Roee Gutman (Brown University)

Full Bayesian Procedure for File Linking to Analyze End-of-Life Medical Costs

End-of-life medical expenses are a significant proportion of all health care expenditures. These costs were studied using costs of services from Medicare claims and cause of death (COD) from death certificates. In the absence of a unique identifier linking the two datasets, common variables identified unique matches for only 33% of deaths. The remaining cases formed cells with multiple cases (32% in cells with an equal number of cases from each file and 35% from cells with an unequal number). We sampled from the joint posterior distribution of model parameters and the permutations that link cases from the two files within each cell. The linking models included the regression of location of death on COD and other parameters, and the regression of cost measures with a monotone missing data pattern on COD and other demographic characteristics. Permutations were sampled by enumerating the exact distribution for small cells and by the Metropolis algorithm for large cells. Sparse matrix data structures enabled efficient calculations despite the large dataset ($\approx$1.7 million cases). The procedure generates m datasets in which the matches between the two files are imputed. The datasets can be analyzed independently and combined using Rubin's Multiple Imputation rule. Our approach can be applied in other file linking applications.