Understanding non-inferiority trials: an introduction

Penelope M. A. Brasher, PhD · Gary Dobson, MD

Received: 14 February 2014 / Accepted: 17 February 2014 / Published online: 8 March 2014
© Canadian Anesthesiologists' Society 2014

Non-inferiority studies are appearing in medical journals with increasing frequency. At first glance, one might think the intent of such a study is to publish negative results or, in the worst case scenario, to create a marketing ploy that would have us consider replacing a tried-and-true treatment with a new one.¹ Nevertheless, these trials can bring value by introducing new treatments that may be therapeutically similar to the standard treatment but with certain advantages. In this issue of the *Journal*, Merry *et al.*² show that clevidipine, a short-acting dihydropyridine calcium channel antagonist, is non-inferior to nitroglycerin (NTG) in the pre-bypass period of patients undergoing cardiac surgery. The information needed for a clinician to assess a non-inferiority trial is highlighted in the context of this study.

The non-inferiority trial attempts to show that the new treatment is not an unacceptably worse alternative to the standard.³ This is done by stating a margin of non-inferiority for the effect of the treatment. The non-inferiority margin describes the amount of reduction in efficacy that will be tolerated. As long as the new treatment is not worse than the standard treatment by this margin, the new treatment would be considered non-inferior.

What is the rationale for performing a non-inferiority trial? In the regulatory setting, the efficacy of a new treatment is usually shown in a placebo-controlled trial. In

some situations, however, it would be unethical to withhold treatment; hence, a non-inferiority trial is used to indirectly show the efficacy of the new treatment. Alternatively, a trial sponsor may have a new treatment that provides only a small advantage over the standard. Showing this superiority would require a very large study, while non-inferiority could be shown with a much smaller sample. In the non-regulatory setting, non-inferiority trials are most often used to show that an alternative treatment with advantages regarding safety, convenience, or cost has an efficacy similar to, or at least not much worse than, the standard treatment.

The first order of business is to decide if a non-inferiority design is reasonable. The reader must question whether the purported advantages of the new treatment over the standard would justify adopting the new treatment should non-inferiority be shown. Typical advantages are price, ease of use, or a reduced number of side effects. If the answer is “no”, then the non-inferiority design should be abandoned. If the answer is “yes”, then the reader will need to consider the next issue, i.e., the assay sensitivity and the consistency of the control treatment. In a non-inferiority trial, two comparisons are made: the explicit comparison of the new treatment with an active control treatment and the implicit comparison of the active control treatment with placebo. Generally, there is no placebo control in these studies so the reader must be convinced that the treatment (active control) against which the new treatment is being compared would provide effective management (i.e., be superior to a placebo) in the context in which it is being used. This is referred to as the assay sensitivity.⁴ If the efficacy of the standard is border-line, then showing that the new treatment is non-inferior is of little value as it may in fact be no different from a placebo. It is also important that previous trials using the standard

P. M. A. Brasher, PhD
Centre for Clinical Epidemiology and Evaluation, Vancouver
Coastal Health Research Institute, Vancouver, BC, Canada

G. Dobson, MD (✉)
Department of Anesthesiology, Peter Lougheed Centre,
University of Calgary, Room 2267, 3500 26th Ave NE, Calgary,
AB T1Y 6J4, Canada
e-mail: gary.dobson@AlbertaHealthServices.ca

Figure The new treatment may be found to be: superior (A), non-inferior but not superior (B,C), non-inferiority not shown (D,F), inferior and non-inferior (E), and both inferior and not non-inferior (G)

Figure Le nouveau traitement peut être jugé: supérieur (A), non-inférieur mais pas supérieur (B, C), non-infériorité non établie (D, F), inférieur et non-inférieur (E), et à la fois inférieur et pas non-inférieur (G)

treatment have shown consistent effects in the patient population under study. The evidence supporting the size and consistency of the effect of the standard treatment vs placebo should be described, and a forest plot of published studies is ideal.

The reader also needs to evaluate the margin of non-inferiority selected by the investigators. In the regulatory setting, two possible approaches are suggested: the fixed margin method and the synthesis method. Each method offers a different approach to address the variability in the estimate of the effect of the active control by incorporating data from historical studies in the non-inferiority study. The interested reader is referred to the paper by Schumi & Wittes¹ and the respective references. In the non-regulatory setting, less formal methods are used to specify the non-inferiority margin. A common approach is to survey a panel of clinicians and ask them to state the decrease in efficacy they would find acceptable given the potential benefits of the new treatment.

