Subha Guha

Department of Biomedical Data Science, Dartmouth College
E-mail: subha.guha@dartmouth.edu

Research Interests

My research focuses on causal inference, transfer learning and data integration, Bayesian and computational methods, and high-dimensional statistical inference, with applications to cancer, genomics and multi-omics data, microbiome data, electronic health records, and other complex observational studies.

Education

Ohio State University, Columbus, Ohio — Ph.D. in Statistics, June 2004

Indian Institute of Technology, Kanpur, India — M.Sc. in Statistics, June 1997

Professional and Academic Experience

Dartmouth College, Department of Biomedical Data Science, Hanover, NH
Professor — **2025 – present**

University of Florida, Department of Biostatistics, Gainesville, FL
Associate Professor — **2018 – 2025**

University of Missouri, Department of Statistics, Columbia, MO
Associate Professor — **2013 – 2017**
Assistant Professor — **2007 – 2013**

Harvard School of Public Health, Department of Biostatistics, Boston, MA
Postdoctoral Research Fellow — **2004 – 2007**

Ohio State University, Department of Statistics, Columbus, OH
Graduate Teaching Assistant — **1999 – 2004**

Research Grants

National Institute of Arthritis and Musculoskeletal and Skin Diseases, *Biomechanics Contributions to Symptoms and Joint Health in Individuals with Rotator Cuff Tears*, R01AR084273, 2024 – 2028. Role: Co-I.

National Institute of Arthritis and Musculoskeletal and Skin Diseases, *Nervous system influences on recovery from painful rotator cuff tears*, R01AR080058, 2023 – 2028. Role: Co-I.

National Cancer Institute, *Causal Machine Learning in Cancer Survival by Integrating Multiple High-Dimensional Observational Studies*, R01CA269398, 2022 – 2026. Role: MPI.

National Cancer Institute, *The Boston Lung Cancer Survivor Cohort*, U01CA209414, 2023 – 2028. Role: Co-I.

Health Resources & Services Administration HIV/AIDS, *Improving Sexually Transmitted Infection Screening and Treatment among People Living with or at Risk for HIV*, U90HA32147, 2020 – 2023. Role: Co-I.

National Science Foundation, *Collaborative Research: New Bayesian Nonparametric Paradigms of Personalized Medicine for Lung Cancer*, DMS-1854003, 2015 – 2020. Role: PI.

National Science Foundation, *Bayesian Mixture Models: Unified Theoretical Frameworks and MCMC Methods*, DMS-0906734, 2009 – 2013. Role: PI.

Department of Health and Human Services, *Statistical Informatics for Cancer Research*, 2009 – 2013. Role: Co-I.

Honors and Awards

Research featured in "Member Highlight" section of the University of Florida Cancer Center newsletter, February 2022.

Top Faculty Achiever of the University of Missouri, 2016.

Faculty Fellowship Award of the Department of Statistics, University of Missouri, August 2014.

Craig Cooley Memorial Prize for scholarly excellence and leadership qualities, 2004, Department of Statistics, Ohio State University.

Thomas and Jean Powers Teaching Award, 2003, Department of Statistics, Ohio State University.

Distinguished teaching assistant, Department of Statistics, Ohio State University, for four consecutive quarters.

Winner of the Student Paper Competition 2002 organized by ASA Statistical Computing and Graphics Sections.

Editorial Work

Associate Editor, *Statistics in Medicine*, Wiley, 2026–present.

Associate Editor, *Statistics in Biosciences*, Springer Nature, 2026–present.

Associate Editor, *Bayesian Analysis*, International Society for Bayesian Analysis, 2020–2025.

Co-edited a book entitled *Statistical Analysis of Microbiome Data*, published by Frontiers in Probability and the Statistical Sciences, Springer Nature, Switzerland (2021).

Academic Service

Grant Review

National Institutes of Health, Analytics and Statistics for Population Research Panel A (ASPA) Study Section (March 2026)
National Institutes of Health, Cancer and Hematologic Disorders (CHD) Study Section (June 2025)
National Institutes of Health, Cancer and Hematologic Disorders (CHD) Study Section (April 2025)
National Institutes of Health, Analytics and Statistics for Population Research Panel A (ASPA) Study Section (June 2024)
National Institutes of Health, Analytics and Statistics for Population Research Panel A (ASPA) Study Section (February 2024)
National Cancer Institute, Cancer Epidemiology Cohort Studies (May 2023)
National Institutes of Health, Neurological, Mental and Behavioral Health (NMBH) Study Section (February 2023)
National Institutes of Health, Biostatistical Methods and Research Design (BMRD) Study Section (June 2017)
National Security Agency, Mathematical Sciences Grant Program (2016)
National Science Foundation, Division of Mathematical Sciences (2015)

