Somnath Datta

Curriculum Vitae (Revised: September 10, 2024)

PERSONAL

Born 1962, Calcutta (now Kolkata), India
US Citizen
Resident of Florida
ORCID: 0000-0003-4381-1842
Visited 129 independent counties (out of ~196) plus 8 others

UF Office
Department of Biostatistics
University of Florida
2004 Mowry Rd., Rm 5226 CTRB
Gainesville, FL 32610
USA
(352) 294-5920
somnath.datta@ufl.edu

Home Office
8817 SW 12th Rd
Gainesville, FL 32607
USA
(502) 554 0899
somndatta@gmail.com
www.somnathdatta.org
www.somnathdatta.net

EDUCATION

- Ph. D. (1988), Statistics and Probability, Michigan State University, East Lansing.
- M. Stat. (1985), Mathematical Statistics and Probability, Indian Statistical Institute, Calcutta.
- B. Stat. (1983), Statistics, Indian Statistical Institute, Calcutta.

ACADEMIC POSITIONS HELD

• 2015 (Fall) – present: Professor (tenured), Preeminence Hire in Genomic Medicine, Department of Biostatistics, University of Florida, Gainesville, FL, USA.

• 2005 (Summer) – 2015 (Summer): Professor (tenured), Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA.

• 1998 (Fall) – 2005 (Spring): Professor, Department of Statistics, University of Georgia, Athens, GA, USA.

• 1993 (Fall) – 1998 (Summer): Associate Professor (tenured), Department of Statistics, University of Georgia, Athens, GA, USA.

• 1988 (Fall) – 1993 (Summer): Assistant Professor, Department of Statistics, University of Georgia, Athens, GA, USA.

ADMINISTRATIVE POSITIONS HELD

• 2009 (May) – 2015 (August): Vice Chair, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA.

• 2008 (Fall) – 2012 (Summer) : Biostatistics PhD Program Director, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA.

• 2005 (Fall) – 2008 (Summer): Biostatistics Graduate Coordinator, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA.

OTHER POSITIONS

• 2016 – : Adjunct Professor, Department of Statistics, University of Florida, Gainesville, FL, USA.

• 2015 – 2020: Adjunct Professor, Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA.

• 2011 – 2015: Associated Faculty, Conn Center for Renewable Energy Research, University of Louisville, Louisville, KY, USA.

• 2006 – 2015: Senior Biostatistician, Christopher and Dana Reeve Foundation NeuroRecovery Network (NRN), University of Louisville, Louisville, KY, USA.

• 2016 – : Senior Biostatistician, Brain Rehabilitation Research Center of Excellence, North Florida/South Georgia VA Medical Center, Gainesville, FL, USA.

SHORT ACADEMIC VISITS

• Southwest Jiaotong University, Chengdu, China, October 2018.

• Department of Probability and Statistics, National University of Uzbekistan, Tashkent, Uzbekistan, May 2016.

• Department of Statistics, Tunghai University, Taichung, Taiwan, December 2013.

• Institute of Mathematical Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia, June 2013.

• Department of Statistics, University of Concepción, Concepción, Chile, January 2013.

• Department of Statistics and OR, University of Vigo, Vigo, Spain, September 2010.

• Department of Mathematics for Science and Technology, University of Minho, Guimarães, Portugal, September 2010.

• Department of Statistics, Southwest Jiaotong University, Chengdu, China, May-June, 2010.

• Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, May 2009.

• School of Public Health, University of Tampere, August 2008, May 2009, June 2011.

• Department of Statistics and Applied Probability, National University of Singapore, December 2004.

• Math Stat Division, Indian Statistical Institute, Calcutta, July 1999.

AWARDS/HONORS

• 2019 **CDC ATSDR 2019 Statistical Science Award**: Best Theoretical Paper, "Multisample adjusted U-Statistics that account for confounding covariates." Satten, G. A., Kong, Maiying, and Datta, Somnath. Statistics in Medicine, 37, 3357–3372 (2018).

• 2018: **President**, International Indian Statistical Association (IISA), www.intindstat.org.

• 2017: **Dean's Citation Paper Award,** College of Public Health and Health Professions, University of Florida.

• 2015: **Preeminence Hire**, University of Florida.

• 2014: Appointed **University Scholar**, University of Louisville.

• 2014: **Outstanding Mentor of Doctoral Students Award** nominee, University of Louisville.

• 2013: **President's Distinguished Faculty Award in Research for Career Achievement**, University of Louisville.

• 2011: **2010-2011 Faculty Favorite**, "An Outstanding Professor Nominated by Students", Delphi Center for Teaching and Learning, University of Louisville.

• 2011: **CDC ATSDR 2011 Statistical Science Award**: Best Theoretical Paper, "Inverse Probability of Censoring Weighted U-statistics for Right-Censored Data with an Application to Testing Hypotheses", Datta, Somnath, Bandyopadhyay, Dipankar and Satten, Glen A., Scandinavian Journal of Statistics, 37, 680-700 (2010).

• 2010: **Elected Fellow, Institute of Mathematical Statistics**.

"For contributions to compound decision theory, bootstrap inference for Markov chains and time series, survival analysis and counting processes, and biostatistics and bioinformatics; and for editorial services to the profession."

• 2010: **Vice-president**: **Forum for Interdisciplinary Mathematics**, 2011-2013.

• 2010: **Provost's Awards for Exemplary Advising** nominee, University of Louisville.

• 2009: **Elected member, International Statistical Institute**.

• 2008: **Vice-president**: **Forum for Interdisciplinary Mathematics**, 2009-2011.

• 2007: **Best Poster Award**, First Place, American Spinal Injury Association, 33rd Annual Scientific Meeting, for the poster "A Multivariate Examination of Temporal Change in BERG Balance Scale Variables for Patients with ASIA C

and D Spinal Cord Injuries" by S. Datta, D. Lorenz, M. Schmidt-Read, E. Ardolino, S. Morrison, and S. J. Harkema.

- 2007: Listed in **Who's Who in America**, 61st Edition.

- 2006: **Elected Fellow**, **American Statistical Association.**

  "For outstanding research in theoretical and applied statistics including decision theory, bootstrap theory, survival analysis and analysis of microarray data."

- 2005: **CDC ATSDR 2005 Statistical Science Award**: Best Application Paper, "Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens" by Satten, G. A., Datta, S., Moura, H., Woolfitt, A., Carvalho, G., De, B. K, Pavlopoulos, A., Carlone, G. M., and Barr, J., Bioinformatics, 20, 3128-3136 (2004).

- 2004: **CDC ATSDR 2004 Statistical Science Award**: Best Theoretical Paper, "Marginal analyses of clustered data when cluster size is informative" by Williamson, J. M., Datta, S. and Satten, G. A, Biometrics, 59, 36-42 (2003).

- 2003: **Snedecor Award** nominee for the paper "Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under stage dependent censoring" by Datta, S. and Satten, G. A., Biometrics, 58, 792-802 (2002).

- 2001: **CDC ATSDR 2001 Statistical Science Award**: Best Theoretical Paper, "A simulate-update algorithm for missing data problems" by Satten, G. A. and Datta, S., Computational Statistics, 15, 243-277 (2000).

- 1999: **CDC ATSDR 1999 Statistical Science Award**: Best Theoretical Paper, "A semiparametric approach to the proportional hazards model for interval censored data", by Satten, G. A. and Datta, S. and Williamson, J. M., Journal of the American Statistical Association, 93, 318-327 (1998).

- 1985-1988: Intermittent **fellowships for merit** throughout in the Ph. D. program at Michigan State University; GPA 4.0/4.0.

- 1986: **Pass with distinction** in Ph. D. prelims at Michigan State University.

- 1980-1985: **First class honors with distinction** in B. Stat. and M. Stat. and many **cash awards** throughout these programs.

RESEARCH

Ph. D. Dissertation Title: "Asymptotically Optimal Bayes Compound and Empirical Bayes Estimators in Exponential Families with Compact Parameter Space" (Professor James F. Hannan, Ph. D. dissertation advisor).

Research Interest (past & present): **Biostatistical Methods**, Bootstrap Methods, Causal Inference, Compound Decision, **Clustered Data**, Clustering and Classification, **Dental and Craniofacial**

Data, Empirical Bayes, **Statistical Genomics** (microarrays, next generation sequencing, single cell RNA seq, microbiome etc), Nonparametrics, Personalized Medicine, Proteomics, Rank Tests, **Survival Analysis and Multistate Models**, Time Series Analysis.

Collaborative: Materials Science, Plant Science, **Rehabilitation Science**.

Mathematics Genealogy (taken from The Mathematics Genealogy Project):
S. Datta → J.F. Hannan → W. Höffding → A. Klose → A.F.K. Wilkens → P.R. Harzer → 1,2
1 C.C. Bruhns → J.F.F. Encke → C.F. **Gauss**
2 W. Scheibner → C.G.J. **Jacobi**

GRADUATE STUDENTS MENTORED (As major professor)

Doctor of Philosophy (Dissertations)

1. **Michael R. Allen**, *"Inference and Bootstrap for Some Linear Time Series Models."* Completed: Summer 1997. University of Georgia. Currently a Full Professor in the Department of Mathematics, Tennessee Technological University, Cookeville, TN, USA.

2. **Sahmyeong Kim** (jointly with I. V. Basawa), *"Inference for Nonlinear Time Series Models via Estimating Functions."* Completed: Spring 1998. University of Georgia. Currently at Department of Applied Statistics, Chung-Ang University, Seoul, S. Korea.

3. **HaiTao Zheng** (jointly with I. V. Basawa), *"Inference for Time Series Models for Count Data."* Completed: Summer 2005. University of Georgia. Currently a Full Professor in the Department of Statistics, Southwest Jiaotong University, Chengdu, China.

4. **Dipankar Bandopadhyay**, *"Novel Nonparametric Methods for Event Time Data."* Completed: Spring 2006. University of Georgia. Currently a Full Professor in the Department of Biostatistics, School of Medicine, Virginia Commonwealth University (VCU), Richmond, VA, USA.

5. **DeSale Habtzghi** (jointly with M. Meyer), "*Maximum Likelihood Based Estimation of Hazard Function under Shape Restrictions and Related Statistical Inference."* Completed: Spring 2006. University of Georgia. Currently an Associate Professor in the Department of Statistics, DePaul University, Chicago, IL, USA.

6. **Ling Lan**, "*Inference for Multistate Models*." Completed: Summer 2008. University of Louisville. Currently a Mathematical Statistician at the Food and Drug Administration (FDA). First employment as an Assistant Professor in the Department of Biostatistics and Epidemiology, Medical College of Georgia, Augusta, GA, USA.

7. **Vasyl Pihur** (jointly with Susmita Datta), "*Statistical Methods for High-Dimensional Genomics Data Analysis."* Completed: Summer 2009. University of Louisville. Currently at Snapchat, Inc. First employment at Department of Biostatistics (Irizarry Lab), Johns Hopkins University, Baltimore, MD, USA.

8. **Jie Fan**, "*Inference for Time to Event and Sojourn Time Data under Right Censoring Using Reweighting Approaches."* Completed: Summer 2010. University of Louisville. Currently Self-employed. First employment at Lombardi Cancer Center, Georgetown University, Washington, D.C., USA.

9. **Doug Lorenz** (Department of Mathematics, jointly with R. Gill), *"Marginal Nonparametric Inference for Waiting Times in Multistage Models: Hypothesis Testing and Regression."* Completed: Spring 2010. University of Louisville. Currently an Associate Professor in the Department of Bioinformatics and Biostatistics, University of Louisville, Louisville, KY, USA.

10. **Farida Mostajabi** (jointly with Susmita Datta), *"Regression Methods for Survival and Multistate Models".* Completed: Summer 2011. University of Louisville. Currently a Lead Quality Analyst & Manager, Transformation (Data Analytics) at Cincinnati Children's Hospital Medical Center, Cincinnati, OH, USA.

11. **Amanda Nicole Ferguson** (jointly with G. Brock), *"Methods and Software for Nonparametric Estimation in Multistate Models."* Completed: Summer 2011. University of Louisville. Currently an Associate Professor of Statistics in the School of Data Science and Analytics, Director of Human Studies Lab, Kennesaw State University, Atlanta, GA, USA.

12. **Sutirtha Chakraborty** (jointly with Susmita Datta), "*Novel Methods based on Regression Techniques to Analyze Multistate Models and High-dimensional Omics Data."* Completed: Summer 2013. University of Louisville. Currently at Abbott Laboratories, India. First employment at Department of Biostatistics, Harvard University (RafaLab), Cambridge, MA, USA.

13. **Joseph Bible**, *"Novel Applications of and Extensions to Linear Regression Methods for the Biomedical and Materials Sciences."* Completed: Spring 2015. University of Louisville. Currently an Associate Professor in the Department of Mathematical Sciences, Clemson University, Clemson, SC, USA.

14. **Hyoyoung Choo-Wosoba**, "*Inference for a Zero-inflated Conway-Maxwell Poisson Regression for Clustered Count Data*". Completed: Spring 2016. University of Louisville. Currently a staff scientist at the National Institutes of Health. First employment at Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health, under Dr. Paul S. Albert, Rockville, MD, USA.

15. **Sandipan Dutta**, "*Some Contributions to Nonparametric and Semiparametric Inference for Clustered and Multistate Data".* Completed: Spring 2016. University of Louisville. Currently an Assistant Professor at Department of Mathematics & Statistics, Old Dominion University, Norkolk, VA, USA.

16. **Chathura Siriwardhana** (jointly with K. B. Kulasekera), "*Single Index Regression Methods for Personalized Treatment Allocation and Multistate Models".* Completed: Spring 2016. University of Louisville. Currently an Assistant Professor at Department of Quantitative Health Sciences, University of Hawaii, Honolulu, HI, USA.

17. **Abdia Yunathan** (jointly with M. Kong and K. B. Kulasekera). *"Propensity Score Based Methods for Estimating the Treatment Effects Based on Observational Studies."* Completed: Summer 2016. University of Louisville. Currently a Postdoctoral Research Fellow at The University of British Columbia, Canada.

18. **Yichen Chen.** *"Using U-Statistics to Compare Grouped Marginal Distributions of Right Censored Event and Waiting Times in Observational Studie."* Completed: Summer 2019. University of Florida. Currently at Department of Global Pediatric Medicine St. Jude Children's Research Hospital, TN, USA.|

19. **Tyler Grimes.** *"Some Contributions to the Differential Network Analysis of –Omics Data."* Completed: Summer 2020. University of Florida. Currently Assistant Professor at Department of Mathematics & Statistics, University of North Florida, Jacksonville, FL, USA.

20. **Tong Kang.** *"Analysis of Non-continuous Zero-inflated Longitudinal and Clustered Data with Application to Dental Studies."* Completed: Spring 2021. Joined Bristol Myers Squibb in July 2021.