After study completion, a confidence interval (usually a 95% confidence interval) is constructed to contrast the two treatments. In the Figure, we show the possible outcomes for a non-inferiority trial where the ratio of the treatment effects summarizes the between group contrast in efficacy. In this example, a ratio < 1 indicates an advantage with the new treatment, and an arbitrary non-inferiority margin of

~1.8 was selected. The new treatment may be found to be: superior (A), non-inferior but not superior (B,C), non-inferiority not shown (D,F), inferior and non-inferior (E), and both inferior and not non-inferior (G).

The results from inferiority trials should include both intention-to-treat (ITT) and per protocol (PP) analyses.⁵ This need is most readily explained by imagining a very poor study in which both the active control and the new treatment were distributed evenly between the study groups. With superiority trials, there would probably be no demonstrable difference and the superiority of the new treatment would be rejected. With a non-inferiority trial, the absence of difference could be interpreted as the new treatment being non-inferior. Non-inferiority must be shown with both ITT and PP analysis.

Returning to the publication in the *Journal*, NTG has long been used to manage blood pressure in the perioperative period, and the authors provide references supporting the use of NTG as an active control. Nevertheless, the original studies predate the references, and as a further decade has passed since the study was completed, the possibility should be considered that changes in perioperative management may have altered the effect of NTG relative to placebo. The potential advantages of clevidipine over NTG include its mechanism of action and metabolism. Clevidipine selectively reduces afterload, has a half-life of approximately one minute with no active metabolites, does not demonstrate tachyphylaxis, and is not associated with rebound hypertension or tachycardia when discontinued.⁶ The sample size calculation was provided and required a minimum of 50 patients per group to meet their defined limits.

The authors' definition of what is unacceptably worse is of central importance, and their choice and rationale was clearly articulated. The primary outcome was the variability in mean arterial pressure (MAP) during the pre-bypass period. The area-time integral of all differences in MAP greater than 5% from the predefined ideal was measured and normalized by hour (mmHg·min·hr⁻¹). The geometric means for the clevidipine and TNG groups were then calculated, and their ratio with 95% confidence intervals was determined. An upper boundary of > 1.5 for this ratio was defined as an unacceptable (not non-inferior) variation in MAP. The rationale for this upper limit was its clinical relevance and the inherent difficulties in managing a vasodilator in a blinded study. Taking into account the stringent limits on MAP that were being imposed, the upper boundary of 1.5 seems reasonable. The reported ratio was 0.97 with an upper boundary of 1.27, and so clevidipine was found to be non-inferior. The ITT and PP analysis were similar and both showed non-inferiority. These were available for the reviewers but were not included in the final draft.

As described by the authors, both the outcome and the perioperative period being reported are different from previous publications. Furthermore, as the focus of this cardiac surgery study was on the pre-bypass period, the results should be relevant to all anesthesiologists. This manuscript contributes to our understanding of the pharmacodynamics of clevidipine and illustrates how non-inferiority trials can be of value.

Comprendre les études de non-infériorité: une introduction

Des études de non-infériorité sont publiées de plus en plus souvent dans les journaux médicaux. On pourrait penser, à première vue, qu'il s'agit d'un moyen pour faire publier des résultats négatifs ou, dans le pire des cas, d'un complot marketing pour nous faire envisager le remplacement d'un traitement testé et confirmé par un nouveau traitement.¹ Ces études peuvent néanmoins être intéressantes en nous présentant de nouveaux traitements qui pourraient être comparables sur le plan thérapeutique, mais offrir des bénéfices par rapport au traitement de référence. Dans ce numéro du *Journal*, Merry *et coll.*² démontrent que la clévidipine, un antagoniste des canaux calcium à courte durée d'action de la famille de la dihydropyridine, n'est pas inférieure à la trinitrine (NTG - nitroglycérine) dans la période précédant un pontage chez les patients devant subir une chirurgie cardiaque. L'information dont le clinicien a besoin pour évaluer une étude de non-infériorité est soulignée dans le contexte de cette étude.

L'étude de non-infériorité essaie de démontrer que le traitement n'est pas une option plus mauvaise, au point d'être inacceptable, par rapport au traitement de référence.³ Cela se fait en définissant une marge de non-infériorité pour l'effet du traitement. La marge de non-infériorité décrit quelle proportion de baisse de l'efficacité sera tolérée. Aussi longtemps que le nouveau traitement n'est pas moins bon que le traitement de référence au-delà de ce seuil, nous considérerions le nouveau traitement comme non-inférieur.

