William Bekerman

bekerman@wharton.upenn.edu • https://willbekerman.github.io/

EDUCATION

University of Pennsylvania, The Wharton School, Philadelphia, PA **Expected May 2027**
Doctor of Philosophy, Statistics
Honors: National Science Foundation Graduate Research Fellow
Advisor: Dylan Small

Cornell University, College of Agriculture and Life Sciences, Ithaca, NY **August 2018 – May 2022**
Bachelor of Science, Biometry & Statistics (summa cum laude)
Honors: Merrill Presidential Scholar; Hunter R. Rawlings III Cornell Presidential Research Scholar
Advisor: Joe Guinness

PUBLICATIONS

Journal Articles

- Mi, X., **Bekerman, W.**, Sims, P.A., Canoll, P.D., & Hu, J. (2024). RZiMM-scRNA: A regularized zero-inflated mixture model framework for single-cell RNA-seq data. *Annals of Applied Statistics, 18(1),* 1-22.
- **Bekerman, W.**, & Guinness, J. (2023). Comparison of CYGNSS and Jason-3 wind speed measurements via Gaussian processes. *Data Science in Science, 2(1),* 2194349.
- **Bekerman, W.**, & Srivastava, M. (2021). Determining decomposition levels for wavelet denoising using sparsity plot. *IEEE Access, 9*, 110582–110591.

Under Review

- **Bekerman, W.,** Dalal, A., del Ninno, C., & Small, D.S. (2024). Planning for gold: Hypothesis screening with split samples for valid powerful testing in matched observational studies. *Major revision at Biometrika.*
- Jin, H. A., **Bekerman, W.,** Small, D.S., & Rabinowitz, A. (2024). An observational study on effects of contact, collision, and non-contact sports participation on cognitive and emotional health. *Submitted.*

Preprints

- **Bekerman, W.,** Bogomolov, M., Heller, R., & Small, D.S. Protocol for an observational study on the effects of paternal alcohol use disorder on children's later life outcomes. *arXiv:2412.15535.*
- Zhang, R., Kong, J., Small, D. S., & **Bekerman, W.** (2025). Protocol for an observational study on the effects of combinations of adverse childhood experiences on adult depression. *arXiv:2502.17679.*

Working Papers

- **Bekerman, W.,** Small, D.S., & Fogarty, C.B. Multivariate one-sided testing via sample splitting in an observational study of the effect of poverty on children's physical fitness.
- **Bekerman, W.,** Tan, K., & Tchetgen Tchetgen, E.J. Non-asymptotic inference for two-stage least squares.
- **Bekerman, W.,** Bogomolov, M., Heller, R., & Small, D.S. Data turnover for flexible inference with split samples.

PRESENTATIONS

- International Conference on Multiple Comparison Procedures, Philadelphia, PA. August 2025.
- Joint Statistical Meetings, Nashville, TN. August 2025.
  - **IMS Hannan Graduate Student Travel Award Recipient.**
- American Causal Inference Conference, Detroit, MI. May 2025.
- ENAR Spring Meeting, New Orleans, LA. March 2025.
- Larry Brown PhD Student Symposium, Philadelphia, PA. November 2024
- International Seminar on Selective Inference, Virtual Presentation. October 2024.

- Penn Selective Inference Reading Group, Philadelphia, PA. October 2024, February 2025.
- Center for Causal Inference, University of Pennsylvania, Philadelphia, PA. September 2024.
- Penn Conference on Big Data in Biomedical & Population Health Sciences, Philadelphia, PA. September 2024.
- American Causal Inference Conference, Seattle, WA. May 2024.
- Biophysical Society Annual Meeting, San Francisco, CA. February 2022.
- Joint Statistical Meetings, Virtual Conference. August 2021.

INDUSTRY EXPERIENCE & SELECTED PROJECTS

Pharmaceutical Consulting **June 2021 – August 2021**

ZS Associates (Decision Analytics Intern)

- Optimized the efficiency of an omni-strategy messaging campaign for a large pharmaceutical company.
- Identified statistically significant engagement and retention discrepancies due to campaign overlap and messaging frequency.
- Developed and presented detailed, action-oriented recommendations during various client meetings.

NFL Big Data Bowl 2020 **February 2020 – April 2020**

Cornell Data Science (Project Lead)

- Predicted the result of rushing plays using game, play, and player data provided by NFL's Next Gen Stats.
- Constructed a convolutional neural network model using the Keras library in Python and achieved a mean absolute error of approximately 2.35 yards and a cumulative rank probability score (CRPS) of approximately 0.0118 (top 1% in competition).
- Implemented a gradient boosting machine and random forest regression model, improving mean absolute error and CRPS by almost 15% compared to a baseline linear regression model.
- Visualized the data provided by NFL's Next Gen Stats and the outcome of each play using JavaScript.

Fake News Challenge **February 2019 – April 2019**

Cornell Data Science

- Built a random forest model based on bag of words features to determine the relevance of an article for the Fake News Challenge.
- Constructed a siamese neural network model to simultaneously analyze the headline and body of an article to discern its stance.
- Achieved a relevance detection accuracy score of over 95% and stance detection accuracy of almost 75% (top ten in Challenge).
- Created interactive visualizations of relevance/stance detection features and LSTM activations overlaid onto source text.
- *** Awarded First Prize ($750) by Sandia National Laboratories at Cornell University's Bits on Our Mind Showcase.

UNDERGRADUATE TEACHING EXPERIENCE

- BTRY 4030 - Linear Models with Matrices. Fall 2021
- BTRY 6020 - Statistical Methods II (Graduate-Level). Spring 2021
- INFO 2950 - Introduction to Data Science. Fall 2020
- PLBIO 2400- Green World, Blue Planet. Fall 2019, Fall 2020, Fall 2021

SERVICE AND LEADERSHIP

Internal

Wharton Statistics Student Seminar Co-Organizer

Mentoring

PhD Students: Zihan Zhu (Penn)
Undergraduate Students: Hannah Jin (Penn, now PhD student at Harvard), Matthew Spivey (Penn), Ruizhe Zhang (Fudan)
Served as peer advisor of students in Biometry & Statistics Major at Cornell

Reviewing

Journal of Biopharmaceutical Statistics, Journal of Quantitative Analysis in Sports

Judging

Long Island Science and Engineering Fair: Intelligent Machines, Mathematics, Software, Robotics

OTHER ACTIVITIES

- Alpha Phi Omega Service Fraternity (Family Head, Leadership/Friendship/Service Award Recipient)
- Cornell Data Science Project Team (Team Lead, Project Lead, Recruitment Chair)

PROFESSIONAL MEMBERSHIPS

- American Statistical Association
- Institute of Mathematical Statistics
- International Biometrics Society
- Society for Causal Inference