21. **Seungjun Ahn** (jointly with Sam Wu)**.** *"Novel Approaches to Differential Network Analysis Based on -Omics Data in Presence of Covariates."* Completed: Spring 2023, Currently an Assistant Professor at Center for Biostatistics, Department of Population Health Science and Policy, Icahn School of Medicine, Mount Sinai, New York, NY, USA.

22. **Samuel Anyaso-Samuel.** *"Advances in Cluster-correlated Data Analysis when Cluster size is Informative."* Completed: Summer 2023. Starting in August at the Division of Cancer Epidemiology & Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA.

23. **Yuting Yang** (jointly with Sam Wu)**.** *"Some Contributions to Biostatistical Methods for Progressive Multistate Models and Two-Stage Designs."* Completed: Summer 2023. Starting in August as an Assistant Professor in the Health Informatics Institute/Pediatrics, University of South Florida, Tampa, FL, USA.

24. **Xiaoxi Zhang** (jointly with Peihua Qiu). Working on *Survival Analysis*. Expected completion: Spring 2025.

25. **Shoumi Sarkar** (jointly with Peihua Qiu). Working on *Non-continuous data modeling*. Expected completion: Spring 2025.

26. **Owen Visser**. Working on *Single cell clustering and related issues*. Expected completion: Spring 2026.

27. **Vincent Mei** (jointly with Sam Wu)**.** Expected completion: Spring 2027.

28. **Victor Agboli.** Expected completion: Spring 2028.

29. **Yulin Li** (jointly with Sam Wu)**.** Expected completion: Spring 2028.

<u>Master of Science</u> (Theses)

1. **Cathleen Gillespie**, *"Intra-Individual Variation in Serum Vitamin A Measures Among Participants in the Third National Health and Nutrition Examination Survey 1988-1994."* Completed: Spring 2002. University of Georgia.

2. **Guiping Yang**, *"A New Bivariate Survival Function Estimator under Random Right Censoring."* Completed: Spring 2005. University of Georgia.

3. **Vasyl Pihur**, *"Weighted Rank Aggregation of Cluster Validation Measures: A Monte Carlo Cross-Entropy Approach."* Completed: Spring 2007. University of Louisville.

4. **Jie Fan** (jointly with G. Brock), *"Imputation Based Statistical Tests for Right Censored Data."* Completed: Summer 2007. University of Louisville.

5. **Bart Brown** (jointly with G. Brock), *"A Novel Method for Reference Interval Estimation Using the Inverted Q-Q Plot."* Completed: Summer 2007. University of Louisville.

6. **Ming Wang** (jointly with M. Kong), *"Analysis for Clustered Longitudinal Data."* Completed: Summer 2008. University of Louisville.

7. **Daniel Riggs**, *"An Investigation of Sliced Inverse Regression with Censored Data."* Completed, Summer 2010. University of Louisville.

8. **Sheng Xu** (jointly with M. Kong), *"Marginal models for Iowa Fluoride Study."* Completed, Spring 2013. University of Louisville.

Master of Science (Capstone projects - all in the Department of Biostatistics, University of Florida)

1. **Candice Sammons**, Created an R package for survival function estimation adjusted for confounders. April 2018.

2. **Dustin Duffy**, Performed survival analysis on a breast cancer dataset containing genomic information. March 2020.

3**. Sumia Tahir**, Performed differential gene expression analysis on a dataset related to Glioblastoma HSR-GBM1 cells. April 2020.

OTHER GRADUATE STUDENTS MENTORED (only if the mentoring is substantial)

1. **Ana Caroline Costa Sá**, Ph.D. student in the Department of Pharmacotherapy and Translational Research, College of Pharmacy, University of Florida (Dr. Julie Johnson, Major Professor), Graduated Spring, 2017.

2. **Mary Gregg,** M.S. student in the Department of Bioinformatics and Biostatistics, University of Louisville (D. Lorenz, Major Professor), Graduated Spring, 2016.

3. **Mary Gregg,** Ph.D. student in the Department of Bioinformatics and Biostatistics, University of Louisville (D. Lorenz, Major Professor), Graduated Summer, 2020.

POST-DOCTORAL RESEARCH ASSOCIATES MENTORED

1. **Grzegorz Boratyn** (jointly with Susmita Datta and John Klein), 2007-2008. Currently a Staff Scientist at NLM, National Institutes of Health, USA.

2. **Partha Dey** (jointly with Krishna Rajan), 2012-2014. Currently Associate Professor in Academy of Technology at Adisaptagram, Hooghly, India.

3. **Alejandro Riveros Walker** (jointly with Susmita Datta), 2016- 2018. Currently Research Assistant Professor at Department of Oral Biology, University of Florida, USA.

VISITING SCHOLARS MENTORED

1. **Nasim Vahabi** (visited from Tarbiat Modares University, Tehran, Iran), worked in Ordinal Data Regression, University of Florida. July 2016-September 2017.

2. **Rustamjon Muradov** (visited from Namangan Institute of Engineering and Technology, Tashkent, Uzbekistan), worked in Bivariate Survival Analysis, October 2021 - January 2022.

MS STATISTICIANS MENTORED

1. **Doug Lorenz**, University of Louisville, 2006-2010.

2. **Zhiguo Chen**, University of Florida, 2016- 2018.

PUBLICATIONS

Erdös Number = 3 (Datta → McCormick → Canfield → Erdös)

Edited Books

2. Datta, S. and Guha, S. ***Statistical Analysis of Microbiome Data***. Springer, 346 pages, ISBN: 978-3-030-73350-6 (Print), ISBN 978-3-030-73351-3 (Online) (2021).

1. Datta, S. and Nettleton, D. ***Statistical Analysis of Next Generation Sequencing Data***, Springer, 432 pages, ISBN: 978-3-319-07211-1 (Print) 978-3-319-07212-8 (Online) (2014).

Articles

* indicates a graduate student author (at the time of work)

195. Zhang, X.*, Datta, S. and Qiu, P. Effective comparison of two potentially crossing hazard rate curves. ***Journal of Statistical Computation and Simulation***, to appear **(2024).**

194. Zhang, X.*, Datta, S. and Qiu, P. Comparing two hazard curves when there is a treatment time-lag effect. ***Statistics in Medicine***, to appear **(2024).**

193. Anyaso-Samuel, S.* and Datta, S. Testing for marginal covariate effect when the subgroup size induced by the covariate is informative. ***Statistical Methods in Medical Research*****,** doi: 10.1177/09622802241254196 **(2024).**

192. Sarkar, S.*, Anyaso-Samuel, S.*, Qiu, P., and Datta, S. Multiblock partial least squares and rank aggregation: Applications to detection of bacteriophages associated with antimicrobial resistance in the presence of potential confounding factors. ***Statistics in Medicine***, appeared online ahead of print, https://doi.org/10.1002/sim.10058 **(2024).**

191. Ahn, S.* and Datta, S. Differential network connectivity analysis for microbiome data adjusted for clinical covariates using jackknife pseudo-values. ***BMC Bioinformatics***, 25, 117 **(2024).**

190. Ahn, S.* and Datta, S. SOHPIE: Statistical approach via pseudo-value information and estimation for differential network analysis of microbiome data. ***Bioinformatics***, 40, 1, btad766, **(2024).**

189. Yang, Y.*, Wu, S. S., and Datta, S. Regression analysis of a future state entry time distribution conditional on a past state occupation in a progressive multistate model. ***Statistical Methods in Medical Research,*** https://doi.org/10.1177/09622802231206473, to appear in print **(2024).**

188. M., Elise F. Grandy, Gebel, L. M. U., Santana, T. M., Rodriguez, A. L. Singh, S. K. , Fernandez, M. I., Dalugdug, J. C., Garcia-Colon, E. M., Lybeshari, K., Alexander, D. R., Maria I. Maura, M. I., Cabrera Gonzalez, M. D., Almeida, C. D., Anyaso-Samuel, S., Datta, S., Schiefer, M. Bilateral subdiaphragmatic vagal nerve stimulation using a novel waveform decreases body weight, food consumption, adiposity, and activity in obesity-prone rats. ***Obesity Surgery***, 34, 1-14 **(2024).** PMCID: PMC10781827

187. Ahn, S.* and Datta, S. PRANA: an R package for differential co-expression network analysis with the presence of additional covariates. ***BMC Genomics,*** 24, 687 **(2023)**.

186. Anyaso-Samuel, S.*, Bandopadhyay, D., and Datta, S. Pseudo-value regression of clustered multistate current status data with informative cluster sizes. ***Statistical Methods in Medical Research***, 32, 1494-1510 **(2023).**

185. Zhang, X.*, Seungjun Ahn, S.*, Qiu, P., and Datta, S. Identification of shared biological features in four different lung cell lines infected with SARS-CoV-2 virus through RNA-Seq analysis. ***Frontiers in Genetics-Computational Genomics***, 14:1235927 **(2023)**.

184. Waid-Ebbs, J. K., Wen, P-S, Grimes, T., Datta, S., Perlstein, W. M., Hammond , C., Daly, J. J. Executive function improvement in response to meta-cognitive training in chronic mTBI/PTSD. ***Frontiers in Rehabilitation Sciences*****,** 4:1189292 **(2023).** PMCID: PMC10360208

183. Anyaso-Samuel, S.* and Datta, S. Adjusting for informative cluster size in pseudo-value based regression approaches with clustered time to event data, ***Statistics in Medicine***, 42, 2162-2178. **(2023).** PMCID: PMC10219850

182. Ahn, S.*, Grimes, T., and Datta, S. A pseudo-value regression approach for differential network analysis of co-expression data. ***BMC Bioinformatics***, 24:8 **(2023).** PMCID: PMC9830718

181. Kang, T.*, Gaskins, J., Levy, S., and Datta, S. Analyzing dental fluorosis data using a novel Bayesian model for clustered longitudinal ordinal outcomes with an inflated category. ***Statistics in Medicine***, 42, 745-760 **(2023).**

180. Siriwardhana, C.*, Kulasekera, K. B., and Datta, S. Selection of the optimal personalized treatment from multiple treatments with right-censored multivariate outcome measures. ***Journal of Applied Statistics****,* to appear **(2023).** doi: 10.1080/02664763.2022.2164759.

179. Gregg, M., Datta, S., and Lorenz, D. htestClust: A package for marginal inference of clustered data under informative cluster size. ***The R Journal*****,** 14, 54-66 **(2022).**

178. Grimes, T.* and Datta, S. A novel probabilistic generator for large-scale gene association networks. ***PLOS One***, 16(11): e0259193. PMCID: PMC8589155 **(2021).**

177. Nevalainen, J., Datta, S., Toppari, J., Ilonen, J., Hyöty, H., Veijola, R., Knip, M., and Virtanen, S. M. Frailty modeling under a selective sampling protocol:an application to type 1 diabetes related autoantibodies. ***Statistics in Medicine,*** 40, 6410-6420 **(2021).**

176. Ahn, S.*, Grimes, T., and Datta, S. The analysis of gene expression data incorporating tumor purity information. ***Frontiers in Genetics,*** 12: 642759 **(2021).**

175. Bottari, S., Lamb, D. G., Murphy, A. J., Porges, E. C., Rieke, J. D., Harciarek, M., Datta, S., and Williamson, J. B. Hyperarousal symptoms and decreased right hemispheric frontolimbic white matter integrity predict poorer sleep quality in combat-exposed veterans, ***Brain Injury***, 35, 922-933 **(2021).**

174. Shen, B.*, Chen, C., Liu, D., Datta, S., Ghahramani, N., Chinchilli, V. M., and Wang, M. Joint modeling of longitudinal data with informative cluster size adjusted for zero-inflation and a dependent terminal event, ***Statistics in Medicine***, 40, 4582-4596 **(2021).**

173. Datta, S., Anyaso-Samuel, S.*, Sachdeva, A.*, and Subharup Guha, S., Metagenomic geolocation prediction using an adaptive ensemble classifier. ***Frontiers in Genetics,*** 12:642282 **(2021).**

172. Guha, S. and Datta, S. A Bayesian approach to restoring the duality between principal components of a distance matrix and operational taxonomic units in microbiome analyses. In ***Statistical Analysis of Microbiome Data,*** Eds: S. Datta and S. Guha, pp 271- 291, Springer **(2021).**

171. Anyaso-Samuel, S.*, Sachdeva, A.* , Subharup Guha, S., and Datta, S. Bioinformatics pre-processing of microbiome data with an application to metagenomic forensics. In ***Statistical Analysis of Microbiome Data,*** Eds: S. Datta and S. Guha, pp 45-78, Springer **(2021).**

170. Williamson, J. B., Lamb, D. G., Porges, E. C., Bottari, S., Woods, A. J., Datta, S., Langer, K., and Cohen, R. A. Cerebral metabolite concentrations are associated with cortical and subcortical volumes and cognition in older adults. ***Frontiers in Frontiers in Aging Neuroscience***, 12: 587104 **(2021).**

169. Kang, T.*, Gaskins, J., Levy, S., and Datta, S. A longitudinal Bayesian mixed effects model with hurdle Conway-Maxwell-Poisson distribution. ***Statistics in Medicine,*** 40, 1336-1356 **(2021)**.

168. Kang, T.*, Levy, S., and Datta, S. Analyzing longitudinal clustered count data with zero inflation: marginal modeling using the Conway-Maxwell-Poisson distribution. ***Biometrical Journal,*** 63, 761-786 **(2021)**. PMCID: PMC9167575

167. Grimes, T.* and Datta, S. SeqNet: an R package for generating gene-gene networks and simulating RNA-seq data. ***Journal of Statistical Software,*** 98: 12 **(2021).**

166. Gregg, M.*, Datta, S., and Lorenz, D. Variance estimation in tests of clustered categorical data with informative cluster size. ***Statistical Methods in Medical Research,*** 29, 3369-3408 **(2020).**

165. Boissoneault, C., Grimes, T.*, Rose, D. Waters, M., Khanna, A., Datta, S., and Daly, J. Innovative long-Dose neurorehabilitation for balance and mobility in chronic stroke: A preliminary case series. ***Brain Sciences***, 20, 555 **(2020).**

164. Siriwardena, C.*, Datta, S., and Kulasekera, K. B. Selection of the optimal personalized treatment from multiple treatments with multivariate outcome measures. ***Journal of Biopharmaceutical Statistics,*** 30, 462-480 **(2020)**.

