World Journal of Cardiology

Submit a Manuscript: http://www.wjgnet.com/esps/
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
DOI: 10.4330/wjc.v7.i1.1

World J Cardiol 2015 January 26; 7(1): 1-5
ISSN 1949-8462 (online)

EDITORIAL

Opportunities and challenges of clinical trials in cardiology using composite primary endpoints

Geraldine Rauch, Bernhard Rauch, Svenja Schüler, Meinhard Kieser

Geraldine Rauch, Svenja Schüler, Meinhard Kieser, Institute of Medical Biometry and Informatics, University of Heidelberg, D-69120 Heidelberg, Germany
Bernhard Rauch, Institut für Herzinfarktforschung, D-67063 Ludwigshafen, Germany
Author contributions: All authors contributed to this manuscript.
Conflict-of-interest: The authors declare no conflicts of interest regarding this manuscript.

Correspondence to: Dr. Geraldine Rauch, Institute of Medical Biometry and Informatics, University of Heidelberg, Im Neuenheimer Feld 305, D-69120 Heidelberg, Germany. rauch@imbi.uni-heidelberg.de
Telephone: +49-6221-561932
Fax: +49-6221-561932
Received: October 27, 2014
Peer-review started: October 27, 2014
First decision: November 27, 2014
Revised: December 19, 2014
Accepted: December 29, 2014
Article in press: January 4, 2015
Published online: January 26, 2015

Abstract

In clinical trials, the primary efficacy endpoint often corresponds to a so-called "composite endpoint". Composite endpoints combine several events of interest within a single outcome variable. Thereby it is intended to enlarge the expected effect size and thereby increase the power of the study. However, composite endpoints also come along with serious challenges and problems. On the one hand, composite endpoints may lead to difficulties during the planning phase of a trial with respect to the sample size calculation, as the expected clinical effect of an intervention on the composite endpoint depends on the effects on its single components and their correlations. This may lead to wrong assumptions on the sample size needed. Too optimistic assumptions on the expected effect may lead to an underpowered of the trial, whereas a too conservatively estimated effect results in an unnecessarily high sample size. On the other hand, the interpretation of composite endpoints may be difficult, as the observed effect of the composite does not necessarily reflect the effects of the single components. Therefore the demonstration of the clinical efficacy of a new intervention by exclusively evaluating the composite endpoint may be misleading. The present paper summarizes results and recommendations of the latest research addressing the above mentioned problems in the planning, analysis and interpretation of clinical trials with composite endpoints, thereby providing a practical guidance for users.

Key words: Composite endpoint; Competing risks; Multiple testing; Time-to-event; Adaptive designs

Core tip: When planning a clinical trial with a composite primary endpoint: (1) Be aware of planning uncertainties when calculating the sample size and incorporate them in an adequate way; (2) Include a multiple testing strategy for an improved interpretation of the study results; (3) Take into account competing risks when analyzing the individual components of a composite endpoint; and (4) Analyze subsequent events in an adequate multi-stage model.

Rauch G, Rauch B, Schüler S, Kieser M. Opportunities and challenges of clinical trials in cardiology using composite primary endpoints. *World J Cardiol* 2015; 7(1): 1-5 Available from: URL: http://www.wjgnet.com/1949-8462/full/v7/i1/1.htm DOI: http://dx.doi.org/10.4330/wjc.v7.i1.1

RATIONALE FOR USING COMPOSITE ENDPOINTS

Clinical trials often focus on event variables as primary efficacy endpoints. In cardiology, "death" is often considered as the outcome of primary interest. However, clinically most relevant event types like "death" may be rare in many clinical conditions under investigation[1]. For example, due to the beneficial effects of modern treatments, patients with cardiovascular events like acute myocardial infarction experience a low mortality in the following years. Therefore, the assessment of differences in the survival curves of several treatment options may be difficult[2]. Using a rare event as primary endpoint results in the need of large sample sizes, a prolonged follow-up, and consequently an increased financial support, which often is not available. Thus, a "relevant and important treatment benefit" as claimed by the ICH E9 Guideline[3] cannot always be achieved by evaluating a single event endpoint, especially if this event type occurs with a low frequency[4]. By combining several types of events in a composite endpoint, the number of expected events is increased thereby intending an enlarged overall treatment effect. In the field of cardiovascular research, apart from death, clinical events like "non-fatal myocardial infarction", "non-fatal stroke", or "cardiovascular hospital admissions" also are of clinical interest and thus included into composite endpoints.

