

Curriculum Vitae

Yihao Tan

Email: yihaotanus@outlook.com; yihaotan@buffalo.edu

Personal Website: YihaoTancn.github.io

EDUCATION

09/2022 - Expected 12/2026

- Phd in Biostatistics | State University of New York at Buffalo, NY, USA
- Thesis: Nonparametric Empirical Bayes Modeling on Spontaneous Reporting System Data for Pharmacovigilance.
- Supervised by Drs. Saptarshi Chakraborty & Marianthi Markatou
- GPA: 3.91/4.0

09/2019 - 06/2022

- Master of Art in Biostatistics | State University of New York at Buffalo, NY, USA
- Thesis: The Performance of Clustering Algorithms Under Measurement Error.
- Supervised by Dr. Marianthi Markatou
- GPA: 3.97/4.0

09/2013 - 07/2017

- Bachelor of Science in Biotechnology | Shenzhen University, Guangdong, China

RESEARCH INTERESTS

Bayesian Analysis, Empirical Bayes, Non-parametric Statistics, Pharmacovigilance, Clustering, Measurement Error, Mixed-type Data, Longitudinal Analysis.

PUBLICATIONS

1. **Tan Y**, Markatou M, Chakraborty S, Flexible Empirical Bayesian Approaches to Pharmacovigilance for Simultaneous Signal Detection and Signal Strength Estimation in

Spontaneous Reporting Systems Data, *Statistics in Medicine* 44, no. 18-19 (2025): e70195, <https://doi.org/10.1002/sim.70195>.

2. **Tan Y**, Markatou M, Chakraborty S. pvEBayes: An R Package for Empirical Bayes Methods in Pharmacovigilance. *arXiv* preprint arXiv:2512.01057. 2025 Nov 30. Submitted to *R Journal*.
3. Wang T, Fan L, He L, **Tan Y**, Zhou Y, Chen L, Wei J, Xie H, Chen P, Demographic Disparities in Musculoskeletal Pain: Exploring Trends and Public Health Strategies for Prevention and Intervention. Submitted to *PLOS One*.
4. Markatou M, Mukhopadhyay R, Good J, **Tan Y**, Dharia A, Kennedy O, Brown L, Talal A, From Innovation to Impact: The CREATE Framework as a Blueprint for Large Language Model (LLM) Adoption in Opioid Treatment Programs (OTPs). Submitted to *Journal of Medical Internet Research*.
5. **Tan Y**, Markatou M, Chakraborty S, A Review of Statistical Methods for Spontaneous Reporting System Data Mining: Signal Detection and Beyond. *arXiv* preprint arXiv:2604.18898. 2026 Apr 21. Submitted to *Statistical Methods in Medical Research*.

IN PROGRESS

1. **Tan Y**, Mukhopadhyay R, Markatou M, Clustering longitudinal mixed-type data using extended KAMILA algorithm. Ongoing research.
2. **Tan Y**, Mukhopadhyay R, Markatou M, Clustering with Quadratic distance. Ongoing research.
3. Dharia A, Markatou M, Good J, **Tan Y**, Mukhopadhyay R, Talal A, Focus group study. Ongoing research.

STATISTICAL SOFTWARE DEVELOPMENT

- **pvEBayes**: Empirical Bayes Methods for Pharmacovigilance, **Creator & Maintainer**
 - An open-source R package offering a suite of empirical Bayes methods for spontaneous reporting systems (SRS) data mining.
 - Provides accessible, well documented R-based front end, while several computation-intensive routines are written in highly efficient custom C++ codes.
 - Published and maintained on [CRAN](#) and [GitHub](#).
 - Currently under review by [rOpenSci](#).

EMPLOYMENT

- **Research Assistant**, SUNY Research Foundation | 1/2025 - current
 - Collaborating on the research project [“Developing Methods for Clustering Longitudinal Mixed-Type Data for Comparative Effectiveness Research”](#) led by

Dr. Marianthi Markatou (Funder: Patient-Centered Outcomes Research Institute (PCORI)).

- Participating in data acquisition design across multiple clinics to ensure sample representativeness, and downstream data processing, for example converting paper-based clinical records into structured datasets.
 - Core analysis team member in focus group study.
 - Participating in research publications.
-
- **Teaching Assistant**, SUNY, University at Buffalo | 2/2023 - 12/2024
 - Supported instruction for graduate and undergraduate courses, including Statistical Analysis I & II (STA527/528), Statistical Data Mining I (STA 545), and Statistical Methods (STA119).
 - **Research Assistant**, SUNY Research Foundation | 2/2021 - 2/2023
 - Digitized and transformed raw, paper-based clinical records into structured, analyzable datasets.
 - Managed the acquisition, cleaning, and processing of FibroScan datasets to prepare them for downstream statistical analysis.
 - Investigated and developed clustering methodologies for handling data with measurement error.

TEACHING

- **Teaching Assistant**, SUNY, University at Buffalo | 2/2023 - 12/2024
 - Supported instruction for graduate and undergraduate courses, including Statistical Analysis I & II (STA527/528), Statistical Data Mining I (STA 545), and Statistical Methods (STA119).
 - Led weekly recitation sessions, guiding students through practical exercises and reviewing lecture concepts.
 - Conducted regular office hours to provide one-on-one mentorship, clarify concepts from the class, and assist with coursework troubleshooting.

CONFERENCE PARTICIPATION

1. **Tan Y**, Markatou M, Chakraborty S, *On the Use of Non-parametric Empirical Bayesian Methods in Pharmacovigilance*. Oral Presentation at the ENAR Spring Meeting. Baltimore, MD. March 2024
2. **Tan Y**, Markatou M, *The Performance of Clustering Algorithms under Measurement Error*. Poster Presentation at the UP-STAT Conference. Buffalo, NY. May 2022.

SERVICE

- Conference Volunteer, UP-STAT Conference, Buffalo, NY, May 2022

AWARD, HONORS & AFFILIATIONS

- **Student Member**, Eastern North American Region (ENAR) of the International Biometric Society. Jan 2022 - Present
- **Sidney Adelman Outstanding Master's Student Award**, Department of Biostatistics, School of Public Health and Health Professions, State University of New York at Buffalo, Buffalo, NY. May 2022.
- **Student Member of Delta Omega Honorary Society Gamma Lambda Chapter**, School of Public Health and Health Professions, State University of New York at Buffalo, Buffalo, NY. May 2022.
- **Provincial Second Prize**, China High School Biology Olympiad, Botanical Society of China & China Zoological Society, Nov 2012.

SKILLS

- **Computer:** R, C++, Python, SAS, MATLAB, Java, MySQL.
- **Systems & Infrastructure:** High-Performance Computing (HPC), SLURM, Linux/Unix, Git.