BUWEI HE

4th Year Ph.D Candidate

📞 803-908-3863 ✉ buwei@email.sc.edu in linkedin.com

SUMMARY

Ph.D. in Biostatistics with a solid theoretical foundation in statistics and extensive practical programming experience. Capable of efficiently translating analytical needs into executable data solutions and ensuring high-quality task delivery.

Proficient in R, Python, SAS, etc. Possesses rich interdisciplinary collaboration experience in biostatistical research, with a deep understanding of scientific workflows and paradigms, enabling rapid integration into projects and value contribution. Enjoys sharing knowledge with strong communication skills and teaching enthusiasm, effectively participating in team knowledge transfer and training.

EDUCATION

University of South Carolina
Aug 2022 – May 2026 (anticipated)

Doctor of Philosophy (Ph.D.) in Biostatistics, GPA: 4.0/4.0
Columbia, SC, U.S.

Core course: Mathematical statistics, High dimensional data analysis, Advanced survival analysis, Linear statistical models, Advanced computational statistics, Biostatistical machine learning

Rutgers University
Aug 2019 – June 2021

Master of Science (M.S.) in Biostatistics, GPA: 3.97/4.0
New Brunswick, NJ, U.S.

Core course: Statistical inference, Advanced regression methods, Biostatistics theory, Statistical learning, Categorical/Longitudinal data analysis, Clinical trial

Beijing Forestry University
Sep 2015 – June 2019

Bachelor of Science (B.S.) in Biotechnology, GPA: 3.44/4.0
Beijing, China

Core course: College Mathematics/Physics/Chemistry/Linear Algebra, Experimental Design and Statistical Analysis, Botany, Zoology, Cell Biology, Molecular Biology, Proteomics, Microbiology, Biochemistry, Genetics

EXPERIENCES

University of South Carolina Big Data Health Science Center
Aug 2022 – Present

Graduate Research Assistant advised by Dr. Jiajia Zhang
Columbia, SC, U.S.

- Cleaned and preprocessed statewide EHR data from SCDHEC using deduplication, imputation, and outlier handling to build high-quality datasets.
- Collaborated with interdisciplinary medical researchers to plan studies, develop methods, and formulate statistical strategies.
- Applied traditional (e.g., survival analysis, longitudinal modeling) and advanced methods (e.g., high-dimensional regression, deep learning) to analyze complex health data and generate insights.
- Co-authored multiple peer-reviewed publications and conference presentations.

Fudan University Human Phenome Institute
Aug 2021 – June 2022

Research Assistant advised by Dr. Sijia Wang
Shanghai, China

- Cleaned and performed QC on large-scale WGS and Bead-chip data from a multi-ethnic Chinese cohort. Developed an automated R-based alignment/correction algorithm to enhance data consistency.
- Independently conducted GWAS using genomic and phenotypic data, applying multiple statistical models to identify significant SNP-trait associations.
- Contributed data analysis to the national key project "Survey of Anthropometric Characteristics of Chinese Ethnic Groups".

Rutgers School of Dental Medicine
Jan 2021 – June 2021

Data Quality Analyst advised by Dr. Shou-en Lu
New Brunswick, NJ, U.S.

- Led end-to-end data management for the Opioid Analgesic Reduction Study (OARS), collecting raw data from 9+ clinical centers, performing ETL, and integrating into REDCap.
- Automated daily/weekly clinical trial monitoring using SAS (PROC SQL, MACRO, ODS), tracking recruitment, retention, medication adherence, and SAEs.
- Conducted statistical analyses (e.g., descriptive and trend analysis); translated results into actionable insights to support enrollment strategy and risk management.

Selected Projects

Dissertation project 2: Nonparametric Cure Rate Estimation for the Generalized Odds Rate Mixture Cure Model with Interval-Censored Data   May 2025 – Present
- Proposed a novel method for estimating the cure rate in interval-censored data using the generalized Turnbull estimator.
- Updated likelihood functions and the corresponding EM algorithm for parameter estimation.
- Conducted an extensive simulation study demonstrating improved accuracy and robustness of cure rate estimation in small-sample, interval-censored settings.
- Delivered an abstract presentation at the Eastern North American Region (ENAR) 2026 conference.

Dissertation project 1: Presmoothing-based Cure Rate Estimation for the Semiparametric Accelerate Failure Time Mixture Cure Model   Nov 2024 – Present
- Proposed an improved EM optimization structure using kernel smoothed estimator for cure rate estimation.
- Derived the likelihood function and corresponding asymptotic properties by incorporating the kernel-smoothed estimator and the accelerated failure time-type latency within the framework of our newly proposed optimization structure.
- Conducted a simulation study demonstrating the superior accuracy of parameter estimates achieved by the proposed method in handling small sample sizes and high-dimensional covariates, compared to traditional approaches.
- Illustrated the superiority of the proposed method on a real-world HIV dataset from South Carolina Department of Public Health.

Suppression rate prediction using medical prescriptions deciphered by large language model among people with HIV   Oct 2024 – Jan 2025
- Trained predictive models for HIV viral suppression rates by integrating structured data (lab results) and unstructured data (prescriptions, medical notes) using both traditional (regression-based approaches) and advanced (tree-based models, deep neural networks) methods.
- Queried medical prescriptions by SQL, applied BERT mdoel to extract informative features, and performed dimension reduction on the raw features by principal component analysis.
- Our findings demonstrated enhanced prediction accuracy by incorporating BERT-derived features into the deep neural network.

