- June 12: Nicole Brown Simpson and Ronald Goldman are murdered outside the Simpson home in Los Angeles. O.J. Simpson is later acquitted of the killings, but is held liable in a civil suit.
- June 15: The Lion King is released in theaters, making $422,783,777 in the United States ($951,583,777 worldwide). It is the highest-grossing film of the year.
- June 17: NFL star O.J. Simpson and his friend Al Cowlings flee from police in his white Ford Bronco. The low-speed chase ends at Simpson's Brentwood, Los Angeles, California mansion, where he surrenders. This event pushed black-white division to a new high.
- September 13: President Bill Clinton signs the Federal Assault Weapons Ban, which bans the manufacture of new firearms with certain features for a period of 10 years.
- September 13: President Bill Clinton signs the Violence Against Women Act of 1994 (VAWA). The Act provided $1.6 billion toward investigation and prosecution of violent crimes against women, imposed automatic and mandatory restitution on those convicted, and allowed civil redress in cases prosecutors chose not to prosecute. The Act also established the Office on Violence Against Women within the Department of Justice.
- October 9: US sends forces to Persian Gulf.
- November 4: The first conference devoted entirely to the subject of the commercial potential of the World Wide Web opens in San Francisco. Featured speakers include Marc Andreessen of Netscape, Mark Graham of Pandora Systems, and Ken McCarthy of E-Media.
- November 7: WXYC, the student radio station of the University of North Carolina at Chapel Hill, provides the world's first internet radio broadcast.
- November 16: A Federal judge issues a temporary restraining order prohibiting California from implementing Proposition 187, which would have denied most public services to illegal aliens.

Around the world,

- January 5-6: Serbs' heavy weapons pound Sarajevo.
- April 6: Thousands dead in Rwanda massacre.
- April 29: South Africa holds first interracial national election; Nelson Mandela elected President.
- May 4: Israel signs accord with Palestinians.
- October 17: Israel signs peace treaty with Jordan.
- August 31: IRA declares cease-fire in Northern Ireland.
- October 13: Ulster Protestants declare ceasefire.
- October 4: Aristide returns to Haiti.
- November 9: Aristide forms Government with Prime Minister and full Cabinet.
- December 11: Russians attacks secessionist Republic of Chechnya.

These are the events and issues of 26 years ago. How many of such event didn't repeat itself now? Recently, Patricia (my wife) experienced sciatica, which is the pain that radiates along the path of the sciatic nerve, which branches from her left lower back through her left hip and buttocks and down on her left leg. The pain associated with sciatica is quite severe. It is probably caused by disc herniation pressed directly on the nerve or irritation or inflammation of the sciatic nerve. The pain was severe and she called her pharmacist friend and was advised to take Advil to release the pain and inflammation. After taking Advil for more than a week, she felt uncomfortable in her abdomen and loss of appetite that we thought it was due to the upsetting effect of NSAID (nonsteroidal anti-inflammatory drugs).

Nonsteroidal anti-inflammatory drugs (NSAIDs) are members of a drug class that reduces pain, decreases fever, prevents blood clots, and in higher doses, decreases inflammation. Side effects depend on the specific drug but largely include an increased risk of gastrointestinal ulcers and bleeds, heart attack, and kidney disease. The term nonsteroidal distinguishes these drugs from steroids, which while having a similar eicosanoid-depressing, anti-inflammatory action, have a broad range of other effects. First used in 1960, the term served to distance these medications from steroids, which were particularly stigmatized at the time due to the connotations with anabolic steroid abuse.

NSAIDs work by inhibiting the activity of cyclooxygenase enzymes (COX-1 or COX-2). In cells, these enzymes are involved in the synthesis of key biological mediators, namely prostaglandins, which are involved in inflammation, and thromboxane, which are involved in blood clotting.

