Steve Broll

https://stevebroll.github.io/ • sb2643@cornell.edu

Research interests

High-dimensional models, variable selection, -omics data, longitudinal and time series models, data visualization, batch effect correction, tensor decomposition and regression

Education

2020 – 2025 **Cornell University** – Ithaca, New York
PhD in Statistics
Dissertation: *Variable Selection for High-Dimensional Longitudinal Omics Data with a Continuous or Misclassified Binary Outcome*
Committee: Martin T. Wells, Sumanta Basu, Myung Hee Lee

SMoCS Lab Member (Statistical Modeling of Complex Systems)
AIPrN Predoctoral Fellow (AI and Precision Nutrition)

2020 – 2023 **Cornell University** – Ithaca, New York
MS in Statistics

2016 – 2020 **Texas A&M University** – College Station, Texas
BS in Statistics
Minor in Mathematics

Awards and Fellowships

2024 AI and Precision Nutrition (AIPrN) NIH T32 Predoctoral Fellow

2024 Cornell Graduate School Conference Travel Grant

2023 Cornell Graduate School Conference Travel Grant

2020 Dr. Newton Service Award (Texas A&M University)

2018 Statistics Department Scholarship (Texas A&M University)

2018 First Prize, Texas A&M University Institute of Data Science Undergraduate Competition

2016 President's Endowed Scholarship (Texas A&M University)

Publications

2025 **Variable Selection for High-Dimensional Longitudinal Omics Data with a Continuous or Misclassified Binary Outcome**
Steven Broll.
Cornell University.

2025 **PROLONG: Penalized Regression for Outcome guided Longitudinal Omics analysis with Network and Group constraints**
Steven Broll, Sumanta Basu, Myung Hee Lee, Martin T Wells.
Bioinformatics.

2021 **Interpreting blood GLUcose data with R package iglu**
Steven Broll, Jacek Urbanek, David Buchanan, Elizabeth Chun, John Muschelli, Naresh M Punjabi, Irina Gaynanova.
PLOS One.

Preprints

2025 **Urine metabolomic biomarkers linked to C-reactive protein-interleukin-6 axis in persons living with HIV and tuberculosis**
Andrea B. Doltrario, Myung Hee Lee, **Steven Broll**, Kathryn Dupnik, Vanessa Rouzier, Patrice Severe, Nancy Dorvil, Jean W Pape, Serena P. Koenig, Daniel W. Fitzgerald, Kyu Y. Rhee.
medRxiv.

Research experience

2024 – Present **T32 Predoc Fellow**
Committee: Martin T. Wells, Sumanta Basu, Myung Hee Lee, Saurabh Mehta
Continued work on extensions to PROLONG, with applications to omics and precision nutrition data. Debiased group lasso and inference for treatment interaction modeling. EM algorithm for binary outcome with potential misclassification.

2021 – 2023 **Graduate Research Assistant**
Committee: Martin T. Wells, Sumanta Basu, Myung Hee Lee
Batch effect correction, high-dimensional network modeling and differential expression analysis for proteomic data. Batch effect correction and high-dimensional modeling for metabolomic data from various *Mtb* cohorts. Developed PROLONG model and worked on various extensions.

Teaching experience

Spring 2025 — **Teaching Fellow, STSCI 2000: Essential Statistics and Data Science , eCornell - National Ed Equity Lab (NEEL) Partnership**

Fall 2020 — **Teaching assistant, BTRY 6010: Statistical Methods 1, Cornell University**

Talks and Posters

Variable Selection for High-Dimensional Longitudinal Omics Data with a Continuous or Misclassified Binary Outcome

July 2025 — Cornell University B-Exam

Debiased PROLONG

April 2025 — Cornell University Center for Precision Nutrition and Health

PROLONG: Penalized Regression for Outcome guided Longitudinal Omics analysis with Network and Group constraints

November 2024 — Cornell University Center for Precision Nutrition and Health

August 2024 — Joint Statistical Meetings

March 2024 — International Biometric Society Eastern North American Region (ENAR) Meeting

September 2023 — Cornell Celebration of Statistics and Data Science (Poster)

August 2023 — Cornell University A-Exam

August 2023 — Joint Statistical Meetings

March 2023 — International Biometric Society Eastern North American Region (ENAR) Meeting

Mentorship and service

February 2021 – December 2022 — **Cornell Directed Reading Program**
Met weekly with an undergraduate student each semester, selecting a book or set of papers of mutual interest, covering topics from spatiotemporal modeling, statistical learning, measure theoretic probability, nonparametric regression and GAMs.

September 2017 – May 2020

Statistics Peer Mentor

Developed and maintained relationships with and provide assistance and support for first-year students (focus on a smooth transition, acclimation, and a sense of belonging). Maintained regular contact with five to six assigned students, and served as a positive academic and social role model.

Technical skills

Programming languages

Proficient in: R, Shiny, Quarto

Familiar with: Python, C++, Stan

Software

LaTeX, Git

Selected Coursework

Distribution Theory, Theory of Inference, Theory of Linear Models, Multivariate Analysis, Flexible Regression using R, Statistical Computations, Computationally Intensive Stat Methods I and II, Spatial Data Analysis, Time Series and Spatiotemporal Data Science, Mathematical Statistics I and II, Generalized Linear Models, Advanced Statistical Consulting, Bayesian Statistics and Data Analysis, Principles of Analysis II, Probability I and II, Deep Learning Theory and Applications.