Surgical clinical trials with non-inferiority design: a cross-sectional bibliometric analysis

Chi Shu[1,2,3], Bin Huang[1], Ding Yuan[1], Yi Yang[1], Xiaojiong Du[1], Yazhou He[4], Xin Chen[4], Jichun Zhao[1]

[1]Department of Vascular Surgery, West China Hospital, Sichuan University, Chengdu, China; [2]West China School of Medicine, Sichuan University, Chengdu, China; [3]Department of Vascular Surgery, Toronto General Hospital, Toronto, ON, Canada; [4]Department of Oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, Chengdu, China

Contributions: (I) Conception and design: X Chen, C Shu, B Huang; (II) Administrative support: X Chen, Y Yang; (III) Provision of study materials or patients: X Chen, Y Yang; (IV) Collection and assembly of data: All authors; (V) Data analysis and interpretation: All authors; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Jichun Zhao. Department of Vascular Surgery, West China Hospital, Sichuan University, 37 Guo Xue Alley, Chengdu 610041, China. Email: zhaojc3@163.com; Xin Chen. Department of Oncology, West China School of Public Health and West China Fourth Hospital, Sichuan University, 16 Ren Min South Street, Chengdu 610041, China. Email: liangchaokaiji@foxmail.com.

Background: Wide-spread concerns have been raised about possible bias in published surgical non-inferiority trials. Therefore, we performed a comprehensive bibliometric analysis to identify the existence of bias, and provided recommendations for future non-inferiority trials.

Methods: Databases including MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials were systematically searched (last update on 27 April 2020) to include published phase II and phase III non-inferiority surgical trials. We collected general information and parameters associated with trial design. The association between extracted factors and establishment of non-inferiority was then analyzed.

Results: A total of 347 trials were included in this study. Only 13 (3.7%) trials reported the pre-specified non-inferiority margin in registration, and 99 (28.5%) trials justified margin selection in ultimate trial publications. A significant association was found between industry funding and increased odds of achieving non-inferiority [odds ratio (OR): 1.17, 95% confidence interval (CI): 1.06 to 1.30, P=0.001]. Moreover, trials which had been presented in conferences were less likely to claim non-inferiority (OR: 0.83, 95% CI: 0.69 to 0.99, P=0.035).

Conclusions: Our study was the first quantitative analysis revealing the presence of biases in findings of existing surgical non-inferiority trials, which could possibly mislead surgeons' clinical decision making. We suggest improving reporting of detailed study design especially funding sources as well as margin justification for future trials. We also encourage conference presentation of ongoing trials prior to the ultimate publication.

Keywords: Non-inferiority; design; bias; surgical trial; bibliometric analysis

Introduction

Non-inferiority design has been deployed in a growing number of surgical clinical trials. It is the optimal choice for investigating new surgical procedures which may not present significant clinical superiority but offers certain advantages such as increased cost-efficiency, ease of operation, and reduced invasiveness (1). To date, a few surgical novel techniques, such as the robot-assisted and laparoscopic procedures (2,3), have been recommended by official guidelines based on the findings of non-inferiority trials.

Page 2 of 8

Shu et al. Non-inferiority trials in surgery

Concluding non-inferiority is based on comparison between confidence intervals of treatment effects and pre-defined and clinically acceptable margins, known as non-inferiority margins. One of the most challenging points in non-inferiority design is margin justification since it should balance both clinical and statistical perceptions (4). Theoretically, the probability of establishing non-inferiority should be independent from pre-specified parameters except for the type II error (β) or statistical power under the alternative hypothesis. However, there have been wide-spread concerns regarding the validity of established non-inferiority, especially on account of the arbitrary definition of non-inferiority (5), where biases could stem from (6,7). An earlier systematic review found that even in high-quality journals, non-inferiority design of clinical trials was reported inconsistently and did not follow official recommendations (8). Biased findings of non-inferiority, if approved by guidelines, could potentially mislead surgeons in clinical decision-making and eventually result in patients receiving inferior surgical treatments. However, quantitative evidence is still lacking, leaving this issue unsolved.

