Users' guide to the surgical literature: how to assess a noninferiority trial

Achilleas Thoma, MD, MSc
Forough Farrokhyar, MPhil, PhD
Daniel Waltho, MD
Luis H. Braga, MD, PhD
Sheila Sprague, PhD
Charlie H. Goldsmith, PhD

Accepted July 5, 2017

Correspondence to:

A. Thoma
206 James St South
Hamilton ON L8P 3A9
athoma@mcmaster.ca

DOI: 10.1503/cjs.000317

SUMMARY

A well-planned randomized controlled trial (RCT) is the most optimal study design to determine if a novel surgical intervention is any different than a prevailing one. Traditionally, when we want to show that a new surgical intervention is superior to a standard one, we analyze data from an RCT to see if the null hypothesis of “no difference” can be rejected (i.e., the 2 surgical interventions have the same effect). A noninferiority RCT design seeks to determine whether a new intervention is not worse than a prevailing (standard) one within an acceptable margin of risk or benefit, referred to as the “noninferiority margin.” In the last decade, we have observed an increase in the publication of noninferiority RCTs. This article explores this type of study design and discusses the tools that can be used to appraise such a study.

A well-planned randomized controlled trial (RCT) is the most optimal study design to determine if a novel surgical intervention is any different than a prevailing one. Traditionally, when we want to show that a new surgical intervention is superior to a standard one, we analyze data from an RCT to see if the null hypothesis of “no difference” can be rejected (i.e., the 2 surgical interventions have the same effect).^{1,2} Let's consider a hypothetical RCT that compares laparoscopic to open appendectomy and the outcome measured is a pain score based on a Likert scale from 0 to 10. Suppose it was found that the mean pain score following laparoscopic appendectomy was 7 points and that following open appendectomy was 8 points, and that this 1-point difference was statistically significant. Such a result would be uncommon because it would require a large sample size, but let's accept this for now. Although statistically the result is significant, we do not consider this 1-point difference to have clinical relevance. This type of thinking addresses the concept of minimum clinically important difference (MCID), which describes a threshold that might persuade us to change our surgical practice. The meaningful MCID is usually based on the available best evidence derived from previous systematic reviews, pilot/feasibility studies or clinical judgment based on discussion with experts in the field.

In another hypothetical RCT, the length of stay (LOS) after laparoscopic appendectomy was observed to be 24 hours versus 30 hours after open appendectomy, with a $p < 0.05$. It would be meaningless to conclude that the observed difference of 6 hours is the truth without reporting a confidence interval (CI), as a p value alone does not provide information on the degree of uncertainty (variation) applied in measuring the difference in hospital stay.³ Briefly, a CI provides information regarding the degree of uncertainty associated with the observed difference of 6 hours in hospital stay. It is within the CI that the true difference will likely lie. Let's say that in our hypothetical example the 95% CI for hospitalization time difference of 6 hours was 1–11 hours in favour of the laparoscopic approach. This means we are 95% confident that the true difference lies somewhere between 1 and 11 hours,

which is quite wide, raising uncertainty when a definitive conclusion is made.

In the last decade, we have observed an increase in the publication of noninferiority RCTs. This article explores this type of study design and discusses the tools that can be used to appraise such a study.

CLINICAL SCENARIO

At the last cardiac surgery weekly academic rounds there was a heated exchange between 2 surgeons, who were arguing the merit of ex-vivo heart perfusion compared with cold storage as a means of preserving donor hearts before transplantation. To resolve this dilemma, the division head has assigned you, the newest member of the division, with the task of finding the best evidence to answer this clinical question and report your findings to the group at next week's rounds.