American Statistical Association

Organizer and Chair, invited session entitled *Applications of Deep Learning in Biomedical Research*, Joint Statistical Meetings 2026
Organizer and Chair, invited session entitled *Recent Advances in Statistical Methods for Integrative Analyses of Healthcare Data*, Joint Statistical Meetings 2025
Organizer and Chair, invited session entitled *Recent Advances in Statistical Methods for Microbiome Data*, Joint Statistical Meetings 2021
Organizer and chair, invited session entitled *Innovative Approaches for High-dimensional Omics and Neuroimaging Data*, Joint Statistical Meetings 2019
Chair, contributed session entitled *Modern Statistical Methods for Multi-Scale and Time Series Data*, Joint Statistical Meetings 2017
Committee Member, Student Paper Award of the Statistics in Genetics and Genomics Section, Joint Statistical Meetings 2016
Committee Member, Student Paper Award of the Section on Bayesian Statistical Science, Joint Statistical Meetings 2011 and Joint Statistical Meetings 2013

International Indian Statistical Association

Organizer and chair, invited session entitled *Some Innovative Applications in High-dimensional Data Analysis*, Innovations in Data and Statistical Sciences held in Mumbai, India, December 2019
Organizer, invited session entitled *Novel Approaches for Genomics and Censored Outcome Data* at Conference on Current Trends in Survey Statistics held at Institute of Mathematical Sciences, National University of Singapore, August 2019

International Chinese Statistical Association

Organizer, invited sessions entitled *High-Dimensional Models with Applications in Biomedical Sciences* and *New Advances in Machine Learning and AI*, Joint Meetings of 2025 Taipei International Statistical Symposium and 13th ICSA International Conference, December 2025
Organizer, invited session entitled *Recent Advances in Statistical Methods for Analyzing High-Dimensional Cancer and Disease Surveillance Data* at the 12th ICSA International Conference, Hong Kong, December 2022
Reviewer, student paper competition, ICSA 2022 Applied Statistics Symposium, University of Florida, June 2022
Co-organizer and Chair, invited session entitled *Statistical Methods for High Dimensional Microbiome Data* at the ICSA 2022 Applied Statistics Symposium, University of Florida, June 2022

Computational and Methodological Statistics

Organizer, invited session entitled *Modern Statistical Frontiers for Biomedical, Survival, and High-Dimensional Data* at the 19th International Joint Conference CFE-CMStatistics, London, UK, December 2025
Organizer, invited session entitled *Advanced Semiparametric Methods for Complex Data* at the 8th International Conference on Econometrics and Statistics (EcoSta 2025), Tokyo, Japan, August 2025
Scientific Programme Committee, 18th International Joint Conference CFE-CMStatistics, London, UK, December 2024
Organizer, invited session entitled *Recent Advances in Statistical Methods for Analyzing High-Dimensional Cancer Datasets* at 18th International Joint Conference CFE-CMStatistics, London, UK, December 2024
Organizer, invited session entitled *Recent Advances in Statistical Methods for Precision Medicine* at EcoSta 2022 – 5th International Conference on Econometrics and Statistics, Kyoto, Japan, June 2022
Organizer, invited session entitled *Flexible Bayesian approaches in high-dimensional applications* at CMStatistics 2020 – 13th International Conference of the ERCIM WG on Computational and Methodological Statistics, London, UK, December 2020

International Biometric Society

Organizer and Chair, invited session entitled *Applications of Deep Learning in Biomedical Research*, ENAR Spring Meeting, March 2026
Chair, invited session entitled *Recent Advances in Statistical Methods for Microbiome Data* at the 5th International Conference on Econometrics and Statistics (EcoSta 2022), held virtually, June 2022
Organizer, invited session entitled *Innovative Statistical Approaches for High-dimensional Omic and Microbiomic Data*, ENAR Spring Meeting held virtually, March 2020

International Statistical Institute

Organizer, invited session entitled *High-Dimensional Statistical Analysis in Precision Medicine* at 65th ISI World Statistics Congress 2025, The Hague, The Netherlands, October 2025

International Society for Bayesian Analysis

Organizer, invited session entitled *Innovative Bayesian Approaches for High-dimensional Variable Selection* at 4th Eastern Asia Meeting On Bayesian Statistics (EAC-ISBA), Kobe University, Japan, July 2019
Member, Savage Award Committee, 2016 and 2018

Other Conferences and Workshops

Scientific Program Committee, 15th High-Dimensional Data Analysis (HDDA) Conference, Istanbul, Turkey, August 2026
Organizer, special invited session entitled *Statistical Modeling and Inference for Complex and High-Dimensional Biomedical Data*, 15th High-Dimensional Data Analysis (HDDA) Conference, Istanbul, Turkey, August 2026
Organizer, invited session of invited talks at *Ordered Data Analysis, Models and Health Research Methods: An International Conference in Honor of H.N. Nagaraja for His 60th Birthday*, held at University of Texas at Dallas, March 2014

Dartmouth College

2025–2026 Ad-Hoc Reviewer, Quantitative Biomedical Sciences PhD Admissions Committee

University of Florida

2022–2025 Chair, Qualifying Exam Committee
2020–2025 Member, Qualifying Exam Committee
2020–2021 Director, Biostatistics Consulting Lab
2019–2025 Member, Departmental Awards Committee
2019–2025 Member, Departmental Curriculum Committee
2018–2022 Chair of a Ph.D. Committee
2018–2025 Chair of 8 Masters Committees
2018–2025 Member of 7 Ph.D. Committees (Biostatistics and Astronomy)
2020 Poster judge, PHHP Annual Research Day
2019 UF Marshal for Spring commencement