There are two types of NSAIDs available: nonselective and COX-2 selective. Most NSAIDs are non-selective and inhibit the activity of both COX-1 and COX-2. These NSAIDs, while reducing inflammation, also inhibit platelet aggregation (especially aspirin) and increase the risk of gastrointestinal ulcers/bleeds. COX-2 selective inhibitors have less gastrointestinal side effects but promote thrombosis and substantially increase the risk of heart attack. As a result, COX-2 selective inhibitors are generally contraindicated due to the high risk of undiagnosed vascular disease. These differential effects are due to the different roles and tissue localizations of each COX isoenzyme. By inhibiting physiological COX activity, all NSAIDs increase the risk of kidney disease and through a related mechanism, heart attack. In addition, NSAIDs can blunt the production of erythropoietin resulting in anemia, since hemoglobin needs this hormone to be produced. Prolonged use is dangerous and case studies have shown the health risk with celecoxib. COX-2 selective drugs will be discussed in more details in my report of 2006.

Pat's experience reminds me of a project that I was involved in 1994. In October 1984, an FDA Advisory Committee meeting was held to evaluate the safety concern on UPU (Upper gastrointestinal bleeding, perforation, and ulcer) of several NSAIDs. In Rossi et al (1987) and Hsu (1985), the authors compared the AE reports associated with Piroxicam with seven other nonsteroidal anti-inflammatory drugs (Diflunisal, Ibuprofen, Naproxen, Fenoprofen, Tolmetin, Sulindac and Meclofenamate). Dr. Jiann Ping Hsu was the FDA statistician responsible for postmarketing drug safety analysis. She left FDA for industry around 1986. I was hired to replace her in 1987. Hsu (1985) introduced the incidence rate of the event as the number of ADE reports of the drug divided by the number of prescriptions. After I completed the project of Halcion adverse events, Dr. Rossi brought a copy of the two reports to me and asked me to see if I could analyze the data with adjustment for secular trend and marketing year of the drugs. Therefore, I reanalyzed the data published in Rossi et al (1985). Since the method I used was already reported in my report of year 1993 in the last issue of ICSA Bulletin, I will not repeat it here. I will present only the data and results. The number of UPU events reports of the eight NSAIDs in 1974 to 1985 are given in Table 1. These are the numbers of events related to drugs prescribed. Correspondingly, the number of prescriptions in 1,000 units are given in Table 2. Dividing the number of reports in Table 1 by the number of prescriptions in Table 2, we get the reporting rate of UPU of the drug (i.e., number of reports per prescriptions, ADE reporting rate or generalized incidence density rate). The rate per 100,000 RXs of the drug per each calendar year are given in Table 3. When we compare two drugs such like Piroxicam with Diflunisal, we first stratified the data by the calendar year. We can calculate the overall reporting rates and the ratio of the rate using pooled rates or applying Mantel-Haenszel method on the stratified data as in Table 4. However, when we compare the two drugs marketed at different years, as we pointed out in the early report, there is a secular overall reporting trend as shown in Figure 1 The secular trend is due to changes of the regulatory requirement of what needs to be reported. As a comparison to year 1974, the observed overall reporting rate didn' t change much till year 1978, then started increasing sharply. Therefore,

Table 1. Number of Reports of Upper Gastrointestinal Bleeding Associated with Eight NSAIDS in 1974–1985

	Drug							
Year	Piroxicam	Diflunisal	Ibuprofen	Naproxen	Fenoprofen	Tolmetin	Sulindac	Meclofenamate
1974	—	—	0	—	—	—	—	—
1975	—	—	26	—	—	—	—	—
1976	—	—	26	13	10	15	—	—
1977	—	—	16	15	9	18	—	—
1978	—	—	7	5	4	4	0	—
1979	—	—	2	1	2	1	23	—
1980	—	—	2	15	9	9	65	0
1981	—	—	14	20	5	7	52	3
1982	81	3	11	23	9	14	15	6
1983	151	33	18	44	9	13	18	8
1984	170	14	19	39	11	2	12	2
1985	19	8	22	41	6	6	19	3

