Statistical methods and designs in clinical oncology

Design and analysis of survival data with non-proportional hazards

(9h – 12h30)

Restricted mean survival and hazard ratios

Ludovic Trinquart, Justine Jacot, Sarah C. Conner and Raphaël Porcher
Centre de Recherche Épidémiologie et Statistique Sorbonne Paris Cité (CRESS-UMR1153), Inserm ; Université Paris Descartes ; AP-HP, Hôtel-Dieu, Paris, France.

The hazard ratio (HR) is the preferred measure of treatment effect in oncology clinical trials. However, interpreting the magnitude of survival benefit using HRs may be challenging: HRs are often erroneously interpreted as relative risks and moreover they may vary with time (so-called non-proportional hazards). The restricted mean survival time (RMST) difference and ratio have been suggested as alternative measures of treatment effect.
In this work we aimed to compare empirically the treatment effects measured by the HR and by the difference (and ratio) of RMST in oncology randomized trials. We selected oncology randomized controlled trials from five leading journals during the last 6 months of 2014. After reconstruction of individual patient data for one time-to-event outcome from each trial – preferably the primary outcome – we reanalyzed each trial and compared the treatment effect estimated by the HR with that by the difference (and ratio) of RMST.
We analyzed 54 randomized controlled trials totaling 33,212 patients. The selected outcome was overall survival in 21 (39%) trials. There was evidence of non-proportionality of hazards in 13 (24%) trials. The HR and RMST-based measures were in agreement regarding the statistical significance of the effect, except in one case. The median HR was 0.84 (Q1 to Q3 range, 0.67 to 0.97) and the median difference in RMST was 1.12 months (range, 0.22 to 2.75 months). The average ratio of the HR to the ratio of RMST was 1.11 (95% CI, 1.07 to 1.15), with substantial between-trial variability (I^2 = 86%). Results were consistent by outcome type (overall survival v other outcomes) and whether the proportional hazard assumption held or not.
On average, the HR provided significantly larger treatment effect estimates than the ratio of RMST. The HR may seem large when the absolute effect is small. RMST-based measures may be considered as relevant measures of treatment effect on time-to-event outcomes, be the hazards proportional or not.

Béranger Lueza (Service de Biostatisitque et d'Epidémiologie, Gustave Roussy / Oncostat, CESP INSERM U1018, Villejuif, France): **Moving from the hazard ratio to the difference in restricted mean survival time in IPD meta-analyses**

The difference in restricted mean survival time (rmstD), the area between two survival curves up to a time horizon t*, is often used in cost-effectiveness analyses to estimate the treatment effect in randomized controlled trials. The rmstD is expressed in terms of life years gained and has been suggested as an alternative measure to the common relative measures used to estimate the treatment effect (e.g. hazard ratio), especially in case of non-proportional hazards of death. If one wants to estimate the rmstD from an individual participant data (IPD) meta-analysis, there is a need to take into account the trial effect due to the hierarchical structure of the data. We have compared statistical methods to estimate the rmstD from an IPD meta-analysis of randomized trials. The starting point was a case study (cost-effectiveness analysis) using data from the Meta-Analysis of Radiotherapy in Lung Cancer. This study showed that the investigated methods yielded different estimates for the rmstD and its confidence interval. The choice of a method to estimate the rmstD also impacted on cost-effectiveness results. We then conducted a simulation study to have a better understanding of the properties of the investigated methods. We varied the between-trial heterogeneity for the baseline hazard and for the treatment effect (possibly correlated), the overall treatment effect, the time horizon t*, the number of trials and of patients, the use of fixed or DerSimonian-Laird random effects model, and the proportionality of hazards. We compared the methods in terms of bias, empirical and average standard errors. We would recommend a two-stage method derived from the Kaplan-Meier estimator that formed the best compromise in terms of bias and variance.

Using the average hazard ratio to evaluate treatment effects with non-proportional hazards

Georg Heinze

In situations of non-proportional hazards, an average hazard ratio (AHR) can be computed either by integrating a time-dependent hazard ratio estimate over time, or by direct estimation as proposed by Schemper et al, StatMed 2009 and Wakounig et al, StatMethMedRes 2015. In this talk, we provide an overview of the theory behind the concept of AHR and related measures, and demonstrate the relationship of the AHR with odds of concordance and the two-group concordance probability. We report on empirical results and discuss advantages and disadvantages of the approach. In the era of personalized medicine, AHRs are attractive effect size measures as they summarize the available evidence for treatment decisions at baseline, and – if re-expressed as concordance probabilities – are easily communicable also to patients.

Augmenting the logrank test in the design of clinical trials in which non-proportional hazards of the treatmenteffect may be anticipated.

Mahesh K. B. Parmar and Patrick Royston

MRC Clinical Trials Unit at UCL, Aviation House 125 Kingsway, London, WC2B 6NH, UK.