163. Grimes, T.* and Datta, S. A random graph generation model for transcription networks and nonparametric simulator for RNA-Seq expression data. In ***Computer Data Analysis and Modeling, Stochastics and Data Science, Proceedings of the Twelfth International Conference, Minsk, Sepetember 18-22, 2019,*** Eds: P. Filzmoser and Y. Kharin, pp. 37-42. ISBN 978-985-566-811-5 **(2019)**. NIHMSID: 1647435

162. Siriwardena, C., Datta, S., and Kulasekera, K.B. Personalized treatment selection using data from crossover designs with carry over effects. ***Statistics in Medicine*****,** 38, 5391-5412 **(2019)**.

161. Grimes, T.*, Potter, S. S., and Datta, S. Integrating gene regulatory pathways into differential network analysis of gene expression data. ***Scientific Reports***, 9: 5479 **(2019)**.

160. Yan, X.*, Abdia, Y.*, Datta, S., Kulasekera, K. B., Ugiliweneza, B., Boakye, M., Kong, M. Estimation of average treatment effects among multiple treatment groups by using an ensemble approach. ***Statistics in Medicine,*** 38, 2828-2846 **(2019).**

159. Siriwardena, C.*, Zhao, M., Datta, S., and Kulasekera, K. B. A probability based method for selecting the optimal personalized treatment from multiple treatments. ***Statistical Methods in Medical Research*****,** 28, 749-760 **(2019)**.

158. Chen, Y.* and Datta, S. Adjustments of multi-Sample U-statistics to right censored data and confounding covariates. ***Computational Statistics and Data Analysis,*** 135, 1-14 **(2019)**.

157. Vahabi, N.*, Kazemnejad, A., and Datta, S. A joint overdispersed marginalized random effects model for analyzing two or more longitudinal ordinal responses. ***Statistical Methods in Medical Research****,* 28, 50-69 **(2019)**.

156. Vahabi, N., Kazemnejad, A., and Datta, S. A marginalized overdispersed location scale model for clustered ordinal data. ***Sankhyā,*** 80-B, S103-S134 **(2018)**.

155. Dutta, S.* and Datta, S. Rank based inference for covariate and group effects in clustered data in presence of informative intra-cluster group size. ***Statistics in Medicine***, 37, 4807-4822 **(2018)**.

154. Gregg, M.*, Datta, S., and Lorenz, D. A log-rank test for clustered data with informative within-cluster group size. ***Statistics in Medicine***, 37, 4071–4082 **(2018)**. PMID: 30003565.

153. Satten, G., Kong, M., and Datta, S. Multisample adjusted U-Statistics that account for confounding covariates. ***Statistics in Medicine***, 37, 3357–3372 **(2018)**. PMID: 29923344.

152. Sá, A.C.C.*, Webb, A., Gong, Y., McDonough, C.W., Shahin, M., Datta, S., Langaee, T.Y., Turner, S.T., Beitelshees, A.L., Chapman, A.B., Boerwinkle, E., Gums, J.G., Scherer, S.E, Cooper-DeHoff, R.M., Sadee, W., Johnson, J.A. Blood pressure signature genes and blood pressure response to Thiazide Diuretics: Results from the PEAR and PEAR-2 studies. ***BMC Medical Genomics***, 11, 55 **(2018)**. PMCID: PMC6011347.

151. Grimes, T.L.*, Walker, A.R., Datta, S., Datta, S. Predicting survival times for neuroblastoma patients using RNA-Seq expression profiles. ***Biology Direct***, 13, 11 **(2018)**. PMCID: PMC5977759.

150. Walker, A.R., Grimes, T.L.*, Datta, S., Datta, S. Unraveling bacterial fingerprints of city subways from microbiome 16S gene profiles. ***Biology Direct***, 13, 10 **(2018)**. PMCID: PMC5964687.

149. Choo-Wosoba, H.*, Gaskins, J., Levy, S. M., and Datta, S. A Bayesian approach for analyzing zero-inflated clustered count data with dispersion. ***Statistics in Medicine***, 37, 801-812 **(2018)**. PMCID: PMC5799048.

148. Siriwardena, C.*, Kulasekera, K. B., and Datta, S. Semi-parametric regression of state occupational probability in a multi-state model with right-censored data. ***Lifetime Data Analysis***, 24, 464–491 **(2018)**. PMCID: PMC5816729

147. Pesonen, M.*, Nevalainen, J., Potter, S. S, Datta, S., Datta, S. A Combined PLS and negative binomial regression model for inferring association networks from next-generation sequencing count data. ***IEEE/ACM Transactions on Computational Biology and Bioinformatics***, 15, 760-773 **(2018)**. PMCID: PMC5547023

146. Lorenz, D. J., Levy, S., and Datta, S. Inferring marginal association with paired and unpaired clustered data with applications to dental studies. ***Statistical Methods in Medical Research,*** 27, 1806-1817 **(2018)**. PMCID: PMC5524605.

145. Choo-Wosoba, H.* and Datta, S. Analyzing clustered count data with a cluster specific random effect zero-inflated Conway-Maxwell Poisson distribution. ***Journal of Applied Statistics*****,** 45, 799-814 **(2018).**

144. Sá, A.C.C.*, Webb, A., Gong, Y., McDonough, C. W., Datta, S., Langaee, T. Y., Turner, S. T., Beitelshees, A. L., Chapman, A. B., Boerwinkle, E., Gums, J. G., Scherer, S. E., Cooper-DeHoff, R. M., Sadee, W., Johnson, J. A. Whole transcriptome sequencing analyses reveal molecular markers of blood pressure response to Thiazide Diuretics. ***Scientific Reports***, 7, 16068. **(2017).** doi:10.1038/s41598-017-16343-z. PMCID: PMC5700078.

143. Sikdar, S.*, Datta, S., and Datta, S. EAMA: Empirically adjusted meta-analysis for large-scale simultaneous hypothesis testing in genomic experiments. ***PLOS One***, 12(10): e0187287 **(2017)**.

142. Chakraborty, S.*, Datta, S., and Datta, S. Nonparametric regression of state occupation probabilities in a multistate model. In ***Disease Modelling and Public Health***, Handbook of Statistics, volume 36, ***Chapter 7,*** pp. 175-203, *C.R. Rao, A. S. Rao, S. Pyne, eds.,* Elsevier, Amsterdam, Netherlands, **(2017)**.

141. Vahabi, N.*, Kazemnejad, A., Datta, S. Empirical Bayesian geographical mapping of occupational accidents among Iranian workers. ***Archives of Iranian Medicine***, 20, 5 **(2017).**

140. Sekula, M.*, Datta, S., and Datta, S. *optCluster*: An R package for determining the optimal clustering algorithm. ***Bioinformation*****,** 13, 100-103 **(2017).**

139. Nevalainen, J., Oja, H., and Datta, S. Tests for informative cluster size using a novel balanced bootstrap scheme. ***Statistics in Medicine***, 36, 2630-2640 **(2017)**.

138. Abdia, Y.*, Kulasekera, K. B., Datta, S., Boakye, M., and Kong, M. Propensity scores based methods for estimating average treatment effect and average treatment effect among treated: A comparative study. ***Biometrical Journal***, 59, 967-985 **(2017)**. PMID: 28436047.

137. Dutta, S.*, Datta, S. and Datta, S. Temporal prediction of future state occupation in a multistate model from high dimensional baseline covariates via pseudo-value regression. ***Journal of Statistical Computation and Simulation***, 87, 1363-1378 **(2017)**. PMCID: PMC5524605.

136. Datta, S. Robust regression analysis of longitudinal data under censoring. ***Austrian Journal of Statistics***, 46, 3-11 **(2017)**.

135. Lan, L., Bandyopadhyay, D., and Datta, S. Nonparametric regression in clustered multistate current status data with informative cluster size. ***Statistica Neerlandica,*** 71, 31-57 **(2017)**. PMID: 28436047.

134. Mertens, B. J. A., Datta, S., Brand, R. and Peul, W. Causal effect estimation strategies in a longitudinal study with complex time-varying confounders: A tutorial. ***Statistical Methods in Medical Research,*** 26, 337-355 **(2017)**. PMID: 25147227.

133. Siriwardhana, C.*, Datta, S., and Datta, S. Inter-platform concordance of gene expression data for the prediction of chemical mode of action. ***Biology Direct,*** 11:67 **(2016)**. PMCID: PMC5168706.

132. Sikdar, S.*, Datta, S., and Datta, S. Exploring the importance of cancer pathways by meta-analysis of differential protein expression networks in three different cancers. ***Biology Direct,*** 11:65 **(2016)**. PMCID: PMC5168844.

131. Chakraborty, S.* and Datta, S. A covariate adjusted Mann-Whitney test for comparing two sojourn times under right censoring. In ***Nonparametric Statistics,*** Springer Proceedings in Mathematics & Statistics, Vol. 175, Pages 117-134 (Wenceslao González Manteiga, Juan Romo, Ricardo Cao, eds.), Springer, New York, USA **(2016)**.

130. Dey, P., Bible, J.*, Dey, S. and Datta, S. Applications of feature selection and regression techniques in materials design: A Tutorial. In ***Computational Approaches to Materials Design: Theoretical and Practical Aspects***, Ch 8, (Shubhabrata Datta, ed), IGI Global, Hershey, USA, pp. 224-251 **(2016)**.

129. Bible, J.*, Beck, J. D., and Datta, S. Cluster adjusted regression for displaced subject data (CARDS): marginal inference under potentially informative temporal cluster size profiles. ***Biometrics,*** 72, 441-451 **(2016)**. PMCID: PMC4963003.

128. Dutta, S.* and Datta, S. A novel rank-sum test for clustered data when the number of subjects in a group within a cluster is informative. ***Biometrics,*** 72, 432-440 **(2016)**. PMCID: PMC4870168.

127. Choo-Wosoba, H.*, Levy, S. M., and Datta, S. Marginal regression models for clustered count data based on zero-inflated Conway-Maxwell-Poisson distribution with application to the Iowa Fluoride Study. ***Biometrics,*** 72, 606-618 **(2016)**. PMCID: PMC4948193.

126. Datta, S., Carmen Pardo, M., Scheike, T., and Yuen, K. C. Editorial: Special issue on advances in survival analysis. ***Computational Statistics and Data Analysis,*** 93, 255–256 **(2016).**

125. Datta, S. Empirical Bayes, In ***Wiley StatsRef: Statistics Reference Online***, Wiley, DOI: 10.1002/9781118445112.stat00257.pub2 **(2015)**.

124. Datta, S., Lorenz, D., and Datta, S. Approximate U-statistics for state waiting times under right censoring. In ***Modern Nonparametric, Robust and Multivariate Methods*** (K. Nordhausen, S. Taskinen, Eds.), Springer, pp. 31-46 **(2015)**.

123. Lorenz, D. J.* and Datta, S. A Nonparametric analysis of waiting times from a multistate model using a novel linear hazards model approach. ***Electronic Journal of Statistics***, 9, 419-443 **(2015)**.

122. Kong, M., Xu, S.*, Levy, S., and Datta, S. GEE type inference for clustered zero-inflated negative binomial regression with application to dental caries. ***Computational Statistics and Data Analysis***, 85, 54-66 **(2015)**.

121. Meira-Machado, L., Una-Alvarez, J. and Datta, S. Conditional transition probabilities in a non-Markov Illness-death Model. ***Computational Statistics***, 30, 377-397 **(2015)**.

120. Sikdar, S.*, Choo Wosoba, H.*, Abdia, Y*., Dutta, S.*, Gill, R., Datta, S., and Datta, S. An integrative exploratory analysis of –omics data from the ICGC cancer genomes lung adenocarcinoma study. ***Systems Biomedicine***, 2, 56-64 **(2014)**.

119. Shah, J. S.*, Datta, S. and Datta, S. A multi-loss super regression learner via bagging and rank aggregation with application to survival prediction using proteomics. ***Computational Statistics***, 29, 1749-1767 **(2014)**.

118. Gill, R., Datta, S. and Datta, S. dna: an R package for differential network analysis. ***Bioinformation***, 10, 233-234 **(2014)**. PMCID: PMC4070055.

117. Mitra, R., Gill, R., Datta, S., Datta, S. Statistical analysis of next generation sequencing data: an overview. In ***Statistical Analysis of Next Generation Sequencing Data*** (S. Datta and D. Nettleton, Eds.), Springer, pp. 1-24 **(2014)**.

116. Gill, R., Datta, S. and Datta, S. Bridging in vivo and in vitro data from Japanese Toxicogenomics Project using network analyses. ***Systems Biomedicine***, 2, e28527 **(2014)**.

115. Datta, S. and Beck, J. D. Robust estimation of marginal regression parameters in clustered data. ***Statistical Modelling*** (Special issue on *Statistical Methods in Oral Health Research*), 14, 489-501 **(2014)**. PMCID: PMC4384430.

114. Gill, R., Datta, S. and Datta, S. Differential network analysis in human cancer research. ***Current Pharmaceutical Design*** (Special issue on *Systems Biology*), 20, 4-10 **(2014)**. PMCID: PMC4868626.

113. Nevalainen, J., Datta, S. and Oja, H. Inference on the marginal distribution of clustered data with informative cluster size. ***Statistical Papers*** (Special issue on *Robust Methods for Dependent Data*), 55, 71-92 **(2014)**. PMCID: PMC4394393.

112. Dey, P., Bible, J.*, Datta, S., Broderick, S., Jasinski, J., Sunkara, M., Menon, M. and Rajan, K. Informatics-aided band gap engineering for solar materials. ***Computational Materials Science***, 83, 185–195 **(2014)**.

111. Andriotis, A. N., Mpourmpakis, G., Broderick, S., Rajan, K., Datta, S., Sunkara, M. and Menon, M. Informatics guided discovery of surface structure-chemistry relationships in catalytic nanoparticles. ***Journal of Chemical Physics,*** 140, 094705 **(2014)**. PMID: 24606374

110. Khudhair, M., Melloy, P., Lorenz, D. J., Obanor, F., Aitken, E., Datta, S., Luck, J., Fitzgerald, G., and Chakraborty, S. Fusarium crown rot under continuous cropping of susceptible and partially resistant wheat in microcosms at elevated CO_2. ***Plant Pathology,*** 63, 1033–1043 **(2014)**.

109. Zheng, H., Yang, G.* and Datta, S. A note on nonparametric estimation of a bivariate survival function under right censoring. In ***Contemporary Developments in Statistical Theory*** (S. N. Lahiri, A. Schick, A. Sengupta, T. N. Sriram, Eds.), Springer, pp. 69-84 **(2013)**.

108. Bible, J.*, Datta, S. and Datta, S. Cluster analysis: Finding groups in data. In ***Informatics for Materials Science and Engineering*** (K. Rajan, Ed.), Elsevier, pp 53-70 **(2013)**.