Table 2. Use in 1,000 Prescriptions of the Eight NSAIDS in 1974–1985

	Drug							
Year	Piroxicam	Diflunisal	Ibuprofen	Naproxen	Fenoprofen	Tolmetin	Sulindac	Meclofenamate
1974	—	—	1,104	—	—	—	—	—
1975	—	—	10,844	—	—	—	—	—
1976	—	—	13,650	1,540	822	756	—	—
1977	—	—	12,119	2,703	2029	1800	—	—
1978	—	—	11,238	3,674	2845	2109	1,297	—
1979	—	—	10,686	4,475	2973	2016	10,848	—
1980	—	—	15,942	5,734	3486	2799	8,918	426
1981	—	—	17,784	8,663	3927	2548	8,024	2129
1982	4407	886	18,932	9,931	3721	2711	7,205	1983
1983	7583	2305	19,782	11,461	3279	2449	5,790	1846
1984	8091	2671	20,890	12,521	3187	2583	5,309	1976
1985	8788	2951	21,729	11,656	2838	2561	5,278	2153

Table 3. Reporting Rates (per 100,000 RXs) of UPU Associated with Eight NSAIDS in 1974–1985

	Drug							
Year	Piroxicam	Diflunisal	Ibuprofen	Naproxen	Fenoprofen	Tolmetin	Sulindac	Meclofenamate
1974	—	—	0.00	—	—	—	—	—
1975	—	—	0.24	—	—	—	—	—
1976	—	—	0.19	0.84	1.22	1.82	—	—
1977	—	—	0.13	0.55	0.44	0.89	—	—
1978	—	—	0.06	0.14	0.14	0.19	0.00	—
1979	—	—	0.02	0.02	0.07	0.03	0.21	—
1980	—	—	0.01	0.26	0.26	0.26	0.73	0.00
1981	—	—	0.08	0.23	0.13	0.18	0.65	0.14
1982	1.84	0.34	0.06	0.23	0.24	0.38	0.21	0.30
1983	1.99	1.43	0.09	0.38	0.27	0.40	0.31	0.43
1984	2.10	0.52	0.09	0.31	0.35	0.06	0.23	0.10
1985	0.22	0.27	0.10	0.35	0.21	0.21	0.36	0.14

UPU = upper gastrointestinal bleeding, perforation, and ulcer.

the ratio of the overall reporting rate needs to be considered when stratified the data according to the same marketing year but different calendar year. Using 1974 as the starting point, the ratio can be determined using the best fitting quadratic curve as shown in Figure 1 Since we are comparing all seven NSAIDs with Piroxicam, all the reporting rate is adjusted by r = (fitted all-drug-all-event reporting rate/1982 fitted all-drug-all-event reporting rate). The adjusting factor of Ibuprofen for its marketing years are given in Table 5. For example, Ibuprofen started marketing in 1974, the adjusting factor $r = 0.77/1.46 = 0.55$. By stratifying with the first three marketing year, the ratio of reporting rates of Piroxicam to Ibuprofen and the ratio adjusted by the secular trend are given in Table 6. The ratio of pooling across the three years and Mental-Haenszel odds ratio with secular trend adjustment are given

Table 4. Number of Spontaneous Reports of Piroxicam and Diflunisal, 1982–1984 Data, No Adjustment

	Drug							95% conf. limits	
	Piroxicam			Diflunisal					
Year	# of AEs[a]	RXs[b]	Rate[c]	# of AEs	RXs	Rate	Ratio[d]	Lower	Upper
1982	81	4,407	1.84	3	886	0.34	5.43	1.81	16.25
1983	151	7,583	1.99	33	2305	1.43	1.39	0.96	2.02
1984	170	8,091	2.10	14	2671	0.52	4.01	2.34	6.87
Pooled	402	20,081	2.00	50	5862	0.85	2.35	1.75	3.15
Mantel-Haenszel statistics	—	—	—	—	—	—	2.37	2.05	2.75

[a]Number of adverse events reported.
[b]Number of 1000 prescriptions.
[c]Reporting rate (number of reports per 100,000 prescriptions).
[d]Ratio of reporting rates, i.e., rate of piroxicam/rate diflunisal.