Most often, the composite endpoint is defined as a time-to-first-event variable, where different event types are counted as target events. In some applications, where the time period until the occurrence of an event is not of interest, composite endpoints can also be defined as binary event variables.

In summary, by using composite endpoints the required sample size is usually reduced and the study duration is shortened. Thereby, the use of composite endpoints very often is the only way to realize clinical trials investigating special interventions of interest.

Another important reason to use composite endpoints is when the effect of a new intervention may only be adequately assessed by considering several event variables. For example, atherosclerosis may result in a variety of clinical complications, and a single event endpoint therefore might not be sufficient for an adequate clinical evaluation[5]. Instead of formulating a multiple testing problem for several primary event endpoints, which always results in a loss of power, the ICH E9 Guideline[3] states that a composite outcome "addresses the multiplicity problem without requiring adjustment to the type I error".

CHALLENGES OF USING COMPOSITE ENDPOINTS

Planning and interpreting clinical trials with composite endpoints

Apart from the advantages of composite endpoints as outlined above, there also exist some serious problems and challenges.

In the planning stage of a clinical trial with a composite primary endpoint, calculation of the power may be particularly difficult as the assumed effect of the intervention depends on the effect sizes of the single components and their correlations. However, the level of evidence for these quantities may be low in many applications, as good historical data do not always exist. This complicates the choice of valid parameter assumptions in the planning phase of a study.

Analyzing and interpreting clinical trials with composite endpoints can be challenging as the composite effect as a "net measure" does not necessarily reflect the influence of the new intervention on the individual components[6,7]. Even in case a statistically significant and clinically relevant effect in the composite endpoint is observed, it may happen that the effects for some components are of very different magnitude or even point in opposite directions. As the efficacy of a treatment is usually judged on the composite effect alone, these situations may result in serious misinterpretations. This especially is a problem in case the composite endpoint consists of components of different clinical relevance and the less relevant endpoints refer to the larger effect sizes. The CPMP Guideline "Points to Consider on Multiplicity Issues"[8] therefore recommends to combine only components, which are expected to show effects of similar magnitude and with the same direction. This recommendation, however, may not be realistic in clinical practice. Even in thoroughly planned clinical studies, the initial assumptions about the underlying effect sizes can be wrong. Furthermore, the choice of the components must primarily be guided by their clinical relevance, and similar effects for all relevant event types cannot be expected in many cases.

Competing risks as a source of bias

The individual components of a composite endpoint usually define competing risks. In the presence of competing risks, the event rate of a specific event type also depends on the rates of all competing events[9]. For this reason, the event rates cannot be interpreted without simultaneously reporting all competing event rates. To illustrate this concept, assume that a novel therapeutic intervention in patients with cardiovascular disease is associated with a "one year mortality" of 0.3 as compared to 0.5 in the control group. Within the same group of patients, the rate for a "non-fatal myocardial infarction" might be 0.4 in the treatment group but only 0.2 in the control. If the death rates would not have been reported, one might come to the wrong conclusion that the control is superior to the treatment group with respect to "non-fatal myocardial infarction". When looking at the death rates, however, it becomes evident that the lower rate of "non-fatal myocardial infarction" in the control could exclusively be due to the fact that many patients had died before experiencing a (non-fatal) myocardial infarction. Ignoring the competing event scenario therefore may lead to a serious misinterpretation of treatment efficacy.

Therefore, methods taking into account competing events must be applied whenever the components of a composite endpoint are separately analyzed[5,8].

Follow-up beyond the first event

Composite time-to-first-event variables only take into account the first occurring event. This of course does not imply that there are no other subsequent events of interest occurring later. However, in the time to first event analysis these later events are not investigated, thereby leading to a loss of information.

On the other hand, an adequate and meaningful analysis of subsequent events may be a complex and difficult task, as-once a primary event has occurred the risk for all following events usually changes. For the latter reason models only focusing on a certain type of event, but not taking into account whether other events have occurred before, will yield biased results.

An unbiased approach to evaluate subsequent events would be to use more complex multistate models, which investigate all transition hazards between different subsequent event types[9]. The complexity of these models may be very high, and in order to get estimates with reasonable accuracy of all transition probabilities, the required sample size soon becomes unrealistically large. Therefore, for the confirmatory analysis of the composite and its components the time-to-first-event approach should usually be preferred. However, a descriptive presentation of the absolute numbers of all observed events should be provided in addition, keeping in mind that a correct interpretation of these results may be difficult.

