Journal of Clinical Epidemiology 68 (2015) 1144–1151

Journal of Clinical Epidemiology

The choice of the noninferiority margin in clinical trials was driven by baseline risk, type of primary outcome, and benefits of new treatment

Angèle Gayet-Ageron[a,*], Thomas Agoritsas[a,b], Sandrine Rudaz[a], Delphine Courvoisier[a], Thomas Perneger[a]

[a]*Division of Clinical Epidemiology, University Hospitals of Geneva and Faculty of Medicine, 6 Rue Gabrielle Perret-Gentil, 1211 Geneva 14, Switzerland*
[b]*Department of Clinical Epidemiology & Biostatistics, Faculty of Health Science, McMaster University, 1280 Main St. W. Hamilton, Ontario, Canada L8S 4K1*

Accepted 21 January 2015; Published online 28 January 2015

Abstract

Objectives: To explore characteristics of clinical trials that influence the choice of the noninferiority margin (NIM) when planning the trial.

Study Design and Setting: We conducted an experimental survey among corresponding authors of randomized controlled trials indexed in MEDLINE. We described two hypothetical studies and asked the respondents' opinion on the largest loss of effectiveness that is clinically negligible (or the smallest lost of effectiveness that is clinically important in the superiority scenario). We randomly manipulated four study attributes in each vignette, using a factorial design.

Results: A total of 364 researchers participated. The values for NIMs were significantly lower than the differences to be detected in a superiority trial. The NIM was smaller when the primary outcome was mortality compared with treatment failure, when baseline risk in the control arm was lower, and when the advantage of the new treatment was a lower cost compared with having fewer side effects. In contrast, the population age group under study and the difficulty to recruit patients showed no effect on the choice of the NIM.

Conclusion: In our experimental study, the factors associated with lower NIMs were mortality as a primary outcome, low baseline risk, and a less costly new treatment.

Keywords: Noninferiority margin; Noninferiority study design; Clinical trial; Sample size; Randomized controlled trials; Primary outcome; Minimal clinically important difference

1. Introduction

Noninferiority randomized trials are an increasingly popular study design, especially in the field of oncology, infectious diseases, or cardiovascular diseases [1,2]. The main purpose of such trials is to demonstrate that a new treatment is not substantially less effective than an existing treatment, while providing an additional advantage (in terms of convenience, burden of treatment, side effects, cost, and so forth.) [3]. A key challenge for such studies is to define what "not substantially less effective" means. This translates as the noninferiority margin (NIM): the largest loss of effectiveness that is clinically negligible to establish noninferiority [4].

A recent review article has provided some key points to consider when reading and interpreting the results of noninferiority trials [5]. Nevertheless, there is no formal consensus on how to determine the NIM when planning or interpreting a clinical trial. Some experts would base the NIM on the benefit of the standard treatment vs. placebo or even on the lower confidence limit of the benefit's estimate. But this reasoning can lead to conclusions that are difficult to justify; for example, the more effective the standard treatment, the larger the loss of effectiveness that would be accepted in the noninferiority trials [6]. Others might select the NIM depending on the severity of the primary endpoint, with smaller NIM when mortality or serious adverse events are used [7]. The Food and Drug Administration (FDA) has proposed a loss of effectiveness of 10% (in absolute terms) as compatible with noninferiority for

Conflict of interest: None.

Funding: Research and Development Fund of the University of Geneva Hospitals (8-2011-I).

* Corresponding author. Tel.: +41-2-2372-9027; fax: +41-2-2372-9035.

E-mail address: angele.gayet-ageron@hcuge.ch (A. Gayet-Ageron).

http://dx.doi.org/10.1016/j.jclinepi.2015.01.017
0895-4356/

What is new?

Key findings

- When they need to determine a noninferiority margin (NIM), researchers take into consideration the magnitude of the baseline risk with standard treatment, the type of primary outcome studied, and the additional advantages provided by the new treatment. Their reasoning was neither influenced by the existence of logistical constraints nor by the population age group.