FINDING THE EVIDENCE

To identify the best evidence and inform your colleagues you begin by conducting a literature search according to the "Users' guide to the surgical literature: how to perform a high-quality literature search."⁴ You follow the PICOT format, which serves as the starting point for identification of important key words used in the search process:⁵

- Population: heart transplant patients
- Intervention: heart transplantation with ex-vivo perfusion
- Comparison: heart transplantation with cold storage
- Outcome: patient survival and graft survival
- Time horizon: 30 days after transplant

You then conduct a literature search in PubMed Clinical Queries using the search terms "heart transplantation" AND "ex-vivo perfusion" AND "cold storage," using the "Therapy" and "Broad" filters. You identify 10 articles: 7 ex-vivo human/animal studies,⁶⁻¹² 1 nonrandomized clinical study,¹³ 1 review¹⁴ and 1 RCT.¹⁵ The RCT addresses your research question and has the benefit of being level-I evidence.¹⁶ However, when reading the article, you are perplexed that it is labelled as a "randomized noninferiority trial."

NONINFERIORITY RCT DESIGN

A noninferiority RCT design seeks to determine whether a new intervention is not worse than a prevailing (standard) one within an acceptable margin of risk or benefit, referred to as the noninferiority margin.¹⁷⁻²⁰ It is usually assumed that the standard intervention has been shown to have better (superior) clinical effect than a placebo or an earlier intervention. The new intervention is considered to be noninferior to the standard one when it is shown to have reduced costs, have fewer adverse effects (harm), be less invasive, and be of greater convenience. In trials that investigate noninferiority, the null hypothesis is not symmetric. The new intervention will be proven noninferior if it is

similar to the standard intervention, but not beyond the margin of noninferiority for a specified outcome measure. If the new intervention is found to be superior, it is an additional benefit. Tests of noninferiority should be linked to the predefined noninferiority margin and predefined α . An α of 0.025 for a 1-sided noninferiority hypothesis is equivalent to the 1-sided 97.5% CI, as an α of 0.05 for a 2-sided hypothesis is equivalent to a 2-sided 95% CI.¹⁷

Suppose the hospital administrators would like to expedite surgical patients' hospital discharge with the adoption of this new surgical approach if it is proven to be noninferior to the standard procedure within a 3-hour noninferiority margin.

Figure 1 presents some possible scenarios for noninferiority trials observing mean differences in hospital stay following laparoscopic and open approaches. Scenario A shows that laparoscopic surgery is superior to open surgery, as the CI lies to the left of the no difference line (zero). In scenario B, the CI includes the threshold of noninferiority. It means that noninferiority is not shown, as the true difference in hospital stay could be worse than the 3-hour predefined noninferiority margin for laparoscopic surgery. Scenarios C and D show that laparoscopic surgery is noninferior to open surgery because the upper confidence limit lies to the left of the 3-hour noninferiority margin in hospital stay. Scenario D shows that laparoscopic surgery is definitely noninferior to open surgery, as the CI lies to the left of the noninferiority margin and also excludes zero line of no difference. In scenario E, the laparoscopic surgery is definitely inferior to open surgery with respect to hospital stay, as the lower bound of the CI lies to the right of the noninferiority margin. Such a scenario is less likely, as it requires a very large sample size.¹⁷

The choice of noninferiority margin requires sound clinical judgment.¹⁸ The noninferiority margin should be the smallest clinically meaningful difference between the 2 surgical interventions. In general, margins for mortality or serious adverse events should be more stringent than those for symptom control or quality of life.¹⁸ Many experts have stipulated that the noninferiority margin for efficacy outcomes should be no more than 50%, and preferably no more than 20% of the treatment effect for the standard treatment, as established in placebo-controlled superiority RCTs.^{18,19} Unfortunately no validated rules exist for calculating the noninferiority margin, and many trials use margins that statisticians consider to be too liberal.²¹ It is important that, whenever possible, this margin be validated by published expert consensus²² and not left to the sole discretion of the investigators or sponsors of the study.¹⁸

RETURNING TO THE CLINICAL SCENARIO

The article you identified is a prospective, open-label, multicentre, randomized noninferiority trial, by Ardehali and colleagues¹⁵ that took place at 10 heart transplant centres in the United States and Europe. Eligible heart transplant patients were randomly assigned to receive either

donor hearts preserved with the organ care system (OCS; ex-vivo heart perfusion) or standard cold storage (SCS). The key methodological characteristics of the study are summarized in Figure 2 and Table 1.¹⁵

To effectively appraise a noninferiority surgical RCT, you use a similar framework to that of previous users' guide articles (Box 1).^{1,3,23}

Are the results valid?