University of Missouri

2012–2017	Chair of 4 Ph.D. Committees
2007–2017	Member of 15 Ph.D. and 4 Master's Thesis Committees
2016–2017	Chair, Qualifying Exam Committee
2016–2017	Chair, Computing Committee
2012–2013	University of Missouri Research Board
2011–2015	Graduate Admissions Committee
2011	Chair, Bayesian Discussion Group
2011–2014	Faculty Search Committee
2010–2012	Advisor, Minor in Psychological Statistics and Methods
2009–2014	Member, Qualifying/Preliminary Exam Committee
2008–2011	Seminar Series Chair
2008	Co-organizer, Winemiller Conference

Professional Memberships

American Statistical Association
International Biometric Society
International Chinese Statistical Association
International Indian Statistical Association
International Society for Bayesian Analysis
Institute of Mathematical Statistics

Journal Reviewer for

American Journal of Epidemiology; American Journal of Mathematical and Management Sciences; Australian and New Zealand Journal of Statistics; Bayesian Analysis; Bioinformatics; BMC Bioinformatics; Biometrics; Biostatistics; Communications in Statistics - Simulation and Computation; Computational Statistics and Data Analysis; Journal of the American Statistical Association; Journal of Applied Statistics; Journal of Computational and Graphical Statistics; Journal of the Royal Statistical Society Series B; Journal of Statistical Computation and Simulation; Quality Technology and Quantitative Management; Statistical Methods and Applications; Statistica Sinica; Statistics in Medicine; Wiley Interdisciplinary Reviews: Computational Statistics

Publications

1. Shetty, S., Luo, Y., Thomas, A., Gamil, K., Saxena, I., Guha, S., Vohra, R., and Lott, D. J. (2026). Effect of exercise training on clinical and physiological variables in adults with myotonic dystrophy type 1: A systematic review and meta-analysis. *Journal of Neurology*, **273**, Article 266.
2. Guha, S. and Li, Y. (2026). Bayesian estimation of propensity scores for integrating multiple cohorts with high-dimensional covariates. *Statistics in Biosciences*, **18**, 46–67.

3. Mickle, Angela M., et al. (2025) Applying a Novel Whole Person Approach Findings Show Sex Differences in a Measure of Allostatic Load in Diverse Mid-Older Adults with Chronic Pain Associated with Risk for Knee Osteoarthritis. *The Journal of Pain*, 29.
4. Gu, C., Baladandayuthapani, V., and **Guha, S.** (2025) Nonparametric Bayes Differential Analysis of Multigroup DNA Methylation Data. *Bayesian Analysis*, 20, 489–518. *A preliminary version of the paper was recognized through an SBSS Student Paper Competition Award to Chiyu Gu.*
5. **Guha, S.** and Qiu, P. (2025) Bayesian Pairwise Comparison of High-Dimensional Images. *Journal of Computational and Graphical Statistics*, **34**(4), 1356–1365.
6. Yan, D. and **Guha, S.** (2024) A Clustering Approach to Integrative Analyses of Multiomic Cancer Data. *Journal of Applied Statistics*, 52(8), pp.1539–1560.
7. Manavalan, P. et al. (2024) Analysis of the Acceptability of Multi-Level Sexual Health Interventions and Sexually Transmitted Infection Screening and Testing Among Persons With HIV Across Three Clinical Sites in Florida. *Journal of Acquired Immune Deficiency Syndromes*, 10-1097.
8. Shetty, S., Luo, Y., Thomas, A., **Guha, S.**, and Lott, D. (2024) Effect of Exercise Training on Clinical and Physiological Variables in Adults with Myotonic Dystrophy Type 1: A Systematic Review Protocol. *MethodsX*, 13, 102957.
9. **Guha, S.** and Li, Y. (2024) Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes. *Biometrics*, 80, ujae070. *R package is available at https://CRAN.R-project.org/package=WMAP.*
10. Song, J., **Guha, S.**, and Li, Y. (2023) Bayesian Inference for High Dimensional Cox Models with Gaussian and Diffused-Gamma Priors: A Case Study of Mortality in COVID-19 Patients Admitted to the ICU. *Statistics in Biosciences*, 1–29.
11. Banks, D. and **Guha, S.** (2023) *Contributed Discussion: What is the probability of realizing a distribution from a stick-breaking process that falls outside an ϵ-ball on the base measure? Bayesian Analysis*, 18, 352–353.
12. **Guha, S.**, Jung, R. and Dunson, D. (2022) Predicting Phenotypes from Brain Connection Structure. *Journal of the Royal Statistical Society: Series C*, 71, 639–668.
13. Datta, S. and **Guha, S.** (Eds.) (2021) *Statistical Analysis of Microbiome Data*, Frontiers in Probability and the Statistical Sciences, Springer Nature, Switzerland.
14. Sachdeva, A., Ahn, C., Tiwari, R. and **Guha, S.** (2022) A Novel Approach to Augment Single-Arm Clinical Studies with Real-World Data. *Journal of Biopharmaceutical Statistics*, 27, 1–17.
15. Anyaso-Samuel, S., Sachdeva, A., **Guha, S.**, and Datta, S. (2021) Metagenomic Geolocation Prediction Using an Adaptive Ensemble Classifier. *Frontiers in Genetics, section Computational Genomics*, 12:642282.
16. Anyaso-Samuel, S., Sachdeva, A., **Guha, S.**, and Datta, S. (2021) Bioinformatics Preprocessing of Microbiome Data with an Application to Metagenomic Forensics. *Statistical Analysis of Microbiome Data* (eds. Datta, S. and Guha, S.), Springer Nature, Switzerland.
17. **Guha, S.**, and Datta, S. (2021) A Bayesian Approach to Restoring the Duality between Principal Components of a Distance Matrix and Operational Taxonomic Units in Microbiome Analyses. *Statistical Analysis of Microbiome Data* (eds. Datta, S. and Guha, S.), Springer Nature, Switzerland.
18. **Guha, S.** and Ghosh, S. K. (2020) Probabilistic Detection and Estimation of Conic Sections