NEW BIOMETRICAL METHODS FOR THE INTERPRETATION OF COMPOSITE ENDPOINTS

Overcoming uncertainties in sample size calculation

The standard approach to take account of planning uncertainties is the use of group-sequential or adaptive study designs. These designs allow stopping a trial at an interim stage due to an early demonstration of efficacy or due to futility. Whereas for group-sequential designs the number of interim analyses and the corresponding time points must be strictly planned in advance, adaptive designs additionally allow to change design parameters within an ongoing trial while still controlling the type I error rate.

A standard group-sequential design with one interim analysis (*e.g.*, after inclusion of 50% of the total study population) only offers two options-either to stop the study at interim or to continue the study until the full number of patients specified in the planning stage has been recruited[10-12].

In contrast, when using an adaptive design with one interim analysis, the sample size for the second stage can be recalculated based on the observed treatment effect at interim. If the observed effect at interim is large but not yet significant, only a small sample size for the second stage is needed, whereas the additionally required sample size is large, if the effect observed at interim is small. Moreover, it is possible to incorporate predefined stopping-for-futility rules in such designs, allowing to stop the study early with the acceptance of the null hypothesis whenever, on the basis of the interim data, the primary study goal becomes unrealistic. Thereby the number of patients being exposed to an ineffective treatment can be limited, and time and financial resources can be saved.

Another way to deal with uncertainties in the study planning assumptions is to use a more flexible power approach for sample size calculation. While the classical power is defined as the probability to reject the null hypothesis under a fixed parameter constellation of the alternative hypothesis, Rauch *et al*[13] proposed a so-called "expected power", which is defined as a weighted average over the classical power for different parameter constellations. Thereby, parameter constellations assumed to be more realistic in the planning stage of a study are assigned a higher weight, whereas other, less realistic assumptions are down-weighted. If there is no preexisting evidence available at all, equal weights for all possible parameter constellations might be assigned. The weights, which are defined by prior distributions, thus reflect the level of evidence or uncertainty in the planning stage. Calculating the sample size based on the "expected power" therefore defines a more robust approach in the common case of uncertain planning assumptions. The "expected power" can also be interpreted as a semi-Bayesian power approach[14].

Improving the interpretation of study results

The interpretation of study results may become difficult as the effect of the intervention under investigation on the composite endpoint does not necessarily reflect its effects on the single components. A possible solution of this problem would be to incorporate the (most important) components within the confirmatory test strategy by a multiple testing problem. However, this approach might seem to be contradictory as one main rationale for the use of a composite endpoint was to avoid multiplicity. A multiple testing problem always comes along with a certain loss in power resulting in an increase in sample size. The aim therefore is to create an adequate compromise by a multiple testing procedure, which mainly focuses on the composite endpoint but additionally gives some confirmatory evidence (at least) on the most important components.

In the literature, there exist a variety of either simple but also of more sophisticated multiple testing procedures, which can be applied to evaluate composite endpoints and their components. Simple applicable multiple testing strategies include the Bonferroni-Holm approach[15] or the sequential testing approach for hierarchically ordered hypotheses[16]. The application of at least a simple multiple testing strategy, which allow to address the components in a confirmatory way, is generally recommended. Even, if the trial is powered to

assess only the composite endpoint, these methods often allow a gain in information without increasing the sample size[17].

There also exist a variety of more sophisticated multiple testing procedures, which can be applied to provide sufficient power for the composite as well as for the (most relevant) components. So called "sequentially-rejective methods" represent extensions of the simple approaches outlined above. The underlying idea is to use an optimal splitting of the global significance level to test the individual hypotheses corresponding to the composite and the components. By "recovering" local levels of rejected hypotheses, the power loss due to multiplicity can be limited. Moreover, the test hypotheses for the components may be formulated less strictly than for the composite. For example, if the treatment under investigation already exhibits a significant and relevant effect on the composite, it might be sufficient to demonstrate in addition that the most relevant component is not adversely affected.

The application of sequentially-rejective multiple testing strategies in the evaluation of composite endpoints and their components has to be combined with the methodology for competing risks in order to provide an unbiased analysis and to prevent misinterpretations[18,19].

These methods can be further improved, if the correlation between the test statistics is taken into account. As an event referring to a single component always corresponds to an event in the composite endpoint, the test statistic of the composite and its components are usually highly correlated. By incorporating the information of the underlying correlation, the local significance levels of a multiple testing problem can be chosen less stringent, and the power loss often can be markedly decreased. These two approaches have been investigated recently by Rauch *et al*[20,21].

