Journal of Clinical Epidemiology 76 (2016) 155–165

Journal of Clinical Epidemiology

Simultaneous sequential monitoring of efficacy and safety led to masking of effects

Rik van Eekelen*, Esther de Hoop, Ingeborg van der Tweel

Department of Biostatistics & Research Support, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, PO Box 85060, 3508 AB Utrecht, The Netherlands

Accepted 24 February 2016; Published online 2 March 2016

Abstract

Objective: Usually, sequential designs for clinical trials are applied on the primary (=efficacy) outcome. In practice, other outcomes (e.g., safety) will also be monitored and influence the decision whether to stop a trial early. Implications of simultaneous monitoring on trial decision making are yet unclear. This study examines what happens to the type I error, power, and required sample sizes when one efficacy outcome and one correlated safety outcome are monitored simultaneously using sequential designs.

Study Design and Setting: We conducted a simulation study in the framework of a two-arm parallel clinical trial. Interim analyses on two outcomes were performed independently and simultaneously on the same data sets using four sequential monitoring designs, including O'Brien-Fleming and Triangular Test boundaries. Simulations differed in values for correlations and true effect sizes.

Results: When an effect was present in both outcomes, competition was introduced, which decreased power (e.g., from 80% to 60%). Futility boundaries for the efficacy outcome reduced overall type I errors as well as power for the safety outcome.

Conclusion: Monitoring two correlated outcomes, given that both are essential for early trial termination, leads to masking of true effects. Careful consideration of scenarios must be taken into account when designing sequential trials. Simulation results can help guide trial design.

Keywords: Randomized controlled trial; Simulation; Sequential designs; Multiple outcomes; O'Brien-Fleming; Triangular Test

1. Introduction

Randomized controlled trials (RCTs) are widely recognized as the highest level of evidence in medical studies and the gold standard in causal therapeutic research [1]. A sequential design can be used to maximize trial efficiency [2]. Sequential designs use interim analyses to examine data that have been collected so far. These designs control type I error, which is known to be inflated because of multiple testing [3,4]. If an interim analysis shows a significant effect, both ethical reasoning and economic reasoning suggest that the trial should be stopped before recruitment completion because the research question has been answered [5].

Sequential designs are usually applied on the primary (efficacy) outcome and implicitly assume that only this outcome is being formally monitored at the time. In practice, this will be highly unlikely: aside from secondary outcomes, monitoring at least one safety end point is frequently desirable in phase III clinical trials [1,6–9]. Efficacy monitoring can be closely related to safety monitoring as these outcomes are often correlated (e.g., cardiovascular events and cardiovascular mortality). When unforeseen risks are observed, an ongoing phase III trial could be adapted or even terminated because safeguarding of participants is of utmost importance.

For this purpose, independent Data and Safety Monitoring Boards (DSMBs) examine results from interim analyses and give recommendations concerning the trial [5,10]. If interim analyses show a similar effect of treatment on both efficacy and safety, be it positive or negative, the decision to stop will be straightforward. However, in case of opposite effects, DSMBs need to be decisive between continuing the trial to accumulate more information on treatment efficacy, or choose for the safety of the participants when results show trends toward harmful effects [5,10–16].

Conflicts of interest: None.

* Corresponding author. Tel.: +31 20 566 4127.
E-mail address: r.vaneekelen@umcutrecht.nl (R. van Eekelen).

http://dx.doi.org/10.1016/j.jclinepi.2016.02.024
0895-4356/

What is new?

- Sequential designs are based and applied on the primary (efficacy) outcome only. In practice, multiple outcomes (e.g., safety) will be monitored in a trial alongside this primary outcome. It is unclear how this multiplicity issue, conducting multiple analyses on multiple correlated outcomes, influences decision making when stopping a trial early.

Key findings

- Our simulation study indicated that when an effect was present in both outcomes, competition was introduced, which decreased power.

What this adds to what was known?

- Early stopping of a sequential trial lowers the amount of information available compared to what was planned, which lowers the probability of researchers finding a beneficial effect or safety issue for multiple outcomes. This article depicts that this phenomenon is more likely to occur than we thought.

What is the implication and what should change now?

- RCTs stopped early have a realistic chance of missing effects. Careful consideration of prespecified scenarios must be taken into account when designing sequential trials. Our simulation results and discussion can help guide trial design.

To monitor at least two important outcomes in a trial, using at least two sequential designs simultaneously might be the most intuitive. Clinical trials with such methodology have been reported in literature but remain few in numbers, probably because of multiplicity issues regarding multiple looks at multiple outcomes [9,14,17−22]. Currently, clinical trials typically monitor secondary outcomes in a descriptive way only, without formal testing, and conduct standard statistical tests after trial completion [6,7,12,23]. However, formal statistical inference on secondary outcomes is commonly desirable, especially when descriptive statistics suggest large differences between treatment groups [10]. Several authors have constructed statistical tests for bivariate sequential analyses or extended existing tests to a sequential setting [11,14,24,25]. Unfortunately, bivariate sequential methods have not been implemented in software, as far as we know, or their potential researched thoroughly. Sometimes, the primary outcome represents both efficacy (in case of superiority) and safety (in case of inferiority), for example overall mortality, in which case a two-sided test will suffice. Furthermore, overall tests such as Hotelling's T^2 and the Wald test can be used, possibly on composite outcomes [6,8,19,26]. It is important to note that all mentioned methods aside from bivariate analyses, including viewing descriptive statistics only, can potentially influence decision making of the DSMB and can therefore be considered an informal way of monitoring multiple outcomes simultaneously. Moreover, a recent systematic review comparing published trial papers with their protocols indicates that 80% of the reviewed trials that stopped early did so without specifying interim analyses in their protocol [27]. Note that to prevent any data-driven conclusion, Good Clinical (and Statistical) Practice requires to prespecify interim analyses in the protocol. Another article summarizes monitoring as this frequently occurs in practice and suggests that simultaneous monitoring of efficacy and safety with statistical inference is generally accepted and simply considered a part of modern clinical trials [15].

Although a sequential design is based on the primary (efficacy) outcome, it is apparent that the decision to terminate the study early might be based on many other outcomes or aspects in the trial and thus, in practice, multivariate monitoring takes place. Previous work describes the possible influence of a secondary outcome on trial decision making [26]. However, it was assumed that both outcomes represent efficacy and are uncorrelated. Furthermore, the trial would not be stopped early unless both variables reached a conclusion and only overall power was reported. In this article, we are interested in exploring in more detail what the implications are of sequentially monitoring one efficacy outcome and one possibly correlated safety outcome simultaneously on trial decision making. Therefore, we performed a simulation study to quantify the multiplicity issue in terms of type I error, power, and required sample sizes.

1.1. Motivating example

We will use the Long-Term Intervention on Fractures with Tibolone (LIFT) trial study as an illustration [12]. This study's main goal was to research the effect of the synthetic drug tibolone on the incidence of vertebral fractures in postmenopausal women. Stroke and cancer incidence were measured among others as safety outcomes. Interim analyses showed a nonsignificant decrease in vertebral fractures and an increase in stroke events in the active group (tibolone) compared to the control group (vitamin D): an example of the situation of opposite effects.

What are likely implications on decision making in the LIFT trial when statistical tests would be performed on stroke and vertebral fractures simultaneously? This principle was used as the foundation for our research aim.

2. Methods

In the following, we describe the framework for the simulation study including sequential monitoring, designs,