Did the novel and standard surgical interventions start with similar prognostic factors?

As with the more commonly seen superiority RCT, the non-inferiority RCT is expected to minimize the risk of bias by ensuring concealment of randomization, balance between known and unknown prognostic factors, blinding of patients, surgeons and outcome assessors to treatment allocation, and complete follow-up of all patients. In reviewing the noninferiority trial by Ardehali and colleagues,¹⁵ you see that an independent biostatistician prepared sealed and masked randomization envelopes, which were assigned to the research trial sites. The investigators, however, did not report if the envelopes were opened sequentially and one at a time. Patients, investigators and medical personnel were not blinded to group allocation. They chose an open-label design because the method of donor heart preservation made blinding of medical staff infeasible. In reviewing Table 1 of their article, you see no glaring differences in main demographic characteristics of patients assigned to the 2 competing approaches; these characteristics included age, sex, height, body mass index (BMI), diagnosis of cardiomyopathy of the recipient patients, and the cause of death of the donor patients.

There was, however, some imbalance in the preservation time before the heart transplantation. The preservation time was longer in the OCS group than in the SCS group (324 ± 79 min v. 195 ± 65 min, $p < 0.001$); however, the mean total ischemia time was significantly shorter in the OCS group than in the SCS group (113 ± 27 min v. 195 ± 65 min, $p < 0.001$).

Was the prognostic balance between the 2 surgical groups maintained as the RCT progressed?

As the heart transplantation is a definitive procedure, there is probably little room to provide differential care to affect the prognostic balance after the event. Figure 2 of the article by Ardehali and colleagues¹⁵ shows the flow of the patients in the 2 groups. It details the results of the randomization protocol, wherein 130 patients were randomly assigned to either group: 67 to OCS and 63 to SCS. There appears to have been deviation of the protocol in 2 patients in the OCS and 5 in the SCS groups. It is important to note that 2 patients in the OCS group and 1 patient in the SCS group crossed over (i.e., these patients were transplanted using the other respective system).

Ardehali and colleagues¹⁵ did not report details regarding postoperative care, so you do not know if there was differential care between the 2 groups. Therefore, you cannot conclude with any certainty whether the 2 groups were balanced in this regard.

Did the investigators guard against an unwarranted conclusion of noninferiority?

In the present noninferiority trial, the investigators declared the noninferiority margin (Δ) to be 0.10 (10%). Unfortunately, they did not provide any evidence to

Fig. 1. Possible scenarios for observed mean differences in hospital stay between laparoscopic and open surgery in noninferiority designs. The dotted line represents the noninferiority margin. Reproduced from the study by Piaggio and colleagues¹⁷ with permission from the American Medical Association (license no. 4003080684018).

support this difference, which leads you to wonder if a smaller noninferiority margin (e.g., 5%) could have, or indeed should have, been accepted.

Did the investigators analyze patients according to the surgical treatment they received, as well as to the groups to which they were assigned?

The purpose of randomization is to ensure that prognostic factors are balanced between the surgical interventions. Patients who do not adhere to the allocated treatment, as

in the study protocol, may have a different prognosis than those who do.²⁴ Omission of patients who do not adhere to the novel intervention is likely to bias results toward over-estimation of treatment effects in a superiority trial. An intention-to-treat analysis, wherein patients are analyzed according to the group they were assigned, provides an unbiased estimate of the treatment effectiveness, irrespective of their adherence to the study protocol.

