Empirical power of *CoxCombo* test under uncertain proportional hazards: A simulation study

Jordi Cortés Martínez[1], *Marta Bofill Roig*[2], *Guadalupe Gómez Melis*[3]

[1]jordi.cortes-martinez@upc.edu, Statistics & Operations Research, Universitat Politècnica de Catalunya
[2]marta.bofillroig@meduniwien.ac.at, Center for Medical Data Science, Medical University of Vienna
[3]lupe.gomez@upc.edu, Statistics & Operations Research, Universitat Politècnica de Catalunya

The proportional hazard (PH) premise is assumed in most of the clinical trials with time-to-event endpoints. Under this assumption, the Cox model and the logrank test lead to the most powerful test for comparing survival curves to prove the treatment effect. However, this premise may not always hold true. Alternative approaches, such as the modestly weighted logrank test [1] or the maxCombo test [2] have been proposed to be used in cases where proportionality may not be fulfilled. These approaches are, however, based on weighted logrank tests and lack to provide a clear measure of the effect.

In this work, we propose a new statistical test, *CoxCombo*, which combines several treatment effect measures coming from different Cox models. The goal is to compare the efficacy of an experimental treatment against a control treatment in a trial with a primary survival endpoint. We consider the three Wald statistics Z_C, Z_{WC}, and Z_A, obtained by using a weighted estimation in Cox regression [3], corresponding to testing the effect by means of: simple hazard ratio, average hazard ratio, and average regression effects, respectively. We define then the *CoxCombo* statistic as $Z_{CC} = max\{Z_C, Z_{WC}, Z_A\}$, and calculate its p-value by taking the multivariate distribution of the tests into account. We compare the performance of the proposed test, modestly weighted logrank test and *maxCombo* in terms of the power and type 1 error through a simulation study.

We consider four different scenarios of treatment effect behavior over time: 1) PH; 2) early effect; 3) cross effect; and 4) delayed effect. The last three scenarios imply non-proportional hazards. Piecewise exponential distributions are used to model changes in treatment effect over time. We will use the statistical software R (version 4.1.2) to perform the simulations. Specifically, the `nphRCT`, `coxphw`, and `multicomp` packages will be used to simulate the survival data; to calculate the Wald statistics; and to handle the multiplicity, respectively. In this talk, we discuss the simulation results and the potential of the CoxCombo test as an alternative to the other proposed methods.

1. Magirr D, Burman CF. Modestly weighted logrank tests. Stat Med. 2019;38:3782-90.

2. Lin RS, Lin J, Roychoudhury S, et al. Alternative Analysis Methods for Time to Event Endpoints Under Nonproportional Hazards: A Comparative Analysis, Stat. Biopharm. 2020;12:187-98

3. Schemper M, Wakounig S, Heinze G. The estimation of average hazard ratio by weighted Cox regression. Stat Med. 2009;28,2473-2489.

Keywords: Non-proportional hazards, CoxCombo, MaxCombo