Zhou (Joe) Lan

Residence: Boston Metropolitan Area
E-mail: zhou.joe.lan@gmail.com ✻ *Telephone number:* +1 (404) 834-2768
Website: https://lanzhoubwh.github.io/
Google Scholar: https://scholar.google.com/citations?user=fB87fIgAAAAJ&hl=en

Key Expertise

Large Database Digital Health Analytics — Using real-world evidence data, including registry data, claims data, electronic health records (EHRs), and biobank data, for evaluating treatment effectiveness, patient trajectory modeling, and biomarker investigation.

Clinical Trials Design and Analytics — Supporting late-phase clinical studies, including trial design, statistical analysis, and trial endpoint evaluation using rigorous statistical methodology.

Methodological Innovations in Omics Data — Statistical and machine learning innovations in omics data for biomarker investigation using advanced machine learning methods.

Education

North Carolina State University — *Department of Statistics*
PhD in Statistics — *Aug 2015 – Aug 2019*

Georgia Institute of Technology — *Department of Mathematics*
Master of Science in Statistics — *May 2012 – Aug 2015*

Zhejiang University — *College of Science*
Bachelor's Degree — *Aug 2008 – May 2012*

Work Experience

Sanofi — *Jan 2026 – Present*
Associate Director, Real World Evidence Analytics — *Cambridge, MA*

- Leading the development of study protocols, statistical analysis plans (SAPs), and advanced analytics for observational studies based on claims and EHR data to evaluate treatment effectiveness and safety, including patient trajectory modeling and longitudinal outcome analyses across Sanofi therapeutic areas.
- Integrating novel health informatics techniques, including trial emulation and machine learning-based phenotyping, with advanced statistical methodologies, including causal inference and predictive modeling, to strengthen the RWE team's methodological toolkit.
- Managing team activities, including work planning and prioritization, technical mentorship, cross-functional collaboration, stakeholder communication, and timely delivery of high-quality analytical outputs.

Brigham and Women's Hospital, Harvard Medical School — *Oct 2022 – Jan 2026*
Investigator; Faculty Member in Medicine — *Boston, MA*

- Served as a lead trial biostatistician in late-phase clinical trials and interventional studies, with responsibilities including protocol development, study design, primary and secondary analyses, post-hoc analyses, and publications.
- Led study protocol development, statistical analysis planning, and advanced analytics for interventional and observational studies based on EHR databases to evaluate biomarker predictive performance in cancer imaging, genetics, and cardiovascular disease.

- Led statistical and digital innovations in imaging biomarker investigation for neurological disease using machine learning methods.

Center for Outcomes Research & Evaluation, Yale School of Medicine *Jul 2020 – Sep 2022*
Faculty Statistician *New Haven, CT*

- Provided statistical support for Medicare claims analytics, healthcare quality measurement, and regulatory-grade modeling.
- Delivered statistical education and consultation for non-statistical members of the center, including physicians and trainees.

Eli Lilly and Company *May 2018 – Aug 2018*
Ph.D. Statistics Intern *Indianapolos, IN*

- Developed patient classification based on Bayesian non-parametric clustering

Technical Skills

Programming Languages & Tools	Python, R, C/C++, Bash, SAS, SQL, Slurm Workload Manager
Statistics & Machine Learning	Survival Analysis, Bayesian Modeling, Probabilistic Graphical Models, Quantile Regression, Spatial Statistics, High-Dimensional Data Analysis, Causal Inference
Computation	High-Performance Computing, Parallel Computing, Algorithm Development
Therapeutic Areas	Radiology, Neurology, Cardiovascular Medicine, Oncology, Endocrinology

Databases & Data Sources with Hands-On Expertise

Claims & EHR Databases	MarketScan, STATinMED, TriNetX
Biobank Databases	Human Connectome Project, UK Biobank
Registry Databases	National Inpatient Sample, Boston Scientific ALTITUDE Database, NCDR ICD Registry, Centers for Medicare and Medicaid Services data

