

Longitudinal surrogate marker estimation

Assume that we are interested in examining data from a randomized study that examines the effect of a treatment or intervention in terms of some primary outcome of interest, Y . Patients are randomly assigned at baseline to treatment or control and we assume that Y is fully observed. Let $X(\cdot)$ be a longitudinal surrogate marker measured repeatedly over time from baseline until the end of the study, time t ; we may only observe $X(\cdot)$ with error at a few, irregularly spaced times $X_{ij} = X(t_{ij}) + \epsilon_{ij}$. We are interested in estimating the proportion of treatment effect on the primary outcome that is explained by the treatment effect on the longitudinal surrogate marker. Previous work has tended to focus on estimating this quantity in settings where the surrogate marker is measured only at a single time point and on using restrictive parametric models to obtain the estimate. We propose an estimation method to obtain this quantity when the surrogate marker measurements are repeated (and potentially irregular) within a flexible modeling framework.