What this adds to what was known?

- There is no consensus on how to determine the NIM when planning or interpreting a clinical trial, except one statistical rule from the Food and Drug Administration, which lacks generalizability and is not informed by empirical evidence. Here, we experimentally verified that some predefined features of a trial influenced the choice of an NIM.

What is the implication and what should change now?

- The report of the results from noninferiority trials should include a justification of the choice for a specific NIM. Particularly, the factors considered in the determination of the largest loss of effectiveness that is clinically negligible need to be clearly explained.

anti-infectious or antiretroviral therapies [8,9]. However, such a simple rule is not easily applicable to all situations and is not related to the advantages of the new treatment. More recently, the FDA and the European Medicines Agency have proposed to select the value that preserves at least 50% of the treatment effect of the standard compared with the placebo (or the previous standard), but these rules are not applicable to all situations and could lead to unreasonably wide NIMs [10,11]. In many published reports of noninferiority trials, the NIM is stated but not justified [12]. A previous study found that NIMs vary between medical specialties [13], but the reasons for this phenomenon are unclear. As a result, little is known about the reasoning that researchers use in selecting the NIM. Furthermore, whether researcher-selected NIMs reflect patients' priorities is unclear [14,15].

Several arguments can be used to justify the NIM [16,17]. One candidate factor is the baseline risk in the control arm of the trial. Investigators or clinicians may feel that a given loss of effectiveness—say, 5% increase in mortality—is more acceptable if the baseline risk is high rather than low (whether this is justified is another issue). Another candidate factor that may influence the NIM is the type of outcome: a 5% increase in mortality may be less acceptable than a 5% increase in a less serious event, such as heart failure exacerbation. Indeed, we showed in a previous exploratory study that the NIM was significantly lower when mortality was the primary outcome [16]. Furthermore, if the treatment reduces mortality, the life expectancy of the patient also plays an important role, as one may be less inclined to accept an increase in mortality if the patient has many years of life left. Because both mortality and average life expectancy may vary by medical specialty, customary NIMs may well vary as well across specialties [13]. Another candidate factor that may affect the NIM is the type of advantage postulated with the new treatment (e.g., fewer side effects vs. lower cost). Finally, investigators may be willing to accept a larger loss of efficacy, which implies a smaller sample size for the trial, if they anticipate recruitment difficulties, although this factor is hardly legitimate from the perspectives of clinical decision making.

In this study, we explored the researchers' reasoning in selecting the NIM through an experimental survey among a self-selected sample of corresponding authors who have published the results of a randomized controlled trial between 2010 and 2012. Our aim was to assess the association between predefined factors and the NIM by exposing trialists to the clinical vignettes in which we randomly manipulated four study attributes in a factorial design.

2. Methods

2.1. Study design, participants, and eligibility criteria

We conducted a cross-sectional study among a convenience sample of corresponding authors who have published the results of a randomized controlled trial recorded in MEDLINE between January 1, 2010, and December 31, 2010. Because of the very low response rate (<5%) at the start of the survey, we extended the search to abstracts published in 2011 and 2012. We identified randomized controlled trials using the search query "randomized controlled trial" OR ("randomized" AND "controlled" AND "trial") and retrieved the corresponding author's email address when available. Exclusion criteria were ancillary analyses of previously published studies, review articles, or nonhuman research. Because it carried minimal risk, the project was exempted from formal review by the institutional research ethics committee.

2.2. Electronic questionnaire and random attribution of the version

We created an electronic survey using Limesurvey (LimeSurvey Project, Hamburg, Germany). We pretested the questionnaire among eight volunteers (both clinicians and researchers) to assess validity and clarity of each clinical scenario and also to choose the best modality of answers regarding the vignette. The questionnaire was divided into three sections: the first section explored the respondent's