Representative Projects

Observational Studies Based on Claims/EHR Data for Epidemiological Investigation

- **Project 1 (Sanofi): Treatment Patterns for Relapsing-Remitting Multiple Sclerosis (RRMS) Therapy:**
  - Built and validated patient cohorts from the large **MarketScan** database to study therapy patterns among new users of relapsing-remitting multiple sclerosis (RRMS) therapies.
  - Constructed and visualized treatment trajectories to analyze treatment duration and switching patterns.
  - Performed unsupervised clustering of treatment patterns using K-means based on the number of switches and time to terminal treatment.
- **Project 2 (BWH/HMS): Real-World Evaluation of Statin Therapy Effectiveness in Patients with Diabetes:**
  - Built and validated patient cohorts from large **EHR repositories** to study statin initiation and treatment patterns in diabetes care.

  - Applied **advanced causal inference** methods, including propensity score weighting, mediation modeling, and competing-risks Cox survival models, to evaluate cardiovascular outcomes.
  - Led statistical methodology and results interpretation; contributed methods and results sections for peer-reviewed publications.
    * Shah, Nisarg, Zhou Lan, C. Justin Brown, Seth S. Martin, and Alexander Turchin, "Impact of Statin Nonacceptance on Cardiovascular Outcomes in Patients With Diabetes," *Journal of the American Heart Association* **14**, 11 (2025). DOI: 10.1161/JAHA.124.040464
    * C. Justin Brown, PharmD; Lee-Shing Chang, MD; Naoshi Hosomura, DDS, DMSc, MBA; Shervin Malmasi, PhD; Fritha Morrison, PhD; Maria Shubina, ScD; Zhou Lan, PhD; Alexander Turchin, MD, MS, "Assessment of Sex Disparities in Nonacceptance of Statin Therapy and Low-Density Lipoprotein Cholesterol Levels Among Patients at High Cardiovascular Risk," *JAMA Network Open* **6**, 2 (2023). DOI: 10.1001/jamanetworkopen.2023.1047

- **Project 3 (BWH/HMS): Causal Survival Analysis for Cardiovascular Events in Patients Treated with Sulfonylureas: A Target Trial Emulation**
  - Participated a target trial process using observational data from 12 health systems and insurance plans across the US for people with T2D and hyperglycemia.
  - Performed causal survival analysis to evaluate the time-varying treatment effect of individual sulfonylureas (glimepiride, glipizide and glyburide) vs DPP4i.
    * Murray, Eleanor J., Ellen C. Caniglia, and Lucia C. Petito, "Causal survival analysis: a guide to estimating intention-to-treat and per-protocol effects from randomized clinical trials with non-adherence," *Research Methods in Medicine & Health Sciences* **2**, 1 (2021). DOI: 10.1177/2632084320961994
    * Turchin, Alexander, Lucia C. Petito, Emily Hegermiller, Ryan Carnahan, Amanda DeVries, Shikha Goel, Maria C. Lansang, Michael E. McDonnell, Vinaya Nair, Emily Priest, and Vicki J. Willey, "Cardiovascular events in individuals treated with sulfonylureas or dipeptidyl peptidase 4 inhibitors," *JAMA Network Open* **8**, 7 (2025). DOI: 10.1001/jamanetworkopen.2025.23067

- **Project 4 (BWH/HMS): Evaluating Reliability of NLP-Derived Clinical Variables in Real-World Epidemiologic Analyses:**
  - Proposed and led a methodological study evaluating how **NLP extraction errors from clinical notes** affect causal and epidemiologic inference in real-world data research.
  - Designed and executed **Monte Carlo simulation experiments** using high-performance computing to quantify downstream bias under different NLP error scenarios. *Code: `https://github.com/lanzhouBWH/NLP-Downstream-Impact`*
  - Demonstrated the importance of **data validation and error modeling** in clinical AI pipelines to support robust RWE and regulatory-grade analytics.
  - Publications/Reports:
    * Zhou Lan, Alexander Turchin, "Impact of Possible Errors in NLP-Derived Data on Downstream Epidemiologic Analysis," *JAMIA Open* **6**, 4 (2023). DOI: 10.1093/jamiaopen/ooad111

