

XINLEI DENG

10 Lombard Street, Birmingham, United Kingdom B12 0QD
Xinlei.Deng.Apha@gmail.com | +44 (7887) 296 811 | <https://xinleideng.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).

EXPERIENCES

Novartis

London, U.K.

Principal Biostatistician

03/2024-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 the dose justification for the Phase III filing.
- Contribute to working groups to develop an internal strategy for covariate adjustment in clinical trials (R package).
- Initiate, drive and implement **novel methods** and **innovative trial designs** in alignment with the Lead Statistician.
- 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 boards/ethics committees, external consultants and other external parties with oversight.
- Responsible for functional alignment and ensuring line function awareness of status/issues about the assigned trials.
- Explain statistical concepts in manner easily understood by non-statisticians and provide adequate statistical justifications for actions/decisions/statements.
- Establish and maintain sound working relationships and effective communication within the Clinical Trial Team and 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.

National Institutes of Health

Durham, U.S.

Specialist (Geospatial Data)

03/2024-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, shapefiles, NetCDFs, images**) in **R**.
- Used **parallel computing and memory saving** methods to **speed up** the data linkage process by **1000%**.

National Institutes of Health

Durham, U.S.

Postdoctoral Research Fellow

06/2022-03/2024

- 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, Principal Component, and Factor Analysis.
- Built machine learning models including Catboost, XGboost, Multi-layer Perceptron, Explainable boost machine, and Random Forest using cloud and parallel computing in **Python (Scikit-learn, xgboost, interpret, catboost)**.
- 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 and collaborating with multiple Clinical Research Organizations (Labcorp and DLH Corporation).
- Manage weekly reports, present at international conferences, and give invited talks in NIH.

The State University of New York
Research Assistant

Albany, U.S.
07/2019-06/2022

- 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** to predict COVID-19 and reached **90% accuracy**.
- Developed an **innovative combined model**, the two-stage downscaling model for refining exposure assessment and got R01 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 webpages via R shiny** predicting cardiovascular, gastrointestinal, and infectious diseases.
- Conducted **sample size calculation** in R for 5 research projects 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 the significant scientific findings and results to policymakers and managed the funding reports to NIH.

EDUCATION

Ph.D. (Environmental Health Sciences, top 1%) 05/2022
School of Public Health, State University of New York Albany, U.S.

- Relevant Coursework: *Big data management, Biostatistics, Principle of Statistical Inference, Epidemiology*

M.B.B.S. (Preventive Medicine & Biostatistics, top 2%) 05/2019
School of Public Health, Sun Yat-sen University Guangzhou, China

- Relevant Coursework: *Mathematics, Biostatistics, SAS programming, R programming, Epidemiology*

CERTIFICATES

- | | |
|--|------|
| • iOS Mobile Application Development, Meta | 2024 |
| • Version Control, Meta | 2024 |
| • Artificial Intelligence, Machine Learning, and Deep Learning, DeepLearning. AI | 2023 |
| • Database Design and Basic SQL in PostgreSQL, University of Michigan | 2023 |
| • Fine Tune BERT for Text Classification with TensorFlow, DeepLearning. AI | 2023 |

HONORS

- **First prize** in China **Mathematical Contest in Modeling** 2017.
- **Honorable Award** in American **Mathematical Contest in Modeling** 2018.
- **Invited speaker**: Center for Social & Demographic Analysis 40th Anniversary Colloquium; The International Conference on Environment and Human Health: Challenges and Opportunities in the 21st Century.
- Serving as **Reviewer** for **15 journals** and **2 conferences** and **Guest Editor** for **2 journals**.
- Presented in **8 international conferences**.

APPLICATIONS/PACKAGES

Developed R Package

- Deng X, Zhang W, Lin S. Package “APML” An Approach for Machine-Learning Modelling. DOI: [10.13140/RG.2.2.27638.42563](https://doi.org/10.13140/RG.2.2.27638.42563). Published online 2020.
- rSPARCS: <https://cran.r-project.org/web/packages/rSPARCS/index.html>

Developed Online Tools/Webpages (R-shiny):

- Predictive Model for Enteral Nutrition Intolerance in ICU Patients with Sepsis: xdeng3.shinyapps.io/NIPM/
- Predictive Model for Congenital Heart Diseases from a Large Birth Cohort: xdeng3.shinyapps.io/CHD_model/
- Predictive Model for COVID-19 vs Community-Acquired Pneumonia: xdeng3.shinyapps.io/COVID-19/

PUBLICATIONS

Available at Google Scholar: <https://scholar.google.com/citations?user=nLaS8WkAAAAJ&hl=en>