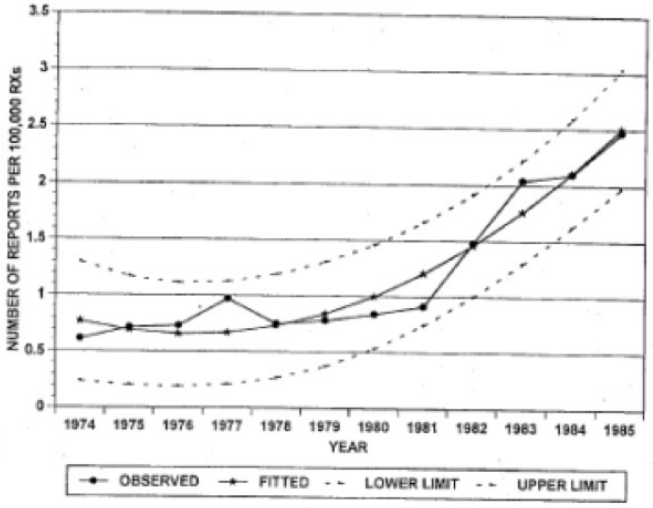


Figure 1: Secular trend of ADE reporting rate of all drugs

in Table 6. The corresponding secular trend adjusting factor of each NSAID are given in Table 7. Finally, the odds ratio of Piroxicam versus all seven NSAIDs combined are given in Table 8. It shows that based on the first three years marketing data, Piroxicam had higher risk of UPU when adjusted for the usage and secular trend (See details in Tsong (1995)).

This was a project done in 1994. The ADE reports are compiled in the FDA ADRS system. The drug usage data needed to be purchased from IMS (International Mathematical System). The name of IMS is probably changed after these many years. Before internet and laptop became popular, it might take months before the data is retrieved and/or compiled from the system. When comparing drugs with the same prescription pattern, ratio of reporting rates is often used as an estimate of the ratio of risk. In modern days, datamining techniques is often used on the ADE reporting data without usage data, we may only determine if a certain drug has disproportional reporting pattern which may not representing the risk of the drug.

Advil is long lasting up to 24-hour effect. The drug label recommends is to take ibuprofen with water to help with swallowing the pill or capsule. Taking it with a meal can slow absorption or how much is ultimately absorbed, but the labeling also allows patients to take it with a meal, particularly if you have stomach irritation after taking it. There are some drugs you should avoid taking ibuprofen with, either because it can alter absorption, exposure, or directly alter the intended pharmacologic effect. For example, avoid concurrent use with aspirin or other anti-inflammatory painkillers. Taking ibuprofen with water will not cancel out the effects

Table 5. Secular Trend Adjusting Factor Reports of Upper Gastrointestinal Bleeding Associated with Ibuprofen in 1974–1985

	Ibuprofen				All drug rep. rate (raw)[b]	All drug rep. rate (fitted)[c]	Adj. factor[d]
Year	AEs	RXs	Market year	Report rate[a]			
1974	0	1,104	1	0.00	0.61	0.77	0.53
1975	26	10,844	2	0.24	0.71	0.69	0.47
1976	26	13,650	3	0.19	0.73	0.65	0.45
1977	16	12,119	4	0.13	0.97	0.67	0.46
1978	7	11,238	5	0.06	0.75	0.73	0.50
1979	2	10,686	6	0.02	0.78	0.84	0.58
1980	2	15,942	7	0.01	0.84	1.00	0.68
1981	14	17,784	8	0.08	0.91	1.21	0.83
1982	11	18,932	9	0.06	1.48	1.46	1.00
1983	18	19,782	10	0.09	2.04	1.76	1.20
1984	19	20,890	11	0.09	2.10	2.11	1.45
1985	22	21,729	12	0.10	2.47	2.51	1.72

[a](AEs/RXs)*100.
[b]All AE reporting rate of all drugs (per 100,000 prescriptions).
[c]Fitted value of (all AE reporting rate of all drugs).
[d]Secular trend adjusted for the 1982 all-drug–all-AE reporting rate, i.e. (fitted all-drug–all-event reporting rate)/(1982 fitted all-drug–all-AE reporting rate).