Ardehali and colleagues¹⁵ conducted both the intention-to-treat and the as-per-protocol analyses and found

Fig. 2. Methodological characteristics of the PROCEED II Trial.¹² *One patient experienced clinically significant ventricular assist device-related complications while waiting for a second donor offer, 1 patient needed a combined heart and kidney transplant while waiting for a second offer; 1 patient became ventilator-dependent on day of transplant, and 1 patient deteriorated and was delisted; 1 donor heart had left-ventricular hypertrophy (> 1.3 cm), and 1 donor was older than 60 years. †Two opened randomization envelopes before confirming availability of the Organ Care System team, 1 organ procurement organization refused to retrieve the Organ Care System because of absence of research consent, and 1 donor heart had a high concentration of serum lactate before retrieval. ‡After turning down initial donor heart offers, 1 recipient had more than 2 sternotomies while waiting for a third offer and 1 recipient withdrew before a second offer. §The randomization card was misread. ¶Deviations in the organ care system group were due to user error in cannulation and 1 recipient receiving unassigned treatment; in the standard cold storage group 1 recipient was enrolled in another pharmaceutical trial and 4 received unassigned treatment. Reproduced from the study by Ardehali and colleagues¹⁵ with permission from Elsevier (license no. 4003081323980).

similar results in both analyses. Therefore, you remain assured that the authors' analysis of the results may be appropriate. However, the authors could have assured readers of the findings by statistically addressing the missing data (e.g., multiple imputations or best and worst case scenario).

Table 1. Primary and secondary outcomes of the noninferiority trial selected for study*

	Group; no. (%)†		Between-group difference; 1-sided 95% UCB or 95% CI	<i>p</i> value
Oucome	OCS	SCS		
Primary outcome (30-d patient and graft survival)				
Intention-to-treat	63/67 (94)	61/63 (97)	2.8 (8.8)	0.45
As-treated	58/62 (94)	64/66 (97)	3.5 (9.6)	0.36
As-per-protocol	56/60 (93)	59/61 (97)	3.4 (9.9)	0.39
Secondary outcome (as-treated population)				
Serious adverse events	8 (13)	9 (14)	1 (−12 to 11)	0.90
Incidence of severe rejection	11 (18)	9 (14)	4 (−8 to 17)	0.52
ICU length of stay, median (IQR), h	147 (107–212)	137 (97–197)	10 (−10 to 42)	0.24
CI = confidence interval; ICU = intensive care unit; IQR = interquartile range; OCS = organ care system; SCS = standard cold storage; UCB = upper confidence bound. *Reproduced from the study by Ardehali and colleagues ¹⁵ with permission from Elsevier (license no. 4003081323980). †Unless indicated otherwise.				

Box 1. Framework for critical appraisal of an article that deals with a surgical noninferiority randomized controlled trial (RCT)

Are the results valid?

- Did the novel and standard surgical intervention groups start with similar prognostic factors?
- Was the prognostic balance between the 2 surgical groups maintained as the RCT progressed?
- Did the investigators guard against an unwarranted conclusion of noninferiority?
- Did the investigators analyze patients according to the surgical treatment they received, as well as to the groups to which they were assigned?
- Did the investigators report a predefined noninferiority margin?
- Did the investigators power the study for test of noninferiority?

What are the results?

- Were all patient-important outcomes considered?
- Were the results precise?
- Were the investigators appropriately interpreting the concept of noninferiority?

How can I apply the results to my patient or clinical practice?

- Were the study patients similar to my patient?
- Were all patient-important outcomes considered?
- Are the likely advantages of the novel surgical treatment worth the potential harm and costs?

Did the investigators report a predefined noninferiority margin?

In a noninferiority RCT it is important that the investigators report the noninferiority margin and the rationale for choosing it. In the statistical analysis section of their methods, Ardehali and colleagues¹⁵ mentioned a 10% noninferiority margin, but provided no rationale for choosing it.

Did the investigators power the study for a test of noninferiority?

