

Frequent-voting independence screening for data of different types or different dimensions*Haeun Moon, Seoul National University*

Modern datasets often include different types of variables with complex features, making variable selection particularly challenging. For example, a measure of dependence with the response variable may not be directly comparable among predictor variables of different types and different dimensions. To address this challenge, this work proposes a frequent-voting based independent screening method for variable selection, which avoids a direct comparison of the dependence measure among different variables. Asymptotic analyses show that the proposed method selects all of the active variables with probability converging to one. We also demonstrate its great finite sample performance through numerical experiments and the application to an ADHD study.

A Multi-Scale Integration of Mechanistic and Statistical Models for RSV Prediction with ILI-Informed Calibration*Saebom Jeon, Sungshin Women's University*

Mechanistic epidemic models provide a principled framework for understanding infectious disease dynamics but often fail to capture interannual variability and short-term fluctuations. In this study, we develop a hybrid statistical-mechanistic framework to enhance a SLIRS-based model of respiratory syncytial virus (RSV) transmission.

First, we propose a season-level calibration method that incorporates influenza-like illness (ILI) information to correct systematic bias in peak amplitude and timing. Second, we model the residuals of the mechanistic model using autoregressive time-series methods, capturing temporal dependence not explained by the underlying epidemiological structure.

Using weekly RSV hospitalization data in South Korea (2015–2020), we show that ILI-informed calibration improves peak-level prediction, while residual-based autoregressive correction significantly reduces both overall and peak-period errors. Interestingly, ILI does not improve short-term predictions when included as an exogenous variable in time-series models, suggesting a separation between seasonal and short-term dynamics.

Our findings emphasize the importance of combining statistical learning techniques with mechanistic models and provide a general framework for improving epidemic prediction.

fSKAT: A Set-Based Association Test for Functional Responses in Genetic Studies*Eunjee Lee, Chungnam National University*

Set-based association testing is widely used in genetic studies, but most existing methods are developed for scalar or multivariate outcomes. In biomedical and neuroimaging studies, however, phenotypes are increasingly observed as functional data, such as signals, imaging profiles, and connectivity curves

indexed by time, space, or anatomical distance. We propose functional SKAT (fSKAT), a score-type variance-component test for assessing the association between a set of genetic variants and a functional response. fSKAT is designed to capture genetic effects that are weak at individual points but accumulate across the functional domain. The method is formulated within a function-on-scalar regression framework and constructs a kernel-based statistic from residual functions estimated under the null model, allowing inference through a mixture-of-chi-square null approximation. Simulation studies show well-calibrated type I error rates and improved power over scalar-summary-based testing approaches. We apply fSKAT to resting-state fMRI and genetic data from the Alzheimer's Disease Neuroimaging Initiative (ADNI), using seed-based functional connectivity curves as a phenotype. The ADNI analysis examines genetic associations with early changes in brain functional connectivity before overt clinical diagnosis and illustrates the utility of fSKAT for complex functional phenotypes.

Sparse and heterogeneous meta-analysis with semiparametric models*Sunyoung Shin, Pohang University of Science and Technology (POSTECH)*

Meta-analysis is a powerful tool to synthesize findings and statistical evidence from independent studies. We consider fixed-effects semiparametric models for meta-analysis, in which different studies may have different values of finite-dimensional parameters of interest such that effect sizes are heterogeneous with the common sparse structure among studies. Nuisance parameter values may also be different. We develop a meta-analysis approach for estimating the finite-dimensional parameters and selecting features solely based on study-specific estimators via regularization employed among studies. The resulting estimators, named sparse and heterogeneous meta-analysis (SHMA) estimators, have estimation and selection consistency without access to individual-level data. Semiparametric efficiency bound is attained when the models are refitted with individual-level data after variable selection. The method is easy to implement with the alternating direction method of multipliers algorithm. This general approach has broad applications including meta-analysis of right-censored data and interval-censored data with Cox proportional hazards models. Simulation studies illustrate the superior performance of our meta-analysis approach over alternative approaches using individual-level data, such as analyses of individual studies and integrative analyses of multiple studies. Analysis of two acute myeloid leukemia survival datasets demonstrates the practical utility of our proposed approach.