from Noisy Data. *Journal of Computational and Graphical Statistics*, 29, 513–522.

19. Yan, D., **Guha, S.**, Ahn, C., and Tiwari, R. (2020) Semiparametric Bayesian Markov Analysis of Personalized Benefit-Risk Assessment. *Annals of Applied Statistics*, 14, 768–788.
20. Drusbosky, L.M, Singh, N. K., Hawkin, K. E., et al (2019) A Genomics-Informed Computational Biology Platform Prospectively Predicts Treatment Responses in AML and MDS Patients. *Blood Advances*, 3(12), 1837–1847.
21. Jha, C., Li, Y. and **Guha, S.** (2017) Semiparametric Bayesian Analysis of High-Dimensional Censored Outcome Data. *Statistical Theory and Related Fields*, 1, 194–204.
22. **Guha, S.** and Baladandayuthapani, V. (2016) A Nonparametric Bayesian Technique for High-Dimensional Regression. *Electronic Journal of Statistics*, 10, 3374–3424.
23. **Guha, S.**, Banerjee, S., Gu, C. and Baladandayuthapani, V. (2015) Nonparametric Variable Selection, Clustering and Prediction for Large Biological Datasets. *Nonparametric Bayesian Inference in Biostatistics* (eds. Mitra, R. and Müller, P.), Springer International Publishing.
24. Cui, S., **Guha, S.**, Ferreira, M. A. R. and Tegge, A. N. (2015) A Hidden Markov Model for Detecting Differentially Expressed Genes from RNA-Seq Data. *Annals of Applied Statistics*, 9, 901–925.
25. **Guha, S.**, Ji, Y., and Baladandayuthapani, V. (2014) Bayesian Disease Classification using Copy Number Data. *Cancer Informatics*, 13(S2), 83–91.
26. **Guha, S.** (2011) Discussion of *Sampling schemes for generalized linear Dirichlet process random effects models* by Kyung, Gill and Casella. *Statistical Methods & Applications*, 20, 291–293.
27. MacEachern, S. N. and **Guha, S.** (2010) Parametric and Semiparametric Hypotheses in the Linear Model. *The Canadian Journal of Statistics*, 39, 165–180.
28. **Guha, S.** (2010) Posterior Simulation in Countable Mixture Models for Large Datasets. *Journal of the American Statistical Association*, 105, 775–786.
29. **Guha, S.** (2010) Bayesian Hidden Markov Modeling of Array CGH Data. *Bayesian Modeling in Bioinformatics* (eds. Dey, D. K., Ghosh, S. and Mallick, B.), Chapman & Hall/CRC.
30. **Guha, S.**, Ryan, L. and Morara, M. (2009) Gauss-Seidel Estimation of Generalized Linear Mixed Models with Application to Poisson Modeling of Spatially Varying Disease Rates. *Journal of Computational and Graphical Statistics*, 18, 818–837.
31. **Guha, S.** (2008) Posterior Simulation in the Generalized Linear Mixed Model with Semiparametric Random Effects. *Journal of Computational and Graphical Statistics*, 17, 410–425.
32. **Guha, S.**, Li, Y. and Neuberg, D. (2008) Bayesian Hidden Markov Modeling of Array CGH Data. *Journal of the American Statistical Association*, 103, 485–497.
33. **Guha, S.** and MacEachern, S. N. (2006) Generalized Post-stratification and Importance Sampling for Subsampled Markov Chain Monte Carlo Estimation. *Journal of the American Statistical Association*, 101, 1175–1184.
34. Li, Y., Tiwari R., and **Guha, S.** (2006) Mixture Cure Survival Models with Dependent Censoring. *Journal of the Royal Statistical Society - Series B*, 69, 285–306.
35. Burden, S., **Guha, S.**, Morgan, G., Ryan, L. Sparks, G. and Young, L. (2005) Spatiotemporal Analysis of Ischemic Heart Disease in NSW, Australia. *Environmental and Ecological Statistics*, 12, 427–448.
36. **Guha, S.**, MacEachern, S. N. and Peruggia, M. (2004) Benchmark Estimation for Markov Chain Monte Carlo Samples. *Journal of Computational and Graphical Statistics*, 13, 683–701.