A completely different approach to improve the interpretation of clinical trials with composite endpoints is to use a weighted combined effect measure, which assigns higher weights to the more important components with the intention that an opposite effect in a relevant component (*e.g.*, "death") is less likely to be masked by a large effect in a component of secondary importance (*e.g.*, "cardiovascular hospital admission"). Recently, Pocock *et al*[22] and Buyse[23] proposed two similar combined effect measures, referred as the "win ratio" and the "proportion in favor of treatment", respectively. Both approaches are based on the same idea: All components are ordered with respect to their clinical relevance. The individual patients are compared between the groups. Based on the component of primary importance, for each comparison the patient with the "better" outcome is determined. In case no unique "winner" can be determined with respect to the most relevant component (*e.g.*, due to censoring, missing values or due to equal performance of both patients), the comparison will be based on the component of secondary importance and so on. This approach intends a higher weighting of the more relevant components, but also allows incorporating subsequent events.

Although this approach appears to be attractive in general, it also has some deficiencies. On the one hand, it can be shown that the weights, which are assigned to the single components, depend on the follow-up and the censoring distribution[24]. Moreover the weights are not standardized, that means they do not sum up to 1. As a consequence, the combined effect measure is not comparable between various studies as required-for example-within the context of meta-analyses. A small effect in the combined measure might thus be due to small effects in the components, but also could be explained by an unfavorable censoring distribution. Therefore, it cannot generally be deduced that these two approaches provide a gain in interpretation.

CONCLUSION

The use of a composite endpoint as primary efficacy variable can provide major advantages compared to a single event endpoint, if the event of primary interest is rare. However, care has to be taken when planning, analyzing and interpreting clinical trials with a composite endpoint as the primary efficacy outcome. The current statistical literature provides a variety of methods to overcome typical challenges arising from the use of composite endpoints thereby strengthening the interpretation of the results of clinical trials and avoiding serious misinterpretations. Now, the time has come to routinely incorporate these new methods into clinical trial applications.

REFERENCES

1 **Ferreira-González I**, Busse JW, Heels-Ansdell D, Montori VM, Akl EA, Bryant DM, Alonso-Coello P, Alonso J, Worster A, Upadhye S, Jaeschke R, Schünemann HJ, Permanyer-Miralda G, Pacheco-Huergo V, Domingo-Salvany A, Wu P, Mills EJ, Guyatt GH. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. *BMJ* 2007; **334**: 786 [PMID: 17403713 DOI: 10.1136/bmj.39136.682083.AE]

2 **Rauch B**, Schiele R, Schneider S, Diller F, Victor N, Gohlke H, Gottwik M, Steinbeck G, Del Castillo U, Sack R, Worth H, Katus H, Spitzer W, Sabin G, Senges J. OMEGA, a randomized, placebo-controlled trial to test the effect of highly purified omega-3 fatty acids on top of modern guideline-adjusted therapy after myocardial infarction. *Circulation* 2010; **122**: 2152-2159 [PMID: 21060071 DOI: 10.1161/CIRCULATIONAHA.110.948562]

3 **European Medicines Agency ICH E9 Guideline.** Statistical principles for clinical trials. Available from: URL: http: //www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002928.pdf

4 **Cannon CP**. Clinical perspectives on the use of composite endpoints. *Control Clin Trials* 1997; **18**: 517-529; discussion 546-549 [PMID: 9408715 DOI: 10.1016/S0197-2456(97)00005-6]

5 **Lubsen J**, Kirwan BA. Combined endpoints: can we use them? *Stat Med* 2002; **21**: 2959-2970 [PMID: 12325112 DOI: 10.1002/sim.1300]

6 **Bethel MA**, Holman R, Haffner SM, Califf RM, Huntsman-Labed A, Hua TA, McMurray J. Determining the most appropriate components for a composite clinical trial outcome.

