I am generally interested in the development of statistical methods with applications to scientific problems in the areas of medicine, epidemiology, public health and quality management. My current research falls into the following themes:

METHODS FOR COMPLEX SURVIVAL DATA

With rapid developments of medical technologies and the advent of new study designs, the complexity of survival data dramatically increases. The complexities given by admixed populations with susceptible and non-susceptible subjects, various study designs, auxiliary covariate information and complex censoring schemes open up many questions and pose challenges for statisticians, while standard methods would not be able to make proper use of such complex data structure and information. Our main recent work in this area focuses on cure rate models, including mixture and non-mixture cure models, with survival data under interval censoring scheme. We have proposed the estimation based on the generalized linear mixed model (GLMM) method and using the Turnbull’s algorithm to account for interval-censoring for mixture cure models. For semiparametric non-mixture cure models, we have established efficient estimation using spline-based method. We also wish to investigate a general class of semiparametric transformation cure models with interval-censored multivariate survival data and establish efficient inference in the presence of auxiliary covariates.

ROBUST ESTIMATION IN FINITE MIXTURE OF GLMMs

The heterogeneity in clustered count data arising from latent subpopulations is a challenging problem in statistical modeling. The finite mixture of GLMMs is proposed as an effective strategy for analyzing clustered count data in this situation. Usually, parameters in such a model are estimated through the residual maximum likelihood estimation which is vulnerable to outliers. To detect if a few outliers or contaminated clusters/subjects present, we have developed influence assessments of clusters or subjects on the model performance using statistical diagnostics. When possible outlying observations/clusters are identified, a robust estimation method is desirable consequently. We then established the minimum Hellinger distance (MHD) estimation, enjoying both robustness and the second order efficiency. Recently, we have proposed a non-parametric maximum likelihood approach embedded with the MHD estimation for a more general model framework with random effects whose joint distribution does not need a parametric form. We are also working on developing the MHD estimation for clustered continuous data and survival data.