37. MacEachern, S. N., Peruggia, M. and **Guha, S.** (2003) Discussion of *A theory of statistical models for Monte Carlo integration* by Kong, McCullagh, Nicolae, Tan and Meng. *Journal of the Royal Statistical Society - Series B*, 65, 612.

Submitted Manuscripts, Preprints, and Work in Progress

1. Yu, J. B., Guha, S., Feyisa, J., Krishnan, A., and Kapadia, N. (2026). The Evolution of Prostate Cancer Presentation in the United States. Brief report in preparation.
2. Sibille, K. T., Tanner, J. J., Mickle, A. M., Guha, S., Garvan, C. S., Lai, S., Vincent, H. K., Schmidt, S., Staud, R., Sorge, R. E., Goodin, B. R., Fillingim, R. B., and Seubert, C. N. (2026). Investigating if an Evidence-Based, Clinically Informed, Quantified Biopsychosocial Assessment Explains Clinical and Biological Differences in Individuals with Chronic Musculoskeletal Pain: A Cross-Sectional Analysis with Data from Two Studies. Manuscript submitted.
3. Guha, S., Xu, M., Reed, C., Priyam, K., and Li, Y. (2026). WMAP: An R package for causal meta-analysis by integrating multiple observational studies. Manuscript submitted.
4. Guha, S., Xu, M., and Li, Y. (2026). Enhancing inference for small cohorts via transfer learning and weighted integration of multiple datasets. Manuscript under revision.
5. Das, P., Song, J., Mohanraj, L., Aberg, K. A., Li, Y., and Guha, S. (2026). Supervised Low-Rank Structure Discovery for Developmental Epigenetic Aging in Ultra-High-Dimensional DNA Methylation Data. Manuscript submitted.

Conference Abstracts

- Feyisa, J., **Guha, S.**, Krishnan, A., Kapadia, N., and Yu, J. B. (2026). The Evolution of Prostate Cancer Presentation. Poster presentation accepted for the 68th Annual Meeting of the American Society for Radiation Oncology (ASTRO).
- **Guha, S.** and Li, Y. Bayesian Estimation of Propensity Scores for Integrating Multiple Cohorts with High-Dimensional Covariates. *EcoSta 2025 Conference, Tokyo, Japan, August 2025.*
- **Guha, S.** and Li, Y. Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes. *IISA 2025 Conference, Lincoln, NE, June 2025.*
- **Guha, S.** and Li, Y. Bayesian Estimation of Propensity Scores for Integrating Multiple Cohorts with High-Dimensional Covariates. *LiDS 2025 Conference, New York, May 2025.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *2024 IISA Conference, December 2024.*
- **Guha, S.** and Li, Y. Bayesian Estimation of Propensity Scores for Integrating Multiple Cohorts with High-Dimensional Covariates. *18th International Joint Conference CFE-CMStatistics, London, UK, December 2024.*
- **Guha, S.** and Li, Y. Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes. *JSM, Portland, OR, August 2024.*
- **Guha, S.** and Li, Y. Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes. *SSC, St. John's, Canada, June 2024.*

- Sachdeva, A., Datta, S., and **Guha, S.** A zero-inflated Bayesian nonparametric approach for identifying differentially abundant taxa in multigroup microbiome data with covariates. *IMS Asia Pacific Rim Meeting, Melbourne, Australia, January 2024.*
- **Guha, S.** and Li, Y. Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes. *ASA Lifetime Data Science Conference, Raleigh, NC, May 2023.*
- Sachdeva, A., Datta, S., and **Guha, S.** A zero-inflated Bayesian nonparametric approach for identifying differentially abundant taxa between multiple groups in microbiome data. *IISA Conference, Bengaluru, India, December 2022.*
- **Guha, S.**, Christiani, D. C., and Li, Y. Unbiased Multigroup Comparisons by Integrating Multiple Observational Studies: a New Concordant Target Population Approach. *EcoSta 2022, June 2022.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *ENAR 2020 Spring Meeting (virtual), March 2020.*
- **Guha, S.** Scalable Methods for Posterior Inference from Big Data. *Bayes Comp 2020, Gainesville, Florida, January 2020.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *Innovations in Data and Statistical Sciences, Mumbai, India, December 2019.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *Current Trends in Survey Statistics, Singapore, August 2019.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *4th EAC-ISBA, Kobe, Japan, July 2019.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *ENAR Spring Meeting, Philadelphia, PA, March 2019.*
- Gu, C., Baladandayuthapani, V., and **Guha, S.** Bayesian Nonparametric Differential Analysis for Dependent Multigroup Data with Application to DNA Methylation Analyses in Cancer. *CFE-CMStatistics, Pisa, Italy, December 2018.*
- **Guha, S.** and Baladandayuthapani, V. A Nonparametric Bayesian Technique For High-Dimensional Regression. *MSAST 2018, Kolkata, India, December 2018.*
- Gu, C., **Guha, S.**, and Baladandayuthapani, V. Scalable Bayesian Nonparametric Learning for High-Dimensional Lung Cancer Genomics Data. *ENAR Spring Meeting, Washington, DC, March 2017.*
- **Guha, S.** and Baladandayuthapani, V. A Nonparametric Bayesian Technique For High-Dimensional Regression. *NBER-NSF SBIES, Washington University in St. Louis, May 2017.*
- **Guha, S.** and Baladandayuthapani, V. A Nonparametric Bayesian Technique for High-Dimensional Regression. *Department of Statistics, North Carolina State University, Raleigh, North Carolina, April 2016.*
- **Guha, S.** and Baladandayuthapani, V. A Nonparametric Bayesian Technique for High-Dimensional Regression. *Joint Statistical Meetings, Seattle, Washington, August 2015.*