Quelle est la justification de la réalisation d'une étude de non-infériorité? Dans le cadre réglementaire, l'efficacité d'un nouveau traitement est habituellement démontrée par une étude contrôlée contre placebo. Dans certains cas, cependant, il ne serait pas éthique de délaissier un traitement; l'étude de non-infériorité sert à démontrer de façon indirecte l'efficacité du nouveau traitement. Le commanditaire d'une étude peut également avoir un nouveau traitement qui n'apporte qu'un modeste avantage

par rapport au traitement de référence. La démonstration de cette supériorité nécessiterait une étude de très grande ampleur, tandis que la non-infériorité pourrait être démontrée avec un échantillon beaucoup plus petit. En dehors du cadre réglementaire, les études de non-infériorité sont le plus souvent utilisées pour démontrer qu'un autre traitement présente des avantages en termes d'innocuité, de commodité ou de coût, et qu'il a une efficacité similaire ou – au minimum – pas très inférieure au traitement de référence.

Le premier point à analyser est de décider si le plan de non-infériorité est raisonnable. Le lecteur doit se poser la question suivante: « les avantages supposés du nouveau traitement par rapport au traitement de référence justifient-ils d'adopter ce nouveau traitement dans le cas où la non-infériorité serait démontrée? » Les avantages habituels sont le prix du traitement, sa facilité d'utilisation et la réduction du nombre d'effets indésirables. Si la réponse est « non », le plan de non-infériorité doit être abandonné. Si la réponse est « oui », le lecteur devra alors considérer le problème suivant, à savoir la sensibilité de l'étude et la pertinence du traitement de référence. Deux comparaisons sont effectuées dans une étude de non-infériorité: la comparaison explicite du nouveau traitement par rapport à un traitement actif de contrôle et la comparaison implicite du traitement actif de contrôle avec un placebo. D'habitude, il n'y a pas de contrôle par placebo dans ces études si bien que le lecteur doit s'assurer que le traitement (le contrôle actif) avec lequel le nouveau traitement est comparé s'avérera efficace (c'est-à-dire sera supérieur au placebo) dans le contexte dans lequel il est utilisé. Cela s'appelle la sensibilité de l'étude.⁴ Si l'efficacité du traitement de référence est limitée, démontrer que le nouveau traitement n'est pas inférieur n'a que peu de valeur car il pourrait ne pas être différent d'un placebo. Il est également important que les études précédentes aient démontré les effets constants du traitement de référence dans la population étudiée. Les données probantes soutenant l'ampleur et la constance de l'effet du traitement de référence contre le placebo doivent être décrites et un graphique en forêt (Forest Plot) des études publiées est un élément idéal.

Le lecteur devra aussi évaluer la marge de non-infériorité choisie par les investigateurs. Deux approches possibles sont proposées dans le cadre réglementaire: la méthode de la marge fixe et la méthode de synthèse. Ces méthodes abordent la variabilité de l'estimation de l'effet du contrôle actif en incorporant des données d'études historiques dans l'étude de non-infériorité, mais de façons différentes. Le lecteur intéressé est renvoyé à l'article de Schumi et Wittes¹ et à ses références. Dans un cadre non réglementaire, des méthodes moins formelles sont souvent utilisées pour la

définition de la marge de non-infériorité. Une approche courante consiste à interroger un panel de cliniciens en leur demandant de définir quelle baisse d'efficacité ils considéreraient acceptable, considérant les avantages potentiels du nouveau traitement.

Après l'achèvement de l'étude, un intervalle de confiance (habituellement un intervalle de confiance à 95 %) est construit pour afficher le contraste entre les traitements. Nous montrons dans la figure les résultats possibles d'une étude de non-infériorité où le ratio des effets thérapeutiques résume le contraste intergroupe pour l'efficacité. Dans cet exemple, un ratio < 1 indique un avantage au profit du nouveau traitement et une marge arbitraire de non-infériorité proche de 1,8 a été sélectionnée. Le nouveau traitement pourrait être déclaré: supérieur (A), non-inférieur mais pas supérieur (B, C), non-infériorité non établie (D, F), inférieur et non-inférieur (E), et à la fois inférieur et pas non-inférieur (G).

