CHANHWA LEE

chanhwa@email.unc.edu | chanhwa-lee | chanhwalee

Expertise

Causal inference (interference, network, observational studies), **Statistical analysis** (high dimensional, survival, nonparametric, precision calculation), **Machine learning** (ensemble learning, TMLE, deep learning), **Statistical genetics** (GWAS, Transcriptomics, Proteomics), **Theoretical statistics** (Semiparametric theory and empirical process).

Education

Ph.D. Biostatistics, University of North Carolina at Chapel Hill *Chapel Hill, NC*

Advisor: Dr. Michael G. Hudgens & Dr. Donglin Zeng *Expected May 2025*

- Special Commendation Award for 1st prize in Doctoral Comprehensive Exam (Theory)
- Korea Foundation for Advanced Studies (KFAS) Doctoral Study Abroad Program Fellowship ($65K)

B.Sc. Statistics and Mathematics, Seoul National University, summa cum laude *Seoul, Korea*

GPA: 4.06 / 4.3 *Aug 2020*

- The Presidential Science Scholarship (Top national scholarship for outstanding STEM students, $44K)

Publications

Published

Li, L., **Lee, C.**, Cruz, D. F., Krovi, S. A., Hudgens, M. G., Cottrell, M. L., & Johnson, L. M. (2022). Reservoir-Style Polymeric Drug Delivery Systems: Empirical and Predictive Models for Implant Design. *Pharmaceuticals*, 15(10), 1226.

In Review

Lee, C., Zeng, D., & Hudgens, M. G. (2023). Efficient Nonparametric Estimation of Stochastic Policy Effects with Clustered Interference. *Journal of the American Statistical Association*.

In Prep

Chen, B., **Lee, C.**, Tapia, A., Reiner, A., Tang, H., Kooperberg, C., Li, Y., & Raffield, L. (2023). Proteome-Wide Association Study Using Cis and Trans SNPS and Applied to Blood Cell and Lipid-Related Traits in the Women's Health Initiative Study.

Alabanza, P., Fischer, W. A., Eron, J. J., Lakshmanane, P., Loftis, A. J., Baric, R. S., Krajewski, T. J., **Lee, C.**, Mollan, K. R., Moreria, F. R., Sheahan, T. P., ... (2023). The Antiviral Mechanism of Action of Molnupiravir in Humans with COVID-19.

Presentations

Contributed Presentations

Aug 2023. **Efficient Nonparametric Estimation of Stochastic Policy Effects with Clustered Interference**, Contributed Papers: Joint Statistical Meeting (JSM 2023), Toronto, Canada.

Mar 2023. **Efficient Nonparametric Estimator of Incremental Propensity Score Effect with Clustered Interference**, Oral-Contributed presentation: Eastern North American Region International Biometric Society (ENAR 2023) Spring Meeting, Nashville, TN.

Honors and Awards

Silver Prize, 38th University Student Contest of Mathematics | Korean Mathematical Society *2019*

Dean's List | College of Natural Sciences, Seoul National University *2014, 2015, 2019*

Bronze Prize, 20th Humantech Paper Award | Samsung Electronics *2014*

2nd Prize, 26th Final Korean Mathematical Olympiad | Korean Mathematical Society *2013*

Silver Prize, 26th Korean Mathematical Olympiad | Korean Mathematical Society *2012*

Research

Causal Inference under Interference using Efficient Nonparametric Estimation

Chapel Hill, NC

Causal Inference Research Lab (CIRL), UNC Chapel Hill *Jan 2022 - Present*

- Developed efficient nonparametric estimation of **causal network effects** under interference based on semiparametric and empirical process theory.
- Used **ensemble of nonparametric and ML models** (spline regression, GAM, boosting, Random Forest, neural net) via SuperLearnear in **R**.
- Paper: *Efficient Nonparametric Estimation of Stochastic Policy Effects with Clustered Interference*.

Drug Release Prediction for Reservoir-Style Polymeric Drug Delivery Systems

Raleigh, NC

Research Triangle Institute International *Mar 2022 - Sep 2022*

- Designed and built drug release prediction model and visualized results in **R**, enabling rational drug implant design without extensive in vitro testing.
- Publication: *Reservoir-Style Polymeric Drug Delivery Systems: Empirical and Predictive Models for Implant Design*.

Fake News Detection using Machine Learning Methods

Chapel Hill, NC

Machine Learning Course Project, UNC Chapel Hill *Aug 2021 - Dec 2021*

- Preprocessed fake news data based on standard **NLP preprocessing procedure** to generate Bag of Words, TF-IDF, and Bigram using **Pandas**.
- Trained ML (SVM, Random Forest, Logistic Regression) and DL (1D CNN, BERT, LSTM, Domain Adaptation) models to build fake news detection model using **scikit-learn**, **PyTorch**, and **Tensorflow**, achieved **91.4% test accuracy**.

Women's Health Initiative Proteome-Wide Association Study

Chapel Hill, NC

Yun Li Statistical Genetics Group, UNC Chapel Hill *Sep 2020 - Aug 2021*

- Analyzed 552 protein levels of 1,002 individuals from Women's Health Initiative data to **identify protein quantitative trait loci (pQTL)** using **EPACTS**.
- Built protein level **prediction model using CV Elastic Net** and investigated cardiovascular diseases related proteins based on predicted levels in **R**.

Growing Student Knowledge Distillation

Seoul, Korea

Deep Learning Course Project, Seoul National University *Sep 2019 - Dec 2019*

- Proposed novel knowledge distillation structure comprised of sequence of CNNs with increasing number of layers, **transferring knowledge from smaller to bigger networks consecutively** using **PyTorch** which resembles a student's cumulative learning process.
- Achieved **89.9% test accuracy** on CIFAR-10 dataset, improvement of 0.2% test accuracy compared to baseline ResNet26.

Professional Experiences

Research Statistics Intern | GlaxoSmithKline (GSK), Philadelphia, US *May 2023 - Aug 2023*

- Built pipeline for high dimensional multi-omics ensemble ML prediction models (BART integrated learner) to identify biomarkers related to cancer cell death drug sensitivity.
- Identified over 100 genomic features (RNA, proteomics, mutation), characterizing genomic profiles of patients who may benefit most from the investigated drug compounds.

Graduate Research Assistant | Center for AIDS Research (CFAR), UNC Chapel Hill *2022 - Present*

- Wrote statistical analysis plans for research grants and provided data analysis.
- Reviewed a paper for Journal of the International AIDS Society.

Teaching Assistant | Dept. of Biostatistics, UNC Chapel Hill *2022*

- BIOS650 Basic Elements of Probability and Statistical Inference

Graduate Research Assistant | Yun Li Statistical Genetics Group, UNC Chapel Hill *2020 - 2021*

Tutoring | Dept. of Statistics, Seoul National University *2019*

- Tutoring undergraduate students on Mathematical Statistics I & II

Technical Skills

Programming Python (Numpy, scikit-learn, PyTorch, TensorFlow), R (dpylr, SuperLearner, ggplot2), Linux, C++, SAS, SQL

Miscellaneous Git, Fotran, PLINK, BCFtools, LaTeX, Markdown, Google Colaboratory, Jupyter Notebook