– **Guha, S.** Nonparametric Variable Selection, Clustering and Prediction for High-Dimensional Regression. *10th Conference on Bayesian Nonparametrics, Raleigh, North Carolina, June 2015.*
– Gu, C., **Guha, S.**, and Baladandayuthapani, V. Scalable Bayesian Nonparametric Learning for High-Dimensional Lung Cancer Genomics Data. *ENAR Spring Meeting, Miami, Florida, March 2015.*
– **Guha, S.** Posterior Simulation in Countable Mixture Models for Large Datasets. *Joint Statistical Meetings, Boston, Massachusetts, August 2014.*
– **Guha, S.** Posterior Simulation in Countable Mixture Models for Large Datasets. *Frontiers in Applied and Computational Mathematics, Newark, New Jersey, October 2012.*
– **Guha, S.** Posterior Simulation in Countable Mixture Models for Large Datasets. *SAMSI PK/PD Program, Research Triangle Park, North Carolina, July 2010.*
– **Guha, S.** Markov Chain Monte Carlo Methods for Mixture Models. *Ninth Valencia International Meeting on Bayesian Statistics / ISBA World Meeting, Alicante, Spain, June 2010.*
– **Guha, S.** Markov Chain Monte Carlo Methods for Mixture Models. *Joint Statistical Meetings, Washington, DC, August 2009.*
– **Guha, S.**, Li, Y., and Neuberg, D. Bayesian Hidden Markov Modeling of Array CGH Data. *Seminar on Bayesian Inference in Econometrics and Statistics, St. Louis, Missouri, May 2009.*
– **Guha, S.** Posterior Simulation in Countable Mixture Models for Large Datasets. *Joint Statistical Meetings, Denver, Colorado, August 2008.*
– **Guha, S.**, Li, Y., and Neuberg, D. Bayesian Hidden Markov Modeling of Array CGH Data. *JSM - Bayesian Computational Methods Session, Seattle, Washington, August 2008.*
– **Guha, S.**, Li, Y., and Neuberg, D. Bayesian Hidden Markov Modeling of Array CGH Data. *Southern Regional Council on Statistics, Charleston, South Carolina, June 2008.*
– **Guha, S.** Generalized Post-stratification and Importance Sampling for Subsampled Markov Chain Monte Carlo Estimation. *Seminar on Bayesian Inference in Econometrics and Statistics, St. Louis, Missouri, May 2005.*
– **Guha, S.**, MacEachern, S. N., and Peruggia, M. Benchmark Estimation for Markov Chain Monte Carlo Samplers. *Joint Statistical Meetings, New York City, New York, August 2002.*

Invited Academic Presentations

Bocconi University, Italy; Department of Mathematical Sciences, University of Texas at Dallas; Department of Biostatistics, Virginia Commonwealth University; Department of Mathematical and Statistical Sciences, Clemson University; Department of Statistics, Iowa State University; Barretos Cancer Hospital, Brazil; Department of Environmental Health, Harvard TH Chan School of Public Health; Department of Mathematical Sciences, University of Nevada, Las Vegas; ENAR 2020 Spring Meeting; Bayes Comp 2020 held in Gainesville, Florida; Conference on Innovations in Data and Statistical Sciences held in Mumbai, India; Conference on Current Trends in Survey Statistics held in Singapore; 4th Eastern Asia Meeting On Bayesian Statistics (EAC-ISBA) held in Kobe, Japan; ENAR Spring Meeting held in Philadelphia, Pennsylvania; Tenth International Triennial Calcutta Symposium on Probability and Statistics held in Kolkata, India; 12th International Interdisciplinary Conference MSAST 2018 held in Kolkata, India; 11th International Conference of the ERCIM WG on Computational and Methodological Statistics held in Pisa, Italy; IISA International Conference in Statistics held in Gainesville, Florida, from May 2018 - present.

Seminar on Bayesian Inference in Econometrics and Statistics (SBIES) conference held in St. Louis, Missouri; Seminar talks at the Epidemiology and Biostatistics Division, University of Illinois at Chicago; Department of Statistics, Temple University; Department of Statistics, North Carolina State University; Department of Statistics and Applied Mathematics, University of California at Santa Cruz; Statistical and Applied Mathematical Sciences Institute (SAMSI) in RTP, North Carolina, from October 2015 - April 2018.