Les résultats des études d'infériorité doivent inclure des analyses de la population en intention de traiter (ITT) en même temps que des analyses per-protocole (PP).⁵ Cette nécessité est le plus facilement expliquée en imaginant une très mauvaise étude dans laquelle le contrôle actif et le nouveau traitement seraient répartis de façon homogène entre les groupes d'études. Avec des études de supériorité, aucune différence ne pourrait probablement être démontrée et la supériorité du nouveau traitement serait rejetée. Dans une étude de non-infériorité, l'absence de différence pourrait être interprétée comme montrant que le nouveau traitement n'est pas inférieur. La non-infériorité doit être démontrée à la fois dans l'analyse ITT et dans l'analyse PP.

Pour revenir à la publication du *Journal*, la NTG est utilisée depuis longtemps pour gérer la pression artérielle en période périopératoire et les auteurs fournissent des références en faveur de l'emploi de la NTG comme contrôle actif. Cependant, les études originales sont antérieures aux références et, dans la mesure où une décennie s'est écoulée depuis la fin de l'étude, il est envisageable que des modifications de la prise en charge périopératoire aient modifié l'effet de la NTG par rapport au placebo. Les avantages potentiels de la clévidipine par rapport à la NTG incluent son mode d'action et son métabolisme. La clévidipine abaisse de façon sélective la postcharge, a une demi-vie d'environ une minute, sans métabolites actifs, ne montre pas de tachyphylaxie, et n'est pas associée à une hypertension rebond ou à une tachycardie après son arrêt.⁶ Le calcul de la taille de l'échantillon a été fourni: il fallait un minimum de 50 patients dans chaque groupe pour satisfaire leurs limites définies.

La définition des auteurs de ce qui est pire et inacceptable a une importance clé et leur choix ainsi que leurs justifications sont exprimés avec une grande cohérence. Le

critère d'évaluation était la variabilité de la pression artérielle moyenne (PAM) au cours de la période précédant le pontage. L'intégrale surface-temps de toutes les différences de PAM supérieures à 5 % par rapport à la valeur idéale prédéfinie a été mesurée et normalisée heure par heure ($\text{mmHg}\cdot\text{min}\cdot\text{h}^{-1}$). Les moyennes géométriques pour les groupes clévidipine et TNG ont alors été calculées et leur rapport avec des intervalles de confiance à 95 % a été déterminé. Une limite supérieure de plus de 1,5 pour ce rapport a été définie comme une variation inacceptable (pas non-inférieur) de la PAM. La justification de cette limite supérieure tenait à sa pertinence clinique et aux difficultés inhérentes à la gestion d'un vasodilatateur dans une étude en aveugle. En prenant en compte les limites strictes qui étaient imposées pour la PAM, une limite supérieure de 1,5 paraît raisonnable. Le ratio rapporté a été de 0,97 avec une limite supérieure de 1,27: en conséquence, la clévidipine a été jugée non-inférieure. Les analyses en ITT et PP ont été similaires et les deux ont fait la démonstration de la non-infériorité. Ces données ont été mises à la disposition des réviseurs mais n'ont pas été incluses dans la version finale de l'article.

Comme l'indiquent les auteurs, le critère d'évaluation et la période périopératoire décrite sont différents de ceux des publications antérieures. En outre, considérant que cette étude de chirurgie cardiaque était focalisée sur la période précédant le pontage, les résultats de l'étude sont pertinents pour tous les anesthésiologistes. Ce manuscrit contribue à notre compréhension de la pharmacodynamie de la clévidipine et illustre comment les études de non-infériorité peuvent avoir de l'intérêt.

Conflicts of interest None declared.

Conflits d'intérêts Aucun déclaré.

References

1. Schumi J, Wittes JT. Through the looking glass: understanding non-inferiority. *Trials* 2011; 12: 106.
2. Merry AF, Avery EG, Nussmeier NA, et al. Clevidipine compared with nitroglycerin for blood pressure control in coronary artery bypass grafting: a randomized double-blind study. *Can J Anesth* 2014; 61: this issue. DOI:10.1007/s12630-014-0131-z.
3. Fleming TR, Odeh-Davis K, Rothmann MD, Li Shen Y. Some essential considerations in the design and conduct of non-inferiority trials. *Clin Trials* 2011; 8: 432-9.
4. Temple R, Ellenberg SS. Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 1: ethical and scientific issues. *Ann Intern Med* 2000; 133: 455-63.
5. Laster LL, Johnson MF. Non-inferiority trials: the 'at least as good as' criterion. *Stat Med* 2003; 22: 187-200.
6. Espina IM, Varon J. Clevidipine: a state-of-the-art antihypertensive drug under the scope. *Expert Opin Pharmacother* 2012; 13: 387-93.