EHR Databases for Evaluating Predictive Performance of Biomarkers

- **Project 1 (CORE/Yale): Digital Cardiac Monitoring and Predictive Modeling for ICD Pacing Risk:**
  - Built study cohorts from multi-institutional cardiac device registry data, including Boston Scientific's ALTITUDE database and the NCDR ICD Registry, by architecting and implementing scalable analytic pipelines.
  - Led statistical modeling to identify **predictors of right ventricular pacing** in patients receiving implantable cardioverter-defibrillators (ICDs), using variable selection and predictive modeling approaches.

- Contributed expertise in **clinical device signal analytics, clinical interpretation, and validation** in collaboration with cardiologists, electrophysiologists, and device-engineering specialists.
- Publication/Reports:
  - Hummel JP, Lan Z, Jones PW, Khera R, Stein K, Curtis JP, Akar J, "Predictors of Right Ventricular Pacing in Patients Undergoing Implantable Defibrillator Placement," *Journal of Cardiovascular Electrophysiology* **36**, 4 (2025).

- **Project 2 (BWH/HMS): Non-Invasive MRI Biomarkers for Tumor Profiling:**
  - Constructed real-world patient cohorts from clinical records to evaluate **clinical, molecular, and radiological predictors** of outcomes in astrocytoma and IDH-mutant glioma.
  - Applied **predictive modeling and survival analysis**, including weighted Cox models for non-proportional hazards, to assess associations with overall survival (OS) and progression-free survival (PFS).
  - Integrated **radiology, genomics, and clinical features** to support precision oncology and treatment response profiling.
  - Performed **multiple imputation** to address missing data and improve robustness of real-world evidence analyses.
  - Publications/Reports:
    - Aleksandra B. Lasica, Zhou Lan, ..., Raymond Y. Huang, Patrick Y. Wen, Gilbert Youssef, "Clinical, Molecular and Radiological Predictors of Prognosis in Newly Diagnosed Astrocytoma, IDH-Mutant, WHO Grade 4," *Neuro-Oncology* **7**, 1 (2025). DOI: 10.1093/neuonc/noaf133
    - Tyler A. Lanman, ..., Zhou Lan, ..., Patrick Y. Wen, L. Nicolas Gonzalez Castro, "Ivosidenib for the Treatment of IDH1-Mutant Glioma, Grades 2 to 4: Tolerability, Predictors of Response, and Outcomes," *Neuro-Oncology Advances* **7**, 1 (2025). DOI: 10.1093/noajnl/vdae227

Clinical Trials Design and Analytics

- **Project 1 (BWH/HMS): Apixaban for Extended Treatment of Provoked Venous Thromboembolism: HI-PRO Trial**
  - Performed **study design** and **primary and secondary endpoint analyses**, generating **regulatory-grade evidence** consistent with clinical trial standards.
  - Contributed to **clinical trial manuscripts and regulatory documentation**, including statistical interpretation and reporting aligned with ICH E9 guidelines.
  - Publications/Reports:
    - Gregory Piazza, Behnood Bikdeli, Arvind K. Pandey, Darsiya Krishnathasan, Candrika D. Khairani, Antoine Bejjani, Ruth H. Morrison, Heather Hogan, Sina Rashedi, Mariana Pfeferman, Junyang Lou, John Fanikos, Nicole Porio, Lisa Rosenbaum, Piotr Sobieszczyk, Zhou Lan, Marie Gerhard-Herman, Umberto Campia, Samuel Z. Goldhaber, for the HI-PRO Trial Investigators, "Apixaban for Extended Treatment of Provoked Venous Thromboembolism: HI-PRO Trial," *New England Journal of Medicine* **393**, (2025). DOI: 10.1056/NEJMoa2509426