107. Chakraborty, S.*, Datta, S. and Datta, S. svapls: An R package to correct for residual expression heterogeneity in gene expression data. ***BMC Bioinformatics***, 14, 236 **(2013)**. PMCID: PMC3733742

106. Mostajabi, F.* and Datta, S. Nonparametric regression of state occupation, entry, exit and waiting times with multistate right censored data. ***Statistics in Medicine,*** 32, 3006-3019 **(2013)**. PMID: 23225570

105. Fan, J.* and Datta, S. Mann-Whitney tests for comparing sojourn time distributions when the transition times are right censored. ***Annals of the Institute of Statistical Mathematics***, 65, 149–166 **(2013)**.

104. Mostajabi, F.*, Datta, S., and Datta, S. Predicting patient survival from proteomic profiles using mass spectrometry data: An empirical study. ***Communications in Statistics - Simulation and Computation***, 42, 485-498 **(2013)**.

103. Habtzghi, D.* and Datta, S. One sample goodness of fit tests in presence of shape restrictions on the hazard rate function. ***Sankhya, Ser B***, 74-B, 171-194 **(2012)**.

102. Ferguson, A. N.*, Datta, S., Brock, G. msSurv, an R package for nonparametric estimation of multistate models. ***Journal of Statistical Software***, 50, 14 **(2012)**.

101. Lorenz, D. J.* and Datta, S. Comparing waiting times in a multistate model: A log-rank approach. ***Journal of Statistical Planning and Inference,*** 142, 2832-2843 **(2012)**.

100. Datta, S., Nevalainen, J. and Oja, H. A general class of signed rank tests for clustered data when the cluster size is potentially informative. ***Journal of Nonparametric Statistics***, 24, 797-808 **(2012)**. PMCID: PMC3467023

99. Chakraborty, S.*, Datta, S. and Datta, S. Surrogate variable analysis using partial least squares (SVA-PLS) in gene expression studies. ***Bioinformatics***, 28, 799-806 **(2012)**. PMID: 22238271

98. Satten, G. A. and Datta, S. Minimum distance type estimation of transformation parameters. ***Journal of Indian Statistical Association*** *(Golden Jubilee Volume)*, 50, 219-239 **(2012)**.

97. Datta, S., Lorenz, D. J., Harkema, S. J. A dynamic longitudinal evaluation of the utility of the Berg Balance Scale in patients with motor incomplete spinal cord injury. ***Archives of Physical Medicine and Rehabilitation***, 93, 1565-1573 **(2012)**. PMID: 22920453

96. Lorenz, D. J., Datta, S., Harkema, S. J. Longitudinal patterns of functional recovery in patients with incomplete spinal cord injury receiving activity-based rehabilitation. ***Archives of Physical Medicine and Rehabilitation***, 93, 1541-1552 **(2012)**. PMID: 22920451

95. Forrest, G. F., Lorenz, D. J., Hutchinson, K., Van Hiel, L., Basso, D. M., Datta, S., Sisto, S. A., Harkema, S. J. Relationships between balance and walking measures at baseline and after locomotor training in incomplete SCI: impact of functional recovery. ***Archives of Physical Medicine and Rehabilitation***, 93, 1553-1564 **(2012)**. PMID: 22920452

94. Datta, S. and Ferguson, A. N.* Nonparametric estimation of marginal temporal functionals in a multistate model. In ***Recent Advances in System Reliability: Signatures, Multi-state Systems and Statistical Inference*** (Ilia Frenkel and Anatoly Lisnianski, Eds.), Springer, London, pp. 219-236 **(2011)**.

93. Lorenz, D. J., Datta, S. and Harkema, S. J. Marginal association measures for clustered data. ***Statistics in Medicine***, 30, 3181-3191 **(2011)**. PMCID: PMC3430979

92. Fan, J.* and Datta, S. Fitting accelerated failure time models to clustered survival data with potentially informative cluster size. ***Computational Statistics & Data Analysis***, 55, 3295-3303 **(2011)**.

91. Pihur, V., Datta, S. and Datta, S. Meta analysis of chronic fatigue syndrome through integration of clinical, gene expression, SNP and proteomic data, ***Bioinformation***, 6, 120-124 **(2011)**. PMCID: PMC3089886

90. Datta, S. and van Houwelingen, H. C. Statistics in biological and medical sciences, Editorial. ***Statistics & Probability Letters***, 81, 715-716 **(2011)**.

89. Wang, M.*, Kong, M. and Datta, S. Inference for marginal linear models with clustered longitudinal data with potentially informative cluster sizes. ***Statistical Methods in Medical Research,*** 20, 347–367 **(2011)**. PMID: 20223781

88. Datta, S., Datta, S., Kim, S., Chakraborty, S. and Gill, R. S. Statistical analyses of next generation sequence data: A partial overview. ***Journal of Proteomics & Bioinformatics***, 3, 183-190 **(2010)**. PMCID: PMC2989618

87. Datta, S, Pihur, V. and Datta, S. An adaptive optimal ensemble classifier via bagging and rank aggregation with applications to high dimensional data, ***BMC Bioinformatics***, 11, 427 **(2010)**. PMCID: PMC2933716

86. Gill, R., Datta, S. and Datta, S. A statistical framework for differential network analysis from microarray data using partial least squares, ***BMC Bioinformatics***, 11, 95 **(2010)**. PMCID: PMC2838870

85. Lan, L.* and Datta, S. Comparison of state occupation, entry, exit and waiting times in K independent multistate models under current status data. ***Statistics in Medicine***, 29, 906 - 914 **(2010)**.

84. Datta, S., Bandyopadhyay, D.* and Satten, G. A. Inverse probability of censoring weighted U-statistics for right censored data with applications. ***Scandinavian Journal of Statistics***, 37, 680–700 **(2010)**.

83. Lan, L.* and Datta, S. Nonparametric estimation of state occupation, entry and exit times with multistate current status data. ***Statistical Methods in Medical Research***, 19, 147-165 **(2010)**.

82. Pihur, V.*, Datta, S. and Datta, S. RankAggreg, an R package for weighted rank aggregation. ***BMC Bioinformatics***, 10, 62 **(2009)**. PMCID: PMC2669484

81. Datta, S. and Datta, S. Computational biology touches all bases. ***Genome Biology***, 10, 303 **(2009)**. PMCID: PMC2688279

80. Datta, S., Lan, L.* and Sundaram, R. Nonparametric estimation of waiting time distributions in a Markov model based on current status data, ***Journal of Statistical Planning and Inference***, 139, 2885-2897 **(2009)**.

79. Pihur, V.*, Brock, G., Datta, S. and Datta, S. Cluster validation for microarray data: An appraisal. In ***Advances in Multivariate Statistical Methods***, (A. SenGupta, ed), Ch 5, 79-94, World Scientific Press. **(2009)**.

78. Datta, S., Lorenz, D. J., Morrison, S., Ardolino, E., Harkema, S. J. A multivariate examination of temporal changes in Berg variables for patients with AIS C and D spinal cord injuries. ***Archives of Physical Medicine and Rehabilitation***, 90, 1208-1217 **(2009)**. PMID: 19577035

77. Pihur, V.*, Datta, S. and Datta, S. Finding cancer genes through meta-analysis of microarray experiments: Rank aggregation via the cross entropy algorithm. ***Genomics***, 92, 400-403 **(2008)**. PMID: 18565726

76. Pihur, V.*, Datta, S. and Datta, S. Reconstruction of genetic association networks from microarray data: A partial least squares approach. ***Bioinformatics***, 24, 561-568 **(2008)**. PMID: 18204062

75. Datta, S. Classification of breast cancer versus normal samples from mass spectrometry profiles using linear discriminant analysis of important features selected by Random Forest. ***Statistical Applications in Genetics and Molecular Biology***, 7 (2), Article 7 **(2008)**. PMID: 18312221

74. Brock, G., Pihur, V.*, Datta, S. and Datta, S. clValid, an R package for cluster validation. ***Journal of Statistical Software***, 25, 4 **(2008)**.

73. Datta, S., Datta, S., Parrish, R. S. and Thompson, C. M. Microarray data analysis, In ***Computational Methods in Biomedical Research***, (R. Khatree and D. Naik, eds.), Chapman & Hall/CRC Biostatistics Series, Volume 24, 1-43 **(2008)**.

72. Datta, S. and Satten, G. A. A signed-rank test for clustered data. ***Biometrics***, 64, 501-507 **(2008)**. PMID: 17970820.

71. Bandyopadhyay, D.* and Datta, S. A novel approach to testing equality of survival distributions when the population marks are missing. ***Journal of Statistical Planning and Inference***, 138, 1722-1732 **(2008)**.

70. Pihur, V., Datta, S. and Datta, S. Understanding Chronic Fatigue Syndrome (CFS) from CAMDA data: A systems biology approach. In ***CAMDA 2007 Proceedings***, online @ http://camda.bioinfo.cipf.es/camda07/agenda/detailed.html **(2007)**.

69. Pihur, V.*, Datta, S. and Datta, S. Weighted rank aggregation of cluster validation measures: A Monte Carlo cross-entropy approach. ***Bioinformatics***, 23, 1607-1615 **(2007)**. PMID: 17483500

68. Boratyn, G. M., Datta, S. and Datta, S. Incorporation of biological knowledge into distance for clustering genes. ***Bioinformation***, 1, 396-405 **(2007)**. PMC1896054

67. Datta, S., Le-Rademacher, J.* and Datta, S. Predicting patient survival from microarray data by accelerated failure time modeling using partial least squares and LASSO. ***Biometrics***, 63, 259-271 **(2007)**. PMID: 17447952

66. Zheng, H.*, Basawa, I. V. and Datta, S. First order random coefficient autoregressive processes. ***Journal of Statistical Planning and Inference***, 173, 212 - 229 **(2007)**.

65. Johnson, S. B.*, Datta, S., Hornung, C. A., Casanova, M. F. Mathematical models of epigenetic influences in Autism: a new perspective based on neuropathological findings. In ***Progress in Autism Research***, (Paul C. Carlisle, ed), Nova Science Publishers, Inc., 101-114, New York: New York **(2007)**.

64. Datta, S., and Datta, S. Combining functional information in selecting clustering algorithms. In ***Proceedings of Interface 2005***, on CD-ROM **(2006)**.

63. Datta, S. and Datta, S. Evaluation of clustering algorithms for gene expression data, ***BMC Bioinformatics***, 7 (Suppl 4), S17, **(2006)**.

62. Datta, S. and Datta, S. Methods for evaluating clustering algorithms for gene expression data using a reference set of functional classes, ***BMC Bioinformatics***, 7, 397 **(2006)**. PMCID: PMC1780133

61. Boratyn, G. M., Datta, S. and Datta, S. Biologically supervised hierarchical clustering algorithms for gene expression data, In ***Proceedings of the 28th IEEE EMBS Annual International Conference***, New York City, USA, 5515-5518 **(2006)**. PMID: 17947147

60. Zheng, H.*, Basawa, I. V. and Datta, S. The p-th order random coefficient autoregressive processes, ***Journal of Time Series Analysis***, 27, 411-440 **(2006)**.

59. Datta, S. and Sundaram, R. Nonparametric marginal estimation in a multistage model using current status data, ***Biometrics***, 62, 829-837 **(2006)**. PMID: 16984326

58. Datta, S. and Datta, S. Validation measures for clustering algorithms incorporating biological information, In ***IEEE Proceedings of International Multi-Symposiums on Computer and Computational Sciences***, (J. Ni, J. Dongarra, Y. Zheng, G. Gu, G. Wolfgang and H. Jin, eds.), 1, 131-135 **(2006)**.

57. Datta, S. Estimating the mean life time using right censored data. ***Statistical Methodology***, 2, 65-69 **(2005)**.

56. Datta, S. and Datta, S. Empirical Bayes screening (EBS) of many p-values with applications to microarray studies, ***Bioinformatics***, 21,1987-1994 **(2005)**. PMID: 15691856

55. Datta, S. and Satten, G. A. Rank-sum tests for clustered data. ***Journal of the American Statistical Association***, 100, 908-915 **(2005)**.

54. Datta, S. Bootstrapping, In ***Encyclopedia of Statistical Sciences***, Second edition, Wiley, **(2005)**.

53. Datta, S. Empirical Bayes methods, In ***Encyclopedia of Statistical Sciences***, Second edition, Wiley, **(2005)**.

52. Satten, G. A., Datta, S., Moura, H., Woolfitt, A., Carvalho, G., De, B. K, Pavlopoulos, A., Carlone, G. M., and Barr, J. Standardization and denoising algorithms for mass spectra to classify whole-organism bacterial specimens, ***Bioinformatics***, 20, 3128-3136 **(2004)**. PMID: 15217815

51. Datta, S. and Datta, S. An empirical Bayes adjustment to multiple p-values for the detection of differentially expressed genes in microarray experiments. In ***Conferences in Research and Practice in Information Technology***, (Y-P. P. Chen, ed.), 29, 155-159 **(2004)**.

50. Datta, S., Satten, G. A., Benos, D. J., Xia, J., Heslin, M., and Datta, S. An empirical Bayes adjustment to increase the sensitivity of detecting differentially expressed genes in microarray experiments, ***Bioinformatics***, 20, 235-242 **(2004)**. PMID: 14734315

49. Satten, G. A. and Datta, S. Marginal Analyses of Multistage Data. In ***Advances in Survival Analysis,*** *Handbook of Statistics* (N. Balakrishnan and C. R. Rao, eds.), 23, 559-574, Elsevier-North Holland **(2004)**.

48. Datta, S. and Datta, S. Comparisons and validation of statistical clustering techniques for microarray gene expression data. ***Bioinformatics***, 19, 459-466 **(2003)**. PMID: 12611800

47. Chakraborty, S. and Datta, S. How will plant pathogens adapt to host plant resistance at elevated CO_2 under a changing climate? ***New Phytologist***, 159, 733-742 **(2003)**.

46. Williamson, J., Datta, S., and Satten, G. A. Marginal analyses of clustered data when cluster size is informative. ***Biometrics***, 59, 36-42 **(2003)**.

45. Datta, S. and Satten, G. A. Estimation of integrated transition hazards and stage occupation probabilities for non-Markov systems under stage dependent censoring. ***Biometrics***, 58, 792-802 **(2002)**. PMID: 12762439

44. Satten, G. A. and Datta, S. Marginal estimation for Multistage models: waiting time distributions and competing risk analyses. ***Statistics in Medicine***, 21, 3-19 **(2002)**. PMID: 11782047

43. Datta, S., Satten, G. A. and Datta, S. Estimation of stage occupation probabilities in multistage models. In ***Advances on Theoretical and Methodological Aspects of Probability and Statistics***, (N. Balakrishnan, ed.), 493-506, Taylor & Francis, New York, NY **(2002)**.

42. Datta, S. and Satten, G. A. Validity of the Aalen-Johansen estimators of stage occupation probabilities and integrated transition hazards for non-Markov models. ***Statistics and Probability Letters***, 55, 403-411 **(2001)**.

