

## Senior Statistical Data Analyst

### Position Summary

The Zhang Lab at Institute for Informatics, Data Science, and Biostatistics (I2DB) focuses on developing and applying innovative causal machine learning to electronic health record (EHR) and claims databases to support clinical and regulatory decision-making. This position is an unparalleled opportunity for the right applicant to make an impact on the future of health care by building ML/AI algorithms to unleash the potential of big data in healthcare. The postdoctoral research associate will take part in challenging projects such as 1) building unsupervised learning models to identify patient subpopulations for precision medicine, 2) developing time-series deep learning and Transformer-based models for disease risk prediction and individualized treatment effect estimation from longitudinal patient trajectory 3) developing federated learning algorithms for causal inference across multiple institutions. This role will also include collaboration with both internal and external clinical researchers to apply methods for generating real-world evidence from EHR and claims data, such as with the Observational Health Data Sciences and Informatics (OHDSI) collaborative network.

### Job Description

#### Primary Duties & Responsibilities

- Develop, implement and evaluate machine learning and deep learning models to assist the research team in addressing clinical questions.
- Write SQL queries to extract relevant data from large-scale databases.
- Preprocess data for model training, including imputing missing values, removing outliers, characterizing dataset, and any other relevant data wrangling.
- Search and review literature in medicine, computer science, and statistics for related research projects.
- Attend weekly lab meetings and meetings with the PI, the collaborators, and the students as needed.
- Performs other duties as assigned.

### Required Qualifications

- Equivalent of Master's degree in computer science, statistics, biomedical informatics, biomedical engineering, or a related quantitative field with one or more years of related experience.
- Demonstrated experience working on projects related to deep learning, representation learning, federated learning, causal inference or other machine learning and artificial intelligence-related modeling.
- Experience with PyTorch, Tensorflow, R/Rstudio, and Linux command line.
- Knowledge on basic probability and statistical inference

### Preferred Qualifications

- Experience interacting with EHR and claims database for healthcare-related research.
- Experience using computing clusters (batch processing), cloud computing services, and containerization.
- Effective verbal, written and interpersonal communication skills.
- Passion for improving clinical decision making and patient care through data-driven analytics.

### Applicant Special Instructions

Please send the following documents to Dr. Linying Zhang at: [linyingz@wustl.edu](mailto:linyingz@wustl.edu).

- A cover letter highlighting research interests and career goals
- CV/Resume
- 2-3 publications or writing samples