Statistical and Machine Learning Innovations in Digital Biomarkers from Omics Data

- **Project 1 (BWH/HMS): Machine Learning and AI Methods for Analyzing Brain White Matter Bundles Using Large Biobank Data: Human Connectome Project–Young Adults Data**
  - Developed a scalable **multimodal neuroimaging analytics pipeline** using *UKFTractography* and *WhiteMatterAnalysis* for over 1,000 participants from the Human Connectome Project.
  - Developed a **novel statistical framework, Fiber Microstructure Quantile Regression**, for white-matter biomarker discovery, addressing high-dimensional tractography, spatial correlation, and heterogeneity across fiber bundles.
  - Applied and evaluated **deep learning techniques** for cerebellar tract segmentation and neuroanatomical mapping.

- Validated digital endpoints, including FA, MD, and number of streamlines, through **numerical studies** for reliability and biomarker robustness.
- Built a reproducible, Docker-based **open-source software pipeline** for tractography analytics and visualization using computer-vision techniques. GitHub: `https://github.com/lanzhouBWH`
- Publications/Reports:
  * Zhou Lan, Yuqian Chen, Jarrett Rushmore, Leo Zekelman, Nikos Makris, Yogesh Rathi, Alexandra J. Golby, Fan Zhang, and Lauren J. O'Donnell, "Fiber Microstructure Quantile Regression: A Novel Statistical Approach for Analyzing White Matter Bundles from Periphery to Core," *Imaging Neuroscience* **3**, 11 (2025). DOI: 10.1162/imag_a_00569
  * Leo R. Zekelman, Suheyla Cetin-Karayumak, Yuqian Chen, Melyssa Almeida, Jon Haitz Legarreta, Jarrett Rushmore, Steve Pieper, Zhou Lan, John E. Desmond, Lissa C. Baird, Nikos Makris, Yogesh Rathi, Fan Zhang, Alexandra J. Golby, Lauren J. O'Donnell, "Consistent Cerebellar Pathway-Cognition Associations Across Pre-Adolescents and Young Adults: A Diffusion MRI Study of 9000+ Participants," *bioRxiv* , (2025). DOI: 10.1101/2025.02.05.636737
  * Legarreta, Jon Haitz, Zhou Lan, Yuqian Chen, Fan Zhang, Edward Yeterian, Nikos Makris, Jarrett Rushmore, Yogesh Rathi, and Lauren J. O'Donnell, "Towards an Informed Choice of Diffusion MRI Image Contrasts for Cerebellar Segmentation," *Human Brain Mapping* **46**, 11 (2025). DOI: 10.1002/hbm.70317

- **Project 2 (BWH/HMS): Magnetic Resonance Spectroscopy-Based Imaging Biomarker Development for Functional Neurological Disorder in Children and Adolescents Using Machine Learning Methods**
  - Developed **NMetNet**, a novel Bayesian network-based statistical framework using Bayesian graphical lasso for identifying **neurometabolic imaging biomarkers** in pediatric FND.
  - Led analysis of high-dimensional MRS data to quantify metabolite interactions and distinguish FND patients from controls using **interpretable network-based machine learning**.
  - Advanced noninvasive biomarker research for pediatric neurological disease, supporting precision diagnostics and early disease characterization.
  - Publications/Reports:
    * Zhou Lan, Sheryl Foster, Molly Charney, Max van Grinsven, Katherine Breedlove, Kasia Kozlowska, Alexander Lin, "Neurometabolic Network for Functional Neurological Disorder in Children and Adolescents," *NeuroImage: Clinical* **46**, (2025). DOI: 10.1016/j.nicl.2025.103767