

Xinlei Deng

Principal Biostatistician

London • Xinlei.Deng.Apha@gmail.com • +44 (7887) 296 811 • <https://xinleidendeng.github.io/>

Summary

- Principal biostatistician at Novartis with 7+ years' experience in statistical modelling and building applications.
- Proficient in multiple coding languages including R/Rshiny, Python, Swift (iOS app), SAS, PostgreSQL, etc.
- Specialized in statistical models, machine learning (Transformer/Densenet), and remote sensing (spatial analysis).

Skills

Biostatistics • Data Science • iOS development • Python • R • SAS • Swift

Positions

Principal Biostatistician

Novartis | Mar 2024 - Present

- Responsible for all statistical tasks on assigned trials (Phase II and Phase III) independently, protocol development in alignment with the development plan, statistical analysis plans, and reporting activities.
- Contribute to planning and execution of exploratory analyses, and/or PK, PK/PD analyses, exploratory biomarker and diagnostic analyses, statistical consultation, and statistical initiative (CAMIS hackathon).
- Conduct the PK/PD simulations and dose-response analyses to support dose justification for submission.
- Contribute to internal strategies for covariate adjustment in clinical trials (R package).
- Initiate, drive and implement novel methods and innovative trial designs with multiple statisticians.
- Provide statistical expertise to support submission activities (Phase II and Phase III) and documents, meetings with and responses to Health Authorities and other drug development activities.
- Contribute to interactions with external review committees, external consultants and other external parties with oversight.
- Responsible for functional alignment about the status/issues of assigned trials.
- Explain statistical concepts in manner easily understood by non-statisticians and provide statistical justifications for actions/statements.
- Establish sound working relationships and effective communication within the Clinical Trial, Biostatistics, & Pharmacometrics team.
- Oversee all Biostatistics resources and deliverables for assigned trials. Ensure timeliness and adequate quality of all Biostatistics deliverables for the assigned trials and/or non-clinical related activities.

Specialist (Geospatial Data)

National Institutes of Health | Mar 2024 - Present

- Proposed the spatial data linkage proposal for REGARDS U.S. nationwide cohort data from 2000-2023.
- Used different APIs tools to pull and preprocess spatial data from U.S. NASA, U.S. EPA, and other institutes.
- Resolved the compatibility of 15+ spatial and temporal resolutions of geospatial datasets.
- Conducted spatial data linkage from 20+ geospatial data sources (rasters, shape files, NetCDFs, images) in R.
- Used parallel computing and memory saving methods to speed up the data linkage process by 1000%.

Postdoctoral Research Fellow

National Institutes of Health | Jun 2022 - Mar 2024

- Designed and led the development of multiple study protocols and Statistical Analysis Plans.
- Fine-tuned transformer (BERT) models for text classification with TensorFlow and TensorFlow Hub.
- Conducted medical imaging analysis using deep learning methods such as ConvNet and Densenet.
- Conducted statistical time-to-event multivariable analysis including Quantile-Based g-Computation, log binomial regression, and COX model in longitudinal nationwide cohort studies in R.
- Addressed the missing value issues by using Last observation carried forward, Multivariate Imputation by Chained Equations algorithm (MICE), and missForest R package.
- Conducted classic statistical analysis such as Propensity Score Matching and Principal Component.
- Built machine learning models including Catboost, XGboost, Multi-layer Perceptron, Explainable boost machine, and Random Forest using cloud and parallel computing in Python.
- Used Shapley value, Boruta selection methods, Bootstrapping AUC, and Selection above Random methods improved predictive performance by 10% percent using Python (SHAP and Boruta).
- Oversaw the code review and ensured planned statistical methods are applicable and optimal.
- Regularly using Git and Github to manage multiple projects and version control.
- Manage weekly reports, present at international conferences, and give invited talks in NIH.

Research Assistant

The State University of New York | Jul 2019 - Jun 2022

- Managed 20-year NY electronic hospitalization claim data (10 TB), NY utility & service data (2 TB), and nation wide meteorological data (2 TB), and COVID-19 data (5 TB) in SQL, SAS, R, and ArcGIS.
- Used large CT scan data and built Deep Learning models (ResNet50) via TensorFlow within Feature Pyramid Network in Python to predict COVID-19 and reached 90% accuracy.
- Developed an innovative two-stage downscaling model for exposure assessment and got NIH funding (\$3 million).
- Conducted statistical analysis including conditional logistic regression, Quantile-Based g-Computation, log binomial regression, COX model, GEE model, and Bayesian spatial-temporal models with INLA inference.
- Developed 3 web-pages via R shiny predicting cardiovascular, gastrointestinal, and infectious diseases.
- Conducted sample size calculation in R for 5 grants and managed budget for 3 large grants (\$4 million)
- Taught sample size and power calculation to graduate students as a guest lecturer every year for three years.
- Delivered significant scientific findings to policymakers and managed the funding reports to Health Authorities.

Projects

Cross-industry covariate adjustment R package and methodology working group

Lead Statistician

Aug 2024 - Present

- Contribute to an established working group to develop an internal strategy for covariate adjustment in clinical trials.
- Support engagement with a cross-industry working group tasked with creating a new open source R package for covariate adjustment, which may become an industry standard.
- Identify gaps in current methodologies and their practical implementations, with a focus on areas that require improvement or innovation
- Contribute to the strategy development concerning the covariate adjustment in clinical trials, aligning with regulatory guidance and industry best practices.
- Contribution to the development of the covadj R Package including developing or adapting R code, conducting reviews and testing, contributing to documentation, and creating example case studies.

Investigation into the composite endpoint (ESSDAI) in Sjögren's Disease, with respect to exposure (PK)

Lead statistician

Mar 2024 - Present

- Supported the submission program in the Sjögren's Disease indication.
- Provided overview of the key sub-scores of ESSDAI and the heterogeneity observed in the study population.
- Utilized Item Response Theory to assess which sub-scores are most "discriminated" and most correlated with the overall disease activity over time when considering treatment effect.
- Conducted comparison of PK-ESSDAI analysis results with results from IRT to provide rationale for dose justification in further submissions.

Education

PHD

School of Public Health, State University of New York

Aug 2019 - May 2022

MBBS

School of Public Health, Sun Yat-Sen University

Jul 2014 - May 2019
