Ha Nguyen

tn325@cornell.edu • (507) 321-0036 • Google Scholar • GitHub

Summary

PhD Candidate in Statistics with a strong mathematical foundation and broad interdisciplinary research experience. Previously a health economics consultant, I led statistical analyses and developed production R packages used across 100+ projects. My current research focuses on high-dimensional statistics, convex optimization, and machine learning, including algorithm development in R/C++ for large-scale network modeling and shape representation for personalized apparel design. Seeking a summer internship in data science or machine learning.

Education

Cornell University, Ithaca, NY — 2022 – Jun 2027 (expected)
PhD in Statistics and Data Science
Topics: High-Dimensional Statistics, Statistical Machine Learning; Graphical Models, Sparse Estimation; Matrix Computation, Convex Optimization
Advisors: Sumanta Basu and Martin T. Wells

St Olaf College, Northfield, MN — 2015 – 2019
BA in Psychology and Mathematics, *magna cum laude*
Advisor: Shelly Dickinson

Work Experience

Analysis Group, Boston, MA — 2019 – 2022
Senior Analyst, Health Economics and Outcomes Research (HEOR)

- Led statistical analyses for numerous clinical and real-world evidence studies, including treatment effect estimation, regression modeling, and survival analysis
- Built core features for a multi-million dollar analytics platform supporting oncology drug pricing decisions, generating significant cost-saving opportunities for the client
- Developed and maintained internal R packages used in 100+ healthcare cases per year to improve the efficiency of statistical data analysis, reducing case budgets by up to 70%

Research Experience

Cornell Statistical Modeling of Complex Systems Lab — 2023 – present
Graduate Student Researcher · Advisor: Sumanta Basu

AutotuneGLASSO: Efficient, Tuning-Free Learning of High-Dimensional Networks

- Developed a tuning-free method for Gaussian graphical model selection that automatically learns nodewise penalties based on feature variances, removing the need for data standardization and manual hyperparameter tuning
- Demonstrated superior estimation accuracy and graph recovery compared to existing tuning methods in both simulations and real-world data; manuscript in preparation
- Built the ATTglasso R package with a high-performance C++ backend, enabling plug-and-play use across domains such as fMRI brain connectivity and financial networks

Algorithmic Garment Design (collaborative)

- Developed spline-based shape representations from 2D coordinate data to model lower-torso curvature in women, achieving high curve-fitting accuracy (median $R^2 > 0.9$)
- Built interpretable supervised models that predict individual curve shapes from body metrics to support personalized fit
- Designing algorithms to reconstruct skirt contours and optimize fit from garment feedback data, applicable to data-driven clothing recommendation systems

Thomas Jefferson University Biostatistics Department — Summer 2018
Student Researcher · Advisor: Benjamin Leiby

- Developed mixed-effects models to analyze longitudinal clinical trial data assessing the impact of a novel cognitive behavioral therapy on comorbid depression in elders with cognitive impairment

- Reviewed literature on missing data mechanisms and implemented imputation strategies to address differential attrition between treatment and control groups

St Olaf Center for Interdisciplinary Research 2017
Research Fellow · Advisor: Shelly Dickinson

- Performed data visualization and hypothesis testing to assess motivational effects of alcohol conditioning in mice

Publications

Google Scholar

Epidemiology and Economic Burden of VHL-Associated Brain and Pancreatic Tumors in the U.S.
Jonasch E, Song Y, Freimark J, Berman R, **Nguyen H**, et al. *Orphanet J Rare Disease*, 2024.

Epidemiology and Economic Burden of VHL-Associated Renal Cell Carcinoma in the U.S.
Jonasch E, Song Y, Freimark J, Berman R, **Nguyen H**, et al. *Clinical Genitourinary Cancer*, 2023.

Disease Progression Rates in Ambulatory Duchenne Muscular Dystrophy by Steroid Type
McDonald C, Marden J, Shieh P, …, **Nguyen H**, et al. *J Comparative Effectiveness Research*, 2023.

Benefits of Autoantibody Enrichment in Early Rheumatoid Arthritis
Michaud K, Conaghan PG, Park SH, …, **Nguyen H**, et al. *Rheumatology and Therapy*, 2023.

Outcomes of Long-Acting Injectables vs. Oral Antipsychotics in U.S. Patients with Schizophrenia
Lin D, Thompson-Leduc P, Ghelerter I, **Nguyen H**, et al. *CNS Drugs*, 2021.

Predicting Remission in Bio-Naïve Crohn's Disease: Vedolizumab vs. Anti-TNFα
Mantzaris G, Yarur A, Wang S, …, **Nguyen H**, et al. *J Crohn's and Colitis*, 2021.

Hospitalization Odds by Antipsychotic Dosing in Schizophrenia: A Network Meta-Analysis
Correll CU, Cook E, Mu F, Ayyagari R, Young J, **Nguyen H**, et al. *Eur Neuropsychopharmacol*, 2021.

Awards & Scholarships

Collaborator, NSF Grant, Cornell Fashion & Body Tech Lab	2025
Statistics Department Summer Research Fellowship, Cornell University	2023, 2024, 2025
Dean's List, St Olaf College	2015 – 2019
Phi Beta Kappa Honor Society	2019
Pi Mu Epsilon Honor Society	2019
1st Place, MinneMUDAC Data Challenge on Voter Turnout Prediction	2018
Center for Interdisciplinary Research Fellowship, St Olaf College	2017

Service & Outreach

Mentor, Cornell Directed Reading on High-Dimensional Statistics	2025
Professional Chair, Cornell Statistics Graduate Society	2023
Co-leader, HEOR Junior Analyst Training Program, Analysis Group	2021 – 2022
President, St Olaf College Sports Analytics Club	2019

Teaching

Lead Teaching Assistant, Cornell University + National Ed Equity Lab 2025
- STSCI 2000: Introductory Statistics and Data Science for underserved high school students

Course Developer, eCornell Online Programs 2023 – 2025
- Developed online courses on Natural Language Processing and Text Analysis in R

Teaching Assistant, Cornell University
- ILRST 2100: Introductory Statistics and Data Science — Fall 2024
- ILRST 2150: Introductory Statistics for Biology — Fall 2023
- ILRST 3110: Probability Models and Inference — Spring 2023
- ILRST 2100: Introductory Statistics — Fall 2022

Skills

Programming: *Fluent:* Python, R · *Familiar:* C/C++, SAS, SQL
Tools: Pandas, R Shiny, Git