# CHANHWA LEE

✉ chanhwalee27@gmail.com  |  🏠 chanhwalee  |  💻 chanhwa-lee  |  in chanhwalee

## Expertise

**Causal inference** (interference, network), **Statistical analysis** (high dimensional, survival, nonparametric), **Coding** (Develop statistical methods in **R** and **Python**), **Machine learning** (ensemble learning, nlp, deep learning), **Theoretical statistics** (Semiparametric efficiency, Gaussian 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)

## Research Experience

### Causal Inference under Interference using Efficient Nonparametric Estimation
*Chapel Hill, NC*

Causal Inference Research Lab, UNC Chapel Hill
*Jan 2022 - Present*
- Developed efficient nonparametric estimation of **causal network effects** under interference based on semiparametric efficiency theory.
- Used **ensemble of nonparametric and ML models** (spline regression, GAM, boosting, Random Forest, neural net) via SuperLearnear in **R**.

### Conditional Average Treatment Effect estimation using Multi Group Gaussian Process
*Chapel Hill, NC*

Didong Li Lab, UNC Chapel Hill
*Aug 2023 - Present*
- Proposed to use Multi Group Gaussian Process to estimate **conditional average treatment effect** to account for heterogeneity among individuals.
- Modeled **interpretable correlation structure between potential outcomes**, allowing maximum likelihood or Bayesian estimation of CATE.

### Fake News Detection using Machine Learning Methods
*Chapel Hill, NC*

COMP 755. 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**.

### 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.

## Publications

**Lee, C.**, Zeng, D., & Hudgens, M. G. (2024). Nonparametric Causal Survival Analysis under Clustered Interference. *Journal of the American Statistical Association*. Under review.

Shook-Sa, B., Zivich, P., **Lee, C.**, ..., & Cole, S. (2024). Double Robust Variance Estimation. *Biometrics*. Under 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*. Under minor revision.

Kilpatrick, K., **Lee, C.**, & Hudgens, M. G. (2023). G-Formula for Observational Studies with Partial Interference, with Application to Bed Net Use on Malaria. *Statistics in Medicine*. Under minor revision.

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 Variants and Applied to Blood Cell and Lipid-Related Traits in the Women's Health Initiative Study. *Genetic Epidemiology*. Under review.

Sheahan, T. P., Laura J. Stevens, L. J., Lakshmanane, P., Krajewski, T. J., **Lee, C.**, ..., & Fischer, W. A. (2023). The Antiviral Mechanism of Action of Molnupiravir in Humans with COVID-19. *Nature Communications*. Under review.

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.

## Technical Skills

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

**Miscellaneous**   Git, PLINK, BCFtools, EPACTS, LaTeX, SLURM, R Shiny, Markdown, Google Colaboratory, Jupyter Notebook, Microsoft Office

## Honors and Awards

| | |
|---|---|
| 2024 | **Early Career Award Finalist, Duke Industry Statistics Symposium 2024**, Dept. Statistics at Duke University |
| 2024 | **Travel Awards, 2024 Winter Workshop on Causal Inference**, Dept. Statistics at University of Florida |
| 2023 | **Travel Awards, Extending Inferences to a New Target Population Workshop**, ICERM at Brown University |
| 2019 | **Silver Prize, 38th University Student Contest of Mathematics**, Korean Mathematical Society |
| 2014,5,9 | **Dean's List**, College of Natural Sciences, Seoul National University |
| 2014 | **Bronze Prize, 20th Humantech Paper Award**, Samsung Electronics |
| 2013 | **2nd Prize, 26th Final Korean Mathematical Olympiad**, Korean Mathematical Society |

## Professional Experience

### Research Statistics Intern | GlaxoSmithKline (GSK), Philadelphia, US
*May 2023 - Aug 2023*
- Built pipeline for **high dimensional multi-omics ensemble ML prediction models** to identify biomarkers related to cancer cell drug sensitivity.

### Graduate Research Assistant | Center for AIDS Research, 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.