The investigators reported in the Methods section of their article that they calculated the 1-sided 95% upper confidence bound based on the normal approximation for the difference between the 2 population proportions. An upper confidence bound less than the 10% noninferiority margin would have rejected the null hypothesis. For the purpose of sample-size calculation, they assumed $\pi_{OCS} = 0.95$ and $\pi_{SOC} = 0.94$. On the basis of these assumptions, use of a normal approximation test and a 1-sided α level of 0.05, inclusion of 54 patients per treatment group would have provided 80% power.

There should be a justification for choosing a superiority versus a noninferiority study design. To some degree the authors justified the choice of the noninferiority design in that the OCS system provides certain benefits, such as the potential of “distant procurement for donor hearts, thus expanding the donor pool” in contrast to the standard cold storage. The justification of the study design is made on the research question asked and the hypothesis — specifically on the clinical advantages of the novel intervention. The measured outcomes play an important role in the sample size calculation through the choice of the MCID. Survival should demand a smaller MCID than, for example, a quality of life (QOL) outcome. The 10% choice as the noninferiority margin seems very liberal, which most surgeons or patients would not accept in a case of life or death. A noninferiority margin of 1%–2% would likely be a better choice. This raises the concern that the study may have been designed originally as a superiority study.

What are the results?

Were all patient-important outcomes considered?

The investigators found that the 30-day patient and heart transplant survival rate (primary outcome) was 94% in the OCS group and 97% in the SCS group ($p = 0.45$). The intention-to-treat analysis (94% v. 97%, $p = 0.36$) and the as-per-protocol analysis (93% v. 97%, $p = 0.39$) supported the overall estimate.

Multiple clinically important outcomes were included, such as graft failure and left and right ventricular dysfunction, with a time horizon of 30 days. Some surgeons may consider this short-term time frame of limited value; a longer follow-up would have been more appropriate. Patient-important

outcomes, such as quality of life, were not considered. A validated patient-reported outcome scale would have provided more information on the merits of the comparative interventions. You note this as a limitation of this noninferiority RCT.

The secondary outcomes — serious adverse events, incidence of severe rejection and median length of stay in the intensive care unit (ICU) — were similar for the 2 approaches. Based on these results, the investigators concluded that the OCS approach was not inferior to the SCS approach. You believe that their conclusion is reasonable based on the results of the study.

Were the results precise?

The precision of the results is normally presented as a confidence interval (CI). In this noninferiority study the authors provided the CI for both the primary outcomes (30-d patient and graft survival) and the secondary outcomes (cardiac-related serious adverse events, incidence of severe rejection and ICU length of stay). They provided this for the intention-to-treat, as-treated and as-per-protocol analyses (Table 1).

Did the investigators appropriately interpret the concept of noninferiority?

The authors reported the 30-day patient and graft survival rates to be 94% in the OCS group and 97% in the SCS group. The patient and graft survival rates were the same, as no repeat heart transplant surgeries were performed. The authors reported that the upper bound of the 95% CI for the percentage differences in the primary effectiveness outcome between the 2 populations was 8.8%, which is less than 10%, so the null hypothesis was rejected in favour of the alternative hypothesis. Based on this finding you concur that noninferiority was shown.

How can I apply the results to my patient or clinical practice?

Were the study patients similar to my patient?

Based on the demographic evidence provided by Ardehali and colleagues¹⁵ in Table 1 of their study (not shown here), in which they reported the mean age, weight, height, BMI and types of cardiomyopathy their patients had as well as the donor characteristics, you conclude that the patients treated in your division would be similar and that, therefore, the study's conclusions are applicable.

Were all patient-important outcomes considered?

The investigators included patient and graft 30-day survival as a primary outcome. A longer survival time horizon (e.g., 1- or 2-yr survival) would have been preferable. The investigators also included 30-day right and left ventricular function and length of stay in the ICU as secondary outcomes. The outcomes research movement in the last 20 years, however, expects clinical investigators to measure

patients' quality of life after medical interventions. The quality of life assessment requires a longer follow-up, and this trial was designed based on immediate and short-term outcome assessment. The authors might have suggested this for future investigation.

