METHODS FOR HANDLING COVARIATE DATA MISSING BY DESIGN IN PROGNOSTIC STUDIES

R. Sutradhar[1], S.B. Bull[2], G.Y. Yi[3], D. Pinnaduwage[4]

[1, 2, 4]*Samuel Lunenfeld Research Institute, University of Toronto, Toronto, Canada*
[3]*University of Waterloo, Waterloo, Canada*

Email: *rsutradh@mshri.on.ca*

Prognostic factors play an important role in the clinical decision-making process. For cases in which patients are collected in a two-phase design, complete covariate information is provided for individuals in the first phase, whereas information on certain prognostic factors may be unavailable for individuals in the second phase. The motivating data arises in an evaluation of molecular genetic prognostic factors in node-negative breast cancer. Individuals in the first phase of the study are measured for both p27 level and Her2 amplification. Patients in the second phase of the study, however, are measured only for their p27 level. Interest lies in determining whether p27 is a significant predictor of node-negative breast cancer outcome, over and above the known prognostic importance of Her2 amplification and other prognostic factors. We investigate a likelihood-based approach for inference in this setting. The likelihood is divided into two components, one comprised of individuals from the first phase, and the other, subjects from the second phase in which a conditional expectation must be evaluated. We conducted simulation studies to compare the performance of various methods under scenarios in which the degree of correlation between the primary and secondary covariates is varied. We found that the likelihood-based approach provides valid inferences whereas ad hoc alternatives can be biased and/or less efficient.