41. Satten, G. A., Datta, S. and Robins, J. M. An estimator for the survival function when data are subject to dependent censoring. ***Statistics and Probability Letters***, 54, 397-403 **(2001)**.

40. Satten, G. A. and Datta, S. The Kaplan-Meier Estimator as an inverse-probability-of-censoring weighted average. ***American Statistician***, 55, 207-210 **(2001)**.

39. Williamson, J. M., Satten, G. A., Hanson, J. A., Weinstock, H., and Datta, S. Analysis of dynamic cohort data. ***American Journal of Epidemiology***, 154, 366-372 **(2001)**. PMID: 11495860

38. Li, G. and Datta, S. A bootstrap approach to nonparametric regression for right censored data. ***Annals of the Institute of Statistical Mathematics***, 53, 708-729 **(2001)**.

37. Datta, S. and Satten, G. A. Estimating future stage entry and occupation probabilities in a multistage model based on randomly right-censored data. ***Statistics and Probability Letters***, 50, 89-95 **(2000)**.

36. Satten, G. A. and Datta, S. A simulate-update algorithm for missing data problems. ***Computational Statistics***, 15, 243-277 **(2000)**.

35. Datta, S. Empirical Bayes estimation with non-identical components. ***Journal of Nonparametric Statistics***, 12, 709-725 **(2000)**.

34. Datta, S., Satten, G. A. and Datta, S. Nonparametric estimation for the three-stage irreversible illness-death model. ***Biometrics***, 56, 841-847 **(2000)**. PMID: 10985224

33. Datta, S., Satten, G. A. and Williamson, J. M. Consistency and asymptotic normality of estimators in a regression model with interval censoring and left truncation. ***Annals of the Institute of Statistical Mathematics***, 52, 160-172 **(2000)**.

32. Satten, G. A., Janssen, R., Busch, M. P., and Datta, S. Validating marker-based incidence estimates in repeatedly screened population. ***Biometrics***, 55, 1224-1227 **(1999)**. PMID: 11315072

31. Allen, M. R.* and Datta, S. Estimation of the index parameter for autoregressive data using the estimated innovations. ***Statistics and Probability Letters***, 41, 315-324 **(1999)**.

30. Satten, G. A. and Datta, S. Kaplan-Meier representation of competing risk estimates. ***Statistics and Probability Letters***, 42, 299-304 **(1999)**.

29. Allen, M.* and Datta, S. A note on bootstrapping M-estimators in ARMA models. ***Journal of Time Series Analysis***, 20, 365-380 **(1999)**.

28. Bagui, S. C. and Datta, S. Some useful properties of the Bayes risk in classification. ***Calcutta Statistical Association Bulletin***, 48, 83-91 **(1998)**.

27. Datta, S., Mathew G. and McCormick, W. P. Nonlinear autoregression with positive innovations. ***Australian & New Zealand Journal of Statistics***, 40, 229-239 **(1998)**.

26. Satten, G. A. and Datta, S. and Williamson, J. M. A semiparametric approach to the proportional hazards model for interval censored data. ***Journal of the American Statistical Association***, 93, 318-327 **(1998)**.

25. Datta, S. and McCormick, W. P. Inference for the tail parameters of a linear process with heavy tailed innovations. ***Annals of the Institute of Statistical Mathematics***, 50, 337-359 **(1998)**.

24. Datta, S. Making the bootstrap work. In ***Frontiers in Probability and Statistics***, (S. P. Mukherjee, S. K. Basu and B. K. Sinha, eds), 119-129, Nasora Publishing, Narosa, New Delhi, India **(1998)**.

23. Datta, S. and Hannan, J. F. A uniform L_1 law of large numbers for functions on a totally bounded metric space. ***Sankhya A***, 59, 167-174 **(1997)**.

22. Datta, S. L_1 density estimation for linear processes. ***Journal of Time Series Analysis***, 18, 375-383 **(1997)**.

21. Datta, S. and Sriram, T. N. A modified bootstrap for autoregression without stationarity. ***Journal of Statistical Planning and Inference***, 59, 19-30 **(1997)**.

20. Datta, S. On asymptotic properties of bootstrap for AR(1) processes. ***Journal of Statistical Planning and Inference***, 53, 361-374 **(1996)**.

19. Datta, S. and McCormick, W. P. Bootstrap inference for a first order autoregression with positive innovations. ***Journal of the American Statistical Association***, 90, 1289-1300 **(1995)**.

18. Datta, S. Limit theory and bootstrap for explosive and partially explosive autoregression. ***Stochastic Processes and Their Applications***, 57, 285-304 **(1995)**.

17. Datta, S. and Sriram, T. N. A modified bootstrap for branching processes with immigration. ***Stochastic Processes and Their Applications***, 56, 275-294 **(1995)**.

16. Datta, S. On a modified bootstrap for certain asymptotically non-normal statistics. ***Statistics and Probability Letters***, 24, 91-98 **(1995)**.

15. Datta, S. A minimax optimal estimator for continuous monotone densities. ***Journal of Statistical Planning and Inference***, 46, 181-193 **(1995)**.

14. Datta, S. Consistency of the mle for a general sequential design problem. ***Sankhya A***, 57, 88-99 **(1995)**.

13. Datta, S. and McCormick, W. P. Some continuous Edgeworth expansions for Markov chains with applications to bootstrap. ***Journal of Multivariate Analysis***, 52, 83-106 **(1995)**.

12. Datta, S. Empirical Bayes estimation in a threshold model. ***Sankhya A***, 54, 106-117 **(1994)**.

11. Basawa, I. V. and Datta, S. Large sample estimation for nested models. ***Journal of the Indian Society of Probability and Statistics***, 1, 19-42 **(1994)**.

10. Datta, S. A solution to the set compound problem with certain non regular components. ***Statistics & Decisions***, 11, 343-355 **(1993)**.

9. Datta, S. and McCormick, W. P. Regeneration based bootstrap for Markov chains. ***Canadian Journal of Statistics***, 21, 181-193 **(1993)**.

8. Datta, S. and McCormick, W. P. On first order Edgeworth expansions for a Markov chain. ***Journal of Multivariate Analysis***, 44, 345-359 (**1993)**.

7. Datta, S. Some non asymptotic bounds for L_1 density estimation using kernels. ***Annals of Statistics***, 20, 1658-1667 **(1992)**.

6. Bhat, B. R. and Datta, S. On the completeness of a family of conditional distributions. ***Statistics and Probability Letters***, 15, 27-30 **(1992)**.

5. Datta, S. A note on continuous Edgeworth expansions and the bootstrap. ***Sankhya A***, 54, 171-182 **(1992)**.

4. Datta, S. and McCormick, W. P. Bootstrap for a finite state Markov chain based on i.i.d. resampling. In ***Exploring the Limits of Bootstrap***, (L. LePage and L. Billard, eds), 77-97 , Wiley, New York, **(1992)**.

3. Datta, S. Nonparametric empirical Bayes estimation with $O(n^{-\frac{1}{2}})$ rate of a truncation parameter. ***Statistics & Decisions***, 9, 45-61 **(1991)**.

2. Datta, S. Asymptotic optimality of Bayes compound estimators in compact exponential families. ***Annals of Statistics***, 19, 354-365 **(1991)**.

1. Datta, S. On the consistency of posterior mixtures and its application. ***Annals of Statistics***, 19, 338-353 **(1991)**.

R PACKAGES

1. *sohpie*, Seungjun Ahn and Somnath Datta, https://cran.r-project.org/web/packages/SOHPIE/index.html, 2023.
Performs differential co-abundance network analysis of microbiome data.

2. *PRANA*, Seungjun Ahn and Somnath Datta, https://cloud.r-project.org/web/packages/PRANA/index.html, 2023
Performs differential network analysis with expression data.

3. *htestClust*, Mary Gregg , Somnath Datta, and Doug Lorenz, https://cran.r-project.org/web/packages/htestClust/index.html, 2022.
A collection of reweighted marginal hypothesis tests for clustered data.

4. *dnapath*, Tyler Grimes, Somnath Datta, https://cran.r-project.org/src/contrib/Archive/dnapath, 2020.
Integrates pathway information into the differential network analysis of two gene expression datasets.

5. *SeqNet*, Tyler Grimes, Somnath Datta, https://cran.r-project.org/web/packages/SeqNet/index.html, 2019.
Generates random gene-gene association networks and simulate RNA-seq data from them.

6. *RankAggreg*, Vasyl Pihur, Somnath Datta, and Susmita Datta, https://cran.r-project.org/web/packages/RankAggreg/index.html, 2018.
Performs aggregation of ordered lists based on the ranks using several different algorithms.

7. *optCluster,* Michael Sekula, Somnath Datta, and Susmita Datta, https://cran.r-project.org/web/packages/optCluster/index.html, 2017.
Package for cluster analysis using statistical and biological validation measures for both continuous and count data.

8. *ClusterRankTest,* Sandipan Dutta and Somnath Datta, https://cran.r-project.org/web/packages/ClusterRankTest/index.html, 2016.
Performs Nonparametric rank based tests for clustered data, especially useful for clusters having informative cluster size and intra-cluster group size.

9. *msSurv*, Nicole Ferguson, Somnath Datta, Guy Brock, https://cran.r-project.org/web/packages/msSurv/index.html, 2015.
Package for nonparametric estimation for right censored, left truncated time to event data in multistate models.

10. *dna,* Ryan Gill, Somnath Datta, Susmita Datta, https://cran.r-project.org/web/packages/dna/index.html, 2014.
Package for conducting differential network analysis from gene expression data.

11. *clValid*, Guy Brock, Vasyl Pihur, Susmita Datta, and Somnath Datta, https://cran.r-project.org/web/packages/clValid/index.html, 2014.
Performs statistical and biological validation of clustering results.

12. *svapls,* Sutirtha Chakraborty, Somnath Datta and Susmita Datta, https://cran.r-project.org/src/contrib/Archive/svapls, 2013.

For accurate identification of genes that are truly differentially expressed over two sample varieties, after adjusting for hidden subject-specific effects of residual heterogeneity.

FIVE MOST CITED PAPERS (Google Scholar, Accessed 9/10/24)
My research are mostly in statistical methods (including some with applications to genomics and oral health research) and statistical theory. As a result, the number of authors in my publications is small and I am usually the senior author.

1. Brock, Pihur, Datta, and Datta (2008), *Journal of Statistical Software*, #74; Count = 1081
2. Datta and Datta (2003), *Bioinformatics,* #48; Count = 529
3. Pihur, Datta, and Datta (2009), *BMC Bioinformatics,* #82; Count = 370
4. Pihur, Datta, and Datta (2007), *Bioinformatics,* #69, Count = 317
5. Chakraborty and Datta (2003), *New Phytologist*, #47, Count = 279

Total citations > 8000, h-index 43, i10-index 113

CONTRIBUTION TO SCIENCE

1. Over the last five years, I have been conducting secondary analysis of periodontal disease and caries data from the Piedmont 65+ Dental Study and the Iowa Fluoride Study, respectively. In the process, I am developing novel statistical methods to account for count data with zero inflation, clustered data and informative cluster size which appear to be present in these types of data.

Representive publications: 130, 129, 128, 115

2. Cluster analysis is being heavily used in diverse area of science as an exploratory tool, notably in high throughput genomic assays such as microarrays and other sequencing platforms. A related area is called classification where the learning is supervised. Over last ten years our group has developed a number of cluster validation measures and bioinformatics tools (e.g., R packages) for the most appropriate cluster analysis for a given dataset and for building optimal classifiers. Several of these papers have been extremely popular as can be seen from their citation counts and the resulting tools are being used by many scientists. I have been a senior author of these publications.

Representive publications: 82, 74, 62, 48

3. Genomic association network analysis is becoming progressively important in biological and medical research since genes typically act in consort. Our team has developed various bioinformatics methods and tools for reverse engineering of genomic networks and for detecting how they change between two or more biological conditions. I have also co-edited a book on statistical methods for analyzing NGS data.

Representive publications: 161, 118, 114, 86

4. We have used combination of statistical modeling and physics based calculation to predict properties of novel materials for various applications such as solar cells and magnetic storage.

Representive publications: 131, 112, 111, 108

5. I had served as the senior biostatistician for the NeuroRecovery Network (NRN) headed by Susan Harkema. Activity-based interventions are emerging as a more successful approach for functional recovery after neurologic injury. The NeuroRecovery Network (NRN), a specialized network of treatment centers providing standardized, activity based therapy for spinal cord injured (SCI) patients. We have developed an effective multi-dimensional scale of functional recovery of NRN patient using the locomotor training. Currently, I am working as the senior biostatistician for the Brain Rehabilitation Research Center at NF/SGVHS Malcom Randall VA Medical Center, Gainesville, FL.

Representive publications: 97, 96, 95, 78

NCBI MyBibliography:

http://www.ncbi.nlm.nih.gov/sites/myncbi/somnath.datta.1/bibliography/42187443/public/?sort=date&direction=ascending

EXTERNAL RESEARCH FUNDING (since 1995)

PI Level

21. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". October 2023 - September 2024, Role: Principal Investigator (VA PIs: D. Clark, R. M. Bauer), 2.64 cal months. $ 99,668; salary support $68,923, plus 100% PhD student support.

20. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". October 2022 - September 2023, Role: Principal Investigator (VA PIs: D. Clark, R. M. Bauer), 2.64 cal months. Salary support $67,795, plus 40% PhD student support. 2.64 cal months. Salary support $68,922, plus 40% PhD student support.

19. **National Institutes of Health**, "Longitudinal Analysis of Iowa Fluoride Study Data, Including at Age Twenty-three", February 2022 - January 2024. Role: Principal Investigator, 2.04 cal months. Salary, student, travel and other support. $325,242.

18. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". October 2021 - September 2022, Role: Principal Investigator (VA PIs: D. Clark, R. M. Bauer), 2.64 cal months. Salary support $67,795, plus 40% PhD student support.

17. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". October 2020 - September 2021, Role: Principal Investigator (VA PIs: D. Clark, R. M. Bauer), 2.64 cal months. Salary support $67,141, plus 40% PhD student support.

16. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research Projects". September 2019 - June 2020, Role: Principal Investigator (VA PIs: D. Clark, J Waid-Ebbs, J Daly), 3.24 cal months. Salary support $67,060, plus 40% PhD student support.

15. **National Institutes of Health**, "A Novel Analysis Plan for the Caries and Fluorosis Data from the Iowa Fluoride Study", August 2018 - January 2022 (NCE). Role: Principal Investigator, 2.4 cal months. Salary, student, travel and other support. $332,299.