To determine the existence of bias, we explored the external factors that influence the establishment of non-inferiority by systematically surveying and analyzing the characteristics of published surgical clinical trials.

We present the following article in accordance with the PRISMA reporting checklist (available at https://dx.doi.org/10.21037/atm-21-2626).

Methods

Search strategy and trial selection

Databases including MEDLINE, Embase, and the Cochrane Central Register of Controlled Trials were systematically searched (last update on 27 April 2020, detailed strategy presented in Table S1) with a limitation to publications in the English language. The search was restricted to clinical trials in MEDLINE and Embase. The registry identifier and references of included studies were also cross-checked for additional trials.

All retrieved records were screened by two reviewers (C.S. and B.H.). We included non-inferiority trials that investigated surgical procedures of treatment purposes in at least one treatment arm based on the recommendations from the PubMed queries (9), and excluded trials regarding the diagnostic, cosmetic, and obstetric procedures (10). The inclusion criteria were as follows: completed or ongoing

trials with published results; trials aiming to prove non-inferiority of a new treatment (procedure, technique, material, and so on) to a conventional one, and at least one treatment was surgical related; trials reporting whether the non-inferiority was established. For multiple publications with the same registry identifier, only the one reporting the ultimate findings of the primary outcome was included. Subgroup and post hoc analyses were not eligible. Any discrepancies were solved by discussion with a senior surgeon (J.Z.) and an epidemiologist (D.Y.).

Data extraction

A standard data extraction of included studies was performed by one author (C.S.) with an Excel form and checked by a second author (X.C.). Discrepancies were reviewed and discussed to reach agreement. Essential characteristics of the eligible studies were abstracted by two reviewers (C.S. and X.C.) independently, including first author, publication year, journal name and impact factor in 2019, single or multi-center trial, trial status (completed, interim, or terminated), trial registry number, surgical specialty (e.g., cardiovascular, digestive, urogenital, orthopedic, and so on), follow-up time (months), primary outcome (e.g., event free survival, surgical success, late luminal loss, etc.), funding source (industry or non-industry), conference presentation, and declaration of competing interests.

We also collected methodologic parameters associated with study design including outcome event rate, 1-sided type I error (α), type II error (β), non-inferiority margin reported as both absolute differences such as rate difference and relative effect sizes such as hazard ratios (HR), odds ratios (OR), and risk ratios (RR), justifications of margin selection and estimated sample size. We evaluated the establishment of non-inferiority by examining whether the upper bound of estimated confidence interval (CI) exceeded the pre-specified non-inferiority margin.

Statistical analysis

We performed descriptive analysis for the extracted general characteristics. In particular, categorical variables were expressed as frequencies, while median and inter-quartile range (IQR) were used for continuous variables. We performed Pearson's Chi-square (χ^2) tests and Mann-Whitney U tests to compare the differences of distribution patterns of categorical and continuous characteristics, respectively, between trials with and without establishing

non-inferiority. A 2-sided P value <0.05 was considered as an indicator for significant association between a certain factor and establishment of non-inferiority. Since the probability of establishing non-inferiority should theoretically only be dependent on the type II error (β) under the alternative hypothesis, any other external factors associated with establishment of non-inferiority would imply potential bias. Notably, to model the effect of non-inferiority margin on reported outcome of non-inferiority, we first transformed margins expressed as rate difference to RRs based on the baseline outcome event rate. With regard to studies using continuous effect estimates such as mean differences as the primary outcome, we standardized the effect estimates with the reported standard deviations (SD), and then transformed the continuous estimates to ORs following the Hasselblad and Hedges' method (11,12). A previous study had shown that HRs, ORs, and RRs can be good numerical approximations of one another (13). Therefore, we took the coefficient scale of log-transformed relative effects (HRs, RRs, and ORs) and investigated their association with ultimate establishment of non-inferiority.

