

Special Invited Session Speaker

Nilanjan Chatterjee, received his Bachelor's and Master's degree from the Indian Statistical Institute, Calcutta and a Ph.D. in statistics from the University of Washington, Seattle in 1999. He served as Chief of the Biostatistics Branch of the Division of Cancer Epidemiology and Genetics (DCEG), National Cancer Institute (NCI) for almost eight years. Dr. Chatterjee now serves as a Bloomberg Distinguished Professor at the Johns Hopkins University with joint appointments at the Bloomberg School of Public Health (Biostatistics) and the School of Medicine (Oncology). He remains a Special Volunteer with DCEG.

Dr. Chatterjee's research focuses on a diverse set of quantitative issues that arise in design, analysis, interpretation and public health translation of modern molecular and genetic epidemiologic studies.

Title: Predictive Model Building through Integration of Information across Disparate Data Sources and Summary-statistics

Abstract: Model building based on classical statistical methods, as well as modern machine learning techniques, typically requires availability of a single adequately large dataset, or multiple harmonized datasets across a group of similar studies. In the future, however, development of complex models incorporating a variety of factors from different domains will require integration of information from disparate data sources, which, individually may have information only on subsets of the explanatory variables of interest. Moreover, information from some studies may only be available through pre-computed summary-statistics, generated under certain forms of "reduced" models. In this talk, I will describe some of our recent efforts towards developing statistical methods for model building through data integration under a semiparametric generalized meta-analysis framework. I will illustrate the unique opportunity data integration methods provide through an application involving the development of a COVID-19 mortality risk calculator through integration of information across diverse datasets.