Table 6. Number of Spontaneous Reports of Piroxicam and Ibuprofen, Ratio Adjusted for Marketing Year, and Estimated Secular Trend of All-Drug–All-AE Reporting Rates

	Drug								95% conf. limits	
	Piroxicam			Ibuprofen						
Market year	AEs[a]	RXs[b]	Reporting rate[c]	AEs	RXs	Reporting rate	Ratio[d]	Adj. ratio[e]	Lower	Upper
1	81	4,407	1.84	0	1,104	0.00	—	—	2.80	
2	151	7,583	1.99	26	10,844	0.24	8.31	3.23	2.13	4.88
3	170	8,091	2.10	26	13,650	0.19	11.03	3.42	2.27	5.16
Pooled	402	20,081	2.00	52	25,598	0.20	9.85	3.48	2.61	4.64
Mantel-Haenszel statistics	—	—	—	—	—	—	10.40	3.61	2.66	4.78

[a]Number of adverse events reported.
[b]Number of 1000 prescriptions.
[c]Reporting rate (number of reports per 100,000 prescriptions).
[d]Ratio of reporting rates, i.e., rate of piroxicam/rate of ibuprofen.
[e]Adj. ratio = [(piroxicam rate)/(Ibuprofen rate)]
*[(adj. factor for ibuprofen)/(adj. factor for piroxicam)].

Table 7. Relative Secular Trend Adjusting Factor[a] of the 8 NSAIDS in Each Marketing Year During 1974–1985 When Comparing to Piroxicam Reporting Rate

Market year	Drug							
	Piroxicam	Diflunisal	Ibuprofen	Naproxen	Fenoprofen	Tolmetin	Sulindac	Meclofenamate
1	1	1	0.530	0.450	0.450	0.450	0.500	0.690
2	1	1	0.388	0.380	0.380	0.380	0.479	0.686
3	1	1	0.310	0.345	0.345	0.345	0.476	0.690

[a](Adj. factor of a given marketing year)/(adj. factor of the corresponding year of piroxicam). For example, for the second marketing year of ibuprofen, relative adj. factor = (adj. factor of 1975)/(adj. factor of 1983) = 0.47/1.20 = 0.388.

of such other drugs if recently taken. According to Pat's pharmacist and therapist friends, many botanical products such as green tea and tamarind, may also have effect for anti-inflammatory. But they contain the acid to increase the damage of Advil. As a Chinese patient, Pat drank too much green tea while taking Advil maybe the cause of stomachache or uncomfortable. So, the uncomfortable disappeared af-

Table 8. Number of Spontaneous Reports of Piroxicam and Other NSAIDS, Stratified by Marketing Year and Adjusted for the Fitted Secular Trend of All-Drug–All-AE Reporting Rates

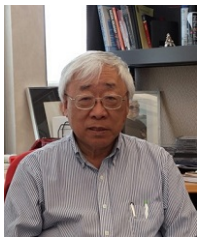
Market year	Drug: Piroxicam AEs[a]	Piroxicam RXs[b]	Piroxicam Reporting rate[c]	Other NSAIDS AEs	Other NSAIDS RXs	Other NSAIDS Reporting rate	Ratio[d]	Adj. ratio[e]	95% conf. limits Lower	95% conf. limits Upper
1	81	4,407	1.84	41	6,831	0.60	3.07	1.71	1.8	4.5
2	151	7,583	1.99	127	32,658	0.39	5.12	2.46	1.9	3.1
3	170	8,091	2.10	124	35,850	0.35	6.07	2.63	2.1	3.3
Pooled	402	20,081	2.00	292	75,339	0.39	5.16	2.40	2.1	2.8
Mantel-Haenszel statistics	—	—	—	—	—	—	4.90	2.31	2.1	2.8

[a]Number of adverse events reported.
[b]Number of 1000 prescriptions.
[c]Reporting rate (number of reports per 100,000 prescriptions).
[d]Ratio of reporting rates, i.e., rate of piroxicam/rate of other NSAIDs.
[e]Adj. ratio = (piroxicam rate)/[(total AEs of other NSAIDs)/(adj. RXs of other NSAIDs)], where adj. RXs of other NSAIDs = sum of [(RXs × relative adj. factor) of each of the other NSAIDs)].

ter switch to drink water or black tea in the days taking Advil.