All statistical analyses were performed using R (version 4.0.2; https://www.R-project.org/).

Results

Selection of studies

A total of 3,312 records were retrieved from the aforementioned three databases. After reviewing titles and abstracts, 746 records were identified for in-depth full-text review. Through cross-checking the trial registry identifier and reference of eligible studies, we enrolled 3 additional studies. At last, 347 non-inferiority surgical clinical trials were included in our study. The flow chart of study selection is presented in *Figure 1*.

General trial characteristics

Basic characteristics of the 347 eligible trials are shown in *Table 1*, with detailed information available in Table S2. Among all the trials, 277 (79.8%) claimed non-inferiority in conclusion. As for methodologic parameters, not much diversity was observed in terms of type I (median 0.05, IQR 0.025–0.05) and type II error (median 0.20, IQR 0.10–0.20); the median sample size was 261 with IQR between 136 and 800; the majority of non-inferiority margins in HR were less than 2 and with a median number of 1.46 (IQR

1.23–2.00). Only 99 (28.5%) trials reported justification for the margin and 58 (58.6%) of them were based on previous trials, while 19 (19.2%) used effect retention method and 16 (16.2%) relied on expert consensus. A total of 204 (58.8%) trials reported method for sample size calculation; of them, 187 (91.7%) were based on previous trials, and only 15 (7.4%) followed instructions from methodologic studies.

Quantitative analysis

As presented in *Table 2*, the essential characteristics were compared between trials with or without establishment of non-inferiority. Among all surgical specialties, cardiovascular related interventions were performed in 157 (56.7%) trials that claimed non-inferiority and 29 (41.4%) trials that failed, which were the highest in both groups. The distribution of surgical specialties was not significantly associated with the establishment of non-inferiority (P=0.09). In trials that achieved non-inferiority, a lower percentage of published protocols (15.9% *vs.* 22.9%) and lower journal impact factor (6.38 *vs.* 8.43) were observed, although no significant difference was detected. A significant association was found between industry funding and increased odds of achieving non-inferiority (OR: 1.17, 95% CI: 1.06 to 1.30, P=0.001). In addition, trials that presented their findings in conferences were significantly less likely to establish non-inferiority (OR: 0.83, 95% CI: 0.69 to 0.99, P=0.035). Regarding parameters associated with trial design, only 13 (3.7%) trials reported the pre-specified margin in registration, and 99 (28.5%) trials justified their selection of non-inferiority margin. No significant associations were identified between the established non-inferiority and other parameters including type I error, type II error, non-inferiority margin, and sample size.

