Imputation of missing longitudinal data using the broken stick model

Stef van Buuren

University of Utrecht

TNO, Netherlands Organisation for Applied Scientific Research

There are two general strategies for the multiple imputation of missing longitudinal data. If the data are collected at similar times, we can organise the data as a broad matrix and impute under the assumption of independence of rows. This approach is simple and flexible but inefficient if there are many time points or if time points differ between observations. The second strategy involves organising the data with time in the rows and imputing it by a mixed model using time as a covariate. This approach has become possible thanks to recent advances in methodology and software but it requires deep methodological and technical expertise. In this presentation, I explore a third strategy, the broken stick model, which combines the simplicity of the broad approach with the technical advantages of the mixed model. The broken stick model transforms data collected at irregular time points to a raster of regular time points. This raster is the same for all observations, so effectively, we replace the irregular data with a simpler set of repeated measures. I will illustrate the essential workings of the method and explain how multiple imputation accommodates for the inherent loss of information induced by the transform.