

Supremum weighted log-rank test and sample size for comparing two-stage adaptive treatment strategies

Abdus S. Wahed and Wentao Feng

Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania 15261, U.S.A.

In two-stage adaptive treatment strategies, patients receive one of the induction treatments followed by a maintenance therapy given that the patient responded to the induction treatment they received. The goal is to compare different treatment strategies resulting from the combination of induction and maintenance treatment. Randomized clinical trials where patients are randomized upfront to different adaptive treatment strategies are limited by the resources it requires in terms of patient recruitment and the statistical power; Particularly when the number of treatment options at various stages are large (> 2). However, sequential multiple assignment randomization trials (SMART, Murphy et al., *Drug Alcohol Depend.* 2007 May; 88(Suppl 2): S1S3.) have been innovated as an alternative to the regular upfront randomized trials. To test for a difference in the effect of different induction and maintenance treatment combinations using the data from a SMART trial, a modified supremum weighted log-rank test is proposed. The test is applied to a data set from a two-stage randomized trial and the results are compared to those obtained using a standard weighted log rank test. A sample size formula is proposed based on the limiting distribution of the supremum weighted log-rank statistic. The sample size formula reduces to the Eng and Kosorok's (Eng and Kosorok, *Biometrics* vol. 61, pp 86-91, 2005) sample-size formula for two-sample supremum log-rank test when there is no second randomization. A connection is also established between the proposed sample size formula and the Schoenfeld's (Schoenfeld, *Biometrics* vol. 39, pp 499-503, 1983) sample size formula. Monte Carlo simulation studies show that the proposed test provides sample sizes which are close to those obtained by standard weighted log-rank test under a proportional hazard alternative. However, the proposed test is more powerful than the standard weighted log-rank test under non-proportional hazard alternatives.