Indiana University-Purdue University Indianapolis

Department of Mathematical Sciences

STATISTICS SEMINAR

12:15pm—1:15pm, Tuesday, March 05, 2024
Zoom Meeting: Meeting ID: 845 0989 4694

Speaker: **Sean McCabe**
Department of Biostatistics and Health Data Science,
Indiana University School of Medicine

Title: **Integrating functional annotations to empower rare variant set by environmental interactions in large biobank datasets**

Abstract:

Analyses of gene environment interactions can provide insights into how the effect of environmental risk factors, such as smoking or the concentration of air pollutants, on health related outcomes can vary based on the mutational patterns in an individual. While analyses of common variants in this setting have been extensively studied, rare-variants have been shown to explain a larger proportion of heritability and are often more deleterious. The development of large biobank datasets can provide hundreds of thousands of samples which can help in the detection of rare-variants and improve statistical power, however these analyses are often still underpowered due to the rarity of the variants. We propose STAARGEI as a way to incorporate external annotation scores to improve the power of rare-variant gene set by environment interactions. STAARGEI incorporates annotation specific weights to construct an omnibus test which combines a SKAT, Burden, and ACAT framework. STAARGEI also uses functional annotations to define the rare-variant sets to be tested and can efficiently handle population structure and sample relatedness through a generalized linear mixed model (GLMM) framework. We applied STAARGEI to a variety of simulation settings and in an analysis of neutrophil count in the UK Biobank dataset using smoking pack years as the interaction variable.

Bio:

Dr. Sean McCabe is an Assistant Professor in the Department of Biostatistics and Health Data Science at Indiana University School of Medicine. Prior to joining the department, Sean was a Postdocoral Research Fellow in the Department of Biostatistics at the Harvard TH

Chan School of Public Health under the guidance of Xihong Lin and received his PhD in Biostatistics from the University of North Carolina at Chapel Hill. Sean's reseach is on developing statistical methods for high-dimensional genomic data and specializes in integrative analyses. His collaborative interests include bioinformatics, cancer research, and clinical trials.