Discussion

Multiple studies have investigated the design, conduct, and interpretation of surgical non-inferiority trials and highlighted the deficiencies such as arbitrary selection of margin and poor quality of reporting (14,15). These studies, however, have only focused on a subspecialty, such as surgical oncology, and were therefore limited by small number of included trials. Therefore, we performed a systematic bibliometric analysis which summarized 347 previously published non-inferiority phase II and III surgical trials.
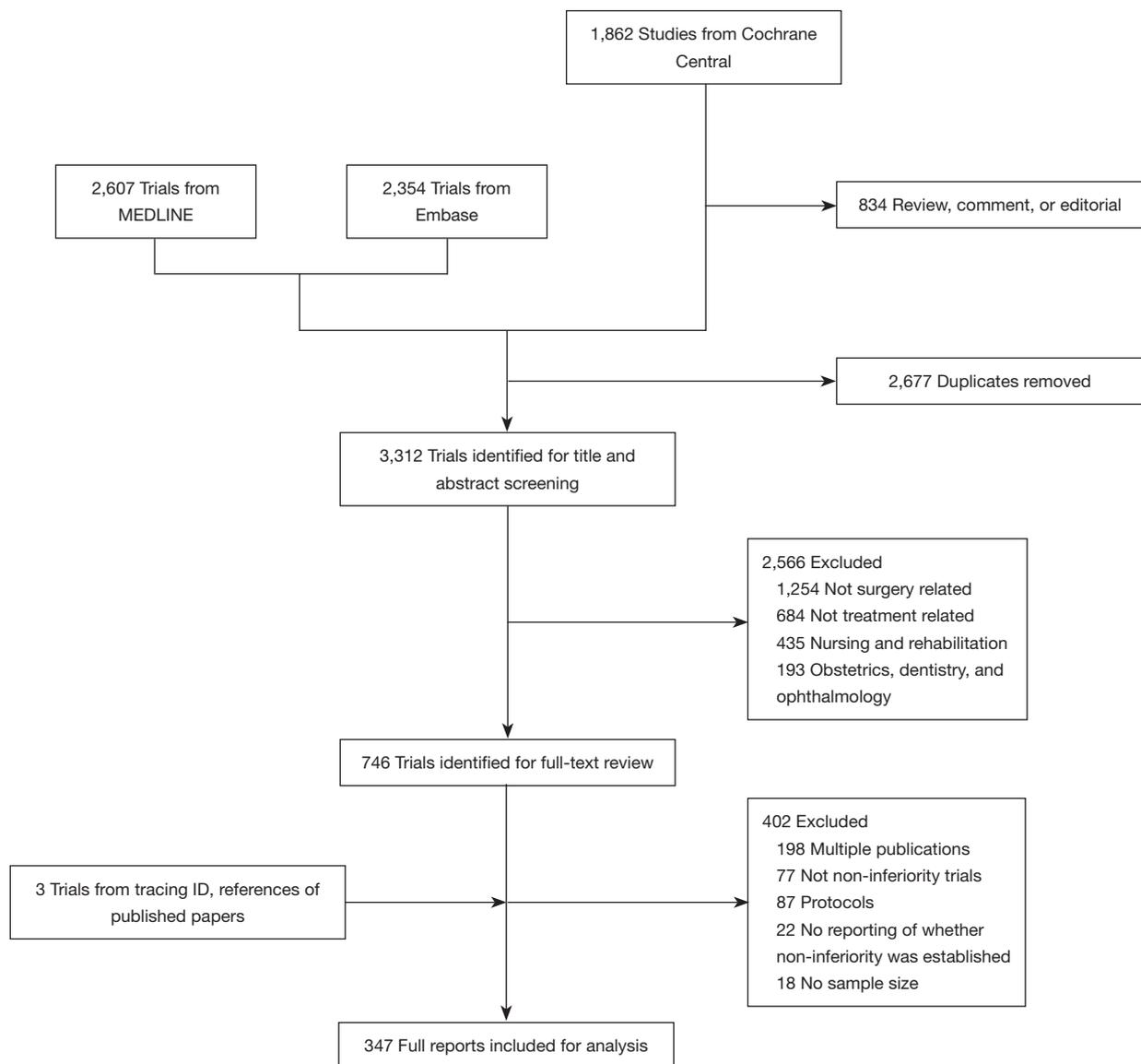
To our best knowledge, this is the first effort that

Figure 1 Flow diagram of study selection process.

quantitatively assessed factors associated with findings of published non-inferiority trials in surgery. We identified industry funding and conference presentation as potential sources of bias in surgical non-inferiority trials. We detected significant industry sponsorship bias which led to the excess establishment of non-inferiority in existing surgical clinical trials, resonating with a previous systematic review which included trials from all disciplines and found that industry-funded trials were more likely to use non-inferiority designs and report "favorable" results (16). To improve transparent reporting, funding sources should be clearly reported both in the trial registration record and the ultimate publication. If an industry-funded trial chooses a product from competing companies as the control arm, a specified statement should be added as part of the competing of interests. We also found that underreporting of trial design and trial results prior to the ultimate publication of trial findings was associated with higher probability of concluding non-inferiority. Based on our findings, conference presentations should be encouraged as it might help preventing possible post-hoc distortion to the original study design. In addition to these biases,

Table 1 Summary of essential characteristics of the 347 included non-inferiority trials