Joint Statistical Meetings held in Seattle, Washington; 10th Conference on Bayesian Nonparametrics, held in Raleigh, North Carolina; ENAR Spring Meeting held in Miami, Florida; Joint Statistical Meetings held in Boston, Massachusetts; Tenth Annual Conference on Frontiers in Applied and Computational Mathematics held at the New Jersey Institute of Technology in Newark, New Jersey; Eighth Workshop on Bayesian Nonparametrics held at Veracruz, Mexico; SAMSI Program on Semiparametric Bayesian Inference: Applications in Pharmacokinetics and Pharmacodynamics held at Research Triangle Park, North Carolina, from July 2010 - September 2015.

Environmental Statistics Retreat 2010 organized by the Department of Biostatistics, Harvard School of Public Health in Rockport, Massachusetts. Seminar talks at the Department of Mathematics and Statistics, University of Missouri - Kansas City; Division of Biostatistics, Washington University in St. Louis; Department of Biostatistics, Harvard School of Public Health; Department of Statistics, Western Michigan University; Department of Biostatistics, MD Anderson Cancer Center, from October 2009 - June 2010.

Seminar talks at the Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health; Department of Statistics, University of Virginia; Department of Statistics, Iowa State University; Department of Epidemiology & Biostatistics, University of Florida; Department of Biostatistics & Medical Informatics, University of Wisconsin-Madison; Department of Statistics, University of California, Davis; Department of Applied Mathematics and Statistics, University of California, Santa Cruz, from January - March 2007.

Seminar talks at the Department of Ambulatory Care and Prevention, Harvard Medical School; Department of Statistics, Temple University; Department of Statistics, University of California, Irvine; Department of Biostatistics, University of California, Los Angeles; Department of Epidemiology and Biostatistics, University of South Carolina; Department of Health Care Policy, Harvard Medical School; Statistics Department, University of Pennsylvania, from January - March 2007.

Racebrook Environmental Statistics Retreat, held in Sheffield, Massachusetts. Seminar talks at the Department of Statistics, Texas A&M University; Statistics Department, Harvard University; Department of Statistics, Texas A&M University; Department of Mathematical Sciences, University of Arkansas; Department of Statistics, Oklahoma State University, from January 2004 - November 2006.

Ph.D. Students

Shiqi Cui. Dissertation topic: *Bayesian Mixture Models for High-Throughput Bioinformatics Applications.* Co-directed with Marco Ferreira. Graduated in December 2014. Working for Google LLC.

Chiyu Gu. Dissertation topic: *Scalable Bayesian Nonparametric Learning for Biomedical Big*

Data. Graduated in May 2018. Working for Bayer Crop Science.

Chetkar Jha. Dissertation topic: *Bayesian Nonparametric Analysis of Multivariate Unordered Categorical data*. Graduated in May 2019. Assistant Professor at Ahmedabad University.

Dongyan Yan. Dissertation topic: *Scalable Computational Algorithms and Massively Parallel Computing for Bayesian Mixture Models*. Graduated in June 2019. Working for Elli Lilly and Company.

Archie Sachdeva. Dissertation topic: *Bayesian Analytical Methods for Microbiome Data*. Graduated in August 2022. Working for Bristol-Myers Squibb Company.

Software

2026 R package *WMAP*, version 1.3.0, accompanying the paper "WMAP: An R Package for Causal Meta-Analysis by Integrating Multiple Observational Studies" by Guha, Xu, Reed, Priyam, and Li (2026, submitted), is available on CRAN at https://CRAN.R-project.org/package=WMAP.

2024 R package to accompany the paper "Causal Meta-Analysis by Integrating Multiple Observational Studies with Multivariate Outcomes" by Guha and Li (2024, *Biometrics*) is posted on GitHub at https://github.com/sguha-lab/WMAP.

2024 R code to accompany the paper "Bayesian Pairwise Comparison of High-Dimensional Images" by Guha and Qiu (2024) is available on GitHub at https://github.com/sguha-lab/sRPM.

2024 R code for implementing the integrative covariate-balancing weighting strategy developed in Guha and Li (*Biometrics*, 2024) is available on GitHub at https://github.com/sguha-lab/causal-meta-analysis-2024.

2024 R code for implementing the B-MSC method of Guha and Li (2024) is available on GitHub at https://github.com/sguha-lab/CRP.

2023 R code for implementing the BayesDiff method of Gu, Baladandayuthapani, and Guha (2023) is available on GitHub at https://github.com/cgz59/BayesDiff.

2022 R code for generating simulated data and fitting the Bayesian Connectomics (BaCon) model class of Guha, Jung, and Dunson (2022) is publicly available at https://github.com/sguha-lab/BaCon/.

2021 Code for fitting microbiome datasets using the Bayesian SVD-type decomposition technique of Guha and Datta (2021) is available at https://github.com/sguha-lab/Microbiome-SVD.

