
One-sample survival tests for non-proportional hazards in oncology clinical trials: a simulation study

Chloé Szurewsky^{*1}, Guosheng Yin², Sarah Jonas³, and Gwénaél Le Teuff³

¹Centre de recherche en épidémiologie et santé des populations – Oncostat, CESP Inserm U1018, Université Paris-Saclay, Villejuif, France – France

²Department of Statistics and Actuarial Science – The University of Hong Kong Pokfulam Road, Hong Kong, R.A.S. chinoise de Hong Kong

³Oncostat – Institut Gustave Roussy, Centre de recherche en épidémiologie et santé des populations – Gustave Roussy.114 rue Édouard Vaillant, 94805 Villejuif Cedex, France

Résumé

In recent years, several one and two-stage single-arm designs for time-to-event trials have been developed to overcome the difficulty to conduct phase II randomized clinical trial in some settings (rare disease) in oncology. Most of these designs rely on the one-sample log-rank test (OSLRT) and its modified version (mOSLRT) that compares the survival curve of an experimental arm to that of an external reference under the proportional hazards (PH) assumption. We propose to adapt the OSLRT and evaluate alternatives when PH does not hold which may be the case for immunotherapies evaluation. We extended Finkelstein's score test developed under PH by using a piecewise exponential model with pre-specified change-points (CPs) for early, middle or delayed relative treatment effect. An accelerated hazards model is used for crossing hazards. As CPs are not a priori known, a two-step approach based on landmark analysis (1st step) is developed to determine the shape of the experimental curve and also to choose the score test (2nd step). We also extended the Restricted Mean Survival Time (RMST-) based test to single-arm trials and constructed combination tests of the developed score tests with the Hochberg correction for multiplicity. The performances (type I error and power) of these tests are evaluated through a simulation study of a phase II with an accrual and a follow-up period of 3 and 4 years, respectively. The reference curve is generated with an exponential distribution with a median survival time of 2 years and no variability. The parameters are sample size of the experimental arm (from 20 to 200 patients), exponential censoring rates (from 0 to 35%) and relative treatment effect (hazard ratio from 0.5 to 1). The simulation study shows that the score tests are as conservative as the OSLRT but more than the mOSLRT. As expected, the score test has the highest power when the data generation matches with the model used to develop the test even when the CPs are misspecified. The 2-step approach only works well with large sample sizes ($n > 100$). The RMST-based test is more powerful than the mOSLRT under non PH only for an early effect with censoring rate less than 15%. Combination tests are very conservative. They give a higher power than the mOSLRT but lower than the right score test under non PH. In conclusion, the score tests for time-to-event trials are efficient in non PH situations when CPs (or approximate values) are known. Further researches needs to be carried out to investigate, for example, the impact of a misspecification of the reference survival curve and multiplicity adjustments for combination tests.

*Intervenant