BACKGROUND:

Most randomized controlled trials with a time-to-event outcome are designed assuming proportional hazards (PH) of thetreatment effect. The sample size calculation is based on a logrank test. However, non-proportional hazards are increasingly common. At analysis, the estimated hazards ratio with a confidence interval is usually presented. The estimate is often obtained from a Cox PH model withtreatment as a covariate. If non-proportional hazards are present, the logrank and equivalent Cox tests may lose power. To safeguard power, we previously suggested a 'joint test' combining the Cox test with a test of non-proportional hazards. Unfortunately, a larger sample size is needed to preserve power under PH. Here, we describe a novel test that unites the Cox test with a permutation test based on restricted mean survival time.

METHODS:

We propose a combined hypothesis test based on a permutation test of the difference in restricted mean survival time across time. The test involves the minimum of the Cox and permutation test P-values. We approximate its null distribution and correct it for correlation between the two P-values. Using extensive simulations, we assess the type 1 error and power of the combined test under several scenarios and compare with other tests. We investigate powering a trial using the combined test.

RESULTS:

The type 1 error of the combined test is close to nominal. Power under proportional hazards is slightly lower than for the Cox test. Enhanced power is available when the treatment difference shows an 'early effect', an initial separation of survival curves which diminishes over time. The power is reduced under a 'late effect', when little or no difference in survival curves is seen for an initial period and then a late separation occurs. We propose a method of powering a trial using the combined test. The 'insurance premium' offered by the combined test to safeguard power under non-PH represents about a single-digit percentage increase in sample size.

CONCLUSIONS:

The combined test increases trial power under an early treatment effect and protects power under other scenarios. Use of restricted mean survival time facilitates testing and displaying a generalized treatment effect.

The Net Chance of a Longer Survival as a Patient-Oriented Measure of Treatment Benefit in Randomized Clinical Trails

Julien Péron

Importance – Time to events or "survival endpoints" are common endpoints in randomized clinical trials. They are usually analyzed under the assumption of proportional hazards (PH), and the treatment effect is reported as a hazard ratio, which is neither an intuitive measure, nor a meaningful one if the PH assumption is not met.

Objective – We show here that a different measure of treatment effect, called "the net chance of a longer survival" is meaningful to patients whether or not the PH assumption applies. The procedure can also be used to permit a comprehensive assessment of several prioritized outcomes between two groups of observations.

Evidence Review – The net chance of a longer survival by at least m months, where m months is considered clinically worthwhile and patient-relevant, is calculated as the probability that a random patient in the experimental arm has a survival longer by at least m months than a random patient in the control group, minus the probability of the opposite situation. The net chance of a longer survival is equal to zero if treatment does not differ from control. It ranges from -100% if all patients in the control group fare better than the patients in the treatment group to +100% in the opposite situation. The net chance of a longer survival can be estimated for different values of m using a statistical technique known as generalized pairwise comparisons. We simulated datasets for realistic trials under various scenarios of proportional and non-proportional survival hazards, and plotted the Kaplan-Meier survival curves as well as the net chance of a longer survival as a function of m.

Findings – When proportional hazards hold, the net chance of a longer survival goes to zero as m increases. In contrast, when treatment effects are delayed or when some patients are cured by treatment, the net chance of a long survival benefit remains high and tends to the cure rate. We present a procedure that estimate the net chance of a longer survival with only a slightly bias even in presence of heavy censoring. We also show how this bias can be corrected when only one time-to-event outcome analyzed.

Conclusions and Relevance - The net chance of a longer survival is an intuitive measure of treatment benefit that has direct relevance to patients and health care providers. It is useful whether or not the assumption of proportional hazards holds in the analysis of survival endpoints.

Keywords: Statistics as topic; treatment effect; survival analysis; randomized controlled trial

Statistical methods and designs in clinical oncology

Designs with biomarkers (13h30 – 17h)

Novel designs for trials with multiple treatments and biomarkers

James Wason, MRC Biostatistics Unit, Cambridge, UK

Response to treatments is often highly heterogeneous. The increased availability of biomarkers and targeted treatments has led to the need for trial designs that efficiently test new treatments in biomarker-stratified patient subgroups. Often new treatments are targeted at a specific biomarker subgroup, but may in fact work in a narrower or broader set of patients.

I will discuss Bayesian adaptive methodology for trials that have multiple treatments and biomarkers. The proposed design incorporates biological hypotheses about links between treatments and biomarker subgroups effects of treatments, but allows alternative links to be formed during the trial. This design has been developed for trials in ovarian cancer and breast cancer. In these trials there was the need for a design that used an adaptive approach to ensure that patients are being allocated to the treatments that are best targeted at their biomarker profile. The statistical properties of this design compares well to alternative approaches available.