14. **U.S. Department of Veterans Affairs IPA**, "Analysis of Traumatic Brain Injury and other Brain Rehabilitation Data". August 2018 - August 2019, Role: Principal Investigator (VA PIs: J Waid-Ebbs, J Daly), 2 cal months. Salary support $59,882, plus 40% PhD student support.

13. **U.S. Department of Veterans Affairs IPA**, "Brain Rehabilitation Research". September 2016 - August 2018, Role: Principal Investigator (VA PI: J Daly), 2.4 cal months. Salary support $115,270, plus MS level statistician support.

12. **National Institutes of Health**, "Exploratory Statistical Analysis of Differential Network Behaviors based on Gene Expression Atlas of Palate Development", August 2016 - July 2020 (NCE). Role: Principal Investigator, 3.6 cal months. Salary, student, travel and other support. $460,449.

11. **National Institutes of Health**, "Novel Statistical Models for Dental Caries", July 2012 - June 2015. Role: Principal Investigator, 3 cal months. Salary, student, travel and other support. $298,551.

10. **National Institutes of Health**, "Rank Tests for Clustered Data with Potentially Informative Cluster Size: Novel Statistical Methods for Analyzing Dental Data", September 2011 - December 2014. Role: Principal Investigator, 1.2 - 2.4 cal months. Salary, student, travel and other support. $313,657.

9. **National Science Foundation**, "SOLAR: New Materials Search for Solar Energy Conversion to Fuels", September 2011 - August 2015. Role: Co-Principal Investigator; Awarded jointly with M. Sunkara (Louisville), M. Menon (Kentucky) and K. Rajan (Iowa State). 1 cal month. Salary, post-doc, student, travel and other support. $1,100,000.

8. **National Security Agency**, "Nonparametric Regression of State Occupation Probabilities, State Entry, Exit and Waiting Time Distributions in a Multistate Model", March 2011 - February 2013. Role: Principal Investigator, 1 cal month. Salary, student and travel support. $73,805.

7. **National Science Foundation**, "Theory and Applications of U-statistics for Multistate Models under Censoring", July 2007 - June 2011. Role: Principal Investigator, 1 cal month. Salary, student and travel. $100,782.

6. **National Security Agency**, "Nonparametric Inference in Censored Data Problems", Jan 2005 - Dec 2006, Role: Principal Investigator, 1 cal month. Salary, student, travel and computing support.

5. **Centers for Disease Control and Prevention**, "Problems in Genetic Epidemiology", June 2001 - May 2005, Role: Principal Investigator, 3 cal months. Salary support.

4. **National Security Agency**, "Large Sample Theory of Inverse Probability of Censoring Weighted Estimation in Multistage and Mixed Linear Models", February 2003 - January 2005, Role: Principal Investigator, 1 cal month. Salary, travel and computing support.

3. **Centers for Disease Control and Prevention**, "Analysis of Complex Survival Data", February 1997 - August 2000, Role: Principal Investigator, 3 cal months. Salary support.

2. **National Security Agency**, "Inference, Bootstrap and Curve Estimation for Time Series Data", April 1996 -March 1998, Role: Principal Investigator, 1 cal month. Salary, travel and computing support.

1. **National Science Foundation**, "Mathematical Sciences Computing Research Environments", August 1995 - July 1996, Role: Co-Principal Investigator (awarded jointly with L. Billard and T. N. Sriram). Computing support.

Non PI Level

9. **National Institutes of Health,** "WISE II - Obesity and Type-2 Diabetes: Bariatric Surgery Effects on Brain Function", PIs. R. Cohen, E. Porges, J. Williamson, March 2024- December 31, 2028. Role: Co-Investigator, 0.6 cal months. Salary support.

8. **National Institutes of Health,** "Optimizing AAV Vectors for Central Nervous System Transduction", PI: C. D Heldermon, August 2017-May 2022. Role: Co-Investigator, 0.6 cal months. Salary support.

7. **National Institutes of Health,** "Assessment of Locomotor Potential Following Stroke", PI: C. Patten, September 2015 - August 2018. Role: Co-Investigator, 1.2 cal months. Salary support.

6. **National Institutes of Health,** "Interactions between Microglia and Dopaminergic Neurons Regulates Dopamine Neurotransmission", PI: H. Khoshbouei, March 2017-February 2019, Role: Co-Investigator, 0.6 cal months. Salary support.

5. **Christopher and Dana Reeve Foundation**, "Development of Neural Recovery Rehabilation and Research Centers", PI: S. Harkema, August 2006 - November 2015, Role: Senior Biostatistician, 0.3 cal months - 4.8 cal months. Salary support.

4. **National Institutes of Health,** "Gross Morphological Correlates to the Minicolumnopathy of Autism", PI: M. Casanova, September 2009 - August 2011. Role: Co-Investigator, 1.2 cal months. Salary support.

3. **National Institutes of Health,** "Plasticity of Human Spinal Neural Networks After Injury", PI: S. Harkema, January 2007 - March 2009, Role: Principal Statistician, 1.2 cal months. Salary support.

2. **National Institutes of Health,** "Outcomes of Teacher Training on Autism", PI: L. Ruble, 2005 - 2008, Role: co-Investigator, 0.6 cal months. Salary support.

1. **National Institutes of Health,** "Efficient Estimation Methods for Censored Survival Data", PI: S. Subramanian, April 2004 - March 2007, Role: Consultant. Flat Fee.

OTHER FUNDING (since 1995)

4. **National Science Foundation**, "Symposium on Statistical Innovation in the Era of Artificial Intelligence and Data Science" (jointly with S. Wu), May 2022, conference grant.

3. **National Science Foundation**, "IISA 2018: From Data to Knowledge, Working for a Better World" (jointly with S. Basu and D. Bandyopadhyay), May 2018, conference grant.

2. **National Science Foundation**, "Statistical Inference for Biomedical Big Data: Theory, Methods and Tools" (jointly with F. Liang, P. Qiu and F. Zou), April 2017, conference grant.

1. **Elsevier,** Editorial Contract, "Statistics & Probability Letters", June 2007 - June 2011. Role: Co-Editor-in-Chief. Honorarium, student, travel and other support.

INVITED TALKS: Conferences

111. Fifth International Workshop on Statistical Analyses of Multi-Outcome Data, July 9-10, 2024, Salzburg, Austria. "Specialized Statistical Analyses of Iowa Fluoride Study Data".

110. Small Area Estimation, Surveys and Data Science, SAE 2023-2024, June 3-7, 2024, Lima, Peru. "Differential co-abundance network analyses for microbiome data adjusted for clinical covariates using jackknife pseudo-values".

109. IMS-APRM 2024, January 4 -7, Melbourne, Australia. "A pseudo-value regression approach for differential network analysis of co-expression data".

108. 2023 ASA Lifetime Data Science Conference, June 1-2, 2023, Raleigh, North Carolina, USA. "Regression Analysis of a Future State Entry Time Distribution Conditional on a Past State Occupation in a Progressive Multistate Model ".

107. 35th Panhellenic – 1st International Statistics Conference organized by the
University of West Attica and the Greek Statistical Institute, May 25-28, 2023, online.
"Regression analysis of a future state entry time distribution conditional on a past state occupation under right censoring and current status data".

106. International Conference on Statistics (ICS 2023), March 20-23, 2023, Mumbai, India, "Marginal models for clustered data and the issue of informative cluster and subgroup sizes".

105. International Indian Statistical Association Conference, December 26th-30th, 2022, Bengaluru, India, "Regression analysis of a future state entry time distribution conditional on a past state occupation".

104. International Symposium on Probability and Statistics: New Frontiers, 12th to 14th August, 2022, online, Kolkata, India, "Regression analysis of a future state entry time distribution conditional on a past state occupation".

103. ISNPS2022: International Symposium on NonParametric Statistics, Paphos, Cyprus, June 20-24, 2022, "Regression analysis of a future state entry time distribution conditional on a past state occupation".

102. LinStat 2020 Conference. Hybrid format, Bedlewo, Polland, August 30 to September 3, 2021, "Marginal inference under informative subgroup size induced by a subject level covariate".

101. 2019 International Indian Statistical Association Conference, Mumbai, India, December 26-30, 2019, "Personalized treatment selection for joint optimization of survival and other outcomes".

100. 11th ICSA International Conference, Hangzhou, China, December 20-22, 2019, "Personalized treatment selection for joint optimization of survival and other outcomes".

99. Fifth Scientific Applied Conference: Statistics and its Applications, Tashkent, Uzbekistan, 17-18 October, 2019, "Some recent statistical approaches to personalized medicine". Plenary Talk.

98. XII International Conference CDAM 2019, Minsk, Belarus, 18-22 September, 2019, "A random graph generation model for transcription networks and nonparametric simulator for RNA-Seq expression data". Plenary Talk.

97. 62nd World Statistics Congress, ISI-WSC 2019, Kuala Lampur, Malaysia, 18-23 August, 2019, "Personalized treatment selection for joint optimization of survival and other outcomes". Special Topics Session.

96. 2019 ICSA Applied Statistics Symposium, Raleigh, NC, USA, June 9-12, 2019. "Personalized treatment selection for joint optimization of survival and other outcomes".

95. International Conference on Computer Age Statistics, Savitribai Phule Pune University, Pune, India, January 3-5, 2019. "Semi-parametric regression of state occupational probability in a multistate model with right-censored data".

94. Second International Conference of Mathematics, Statistics & Information Technology, Tanta University, Egypt, December 18-20, 2018. "Semi-parametric regression of state occupational probability in a multistate model with right-censored data". Plenary Talk.

93. Workshop on Statistical and Computational Challenges in Precision Medicine, IMA, Mineapolis, November 7-9, 2018. "Selection of the Optimal Personalized Treatment from Multiple Treatments with Multivariate Outcome Measures".

92. Michigan State Symposium on Mathematical Statistics and Applications, East Lansing, MI, USA, September 14-16, 2018. "Adjustments of Mann-Whitney U-Statistics for Comparing Event Time Distributions in Observational Studies".

91. ICSA China Conference, Qingdao, China, July 2-5, 2018. "Semi-parametric Regression of State Occupation Probabilities in a Multi-state Model with Right-censored Data".

90. ICSA Applied Statistics Symposium, New Brunswick, NJ, USA, June 14-17, 2018. "A Log Rank Test for Clustered Data with Informative Within-cluster Group Size".

89. 4th Conference of the International Society for Nonparametric Statistics, Salerno, Italy, June 11-15, 2018. "Adjustments of Mann-Whitney U-Statistics for Comparing Event Time Distributions in Observational Studies".

88. 2018 International Indian Statistical Association Conference, Gainesville, FL, USA, May 17-20, 2018. "Robust Regression Analysis of Temporal Data under Censoring".

87. 5th African International Conference on Statistics, March 19–22, University of Botswana, Gaborone, Botswana. "Differential Network Analysis using Next Generation Sequencing Data". Keynote Talk.

86. PCM 125: International Conference in Statistics and Probability, January 2-4, Indian Statistical Institute, Kolkata, India. "Multi-Sample Adjusted U-Statistics that Account for Confounding Covariates".

85. 2017 International Indian Statistical Association Conference, Hyderabad, India, December 28-30, 2017. "Issues of Informative Cluster Size and Beyond in the Context of Rank Based Inferencee". Special Invited Talk.

84. International Conference on Linear Algebra and its Applications, Manipal, India, December 11-15, 2017. "A Combined PLS and Negative Binomial Regression Model for Inferring Association Networks from Next-generation Sequencing Count Data".

83. 61th World Statistics Congress – ISI2017, Marrakech, Morocco, July 16–21, 2017. "Analyzing Clustered Count Data with Cluster-specific Random Effect Zero-inflated Conway-Maxwell-Poisson Distribution". STS Invited Talk.

82. Workshop on Statistical Inference for Biomedical Big Data, Gainesville, USA, April 7-8, 2017, "Network Analysis of Next-Generation Sequencing Count Data".

81. International Statistical Institute Regional Statistics Conference, Bali, Indonesia, March 20-24, 2017. "Network Analysis of Next-Generation Sequencing Count Data".

80. 10th ICSA International Conference on Global Growth of Modern Statistics, Shanghai, China, December 19-22, 2016. "Marginal Regression Models for Clustered Count Data Based on Zero-inflated Conway-Maxwell-Poisson Distribution with Applications".

79. 11th International Conference Computer Data Analysis & Modeling 2016, Minsk, Belarus, September, 6-10. "A Rank-sum Test for Clustered Data when the Number of Subjects in a Group within a Cluster is Informative". Plenary Talk.

78. 3rd conference of the International Society for Non-Parametric Statistics (ISNPS), Avignon, France, June 11-16, 2016. "A Rank-sum test for Clustered Data when the Number of Subjects in a Group within a Cluster is Informative".

77. Frontiers in Applied and Computational Mathematics, FACM 2016, Minisymposium IX-Biostatics II, NJIT, Newark, June 3-4, 2016. "Multi-Sample Adjusted U-Statistics that Account for Confounding Covariates".

76. The 1st International Statistical Conference in Croatia, Zagreb, Croatia May, 5-6, 2016. "Inference for Clustered Count Data Based on Zero-inflated Conway-Maxwell-Poisson Distribution with Application to the Iowa Fluoride Study".

75. Ninth International Triennial Calcutta Symposium, Kolkata, December 28 - 31, 2015. "Marginal and joint Regression Models for Clustered data Inference when the Cluster Size is Potentially Informative."

74. 9th International Conference on Mathematical Sciences for Advancement of Science and Technology, Kolkata, December 21-23, 2015. "Inference for Clustered Count Data Based on Zero-inflated Conway-Maxwell-Poisson Distribution with Application to the Iowa Fluoride Study."

73. 60th World Statistics Congress – ISI2015, Rio de Janeiro, Brazil, July 26–31, 2015. "Inference for Clustered Count Data Based on Zero-inflated Conway-Maxwell-Poisson Distribution with Application to the Iowa Fluoride Study." STS Invited Talk.

72. 7th International Conference on Probability and Statistics (PROBASTAT 2015), Smolenice Castle, Slovakia, June 29 - July 3, 2015. "Marginal and Joint Regression Models for Clustered Data Inference when the Cluster Size is Potentially Informative." Plenary Talk.

71. XIV EMR Brazilian School of Regression Models, Campinas, Brazil, March 2-5, 2015. "Marginal Regression Models for Clustered Data Inference When the Cluster Size is Potentially Informative." Plenary Talk.

70. XIV EMR Brazilian School of Regression Models, Campinas, Brazil, March 2-5, 2015. "Ensemble Regression".

69. XIII CLAPEM-2014, Latin American Congress of Probability and Mathematical Statistics, Cartagena de Indias, Colombia, September 22-26, 2014. "Inverse Probability of Censoring Weighted U-statistics for Right Censored Data."