Am Heart J 2008; **156**: 633-640 [PMID: 18926145 DOI: 10.1016/j.ahj.2008.05.018]
7 **Freemantle N**, Calvert M. Composite and surrogate outcomes in randomised controlled trials. *BMJ* 2007; **334**: 756-757 [PMID: 17431231 DOI: 10.1136/bmj.39176.461227.80]
8 **European Medicines Agency Committee For Proprietary Medicinal Products (CPMP).** Points to consider on multiplicity issues in clinical trials. Available from: URL: http: //www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003640.pdf
9 **Beyersmann J,** Alligniol A, Schumacher M. Competing risks and multistate models with R. New York: Springer-Verlag, 2012
10 **Jennison C,** Turnbull BW. Group sequential methods with applications to clinical trials. USA: Chapman & Hall/CRC, 2000
11 **Bauer P**, Köhne K. Evaluation of experiments with adaptive interim analyses. *Biometrics* 1994; **50**: 1029-1041 [PMID: 7786985 DOI: 10.2307/2533441]
12 **Wassmer G**. Planning and analyzing adaptive group sequential survival trials. *Biom J* 2006; **48**: 714-729 [PMID: 16972724 DOI: 10.1002/bimj.200510190]
13 **Rauch G,** Kieser M. An expected power approach for the assessment of composite endpoints and their components. *Comput Stat Data An* 2013; **60**: 111-122 [DOI: 10.1016/j.csda.2012.11.001]
14 **Daimon T**. Bayesian sample size calculations for a non-inferiority test of two proportions in clinical trials. *Contemp Clin Trials* 2008; **29**: 507-516 [PMID: 18201944 DOI: 10.1016/j.cct.2007.12.001]
15 **Holm S.** A simple sequentially rejective multiple test procedure. *Scand J Stat* 1979; **6**: 65-70 [DOI: 10.2307/4615733]
16 **Westfall PH,** Krishen A. Optimally weighted, fixed sequence and gatekeeper multiple testing procedures. *J Stat Plan Inference* 2001; **99**: 25-40 [DOI: 10.1016/S0378-3758(01)00077-5]
17 **Schüler S**, Mucha A, Doherty P, Kieser M, Rauch G. Easily applicable multiple testing procedures to improve the interpretation of clinical trials with composite endpoints. *Int J Cardiol* 2014; **175**: 126-132 [PMID: 24861257 DOI: 10.1016/j.ijcard.2014.04.267]
18 **Rauch G**, Beyersmann J. Planning and evaluating clinical trials with composite time-to-first-event endpoints in a competing risk framework. *Stat Med* 2013; **32**: 3595-3608 [PMID: 23553898 DOI: 10.1002/sim.5798]
19 **Rauch G**, Kieser M, Ulrich S, Doherty P, Rauch B, Schneider S, Riemer T, Senges J. Competing time-to-event endpoints in cardiology trials: a simulation study to illustrate the importance of an adequate statistical analysis. *Eur J Prev Cardiol* 2014; **21**: 74-80 [PMID: 22964966 DOI: 10.1177/2047487312460518]
20 **Rauch G,** Wirths M, Kieser M. Consistency-adjusted alpha allocation methods for a time-to-event analysis of composite endpoints. *Comput Stat Data An* 2014; **75**: 151-161 [DOI: 10.1016/j.csda.2014.01.017]
21 **Rauch G**, Kieser M. Multiplicity adjustment for composite binary endpoints. *Methods Inf Med* 2012; **51**: 309-317 [PMID: 22525969 DOI: 10.3414/ME11-01-0044]
22 **Pocock SJ**, Ariti CA, Collier TJ, Wang D. The win ratio: a new approach to the analysis of composite endpoints in clinical trials based on clinical priorities. *Eur Heart J* 2012; **33**: 176-182 [PMID: 21900289 DOI: 10.1093/eurheartj/ehr352]
23 **Buyse M**. Generalized pairwise comparisons of prioritized outcomes in the two-sample problem. *Stat Med* 2010; **29**: 3245-3257 [PMID: 21170918 DOI: 10.1002/sim.3923]
24 **Rauch G**, Jahn-Eimermacher A, Brannath W, Kieser M. Opportunities and challenges of combined effect measures based on prioritized outcomes. *Stat Med* 2014; **33**: 1104-1120 [PMID: 24122841 DOI: 10.1002/sim.6010]

P- Reviewer: Amiya E, Iacoviello M, Teragawa H
S- Editor: Ji FF **L- Editor**: A **E- Editor**: Lu YJ

Published by **Baishideng Publishing Group Inc**
8226 Regency Drive, Pleasanton, CA 94588, USA
Telephone: +1-925-223-8242
Fax: +1-925-223-8243
E-mail: bpgoffice@wjgnet.com
Help Desk: http://www.wjgnet.com/esps/helpdesk.aspx
http://www.wjgnet.com