References:

1. Rossi, AC, Hsu JP, Faich GA: Ulcerogenicity of piroxicam: Analysis of spontaneously reported data. Br. Med. J. 294:147-150, 1987.
2. Hsu, JP: Refinement of the methods of analysis for the NSAIDS study. CDER Division of Biometrics Memorandum, JP: Refinement of the methods of analysis for the NSAIDS study. CDER Division of Biometrics Memorandum, FDA, Rockville, MD, 1985.
3. Tsong, YT: Comparing reporting rates of adverse events between drugs with adjustment for year of marketing and secular trends in total reporting. J. of Biopharm. Statist. 5(1): 95-114 (1995).



Yi Tsong, Ph.D.
Division Director
CDER/OTS/OB/DBVI
U.S. Food and Drug Administration

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:

2021 ICSA Applied Statistics Symposium

The 2021 ICSA Applied Statistics Symposium will be held virtually on September 12-15, 2021.

2022 ICSA Applied Statistics Symposium

The 2022 ICSA Applied Statistics Symposium will be held in Gainesville, Florida. More detailed information will be shared later.

2021 ICSA China Conference (Postponed)

The 2021 ICSA China Conference will be held at Xian University of Finance and Economics, Xian, China. This meeting is postponed, and the date will be announced later. For information, please contact Scientific Program Committee Co-Chairs Professor Yingying Fan at fanyingy@marshall.usc.edu and Professor Chunjie Wang at wangchunjie@ccut.edu.cn.

2022 ICSA China Conference

July 1-4, 2022
The 2022 ICSA China Conference will be held at Chengdu from July 1 to July 4, 2022, co-sponsored by Southwest Jiaotong University (SWJTU).

The 12th ICSA International Conference

December 18 - 20 2022
The 12th ICSA International Conference will be held at the Chinese University of Hong Kong from December 18 to December 20 2022.

ICSA Co-sponsored Meetings:

Duke-Industry Statistics Symposium

April 21-23, 2021
Given evolving public concerns regarding COVID-19 and the potential travel uncertainties, the annual Duke-Industry Statistical Symposium (DISS) will be postponed. The symposium will be held from April 21-23, 2021 virtually. For further details regarding this update, please click this link: https://sites.duke.edu/diss.

The theme of the symposium is "Emerging Initiatives in Pharmaceutical Development: Methodology and Regulatory Perspectives." The first day will be devoted to six short courses. The second day and the third day morning are consisted of keynote speeches and 25 parallel sessions. The symposium was established 8 years ago to discuss challenging issues and recent advances related to the clinical development of drugs, biologics and devices and to promote research and collaboration among statisticians from industry, academia, and regulatory agencies.

The 8th Workshop on Biostatistics and Bioinformatics

Postponed to Spring, 2021
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 & Bioinformatics research areas. The workshop will provide the opportunity for faculty and graduate 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/2020Workshop/.

The 63rd ISI World Statistics Congress 2021

July 11-16, 2021
The World Statistics Congress 2021 will be held vir-

tually in July 2021. More information can be found on the ISI 2021 website www.isi2021.org.

The 77^{th} Annual Deming Conference on Applied Statistics

December 6-10, 2021
The 77^{th} Annual Deming Conference on Applied Statistics will be held from Monday Dec. 6 to Wednesday Dec. 8, 2021, followed by two parallel 2-day short courses on Thursday Dec. 9 and Friday Dec. 10 at the state-of-the-art Tropicana Casino and Resort, Havana Tower, Atlantic City, NJ. The purpose of the 3-day Deming Conference on Applied Statistics is to provide a learning experience on recent developments in statistical methodologies in biopharmaceutical applications. The conference is composed of twelve three-hour tutorials on current topics in applied biopharmaceutical statistic and FDA regulations, and a one-hour distinguished keynote speaker on each of the 3 days of the conference. For more information about the conference, please refer to the last page of ICSA Bulletin, and also visit https://demingconference.org/ or email Dr. Din Chen, Deming Publicity Chair, at din@demingconference.org.