Impact of integrase strand transfer inhibitors on cardiovascular disease in people with HIV
                                    Mar 2024 – Dec 2024
- Developed an R algorithm to transform raw medical prescription data into standardized representations, reflecting longitudinal ART regimen patterns.
- Applied the time dependent Cox proportional-hazards model and the accelerated failure time model to assess the effect of integrase strand transfer inhibitors (INSTIs) on cardiovascular disease.
- Our findings highlighted the cardioprotective effects of INSTI-based ART regimens.
- Results were presented at the 2025 American Public Health Association (APHA) conference and published in Annals of Epidemiology.

A longitudinal analysis of the association between substance use disorders and sustained viral suppression among people with HIV in South Carolina   Jun 2023 – Mar 2024
- Led and wrote the data analysis protocol in collaboration with PIs.
- Proposed a new longitudinal sustained viral suppression definition based on EHR dynamic viral load test results.
- Constructed a generalized linear mixed model (GLMM) to evaluate the effects of substance use disorders on sustained viral suppression among people with HIV.
- Revealed a significant risk associated with substance use disorders on sustained HIV viral suppression. Published the results in peer-reviewed journal, AIDS.

Skills

Professional Skills: Survival analysis, Mixture cure model, High-dimensional data analysis, Generalized linear mixed model, Genome-wide association study, Machine learning, Data visualization

Programming Skills: R (tidyverse, data.table, ggplot2, lme4, survival, smcure, ncvsurv, survML), Python (keras, tensorflow), SAS, Linux (plink), LaTeX

Practical Skills: Strong knowledge of statistics, Effective oral and written communication skills, Independent research and teamwork ability

Certifications: SAS Certified Specialist: Base Programming Using SAS 9.4

Honors & Awards

SmartState Center for Healthcare Quality 2024 – 2025 Junior Scholar   **Sep 2024**
University of South Carolina, Columbia, SC, U.S.

Delta Omega Honorary Society in Public Health   **May 2021**
Rutgers School of Public Health, New Brunswick, NJ, U.S.

Publications

[1]   **Buwei He**, Bankole Olatosi, Jiajia Zhang, Sharon Weissman, Xiaoming Li, and Xueying Yang. "Impact of Integrase Strand Transfer Inhibitors on Cardiovascular Disease in People With HIV". en. In: *Annals of Epidemiology* (Nov. 2025), S1047279725003370. ISSN: 10472797. DOI: 10.1016/j.annepidem.2025.11.006.

[2]   Xueying Yang, Ruilie Cai, **Buwei He**, Sharon Weissman, Bankole Olatosi, Xiaoming Li, and Jiajia Zhang. "Disruptions of healthcare visits and viral suppression for people living with HIV during the COVID-19 pandemic in the US". In: *BMC Infectious Diseases* 25.1 (Oct. 2025), p. 1295. ISSN: 1471-2334. DOI: 10.1186/s12879-025-11682-z.

[3]   Atena Pasha, Shan Qiao, Jiajia Zhang, Ruilie Cai, **Buwei He**, Xueying Yang, Chen Liang, Sharon Weissman, and Xiaoming Li. "Changes in mental health care utilisation before and during the COVID-19 pandemic among people living with HIV in the USA: a retrospective cohort study using the All of Us dataset". In: *BMJ Public Health* 3.2 (2025), e002173. ISSN: 2753-4294. DOI: 10.1136/bmjph-2024-002173.

[4]   **Buwei He**, Shujie Chen, Xueying Yang, Bankole Olatosi, Sharon Weissman, Xiaoming Li, and Jiajia Zhang. "Association Between Substance Use Disorders and Sustained Viral Suppression: A Longitudinal Analysis Among People with HIV in South Carolina". In: *AIDS* (Nov. 2024). ISSN: 0269-9370, 1473-5571. DOI: 10.1097/QAD.0000000000004077.

[5]   Fanghui Shi, Ruilie Cai, **Buwei He**, Xiaoming Li, Xueying Yang, Sharon Weissman, Bankole Olatosi, and Jiajia Zhang. "Sexual Orientation, Gender Identity and Virologic Failure Among People with HIV: A Cohort Study in All of US Research Program". In: *BMC Public Health* 24.1 (2024), p. 2091. DOI: 10.1186/s12889-024-19559-7.

[6]   Xiaotong Yang, Bing Geng, Changxiong Zhu, Hongna Li, **Buwei He**, and Hui Guo. "Fermentation performance optimization in an ectopic fermentation system". In: *Bioresource Technology* 260 (2018), pp. 329–337. ISSN: 0960-8524. DOI: https://doi.org/10.1016/j.biortech.2018.03.101.

[7]   Jihang Jiang, Xiawei Peng, Zhenxin Yan, **Buwei He**, Changxiong Zhu, Hui Guo, and Bing Geng. "Isolation and Identification of Potassium-Solubilizing Bacteria from Rhizosphere Soil of Apple Tree". In: *Chinese Journal of Agrometeorology* 38.11 (2017), pp. 738–748. DOI: 10.3969/j.issn.1000-6362.2017.11.006.

[8]   "Psychological wellbeing and correlates of People Living with HIV During the COVID-19 Pandemic: Based on All of Us Data". Prepared for submission.

Conference Contributions

[1]   Sexual orientation, gender identity and virologic failure among people with HIV: a cohort study in all of US research program [poster]
National LGBTQ Health Conference, Atlanta, GA, U.S., 2024

[2]   Association between substance use disorder and sustained viral suppression among people with HIV [poster]
Conference on Retroviruses and Opportunistic Infections (CROI), Denver, CO, U.S., 2024

[3]   Utilizing All of Us data to examine the impact of COVID-19 pandemic on mental health among people living with HIV [poster]
12th International AIDS Society (IAS) Conference on HIV Science, Brisbane, Australia, 2023

[4]   Impact of integrase strand transfer inhibitors on cardiovascular disease in people with HIV [poster]
American Public Health Association (APHA), Washington, D.C., U.S., 2025