I will also briefly discuss future methodology work on adding in new treatments to these types of trials and making use of high-dimensional biomarker data in clinical trials.

Design and analysis of umbrella and basket trials

Axel Benner and Christina Beisel

Division of Biostatistics, German Cancer Research Center, Heidelberg, Germany benner@dkfz.de

In recent years, numerous approaches for biomarker-based clinical trials have been developed and applied. One of these developments are clinical trials, that aim to investigate multiple biomarkers simultaneously in independent subtrials. Multiple biomarker trials can be divided into two main categories: basket trials and umbrella trials. While umbrella trials focus on one specic disease or tumor type that is divided into biomarker-defined subtypes, basket trials include patients solely based on their biomarkers irrespective of histology. We compare basket and umbrella trials and examine their advantages and dis- advantages. Furthermore, we investigate possibilities for design and analysis for these two trials concepts and review a few real life examples where these trial concepts are already being applied. For trials that have already been completed we also discuss some issues they encountered and lessons learned from these trials. Finally, we draw conclusions on possibilities to further improve the design and analysis for these trial concepts in the future.

Causal-inference approaches to evaluation of surrogate endpoints

Pr Tomasz Burzykovski, Hasselt University, Hasselt, Belgium.

« Initially, surrogate-endpoint evaluation methods were developed within the frequentist framework. In particular, the so-called meta-analytic approach, based on the analysis of association between treatment effects on the surrogate and clinical endpoints estimated in multiple clinical trials, was proposed and proved its practical value. In recent years, methods using concepts from the causal-inference framework have become a focus of research. In the presentation we will discuss the conceptual framework underlying the causal-inference-based approaches. Additionally, we will describe the relationship between the meta-analytic and causal-inference approaches.”

Evaluation of treatment effect with paired failure times in a single-arm phase II trial in oncology

Matthieu Texier

In early phase clinical trials in oncology, for cytotoxic drugs, the efficacy is typically evaluated based on the tumor shrinkage. However, this criterion is not always appropriate for more recent cytostatic agents, and alternative endpoints have been proposed. The Growth Modulation Index (GMI), defined as the ratio between the times to progression in two successive treatment lines, has been proposed for a single arm phase II trial. The treatment effect is evaluated by estimating the rate of patients having their GMI superior to a threshold. We investigated two methods to estimate this rate: a parametric one based on the distribution of the times to progression and a non-parametric one, based on a midrank estimator. Through simulations, we studied their operating characteristics and the impact of different design parameters (censoring, dependence, distribution) on these estimators. We show that the non-parametric estimator slightly underestimated the rate and had slightly overconservative confidence intervals in some cases. Conversely, the parametric estimator overestimated the rate and had anticonservative confidence in some cases. The non-parametric method appeared to be more robust to censoring than the parametric one.

We illustrate the two methods with an application to an advanced colorectal cancer dataset. In conclusion, we recommend the non-parametric method as primary analysis but the parametric method can be used as a supplementary tool.

The use of tumor dynamics and new lesions to predict survival with multivariate joint frailty models.

Agnieszka Król[1], Christophe Tournigand[2], Stefan Michiels[3] and Virginie Rondeau[1]

[1]Research Center INSERM U1219, Biostatistical team, Université Bordeaux, France

[2]Hôpital Henri Mondor, Créteil, France

[3]Service de Biostatistique et d'Epidémiologie, Gustave Roussy, University Paris-Saclay, University Paris-Sud, CESP, INSERM U1018, Villejuif, France

Abstract: The RECIST criteria are used as standard guidelines for the clinical evaluation of cancer treatments. The assessment is based on the anatomical tumor burden: change in size of target lesions and evolution of non-target lesions, ie. appearance of new lesions and progression of non-target lesions. Despite indisputable advantages of this standard tool, RECIST are subject to some limitations such as categorization of continuous tumor size or negligence of its longitudinal trajectory. In particular, it is of interest to capture its nonlinear shape heterogeneous among patients and model it simultaneously with recurrent events of non-target disease evolution and overall survival. This complex analysis is achievable using multivariate joint frailty models for longitudinal and survival data. We propose a new mechanistic joint frailty model for longitudinal data, recurrent events and a terminal event. In the model, the tumor size trajectory is described using an ordinary differential equation (ODE) that accounts for the natural growth and treatment induced decline. We perform a simulation study to validate the proposed method and apply the model to a real dataset of a phase III clinical trial for metastatic colorectal cancer. In the results of the analysis, we determine on which component: tumor size, NTL or death, the treatment acts mostly and evaluate the predictive abilities of the components for death. We compare this model with models that consider parametric functions or splines for the SLD trajectory in terms of goodness-of- t and predictive accuracy.

Keywords: Joint models, Longitudinal data, Ordinary differential equation, Survival analysis, Tumor measurement