Computational Statistics and Data Analysis 75 (2014) 151–161

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

Consistency-adjusted alpha allocation methods for a time-to-event analysis of composite endpoints

G. Rauch *, M. Wirths, M. Kieser

Institute of Medical Biometry and Informatics, University of Heidelberg, Germany

ARTICLE INFO

Article history:
Received 21 July 2013
Received in revised form 21 January 2014
Accepted 21 January 2014
Available online 6 February 2014

Keywords:
Clinical trial
Composite endpoint
Multiple testing
Bootstrap

ABSTRACT

Composite endpoints are often used as primary efficacy endpoints, particularly in the field of oncology and cardiology. These endpoints combine several time-to-event variables of interest within a single time-to-first-event variable. Thereby, it is intended to enlarge the expected effect size and thus to increase the power of the clinical trial. However, the interpretation of composite endpoints can be difficult, as the observed effect for the composite does not necessarily reflect the effects of the single components. Therefore, it might not be adequate to judge the efficacy of the new intervention exclusively on the composite effect. Including the most relevant components in an efficacy claim assessed by a confirmatory test strategy could overcome this problem but imposes the problem of multiplicity. Moreover, to show non-inferiority or even superiority of the new intervention with respect to single components is usually not realistic in these settings as the expected individual effects are small. Recently, consistency-adjusted alpha allocation methods were proposed in the literature which can be used and extended to establish a new efficacy claim for a composite endpoint and one main component. The power properties of the new approach are compared to the alternative efficacy claim of proving superiority for the composite and non-inferiority for the main component. Moreover, the methods are illustrated with a clinical trial example. Thereby, the general problem of correlation-adjusted multiple testing procedures is addressed by applying a bootstrapping algorithm to estimate the special correlation structure between a composite endpoint and an individual component in the time-to-event setting.[1]

1. Introduction

Composite endpoints combine several events of interest within a single variable. They are usually defined as time-to-first-event variables, which are evaluated via survival analysis techniques. The main motivation for the use of a composite endpoint is to increase power by enlarging the number of expected events. Often, this is the only possible solution to make a clinical trial feasible in terms of study duration, sample size and costs. On the other hand, many authors commented on possible interpretation problems when using composite endpoints (Bethel et al., 2008; Ferreira-Gonzáles et al., 2007; Freemantle and Calvert, 2007; Freemantle et al., 2003). The problem is that the pooled effect of the composite does not necessarily reflect the effect of the single components. For example, it might be possible that a slight negative effect in

* Correspondence to: Im Neuenheimer Feld 305, 69120 Heidelberg, Germany. Tel.: +49 6221 561932; fax: +49 6221 564195.
E-mail addresses: rauch@imbi.uni-heidelberg.de (G. Rauch), wirths@imbi.uni-heidelberg.de (M. Wirths), meinhard.kieser@imbi.uni-heidelberg.de (M. Kieser).

[1] The R code is available online as supplementary material (see Appendix C).

http://dx.doi.org/10.1016/j.csda.2014.01.017
0167-9473/

a very severe endpoint like death is masked by a large positive effect in a less relevant component, like, for example, hospitalization. In the literature it is therefore recommended not to combine endpoints of different clinical severity (Chi, 2005) or components that are expected to be affected by the intervention in opposite directions (CPMP, 2002). This, however, is usually hard to fulfill in clinical practice as most composite endpoints in oncology or cardiology include death as one component which is always more harmful than other events like tumor progress, metastasis, myocardial infarction, stroke or hospitalization. Moreover, the magnitude and the direction of an intervention effect with respect to an individual component cannot always be predicted correctly in the planning stage.

As a consequence, in addition to the analysis of the composite endpoint, an evaluation of the individual components is recommended (Bethel et al., 2008; Chi, 2005; CPMP, 2002). In most clinical trials with a composite primary endpoint, the individual components are analyzed descriptively. This, however, does not solve the problem of how to deal with a study result that shows superiority of the new intervention with respect to the composite but also shows a negative trend for a severe individual component like death. The composite effect alone might not be an adequate efficacy measure. Although descriptive analyses of components give some supplementary information, they do not justify a proof of efficacy or a proof for the absence of harm. A possible solution is to formulate a multiple test problem involving the composite and at least the most relevant component. This sounds contradictory at first glance, as the motivation for the use of a composite endpoint was to avoid testing single components for which only a low number of events is expected. In fact, it is usually not realistic to show superiority of the new intervention with respect to a single component. A realizable efficacy claim, however, might be to show superiority with respect to the composite and non-inferiority with respect to the most relevant component in an intersection-union test thus proving at least the absence of harm. This approach has been discussed by Huque et al. (2011) and Röhmel et al. (2006).

An alternative strategy might be not to perform formal hypothesis testing for the main component but to introduce a consistency criterion connecting the effect of the composite with the effect of the most relevant component. A possible consistency claim could be, that the component effect should not point in opposite direction to the composite effect. Alosh and Huque recently published several works on consistency-adjusted alpha allocation methods (Huque and Alosh, 2012; Alosh and Huque, 2009) which can be used and extended to address this particular problem (Huque et al., 2011). The approach of consistency-adjusted alpha allocation is an extension of the well-known Bonferroni–Holm method. However, if the first hypothesis exceeds a predefined consistency bound, then testing is stopped with both null hypotheses being accepted. In case the first null hypothesis cannot be rejected but the p-value meets the consistency bound, the second null hypothesis can be tested at an adjusted significance level depending on the correlation between the involved test statistics. Although possible applications to composite endpoints have been shortly addressed by Huque and Alosh (2012) and Huque et al. (2011), there still exist relevant open questions which are answered in this paper. In this work, we addressed three important new issues. The first important task is how to formulate the consistency test problem for the composite endpoint and one main component. We propose to start by comparing the main component effect to a consistency bound that excludes opposite effects and to test the composite subsequently, which is different from the approach presented by Huque and Alosh (2012) and Huque et al. (2011). As the aim here is to show that the main component meets the consistency claim, which is exclusively imposed on the first hypothesis, our proposed ordering of test hypotheses seems intuitive. Note that a consistency criterion to exclude opposite effects is a much weaker claim than a formal superiority test. In the context of clinical trials with composite endpoints, it seems reasonable to impose a strong claim for the composite and a less stringent one for the main component. The second important issue is the application of the consistency-adjusted alpha allocation approach to time-to-event data. The third new aspect is that we address the problem of getting valid correlation estimates between a composite endpoint and a single component by use of a bootstrapping algorithm. We show that the underlying correlation between the test statistics reacts extremely sensitive to distributional parameters that determine the survival functions which makes it difficult to use a priory guesses of the underlying correlation.

Our version of the consistency-adjusted alpha allocation method is compared to the standard intersection-union test combining superiority for the composite and non-inferiority for the component in terms of power and interpretation. Our results are illustrated by simulations and a clinical trial example.

2. Methods

2.1. Local test hypotheses

In order to apply alternative efficacy claims to composite endpoints and their components which will be given as specific multiple testing procedures, it is necessary to formulate the underlying local test hypotheses and test statistics. We focus on a controlled clinical trial with a composite endpoint (CE) consisting of k components. We moreover assume that there exists one main component (MC) which is more relevant and more severe than all other components. In clinical applications this main component will most often be given by death or another fatal event. Note that our assumption to focus on one relevant main component is not really restrictive: In case that there exist several severe main components, it is possible to define the most relevant subcomposite and to apply the methods proposed below to the composite and the main subcomposite.

A confirmatory efficacy claim should thus include the composite and the main component by means of an adequate multiple testing procedure. In the following, the superscripts C and I denote the group affiliation to the control and the