68. International Conference on Survival Analysis in Memory of John P. Klein, June 26 - 27, 2014, Medical College of Wisconsin, Milwaukee, Wisconsin. "A Nonparametric Analysis of Waiting Times from a Multistate Model using a Novel Linear Hazards Model Approach."

67. Second conference of the International Society for Nonparametric Statistics, Cádiz, Spain, June 12-16, 2014. "A Covariate Adjusted Mann-Whitney Test for Comparing Two Sojourn Times Under Right Censoring."

66. 25th Nordic Conference in Mathematical Statistics (Nordstat), Turku, Finland, June 2-6, 2014. "Robust Estimation of Marginal Regression Parameters in Clustered Data."

65. 2013 ICSA International Conference, Hong Kong, China, December 20 - 23, 2013. "A Multi-loss Super Regression Learner (MSRL) with An Application to Survival Prediction Using Proteomics."

64. 10th Applied Statistics 2013 International Conference, Ribno (Bled), Slovenia, September 22 - 25, 2013. KEYNOTE LECTURE. "Clustered Data Inference When the Cluster Size is Potentially Informative."

63. 22nd International Workshop on Matrices and Statistics, Toronto, Canada, August 12-15, 2013. "Robust Regression Analysis of Longitudinal Data Under Censoring."

62. Statistics and Its Interactions with Other Disciplines (SIOD 2013), Ho Chi Minh City, Vietnam, June 5-7, 2013. "A Nonparametric Linear Hazards Model for Waiting Times from a Multistate Model."

61. 2013 Kentucky Workshop on Renewable Energy and Energy Efficiency (RE3), KY International Convention Center (KICC), Short Course, March 24-26, 2013. "Tools for Materials Genome Research."

60. The First International Conference and Summer School in Molecular and Materials Informatics, Melbourne, Australia, February 4-6, 2013. "Regression Approaches to Bandgap Engineering." Keynote Lecture.

59. XIII Chilean Biometric Conference, the XI Chilean Bayesian Conference and the VI Chilean Dental-Statistical Meeting, Concepción, Chile, January 9-11, 2013. KEYNOTE LECTURE. "Statistical Analysis of Piedmont Data."

58. Eighth International Triennial Calcutta Symposium on Probability & Statistics, Kolkata, India, December 27-30, 2012. "Nonparametric Hazard Regression for State Waiting Times in a Multistate Model."

57. The International Conference on Trends and Perspectives in Linear Statistical Inference, LinStat'2012, Bedlewo, Poland, July 16-20, 2012. "Nonparametric Regression for Sojourn Time Distributions in a Multistate Model."

56. First Conference of the International Society for Non-parametric Statistics, Halkidiki, Greece, June 15-19, 2012. "Robust Regression Analysis with Informative Cluster Size."

55. International Workshop on Recent Advances in Time Series Analysis RATS2012, Protaras, Cyprus, June 9-12, 2012. "Robust Regression Analysis of Time Series Data Under Censoring."

54. Statistical Concepts and Methods for the Modern World, Colombo, Sri Lanka, December 28-30, 2011. "Regression for Waiting Time Distributions."

53. 4th International Conference of the ERCIM WG on Computing & Statistics, University of London, UK, December 17-19, 2011. "Nonparametric Regression for Waiting Times from a Multistate Model."

52. 3rd Nordic-Baltic Biometric Conference, Turku, Finland, June 6-9, 2011. "Statistical Analyses of Next Generation Sequence Data ."

51. Applied Stochastic Models and Data Analysis (ASMDA 2011), Rome, Italy, June 7 - 10, 2011. "Mann-Whitney Tests for Comparing Waiting Time Distributions When Transition Times Are Right Censored."

50. Workshop on Statistical Challenges in Life History Analysis at the Centre de Recherches Mathematiques, Montreal, Canada, May 16-19, 2011. "Nonparametric Inference for Multistate Models."

49. DUSDAA, The First International Conference on Theory and Applications of Statistics, Dhaka University, Dhaka, Bangladesh, December 26-29, 2010. "Mann-Whitney Tests for Comparing Waiting Time Distributions When Transition Times Are Right Censored."

48. XXXII National Congress of Statistics and Operations Research and the VI Meeting on Public Statistics, A Coruña, Spain, September 14-17, 2010. "Inference in Accelerated Failure Time Models for Clustered Time to Event Data."

47. LinStat'2010 - International Conference on Trends and Perspectives in Linear Statistical Inference, July 27-31, 2010. "Inference in Accelerated Failure Time Models for Clustered Time to Event Data." Keynote Lecture.

46. Conference on Nonparametric Statistics and Statistical Learning, The Ohio State University, Columbus, OH, May 19 - 22, 2010. "A Class of Signed-Rank Test for Clustered Paired Data When The Cluster Size Is Potentially Informative."

45. Discussant, Session on Current Issues in Statistical Proteomics, ENAR 2010, New Orleans, USA, March 21-24, 2010.

44. The International Symposium on Stochastic Models in Reliability Engineering, Life Sciences, and Operations Management, February 8-11, 2010. "Nonparametric Inference in Multistate Models with Interval Censored Data."

43. VIII IISA Joint Statistical Meeting, Visakhapatnam, India, January 4-8, 2010. "U-statistics for Right Censored Data With Applications."

42. Seventh International Triennial Calcutta Symposium on Probability and Statistics, Kolkata, India, December 28 - 31, 2009. "U-statistics for Right Censored Data With Applications."

41. Joint Statistical Meetings, August 1-6, 2009. "Nonparametric Inference in Multistate Models with Interval Censored Data."

40. First IMS-Pacific Rim Meeting, Discussant for an invited session on "Statistics in Health Sciences", Seoul, June 28-July 1, 2009.

39. Symposium on New Directions in Asymptotic Statistics, University of Georgia, Athens, May 15-16, 2009. "Rank Tests for Clustered Data."

38. Winemiller 2008: Conference on Survival Analysis and Its Applications, October 16-18, 2008, Columbia, Missouri. "Nonparametric Inference for State Waiting Times in Multistate Models."

37. Joint Statistical Meetings, August 3 - 7, 2008, Denver, Colorado. "A Signed-Rank Test for Clustered Data."

36. Nonparametric Statistics and Mixture Models: Past, Present and Future, May 22-25, 2008, State College, PA. "Re-weighted U-statistics for Censored Data."

35. Conference on Recent Advances in Statistics - In honor of Hira Koul's 65th birthday, May 15-17, 2008, E. Lansing, MI. "The Re-weighting Approach in Survival Analysis."

34. ENAR 2008, Arlington, Virginia. March 16-19, 2008. "Nonparametric Estimation of State Waiting Time Distributions in a Markov Multistate Model."

33. Discussant, Session on Multistate Models under Complex Censoring, JSM 2007, July 29, 2007, Salt Lake City, UT, USA.

32. Discussant, Session on Interval Censored Data, ENAR 2007, March 12, 2007, Atlanta, GA, USA.

31. Classification Competition on Clinical Mass Spectrometry Proteomic Diagnosis Data: Presentation of Results, Leiden University Medical Center, March 1, 2007, Leiden, The Netherlands.

30. International Conference on Statistics, Probability and Related Areas by IISA, January 2-5, 2007, Cochin, India. "Predicting Patient Survival from Microarray Data by Accelerated Failure Time Modeling using Partial Least Squares and LASSO."

29. International Conference on Multivariate Statistical Methods, Dec 28-29, 2006, Kolkata, India. "Statistical Classification of Autism Spectrum Disorders: A Case Study."

28. Discussant, Session on Genomics & Proteomics, International Biometric Society Conference IBC 2006, Montreal, Canada, July, 2006.

27. International Multi-Symposiums on Computer and Computational Sciences (IMSCCS|06), June 20-24, 2006, Zhejiang University, Hangzhou, China. "Combining Functional Information in Validation of Statistical Clustering."

26. SCMA 2005 / FIM XII, Twelfth International Conference on Statistics, Combinatorics, Mathematics and Applications, December 2-4, 2005, Auburn University, Auburn, AL, USA. "Estimation of Stage Occupation Probabilities in a Multistage Model with Current Status Data."

25. Workshop on Statistical Analysis of Complex Event History Data, Norwegian Academy of Science and Letters, August 31-September 2, 2005, Oslo, Norway. "Nonparametric Marginal Estimation in a Multistage Model with Current Status Data."

24. Joint Annual Meeting of the Interface and the Classification Society of North America, June 8, 2005 - June 12, 2005, Washington University School of Medicine, St. Louis, MO. "Standardization and De-noising Algorithms for Mass Spectra to Classify Whole-Organism Bacterial Specimens."

23. International Conference on Future of Statistical Theory, Practice and Education, December 29, 2004 - January 1, 2005, Hyderabad, India. "Standardization and De-noising Algorithms for Mass Spectra to Classify Whole-Organism Bacterial Specimens."

22. Eleventh International Conference on Interdisciplinary Mathematical and Statistical Techniques, SCRA 2004, December 27-29, 2004, Lucknow, India. "Standardization and De-noising Algorithms for Mass Spectra to Classify Whole-Organism Bacterial Specimens."

21. International Conference on Statistics in Health Sciences, June 23-23, 2004, Nantes, France. "Standardization and De-noising Algorithms for Mass Spectra to Classify Whole-Organism Bacterial Specimens."

20. IISA Conference, May 2004, University of Georgia, Athens, USA. "Nonparametric Marginal Estimation in a Multistage Model Using Interval Censored Data."

19. International Conference on Reliability and Survival Analysis 2003, May 2003, Department of Statistics, University of South Carolina, Columbia, USA. "Marginal Estimation in Multistage Models Using Current Status Data."

18. SCRA 2002, International Conference on Statistics, Combinatorics and Related Areas and the Ninth International Conference of the Forum for Interdisciplinary Mathematics, December 2002, Allahabad, India. "Estimation of Stage Occupation Probabilities in Multistage Systems under Current Status Data."

17. International Conference on Current Advances and Trends in Nonparametric Statistics, July 2002, Crete, Greece. "Nonparametric Estimation of Stage Occupation Probabilities in Multistage Models under Censoring."

16. IISA International Conference on Statistics, Probability and Related Areas, June 2002, Dekalb, Illinois, USA. "Detection of Differentially Expressed Genes in Microarray Experiments."

15. SCRA 2001, International Conference on Statistics, Combinatorics, and Related Areas, December 2001, Wollongong, Australia.

14. IISA-JSM-INDIA 2000-2001, International conference on Statistics and Probability, December 2000-January 2001, New Delhi, India.

13. Sixth International Conference on Statistics, Combinatorics, and Related Areas, December 1999, Mobile, Alabama, USA.

12. ENAR Spring Meeting, March, 1999, Atlanta, Georgia.

11. IISA Conference, October, 1998, McMaster University, Hamilton, Canada.

10. Conference in honor of Jim Hannan, May 1998, Michigan State University, East Lansing, MI, USA.

9. Special Session on Applied Probability, AMS meeting, October, 1996, Chattanooga, TN, USA.

8. Symposium on Estimating Functions, March 1996, Athens, Georgia, USA.

7. SRCOS/ASA Summer Research Conference (Discussion Leader), June 1995, Indialantic, Florida.

6. INFORMS Applied Probability Conference, June 1995, Atlanta, Georgia, USA.

5. IMS, ENAR Joint Spring Meeting, March, 1995, Birmingham, Alabama, USA.

4. Second International Triennial Calcutta Symposium on Probability and Statistics, December 1994, Calcutta, India.

3. First IMS North American New Researchers Meeting, August 1993, Berkeley, California.

2. The Third Canadian Conference in Applied Statistics, May 1991, Statistics Canada, Montreal, Canada.

1. 214 IMS Meeting (special topic Bootstrap), May 1990, East Lansing, USA. "Bootstrap for a Finite State Markov Chain."

INVITED TALKS: Colloquia

48. Department of Biostatistics, Virginia Commonwealth University, Richmond, VA, USA, September 2024.

47. Department of Mathematics and Statistics, Florida International University, Miami, FL, USA, February 2021.

46. Biostatistics Branch, National Cancer Institute, Rockville, MD, USA, June 2019.

45. Applied Statistics Unit, Indian Statistical Institute, Kolkata, India, July 2018.

44. Department of Biostatistics, Louisiana State University, New Orleans, November 2016.

43. Faculty of Mathematics, National University of Uzbekistan, Tashkent, Uzbekistan, May 2016.

42. Department of Statistics, Florida State University, Tallahassee, FL, USA, February 2016.

41. Applied Statistics Unit, Indian Statistical Institute, Kolkata, India, December 2015.

40. Department of Statistics, University of Florida, Gainesville, FL, USA, November 2015.

39. Department of Biostatistics, University of Florida, Gainesville, FL, USA, October 2014.

38. Department of Statistics, Tunghai University, Taichung, Taiwan, December 2013.

37. Institute of Mathematical Sciences, Faculty of Science, University of Malaya, June 2013.

36. Multiple seminars, Department of Statistics, University of Concepción, Chile, January 14-15, 2013.

35. Departments of Statistics and Biostatistics, University of Kentucky, Lexington, KY, USA, September 2012.

34. Department of Epidemiology & Biostatistics, Drexel University, Philadelphia, USA, March 15, 2012.

33. Department of Statistics, University of Georgia, Athens, GA, USA, July 15, 2011.

32. Steklov Mathematical Institute of Academy of Sciences, St. Petersburg, Russia, June 17, 2011.

31. School of Public Health, University of Tampere, Tampere, Finland, June 13, 2011.

30. Department of Statistics and OR, University of Vigo, Vigo, Spain, September 13, 2010.

29. Biostatistics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA, September 15, 2009.

28. Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Leiden, The Netherlands, May 12, 2009.