IMS Asia Pacific Rim Meeting

Postponed to January 5-8, 2022
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/ for details. Firm dates will be announced at a later date.

Online Training and Seminars:

ICSA Online Training

Online training serves as a viable alternative to traditional continuing education options, e.g., to short courses offered at biostatistical conferences. Over the past year, the ASA Biopharmaceutical Section has been working on creating an online training program aimed at clinical trial statisticians and set up a pilot program, which includes half-day and full-day courses on key topics in biopharmaceutical statistics:

- Analysis of Longitudinal and Incomplete Data
- Multiplicity Issues in Clinical Trials
- Analysis of Surrogate Endpoints in Clinical Trials
- The section has received much positive feedback from industry and academic statisticians. Clinical trial statisticians who took advantage of the online training program emphasized that this program is convenient, inexpensive and quite flexible.

A similar online training program has been set up for ICSA members. As a member of the ICSA, you will receive a 50% discount when you sign up for any course included in the program. The online training courses are based on professionally recorded videos using a format similar to that used in YouTube videos. The videos can be accessed 24/7 on a computer or even on a smartphone. The cost of online training is low compared to traditional training, and it can be further reduced by using a group-training format. Up to 25 people can view an online training course with a single registration, which lowers the cost of online training to about $20-25 per person for full-day courses and $10-15 per person for half-day courses.

For more information about the online training program and to sign up for the individual online courses, please visit this web page: http://sprm.com/icsa/.

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

Fall Series: The Philosophy of Data Science

The series is aimed at incoming statistics and data science students (but will be of significant interest to the general statistics/data science community). The topics will focus on how scientific reasoning is essential to the practice of data science.

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

77th Annual Deming Conference on Applied Statistics

December 6-10, 2021; Atlantic City, NJ
(https://demingconference.org)

The 77th Annual Deming Conference on Applied Statistics will be held from Monday Dec. 6 to Wednesday Dec. 8, 2021, followed by two parallel 2-day short courses on Thursday Dec. 9 and Friday Dec. 10 at the state-of-the-art Tropicana Casino and Resort, Havana Tower, Atlantic City, NJ.

The purpose of the 3-day Deming Conference on Applied Statistics is to provide a learning experience on recent developments in statistical methodologies in biopharmaceutical applications. The conference is composed of twelve three-hour tutorials on current topics in applied biopharmaceutical statistic and FDA regulations, and a one-hour distinguished keynote speaker on each of the 3 days of the conference. The books, on which these sessions are based, are available for sale at an approximately 40% discount. Attendees will receive hard copy program proceedings of the presentations.

There will be poster sessions. Early registrants who submit a poster presentation will receive a $150 discount. For poster submission, please contact "Deming Poster Chair": Dr. Pinggao Zhang at email: pinggao.zhang@takeda.com.

There will be student scholar presentations. For a student scholar application, please contact "Deming Scholar Chair": Dr. Sofia Paul at email: sofia.x.paul@gsk.com.

The conference is sponsored by the American Statistical Association Biopharmaceutical Section and the International Chinese Statistical Association. Walter Young has chaired this conference for 52 consecutive years. The program committee include: Alfred Balch, Joseph Borden, Ivan Chan, (Din) Ding-Geng Chen, Kalyan Ghosh, Satish Laroia, Sofia Paul, Manoj Patel, Naitee Ting, Bill Wang, Wenjin Wang, Yibin Wang, Li-an Xu, Walter Young and Pinggao Zhang.

Registration is expected to open by mid-August and one page program should be available by that time. For more information about the conference, please visit https://demingconference.org/ or email Din Chen, Deming Publicity Chair, at din@demingconference.org.