JOURNAL OF THE AMERICAN COLLEGE OF CARDIOLOGY
VOL. 66, NO. 25, 2015

ISSN 0735-1097/$36.00
PUBLISHED BY ELSEVIER INC.
http://dx.doi.org/10.1016/j.jacc.2015.10.051

THE PRESENT AND FUTURE

STATE-OF-THE-ART REVIEW

Challenging Issues in Clinical Trial Design

Part 4 of a 4-Part Series on Statistics for Clinical Trials

Stuart J. Pocock, PhD,* Tim C. Clayton, MSc,* Gregg W. Stone, MD†

ABSTRACT

As a sequel to last week's paper on the fundamentals of clinical trial design, this paper tackles related controversial issues: noninferiority trials, the value of factorial designs, the importance and challenges of strategy trials, Data Monitoring Committees (including when to stop a trial early), and the role of adaptive designs. All topics are illustrated by relevant examples from cardiology trials. (J Am Coll Cardiol 2015;66:2886–98) © 2015 by the American College of Cardiology Foundation.

Randomized controlled trials are the cornerstone of clinical guidelines informing best therapeutic practices; however, their design and interpretation may be complex and nuanced. This review explores challenging issues that may arise and builds on the fundamentals of trial design covered in last week's paper.

Specifically, we offer guidance on how to design and interpret noninferiority trials where the goal is to demonstrate that the efficacy of a new treatment is as good as that achieved with a standard treatment.

Factorial trials, where 2 (or more) therapeutic issues are simultaneously evaluated in the same study, present an interesting opportunity that should be considered more often in cardiology research.

Trials that compare substantially different alternative treatment strategies can be of great value in enhancing good patient management, and we present guidance on the topic to stimulate greater interest in overcoming the difficulties in undertaking such pragmatic studies.

All major cardiology trials have both ethical and practical needs for data monitoring of the accumulating evidence over time. We provide insights into how Data Monitoring Committees (DMCs) should function, offering statistical guidelines and practical decision-making considerations as to when to stop a trial early.

Finally, there is a growing interest in adaptive designs, but few instances of their implementation in cardiology trials. We focus on adaptive sample size re-estimation and enrichment strategies, with guidance on when and how they may be used.

All of these issues are illustrated by experiences from actual cardiology trials, demonstrating the real-world implications of trial design decisions.

NONINFERIORITY TRIALS

Increasingly, major trials are conducted to see if the efficacy of a new treatment is as good as a standard treatment (1–3). The new treatment usually has some other advantage (e.g., fewer side effects, ease of administration, lower cost), making it worthwhile to demonstrate noninferiority in respect to efficacy.

The standard approach to designing a noninferiority trial is to pre-define a noninferiority margin, commonly called delta, for the primary endpoint. This is the smallest treatment difference, which, if true, would mean that the new treatment is declared inferior. This is on the basis of the belief that any difference smaller than this would constitute clinically accepted grounds of "therapeutic interchangeability" (4). The trial's conclusions then depend on where the 95% confidence interval (CI) for

Listen to this manuscript's audio summary by *JACC* Editor-in-Chief Dr. Valentin Fuster.

From the *Department of Medical Statistics, London School of Hygiene and Tropical Medicine, London, United Kingdom; and the †Columbia University Medical Center, New York-Presbyterian Hospital, and the Cardiovascular Research Foundation, New York, New York. The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

Manuscript received September 24, 2015; revised manuscript received October 25, 2015, accepted October 25, 2015.

the treatment difference ends up in relation to this margin. If the upper bound of the 2-sided 95% CI is less than delta, one can claim evidence that the new treatment is noninferior.

For instance, the ACUITY (Acute Catheterization and Urgent Intervention Triage Strategy) trial compared bivalirudin with the standard treatment of heparin plus a glycoprotein IIb/IIIa inhibitor in patients with acute coronary syndrome (ACS) for 30-day composite ischemia (death, myocardial infarction [MI], or revascularization) (5). The noninferiority margin was set at a relative risk of 1.25. The trial's findings revealed composite ischemia rates of 7.8% and 7.3% in the bivalirudin and control groups, respectively, with relative risk: 1.08; 95% CI: 0.93 to 1.24. Because the upper bound of the CI of 1.24 was less than the pre-declared delta of 1.25, one can conclude that there is evidence of noninferiority. The reason this matters is that bivalirudin also had a markedly lower risk of major bleeding, an important consideration when choosing between antithrombin therapies.