27. School of Public Health, University of Tampere, Tampere, Finland, May 6, 2009.

26. Department of Statistics, University of California, Davis, November, 2008.

25. Department of Statistics and Probability, Michigan State University, E. Lansing, MI, USA, March 2007.

24. Department of Statistics, University of Kentucky, Lexington, KY, USA, October 2005.

23. ASA Kentucky Chapter, Frankfort, KY, USA, September 2005.

22. Department of Statistics and Applied Probability, National University of Singapore, Singapore, December 2004.

21. Department of Bioinformatics and Biostatistics, University of Louisville, KY, USA, November 2004.

20. Department of Biostatistics, University of Minnesota, MN, USA, March 2004.

19. CHEDA user group, BimCore and Department of Biostatistics, Emory University, Atlanta, GA, USA, March 2004.

18. Department of Mathematics, Univ. of N. Carolina, Charlotte, NC, USA, April 2003.

17. Department of Biostatistics, Emory University, March 2003.

16. Department of Statistics, Univ. of S. Carolina, Columbia, SC, USA, October 2001.

15. Department of Biostatistics, Univ. of Alabama, Birmingham, AL, USA, August 2001.

14. School of Industrial and Systems Engineering, Georgia Tech., Atlanta, GA, USA, April 2001.

13. Indian Statistical Institute, Calcutta, India, July 1999.

12. Department of Statistics, Texas A&M University, College Station, TX, USA, May 1996.

11. Department of Statistics, Univ. of North Carolina, Chapel Hill, NC, USA, April 1995.

10. Department of Statistics, SUNY at Buffalo, Buffalo, NY, USA, February 1995.

9. Department of Mathematics, Univ. of North Carolina, Charlotte, NC, USA, February 1995.

8. Division of Statistics and Mathematics, Indian Statistical Institute, Calcutta, India, September 1992.

7. Computer Science Unit, Indian Statistical Institute, Calcutta, India, August 1992.

6. Department of Statistics, Iowa State University, Ames, IA, USA, September 1989.

5. Department of Statistics, University of Wisconsin, Madison, WI, USA, September 1989.

4. Department of Statistics and Probability, Michigan State University, East Lansing, MI, USA, June 1989.

3. Department of Statistics, University of Georgia, Athens, GA, USA, February 1988.

2. Department of Statistics, Purdue University, West Lafayette, IN, USA, February 1988.

1. Department of Mathematics, McGill University, Montreal, Canada, January 1988.

REFEREED ORAL PRESENTATIONS

5. 14th Annual International Conference on Critical Assesment of Massive Data Analysis (CAMDA2015), Dublin, Ireland, July 10-11, 2015.

4. MCP 2009: The 6th International Conference on Multiple Comparison Procedures, Tokyo, Japan, 2009.

3. 29th Annual Conference of the International Society of Clinical Biostatistics, August 17-21, 2008, Copenhagen, Denmark.

2. CAMDA 2007, December 13-14, Valencia, Spain.

1. The Second Asia Pacific Bioinformatics Conference 18-22 Jan, 2004, Dunedin, New Zealand.

TOPIC CONTRIBUTED/CONTRIBUTED

15. JSM, Portland, August 2024, Topic Contributed (oral).

14. JSM, Philadelphia, August 2020, Topic Contributed (oral).

13. JSM, Vancouver, Canada, July 2018. Topic Contributed (oral).

12. JSM, Baltimore, August 2017. Topic Contributed (oral).

11. JSM, Chicago, August 2016. Topic Contributed (oral).

10. JSM, Vancouver, Canada, August 2010. Topic Contributed (oral).

9. ISMB, Vienna, Austria, July 2007 (poster).

8. Research Louisville, Louisville, October 2005 (poster).

7. JSM 2004, Toronto, August 2004 (oral).

6. ENAR Spring Meeting, Pittsburgh, March 2004 (oral).

5. IBS Meeting, Cape Town, South Africa, December 1998 (oral).

4. IBS Meeting, Amsterdam, The Netherlands, July 1996 (oral).

3. 56th IMS Annual Meeting, San Francisco, August 1993 (oral).

2. Second International Symposium on Probability and Its Applications, Bloomington, March 1993 (oral).

1. Special Contributed Session, 5th Purdue Symposium on Statistical Decision Theory and Related Topics, W. Lafayette, Indiana, June 1992 (oral).

WORKSHOPS/MEETINGS ATTENDED

14. Workshop on Applications-Driven Geometric Functional Data Analysis, FSU, October, 2017.

13. FaceBase 2017 Annual Meeting in Boston, USA, May, 2017.

12. AAAS Annual Meeting, San Jose, CA, USA, February 2015.

11. FaceBase 2015 Annual Meeting, Marina del Rey, CA, USA, January 2015.

10. CAMDA 2014, Boston, USA, July 2014.

9. Joint Statistical Meetings, Montreal, Canada, August, 2013.

8. Joint Statistical Meetings, San Diego, CA, July-August, 2012.

7. Conference on New Statistical Methods for Next Generation Sequencing Data Analysis, Ames, Iowa, May 11, 2012.

6. Conference on Data Analysis, Santa Fe, NM, February-March, 2012.

5. Rocky '08, 6th Annual Rocky Mountain Bioinformatics Conference, Snowmass, CO, December, 2008.

4. NIEHS SNPs Workshop, Brown Hotel, Louisville, KY, January 2008.

3. UT-ORNL-KBRIN Bioinformatics Summit 2008, Lake Barkley State Park, KY, April, 2008.

2. UT-ORNL-KBRIN Bioinformatics Summit 2006, Lake Barkley State Park, KY, April, 2006.

1. Symposium on Multivariate Analysis, Hong Kong Baptist College, Hong Kong, March 1992.

EDITORIAL WORK

Current

• Book Series **Editor-in-Chief**, ***Frontiers in Statistical Sciences and Probability***, Springer, 2012-current. Nine volumes in the series to date.

• **Associate Editor**, ***BMC Bioinformatics***, Springer Nature, 2010-current.

• **Associate Editor**, ***Bioinformatics and Biology Insights***, Sage, 2024-current.

Past

• **Editor-in-Chief** (co with H. Koul), ***Statistics & Probability Letters***, Elsevier, 2007-2012 (Handled over 1000 papers).

• **Guest Editor** (co with María del Carmen Pardo Llorente, Thomas Scheike, and Kam C. Yuen), Special Issue on "Survival Analysis". ***Computational Statistics & Data Analysis***, Elsevier, 2014-2015.

• **Guest Editor** (co with H. van Houwelingen), Special Issue on "Statistics in Biological and Medical Sciences". ***Statistics & Probability Letters***, Elsevier, 2010-2011.

• **Associate Editor**, ***The American Statistician***, American Statistical Association/Taylor & Francis, 2005- 2012.

• **Co-Editor**, *Sankhyā*, Indian Statistical Institute/Springer, 2001-2007.

• **Associate Editor**, ***Communications in Statistics-Theory & Methods***, Taylor & Francis, 2002-2020.

• **Associate Editor**, ***Communications in Statistics-Simulation and Computation***, Taylor & Francis, 2002-2020.

• **Associate Editor**, ***Communications in Statistics-Data Analysis and Applications***, Taylor & Francis, 2016-2020.

TEACHING

At University of Florida (2016 -):

• PHC 7066: Large Sample Theory. A core course for the doctoral students in biostatistics.

At University of Louisville (2005 - 2015):

• PHST 762: Advanced Statistical Inference. PhD (Biostatistics concentration) core course.

• PHST 783: Advanced Survival Analysis. PhD (Biostatistics concentration) core course.

• PHST 780: Advanced Nonparametrics. PhD (Biostatistics concentration) elective course.

• Numerous Independent Study courses.

• Co-taught a short course titled "Tools for Materials Genome Research" at 2013 Kentucky Workshop on Renewable Energy and Energy Efficiency, KY International Convention Center on March 24, 2013.

At University of Georgia (1988 - 2005):

• STA 2000: Elementary Statistics. Large lecture format (150--250 students).

• STA 8530: Advanced Statistical Inference 1. Ph. D. core course.

• STA 8540: Advanced Statistical Inference 2. Ph. D. core course.

• STA 8550: Asymptotic Inference. Ph. D. level. Books used Asymptotic Statistics by van der Vaart and Approximation Theorems of Mathematical Statistics by Serfling.

• STA 8570: Statistical Decision Theory. Books used Statistical Decision Theory by Berger and Mathematical Statistics: A Decision Theoretic Approach by Ferguson.

• STA 8650: Bootstrapping Techniques. Books used The Jackknife, the Bootstrap and Other Resampling Plans by Efron and The Bootstrap and Edgeworth Expansion by Hall.

• STA 9270/80: Supervised Statistical Consulting. Students get real life experience in Statistical Consulting.

• STA 3330: Advanced Applications and Computing. Book used Modern Applied Statistics with S, 4th Edn., by W. N. Venables and B. D. Ripley.

• STA 8990: Special Topics in Statistics. A course in advanced survival analysis offered to the Ph.D. students. Book used Statistical Models Based on Counting Processes by Andersen, Gill, Borgan and Keiding.

• STA 4/6380: Survival Analysis. An introductory course in Survival Analysis.

• STA 4/6240: Sampling and Survey Methods. An introductory course in sampling.

SERVICE

Scholarly Journal Refereeing

• Referee for *Annals of Statistics*, *Biometrics*, *Biometrika*, *Bioinformatics*, *BMC Bioinformatics*, *Communications in Statistics, Journal of American Statistical Association, Journal of Multivariate Analysis, Journal of Nonparametric Statistics, Journal of Statistical Planning and Inference, Lifetime Data Analysis, Mathematical Methods in Statistics, Sankhya, Scandinavian Journal of Statistics, Statistics in Medicine, Statistics & Decisions, Statistics and Probability Letters, Statistical Methodology* and many other journals.

Grant Reviews (selected)

• National Research Foundation South Africa, January 2024.

• Grant Reviewer, FONDECYT Regular Competition 2020, Chile, 2019.

• Grant Reviewer, United States-Israel Binational Science Foundation Research Proposal, Israel, 2017.

• Grant Reviewer, eScience Enabling Technologies Research Proposal, Vienna, Austria, 2016.

• Grant Reviewer, Austrian Science Fund, The Netherlands, 2014.

• Grant Reviewer, Medical Research Council, United Kingdom, 2014.

• Grant Reviewer, Human Frontier Science Program, France, 2012.

• Grant reviewer for Portuguese Foundation for Science and Technology, 2010.

• Grant Review Panel member, Special Emphasis Panel ZCA1 TCRB-Y (J1) S, Bridging the Gap Between Cancer Mechanism and Population Science, NCI, National Institutes of Health, October 2015.

• Grant Review Panel member, Special Emphasis Panel ZRG1 MOSS-F (02) M, NIDCR, National Institutes of Health, April 2014.

• Temporary member, Oral, Dental and Craniofacial Sciences Study Section, National Institutes of Health, February 2014.

• Grant Review Panel member, Special Emphasis Panel ZRG1 MOSS-C (80) S, Pathophysiology and Clinical Studies of Osteonecrosis of the Jaw, NIDCR, National Institutes of Health, February 2014.

• Grant Review Panel member, Special Emphasis Panel of NIDCR, National Institutes of Health, 2012.

• Grant Review Panel member, Integrative Cancer Biology and Tumor Microenvironment, NCI, National Institutes of Health, 2010.

• Grant Review Panel member, Integrated Cancer Biology, NCI, National Institutes of Health, 2004.

• Grant Review Panel member, National Science Foundation, Statistics, 2008.

• Grant reviewer for National Science Foundation - multiple occasions.

Other Reviews

- Scientific Sessions Reviewer for the 2023 and 2024 AAAS Annual Meetings.
- Reviewer for Mathematical Reviews (multiple occasions).
- Book proposal reviewer for Springer (multiple occasions).
- External evaluator for numerous promotion and tenure cases.
- External evaluator for overseas PhD dissertations.

Conference Organization

• Invited Session Organizer, IMS-APRM 2024, January 4 -7, Melbourne, Australia.

• International Program Committee member, ICSA 2023 China Conference, June 30 – July 3, 2023, Chengdu, China.

• Co-Program Chair, ICSA 2022 Applied Statistics Symposium, June 19-22, 2022, University of Florida, Gainesville, FL, USA.

• Scientific Program Committee member, Conference on Current Trends in Survey Statistics 2019, IMS and National University of Singapore, August 13-16, 2019, Singapore.

• Invited session, 4th Conference of the International Society for Nonparametric Statistics, June 11-15, 2018, Salerno, Italy.

• IISA Annual Conference 2018, Organizer, May 17-20, 2018, University of Florida, Gainesville, FL, USA.

• International Program Committee member, Second USA - Uzbekistan Conference on Analysis and Mathematical Physics, August, 2017, Urgench State University, Uzbekistan.

• Topic Contributed Session, 2017 Joint Statistical Meetings, July 29 - August 3, Baltimore, Maryland.

• Organizer, Special Topics Session, 60th World Statistics Congress – ISI2017, July 2017, Marrakech, Morocco.

• Co-organizer, Biostatistics Workshop: Statistical Inference for Biomedical Big Data, April 7 & 8, 2017, University of Florida, Gainesville, Florida, USA.

• Organizing Committee member, 15th Annual International Conference on Critical Assesment of Massive Data Analysis, July 2016, Orlando, Florida, USA.

• Organizer, Invited session Nonparametric Methods in Biostatistics, 3rd conference of the International Society for Non-Parametric Statistics (ISNPS), June 2016, Avignon, France.

• Member, International Scientific Program Committee, The 1st International Statistical Conference in Croatia – ISCCRO'16, May, 2016, Zagreb, Croatia.

• Organizer, Special Topics Session, 60th World Statistics Congress – ISI2015, July 2015, Rio de Janeiro, Brazil.

• Scientific Committee Member, XIV EMR, Brazilian School of Regression Models, March 2015, Campinas, Brazil.

• Organizer, Invited session on Big Data regression, XIV EMR, Brazilian School of Regression Models, March 2015, Campinas, Brazil.

• Organizer, Invited session on Bioinformatics and Biostatistics, IASSL-2014, Colombo, Sri Lanka.

• Organizer, Invited session on Interval Censoring, IBC 2010, Florianópolis, Brazil.

• Organizer, Invited session on Proteomics, IBC 2008, Dublin, Ireland.

• Organizer, Invited session on Multistate Models, JSM 2007, Salt Lake City, Utah.

• Chair, Invited session on Statistics in Genomics, JSM 2004, Toronto, Canada.

PROFESSIONAL MEMBERSHIP

American Association for the Advancement of Science (Section U/A), American Statistical Association (life), International Statistical Institute, Institute of Mathematical Statistics, International Indian Statistical Association (life), Forum for Interdisciplinary Mathematics (life), International Society for Nonparametric Statistics, International Association for Statistical Computing.

Past: International Society for Clinical Biostatistics, International Biometric Society, International Society for Computational Biology, International Association for Dental Research.

UNIVERSITY SERVICES (selected list)

At University of Florida:

• Chair, Faculty Search Committee, Department of Biostatistics, current.
• Chair, Admissions Committee, Department of Biostatistics, past.
• Member, Curriculum Committee (2015-2017), Awards Committee, Department of Biostatistics, current.
• Member, Research Committee, College of Public Health and Health Professions, past.

At University of Louisville:

• Member, SPHIS Dean's Transition Team.
• Member, Promotion and Tenure Committee, School of Public Health and Information Sciences (Committee Chair for two 3-year terms).
• Member, Curriculum Committee, School of Public Health and Information Sciences.
• Member, Academic Affairs Committee, School of Public Health and Information Sciences.

At University of Georgia:

• Member, Promotion and Tenure Committee, Franklin College of Arts and Science.
• Member, Promotion and Tenure Committee, University of Georgia.
• Member, Faculty Senate.