Are the likely advantages of the novel surgical treatment worth the potential harm and costs?

Although the authors concluded that the novel intervention was noninferior to the standard approach, you should not be rushed to adopt it. The investigators did not report the resource utilization associated with either approach. Many new innovations are costly. The ideal study, therefore, would be one in which resource utilization and costs are captured. Health-related quality of life can also be measured using a utility scale from which quality-adjusted life years (QALYs) can be calculated. The integration of costs and QALYs in a cost-utility analysis can help determine whether the new innovation is cost-effective or not.²⁵

RESOLUTION OF THE SCENARIO

There are consequences to future patients and society if incorrect inferences from a poorly designed and conducted noninferiority RCTs are accepted. It is important to determine if this noninferiority study is really a failed superiority RCT. You can do this by determining whether the authors' noninferiority threshold was appropriate or not. To do so, you review the literature for similar studies to determine the upper boundary for the CI of the primary outcome (30-d mortality and heart transplant survival) and examine the extent to which it exceeds the chosen threshold. If the upper boundary is substantially greater/lower than the threshold chosen by the investigators (10%) you may choose not to adopt this new technology. This is, unfortunately, the case with the RCT by Ardehali and colleagues.¹⁵ Their noninferiority margin of 0.10 (10%) was chosen without supportive documentation.

CONCLUSION

Although, in general, you are happy with the designation of this study as an RCT, you are not persuaded that it met all the criteria for its designation as a noninferiority RCT. Specifically, you are concerned that their noninferiority margin of 0.10 (10%) was chosen without supportive evidence. As a result, you recommend to your colleagues that the study has definite weaknesses. You then offer to review and critique a superiority RCT comparing these approaches to present at next week's rounds.

Affiliations: From the Division of Plastic Surgery, Department of Surgery, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Thoma); the Surgical Outcomes Research Center, Department of Surgery, Faculty of

Health Sciences, McMaster University, Hamilton, Ont. (Thoma, Farrokhyar, Braga, Goldsmith); the Department of Clinical Epidemiology and Biostatistics, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Thoma, Farrokhyar, Sprague, Goldsmith); the Faculty of Medicine, University of Ottawa, Ottawa, Ont. (Waltho); the Division of Urology, Department of Surgery, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Braga); the Faculty of Health Sciences, Simon Fraser University, Burnaby, BC (Goldsmith); and the Division of Orthopedic Surgery, Department of Surgery, Faculty of Health Sciences, McMaster University, Hamilton, Ont. (Sprague).

Competing interests: None declared.

Contributors: All authors designed the study. D. Waltho acquired the data, which F. Farrokhyar, D. Waltho and C. Goldsmith analyzed. A. Thoma, F. Farrokhyar and D. Waltho wrote the article, which all authors reviewed and approved for publication.

