

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

Under Review

- **Bekerman, W.**, Dalal, A., del Ninno, C., & Small, D.S. (2024). Hypothesis selection via sample splitting for valid powerful testing in matched observational studies.
- 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.

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.

PRESENTATIONS

- **Bekerman, W.** (2024, September). Sample splitting for multivariate one-sided testing in an observational study on the fitness of impoverished children. Invited talk upcoming at Center for Causal Inference, University of Pennsylvania, Philadelphia, PA.
- **Bekerman, W.** (2024, May). A split-sampling framework for powerful design of observational studies under unmeasured confounding. Abstract presented at American Causal Inference Conference, Seattle, WA.
- **Bekerman, W.** (2022, February). Enabling dynamics studies of proteins at low concentrations using electron spin resonance. Abstract presented at Biophysical Society Annual Meeting, San Francisco, CA.
- **Bekerman, W.** (2021, August). RZiMM-scRNA: A regularized zero-inflated mixture model framework for single-cell RNA-seq data. Abstract presented at Joint Statistical Meetings, Virtual Conference.

INDUSTRY EXPERIENCES & 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

SELECTED ACTIVITIES

- Alpha Phi Omega Service Fraternity (Family Head, Leadership/Friendship/Service Award Recipient)
- Biometry & Statistics Peer Advisor
- Cornell Data Science Project Team (Team Lead, Project Lead, Recruitment Chair)
- Cornell University Sustainable Design Project Team
- Wharton Bridge Program Peer Mentor
- Wharton Statistics Student Seminar Organizer

PROFESSIONAL MEMBERSHIPS

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