Abstracts

Mixture-based Nonparametric Estimation of Spatial Covariance Functions with Applications to HIV Key Population Size Estimation across Sub-Saharan Africa

Manushi Siriwardana, Penn State

Consistent data on the sizes of key populations, such as female sex workers (FSWs), are often scarce, particularly at the sub-national level. Accurate size estimates are critical to effectively allocate resources and achieve HIV targets. Since FSW population sizes may be spatially correlated across areas, models that account for spatial dependence can improve estimation. An important component of such models is the covariance function, which characterizes the spatial dependence structure of the underlying process. In this work, we study spatial covariance functions to estimate FSW population sizes in Sub-Saharan Africa (SSA). Parametric covariance estimation can suffer from model mis-specification, potentially leading to inefficient or biased predictions. We therefore develop a robust non-parametric approach for estimating the covariance function of a stationary isotropic process in R^d. We focus on a class of covariance functions that are valid in all dimensions, which includes popular kernels such as the exponential and Mat'ern kernels. Leveraging the fact that such covariance functions can be represented as infinite mixtures of scaled Gaussian kernels, we propose two estimation methods: weighted least squares and nonparametric maximum likelihood estimation to estimate the mixing measure of scaled Gaussian kernels. We also develop computationally efficient methods to solve these optimization problems using non-negative least squares and second-order descent updates.

Influenza hospitalization tracking using internet search data and ILI surveillance

Hanqi Shi, Swarthmore College

For influenza control and public health decision-making, timely estimates of influenza hospitalization rates are essential. Unlike traditional influenza surveillance, which is reported with a delay of one to three weeks, Internet search data provide timely signals of ongoing influenza activity. Here we present a statistical framework for real-time tracking of influenza hospitalization rates in the United States. Our approach builds on the ARGO and ARGOX frameworks by combining Google Trends data with lagged influenza-like illness (ILI) surveillance through dynamically trained regularized regression models. Using hospitalization data from the Centers for Disease Control and Prevention's National Healthcare Safety Network (NHSN), our method outperforms models that rely on Internet search data alone. The proposed framework demonstrates that integrating multiple surveillance signals can improve real-time influenza hospitalization tracking and provides a foundation for future joint modeling of ILI activity and hospitalization rates.

The Impact of Recent State Staffing Minimum Mandates on Nursing Home Residents

Angira Mondal, Perelman School of Medicine, University of Pennsylvania

To ensure adequate staffing, some U.S. states have instituted minimum staffing mandates for nursing homes, defined by the Hours Per Resident Day (HPRD). This study explores the impact of these state-level staffing mandates on staffing HPRD and nursing home resident outcomes such as proportion of residents with stage 2 or higher-pressure ulcers, proportion of residents receiving inappropriate antipsychotic medication and the mortality rate. Heterogeneity analysis was completed by dementia status. Using CMS administrative data, we conduct employ staggered difference-in-difference methodology to study the effect of new or updated mandates implemented by 7 states between 2018 to 2023 (CA, AR, MA, CT, NY, RI and PA). We observe that the staffing minimum leads to an increase in the total staffing HPRD, but the average masks considerable state variation, ranging from no effect in AR to significant increased effect in CA. The resident outcomes also generally improve, especially in states where we observe greater staffing increase post the implementation of the regulations. Our findings show that staffing adequacy contributes to poor resident outcomes in US nursing homes.

From Raw Wearable Data to Scientific Inference: A Reproducible Pipeline for Processing Commercial Step Count Data in Aging Cohorts

Nannan Bo, Department of Biostatistics and Bioinformatics, Duke University School of Medicine; Duke Aging Center, Duke University School of Medicine; incoming PhD student, Quantitative Biomedical Sciences Program, Dartmouth College

Commercial wearable devices offer an unprecedented opportunity to objectively measure free-living physical activity at scale. However, raw data generated by consumer devices are rarely analysis-ready. These data are often recorded in epochs (fixed time intervals) and may contain irregular timestamps, missing intervals, and limited wear-time information, making processing decisions critical to downstream scientific inference.

We have developed a transparent, reproducible standard operating procedure (SOP) to transform Garmin epoch-level step-count data into participant-level analytic variables in the STRRIDE-PD Reunion Study, a long-term follow-up of older adults originally enrolled in a supervised exercise intervention trial. The pipeline standardizes timestamps, reconstructs epoch grids, infers wear time, identifies valid observation days, and generates participant-level activity summaries with quality control and sensitivity analyses.

Among 67 participants, the pipeline produced robust estimates of daily step-count metrics across alternative processing criteria. Models using wearable-derived step counts identified associations with cardiometabolic and fitness outcomes that were not detected using self-reported physical activity, demonstrating how measurement approach can influence scientific conclusions.

This work highlights that processing commercial wearable data is not merely data cleaning but a measurement problem requiring transparent, reproducible analytic decisions.