References

1. Thoma A, Farrokhyar F, Bhandari M, et al. Users' guide to the surgical literature. How to assess a randomized controlled trial in surgery. *Can J Surg* 2004;47:200-8.
2. Farrokhyar F, Karanickolas PJ, Thoma A, et al. Randomized controlled trials of surgical interventions. *Ann Surg* 2010;251:409-16.
3. Cadeddu M, Farrokhyar F, Levis C, et al. Users' guide to the surgical literature. Understanding confidence intervals. *Can J Surg* 2012;55:207-11.
4. Waltho DA, Kaur MN, Haynes RB, et al. Users' guide to the surgical literature: how to perform a high quality literature search. *Can J Surg* 2015;58:349-58.
5. Richardson WS, Wilson MC, Nishikawa J, et al. The well-built clinical question: a key to evidence-based decisions. *ACP J Club* 1995;123:A12-3.
6. Van Caenegem O, Beauloye C, Bertrand L, et al. Vanoverschelde JL1. Hypothermic continuous machine perfusion enables preservation of energy charge and functional recovery of heart grafts in an ex vivo model of donation following circulatory death. *Eur J Cardiothorac Surg* 2015;49: 1348-53.
7. Liersch-Nordqvist A, Ingemansson R, Pierre L, et al. Lungs exposed to 1 hour warm ischemia without heparin before harvesting might be suitable candidates for transplantation. *J Cardiothorac Surg* 2015;10:131.
8. Osaki S, Locher MR, Lushaj EB, et al. Functional evaluation of human donation after cardiac death donor hearts using a continuous isolated myocardial perfusion technique: Potential for expansion of the cardiac donor population. *J Thorac Cardiovasc Surg* 2014;148:1123-30.
9. Mulloy DP, Stone ML, Crosby IK, et al. Ex vivo rehabilitation of non-heart-beating donor lungs in preclinical porcine model: delayed perfusion results in superior lung function. *J Thorac Cardiovasc Surg* 2012;144:1208-15.
10. Soares PR, Braga KA, Nepomuceno NA, et al. Comparison between Perfadex and locally manufactured low-potassium dextran solution for pulmonary preservation in an ex vivo isolated lung perfusion model. *Transplant Proc* 2011;43:84-8.
11. Inci I, Ampollini L, Arni S, et al. Ex vivo reconditioning of marginal donor lungs injured by acid aspiration. *J Heart Lung Transplant* 2008;27:1229-36.
12. Kobayashi T, Liu D, Ogawa H, et al. Removal of blood group A/B antigen in organs by ex vivo and in vivo administration of endo-beta-galactosidase (ABase) for ABO-incompatible transplantation. *Transpl Immunol* 2009;20:132-8.
13. Koerner MM, Ghodsizad A, Schulz U, et al. Normothermic ex vivo allograft blood perfusion in clinical heart transplantation. *Heart Surg Forum* 2014;17:E141-5.
14. Minasian SM, Galagudza MM, Dmitriev YV, et al. Preservation of the donor heart: from basic science to clinical studies. *Interact Cardiovasc Thorac Surg* 2015;20:510-9.
15. Ardehali A, Esmailian F, Deng M, et al. Ex-vivo perfusion of donor hearts for human heart transplantation (PROCEED II): a prospective, open-label, multicentre, randomized non-inferiority trial. *Lancet* 2015;385:2577-84.
16. Canadian Task Force on the Periodic Health Examination. Task force report: the periodic health examination. *CMAJ* 1979;121: 1193-254.
17. Piaggio G, Elbourne DR, Pocock SJ, et al. Reporting of noninferiority and equivalence randomized trials — extension of the CONSORT 2010 Statement. *JAMA* 2017;308:2594-604.
18. Scott IA. Noninferiority trials: determining whether alternative treatments are good enough. *Med J Aust* 2009;190:326-30.
19. Kaji AH, Lewis RJ, Mulla SM. Guideline on the choice of the non-inferiority margin. London (UK): Euroean Medicines Agency; 2005.
20. Hahn S. Understanding noninferiority trials. *Korean J Pediatr* 2012; 55:403-7.
21. Lange S, Freitag G. Choice of delta: requirements and reality — results of a systematic review. *Biom J* 2005;47:12-27.
22. Wyrwich KW, Spertus JA, Krenke K, et al. Clinically important differences in health status for patients with heart disease: an expert consensus panel report. *Am Hear J* 2004;147:615-22.
23. Cadeddu M, Farrokhyar F, Thoma A, et al. Users' guide to the surgical literature: how to assess power and sample size. *Can J Surg* 2008;51:476-82.
24. Friedman LM, Furberg CD, DeMets DL, editors. *Fundamentals of clinical trials*, 3rd ed. New York (NY): Springer Verlag; 1998. p. 43-6.
25. Thoma A, McKnight LL. Quality-adjusted life-year as a surgical outcome measure: a primer for plastic surgeons. *Plast Reconstr Surg* 2010;125:1279-87.