A common misunderstanding is that lack of a statistically significant difference between 2 therapies implies that they are equivalent. For instance, the INSIGHT (Intervention as a Goal in Hypertension Treatment) trial compared nifedipine with co-amilozide in hypertension. The authors concluded that the treatments were "equally effective in preventing cardiovascular complications," on the basis of a p value of 0.35 for the primary composite endpoint of cardiovascular (CV) death, MI, heart failure, or stroke (6). But, the observed relative risk of 1.10 had a 95% CI of 0.91 to 1.34. This includes up to a 34% excess risk on nifedipine, making it unwise to conclude that nifedipine is as good as (i.e., noninferior to) co-amilozide.

Figure 1 shows a conceptual plot of how to interpret the results of noninferiority trials. Scenario C (noninferior) indicates what happened in the ACUITY trial. If we suppose that the INSIGHT trial had the same delta, 1.25, then it would have fallen under scenario F (inconclusive). Had more patients been enrolled, the 95% CI would have narrowed, and noninferiority might then have been declared.

Sometimes, the treatment effect (and its delta) is expressed as a difference in percentages, rather than as a relative risk or hazard ratio (the argument being that absolute differences are more clinically relevant than relative risks). For instance, the OPTIMIZE (Optimized Duration of Clopidogrel Therapy Following Treatment With the Zotarolimus-Eluting Stent in Real-World Clinical Practice) trial compared a 3-month versus a 12-month duration of dual antiplatelet therapy after implantation of a zotarolimus-eluting stent (7). For the composite primary endpoint of net adverse clinical events (death, MI, stroke, or major bleed) at 1 year, a 2.7% difference was set as the noninferiority margin. The observed difference was +0.2%, with a 95% CI of −1.5% to +1.9%. Because this excludes the margin of +2.7%, noninferiority of the 3-month duration of treatment was claimed.

This example raises a few issues. When the noninferiority margin is a difference in percentages, it becomes easier (perhaps too easy) to achieve noninferiority if the overall event rate is lower than expected. The OPTIMIZE trial had an anticipated 9% event rate in the control arm, but the observed event rate was 6%. This made the 2.7% margin equivalent to a relative risk margin of 1.45, which is undesirably large. Conversely, if the overall event rate is greater than expected, it may become unreasonably difficult to achieve noninferiority. The opposite considerations of anticipated versus observed event rates apply if a relative risk is chosen for the margin.

Also, the endpoint chosen in the OPTIMIZE trial was not of optimal relevance. The true issue in considering a shorter period of dual antiplatelet treatment concerns the balance between the increased risks of stent thrombosis and MI against the reduced risk of major bleeding. To force these diverse endpoints into a single composite would bias results toward the null. A preferable approach is to prespecify and study separately-powered efficacy and safety endpoints, typically 1 for superiority and 1 for noninferiority. However, a very large sample size may be required to adequately power both the efficacy and safety endpoints.

A composite net adverse clinical events endpoint, consisting of combined safety and efficacy endpoints, has been used in some trials, reflecting the recognition that both types of endpoints (e.g., major bleeding and stent thrombosis) are deleterious and strongly associated with subsequent mortality. However, interpretation of such a combined safety and efficacy endpoint may be challenging, especially if the different components do not have similar effects on patients' well-being or survival. Moreover, because safety and efficacy endpoints often move in different directions (e.g., in response to more potent antithrombotic therapies), their combination in a composite endpoint may mask differences between therapies, making careful examination of each component measure essential.

A key question is the choice of noninferiority margin, which has implications for the required trial size. Power calculations for noninferiority trials (not

ABBREVIATIONS AND ACRONYMS

ACS = acute coronary syndrome

CI = confidence interval

CV = cardiovascular

DMC = Data Monitoring Committee

MI = myocardial infarction

OMT = optimal medical therapy

PCI = percutaneous coronary intervention