Designing clinical trials with survival percentile - Planification d'essais cliniques avec les quantiles de Survie

PhD project

1 Introduction

When evaluating treatment effect, it is common to rely on the hazard ratio, typically by using the ubiquitous Cox model. In the presence of censoring, the hazard rate can be easily estimated from the observed data which makes this model very appealing. In Randomized Clinical Trials (RCT), standard methods already exist to determine the sample size when the estimand is a hazard ratio (typically the hazard ratio for comparing the effect of two treatments) based on either the log-rank test [see 1] or the Cox model [see 2]. However, in cancer studies, some treatments may have a late effect and the proportional hazard assumption imposed by the Cox model is no longer verified. We previously showed that the quantile regression was relevant for estimating the benefit of immunotherapy as compared to chemotherapy on a range of quantiles [see 3]. Thus, we would like to shift from the hazard ratio to the difference in quantile of failure time as estimand because:

(i) it allows for varying treatment effects (across quantiles) which corresponds to the heterogeneity of treatment effect

(ii) quantile regression can accommodate for non-proportional hazards (related to the late effects of immunotherapy).

2 Existing works

Only few statistical tests exist to test the equality of a given quantile between two independent groups from censored data. In the current proposals some issues have been raised. In [4], it has been stressed that the type I error was not well controlled when the distribution types were different in the treatment groups. In [5], the proposed approach does not suffer from this drawback, but they require in turn that the median is reached in each treatment group. Using the pseudo-value framework, the authors in [6] proposed a test that accommodates independent or clustered data and only requires the existence of the pooled survival median. This latter approach is particularly attractive to take into account heterogeneity coming from different data sources. However, all those proposed approaches exhibit a small power in simulation experiments which will impose to use a very large sample size in a RCT to use those tests at a high-power value.

3 Goal of the PhD thesis

The purpose of this PhD thesis is to derive a sample size formula for evaluating treatment effects by comparing pre-specified quantiles in each treatment groups.

1. We will first investigate more deeply the performances of the proposed approaches. In all the simulation settings of the aforementioned papers, the authors did not consider heavy tailed distributions. Since the aim of our approach is to detect long-term treatment effect, it is therefore important to assess the performance of the tests in this setting. Also, a promising method was also developed in [7]. In this paper, the author showed that:

$$\left\{\sqrt{n}(\hat{F}_1^{-1}(p_1) - \hat{G}_1^{-1}(p_1)), \ldots, \sqrt{n}(\hat{F}_J^{-1}(p_J) - \hat{G}_J^{-1}(p_J))\right\}^\top \tag{1}$$

is asymptotically a zero-mean Gaussian process with a covariance matrix that can be consistently estimated. In the above result, $\hat{F}_j$ and $\hat{G}_j$ represent the empirical cumulative distribution functions for two different treatment groups, $p_1, \ldots p_J$ are probability values such that $\hat{F}_j^{-1}(p_j)$, $\hat{G}_j^{-1}(p_j)$ represent the empirical quantiles of order p_j for two different treatment groups.

This approach has not yet been compared to the other proposals. It would be interesting to conduct extensive simulation studies that compare the power of all the different approaches, including the one in [7]. Theoretical expression of the power of the method in [7] is not derived, and since the author needs to make several assumptions for the test to work, it would be relevant to assess how those assumptions might affect the power of the tests. In particular, this method needs an estimator of the density of the distributions which might have an important impact on the resulting power of the method.

2. A statistical test based on a single quantile might be too restrictive as it does not take into account other aspects of the distribution. Also, one does not necessarily know in advance which quantile value to use. On the other hand, group sequential clinical designs is an active research area which consists in comparing quantile effects before the last participant has finished clinical follow-up. This allows for early stopping of the clinical trial if superiority of the treatment effect is assessed or if the data analysis suggests there is little chance of achieving a desired magnitude of treatment effect (futility).

   One major advantage of the work from [7] is that it allows to either test simultaneously different quantiles or to test the same quantile at different analysis times in a group sequential clinical design. More precisely, Formula (1) allows to test the equality of J distinct quantiles if $\hat{F}_1 = \cdots = \hat{F}_J$ and $\hat{G}_1 = \cdots = \hat{G}_J$ but $p_1, \ldots, p_J$ are distincts or to construct group sequential boundaries for testing equality of a single quantile, $F^{-1}(p) = G^{-1}(p)$, if $p_1 = \cdots = p_J$ but the pairs $F_j^{-1}(p) = G_j^{-1}(p)$ come from data obtained at J different analysis times. One major goal would be to derive sample size formulas in a group sequential clinical design based on the method from [7].

