Abstract

Missing values, common in epidemiologic studies, are a major issue in obtaining valid estimates. Simulation studies have suggested that multiple imputation is an attractive method for imputing missing values, but it is relatively complex and requires specialized software. For each of 28 studies in the Asia Pacific Cohort Studies Collaboration, a comparison of eight imputation procedures (unconditional and conditional mean, multiple hot deck, expectation maximization, and four different approaches to multiple imputation) and the naive, complete participant analysis are presented in this paper. Criteria used for comparison were the mean and standard deviation of total cholesterol and the estimated coronary mortality hazard ratio for a one-unit increase in cholesterol. Further sensitivity analyses allowed for systematic over- or underestimation of cholesterol. For 22 studies for which less than 10% of the values for cholesterol were missing, and for the pooled Asia Pacific Cohort Studies Collaboration, all methods gave similar results. For studies with roughly 10-60% missing values, clear differences existed between the methods, in which case past research suggests that multiple imputation is the method of choice. For two studies with over 60% missing values, no imputation method seemed to be satisfactory.