2020 Code for implementing the Bayesian technique of Guha and Ghosh (2020) for inferring unknown conic sections on the basis of noisy data is available at https://github.com/sguha-lab/BayesConics.

2017 R package *NPCluster* implementing VariScan, a nonparametric Bayesian technique for clustering, variable selection, and prediction in high-throughput regression settings. The methodology is developed in Guha and Baladandayuthapani (*EJS*, 2016). Available at https://github.com/suchard-group/NPCluster.

2015 R package *hmmSeq* implementing the Bayesian technique for analyzing RNA-Seq data proposed in Cui, Guha, Ferreira, and Tegge (*AOAS*, 2015). Joint work with the paper's co-authors.

2011 R package *glmmGS* for fitting generalized linear mixed models to massive datasets. Publicly available from CRAN at http://r-forge.r-project.org/projects/glmmgs. Co-developed with Michele Morara, Louise Ryan, and Christopher Paciorek.

2009 The Bayesian hidden Markov model strategy proposed by Guha, Li, and Neuberg (*JASA*, 2008) for array CGH genomics data is implemented in Bioinformatics Toolbox 3.2. It is available at http://www.mathworks.com.

2007 Created software for generating the sample paths of a number of common stochastic processes, verifying their theoretical properties by simulation, and visualizing abstract results such as the martingale central limit theorem. Used in the Department of Biostatistics, Harvard School of Public Health, to teach the graduate-level courses *Analysis of Failure Time Data* and *Probability Theory and Applications I*.

Teaching Experience

University of Florida

PHC 7090 – Advanced Biostatistical Methods I. *Fall 2020-2024.* Level: Graduate. This is the first course of a two-course sequence providing students with advanced knowledge and mathematical details on statistical inference and linear models. Covers Frequentist and Bayesian estimation and inference. It also encompasses estimation and inference for linear models and fitting of regression models.

PHC 6937 – Analysis of Multivariate Data. *Fall 2018, 2020–2022.* Level: Graduate. Covers linear models methodology including simple and multiple regression and analysis of variance including factorial and block designs. The course covers regression for categorical data, random effects models for correlated data, and nonparametric and semiparametric regression.

PHC 6937 – Frontiers in Biostatistics. *Spring 2018–2020.* Level: Graduate. Introduces Biostatistics Masters and Ph.D. students to current issues and methods in modern biostatistics research. Current faculty present selected topics from their current research.

PHC 6937 – Bayesian Biostatistical Methods. *Fall 2019.* Level: Graduate. Equips students with an understanding of the basics of Bayesian statistics, with special emphasis on practical implementation.

University of Missouri

STAT 9710 – Mathematical Statistics I. *Fall 2011–2014.* Level: Graduate. Theory of estimation and tests of hypotheses including sufficiency, completeness and exponential families. Neyman-Pearson lemma, most powerful tests, similarity and invariance. Bayes and minimum variance unbiased estimates. Confidence intervals and ellipsoids.

STAT 9720 – Mathematical Statistics II. *Spring 2012–2017.* Level: Graduate. Asymptotic distributions of maximum likelihood estimators, chi-square and likelihood ratio test statistics. EM algorithm, bootstrap, and introduction to generalized linear models.

STAT 4710/7710 – Introduction to Mathematical Statistics. *Spring 2008–2010, Spring 2012, Spring 2014, Spring 2017, Fall 2010, Fall 2013, Fall 2016, Fall 2017.* Level: Graduate/Undergraduate. Introduction to theory of probability and statistics using concepts and methods of calculus.

STAT 4510/7510 – Applied Statistical Models I. *Spring 2011, Spring 2013, Spring 2015, Fall 2011–2012, Fall 2014, Fall 2016.* Level: Graduate/Advanced Undergraduate. Introduction to applied linear models including regression (simple and multiple, subset selection, estimation and testing) and analysis of variance (fixed and random effects, multifactor models, contrasts, multiple testing).

STAT 4750/7750 – Introduction to Probability Theory. *Spring 2009–2011, Fall 2009–2010, Fall 2017.* Level: Graduate/Advanced Undergraduate. Probability spaces; random variables and their distributions; repeated trials; probability limit theorems.

STAT 4410/7410 – Biostatistics. *Fall 2009.* Level: Graduate/Advanced Undergraduate. Study of statistical techniques for the design and analysis of clinical trials, laboratory studies and epidemiology. Topics include randomization, power and sample size calculation, sequential monitoring, Carcinogenicity bioassay and case-cohort designs.

STAT 4830/7830 – Categorical Data Analysis. *Fall 2008.* Level: Graduate/Advanced Undergraduate. Discrete distributions, frequency data, multinomial data, chi-square and likelihood ratio tests, logistic regression, log linear models, rates, relative risks, random effects, case studies.

Harvard School of Public Health

BIO503 – Programming and Statistical Modeling in R (co-taught with Christopher Paciorek). *Spring 2007.*

BIO283 – Spatial Statistics for Health Research (co-taught with Louise Ryan, Yi Li, Christopher Paciorek). *Fall 2004.*