3. An additional goal is to develop a method that could incorporate data from different clinical trials instead of using only one single clinical trial. This is a common procedure in order to increase the sample size and thus the power of the statistical test. However, this is theoretically challenging as the method needs to deal with the specific dependence structure of such clustered data: the times are correlated among a clinical trial but independent between them. The existing approaches do not account for clustered data except the one in [6]. In this paper, the authors use a pseudo-observations approach (which is based on a jackknife version of the Kaplan-Meier estimator) combined with a pooled quantile estimator. They derive the theoretical distribution of their statistical test under the null assumption of no quantile difference between treatment groups. Therefore, one aim will be to combine works from [6] and [7] to both take into account clustered data and group sequential clinical designs.

4. Finally, we will investigate how to extend those approaches to other endpoints such as the Restricted Mean Survival Time (RMST). Let τ be a predefined time horizon, T the time of interest, the RMST is then defined as $\mathbb{E}[T \wedge \tau] = \int_0^\tau S(t)dt$, where $S(t) = \mathbb{P}[T \geq t]$ is the survival function. The difference in mean survival time quantifies the treatment effect in terms most meaningful to the patients and incorporate the whole information up to the restriction time. Also, as for the difference in terms of quantiles, it is valid under non-proportionality of the treatment effect. In the work from [8], the authors have developed the package "RMSTdesign" that allows sample size calculation for testing the equality of RMST between two groups. It would be interesting to extend this work to allow for several restricting times in a group sequential clinical design. Also, pseudo-value approaches are well suited for estimating the RMST and it would be of interest to develop an approach like in [6] that takes into account clustered data. There's a subtle connection between RMST and percentile of survival because in practice, the choice of a clinically relevant restriction time imposes a percentile of survival. We believe that our proposal will improve our understanding and the implementation of both endpoints in clinical trials.

4 Supervision

- Olivier Bouaziz, maître de conférences HDR, Laboratoire MAP5 (UMR CNRS 8145), Université Paris Cité. olivier.bouaziz@u-paris.fr
- Aurélien Latouche, professeur des universités, Cnam, Institut Curie (Inserm U900), St Cloud. aurelien.latouche@curie.fr

O. Bouaziz and A. Latouche have both contributed to several developments of statistical methods in medical applications. O. Bouaziz is in particular an expert in survival analysis, a topic he has studied both from a theoretical and applied perspective. He has been implied over recent years in several medical collaborations and has worked in particular on malaria, psychiatric disorders, primary immunodeficiency and cancer. A. Latouche has headed 8 Phds in applied Statistics and Medical Statistics. They are co-heading a master internship started in 2023 that initiated this application. The topic of the thesis comprises both theoretical developments and well defined medical needs. The PhD candidate will therefore benefit from the rich experience of her supervisors in those areas.

References

[1] David Schoenfeld. The asymptotic properties of nonparametric tests for comparing survival distributions. *Biometrika*, 68(1):316–319, 1981.

[2] David A Schoenfeld. Sample-size formula for the proportional-hazards regression model. *Biometrics*, pages 499–503, 1983.

[3] Bassirou Mboup, Christophe Le Tourneau, and Aurélien Latouche. Insights for quantifying the long-term benefit of immunotherapy using quantile regression. *JCO Precision Oncology*, 1:173–176, 2021.

[4] Ron Brookmeyer and John Crowley. A k-sample median test for censored data. *Journal of the American Statistical Association*, 77(378):433–440, 1982.

[5] Mohammad H Rahbar, Zhongxue Chen, Sangchoon Jeon, Joseph C Gardiner, and Jing Ning. A nonparametric test for equality of survival medians. *Statistics in medicine*, 31(9):844–854, 2012.

[6] Kwang Woo Ahn and Franco Mendolia. Pseudo-value approach for comparing survival medians for dependent data. *Statistics in medicine*, 33(9):1531–1538, 2014.

[7] Michael R Kosorok. Two-sample quantile tests under general conditions. *Biometrika*, 86(4):909–921, 1999.

[8] Anne Eaton, Terry Therneau, and Jennifer Le-Rademacher. Designing clinical trials with (restricted) mean survival time endpoint: practical considerations. *Clinical Trials*, 17(3):285–294, 2020.