Characteristics	Number (%)
Trial status	
Full report	347
Completed	331 (95.4)
Interim	2 (0.6)
Terminated	14 (4.0)
Publication year	
2016–2020	175 (50.4)
2011–2015	131 (37.8)
2006–2010	37 (10.7)
2003–2005	4 (1.1)
Country	
Europe	159 (45.8)
Asia	93 (26.8)
North America	87 (25.1)
Others	8 (2.3)
Multi-center trials	
Yes	297 (85.6)
No	50 (14.4)
Trial registered	
Yes	287 (82.7)
No	60 (17.3)
Registry institution	
Clinical trial.gov	237 (82.6)
UMIN	12 (4.2)
NTR	12 (4.2)
ISRCTN	10 (3.5)
ChiCTR	5 (1.7)
Others	11 (3.8)
Type of comparison	
Surgery *vs.* surgery	325 (93.7)
Surgery *vs.* medication	22 (6.3)
Comparison between different procedures (surgery *vs.* surgery)	
Stent *vs.* stent	119 (36.6)
Open surgery *vs.* open surgery	102 (31.4)
Intervention *vs.* intervention	71 (21.8)
Open surgery *vs.* intervention	33 (10.2)

Table 1 (*continued*)

Table 1 (*continued*)

Characteristics	Number (%)
Primary endpoint	
Survival	
Event free survival	120 (34.6)
Overall survival	21 (6.1)
Recurrence free survival	14 (4.0)
Disease free survival	7 (2.0)
Surgical success	
Success rate	86 (24.8)
Continuous	
Late luminal loss	51 (14.7)
Score or index	26 (7.5)
Others	22 (6.3)

UMIN, University Hospital Medical Information Network; NTR, Netherlands Trial Registry; ISRCTN, International Standard Randomized Controlled Trial Number Register; ChiCTR, Chinese Clinical Trial Register.

it is worth noting that our study focuses on randomized controlled trials, which may have limited generalizability. Non-inferiority achieved by existing surgical trials should be further validated in the real-world settings due to potentially diverse population (17).

In our study, we found that methodological details of non-inferiority design were severely underreported in current surgical trials. For example, among the 347 eligible trials, only 99 (28.5%) justified their selection of non-inferiority margin, which is comparable to a prior study including trials from all disciplines (6). Poorly justified margin specification could lead to excess achievement of non-inferiority; although in our study, the transformed margin was not associated with establishment of non-inferiority (P=0.81). We thereby call for compulsory reporting of non-inferiority margin and margin justification details in trial registry such as Clinicaltrials.gov and published articles. Any protocol amendment should be documented in detail with caution.

Although no association was observed between surgical specialty and establishment of non-inferiority in our study, potential bias could have been generated, which merits further investigation. In particular, among all included trials of our study, 186 (53.6%) trials investigated cardiovascular and peripheral vascular surgeries, and 57 (16%) trials investigated general surgeries. A prior cross-sectional

Table 2 Characteristics of 347 trials with or without establishment of non-inferiority

Characteristics	Trials with non-inferiority established*		P values**
	Yes (N=277)	No (N=70)	
Surgical specialty, n (%)			
Cardiovascular	157 (56.7)	29 (41.4)	0.09
Digestive	40 (14.4)	17 (24.3)	
Urogenital	36 (13.0)	15 (21.4)	
Orthopedic	26 (9.4)	3 (4.3)	
Other	18 (6.5)	6 (8.6)	
Journal impact factor	6.38 (3.19–23.05)	8.43 (4.56–28.43)	0.19
Follow-up time (months)	12.00 (9.00–12.00)	12.00 (6.75–24.00)	0.67
Protocol published, n (%)			
Yes	44 (15.9)	16 (22.9)	0.15
No	233 (84.1)	54 (77.1)	
Funding type, n (%)			
Non-industry	68 (29.8)	31 (52.5)	0.001
Industry	160 (70.2)	28 (47.5)	
Conference presentation, n (%)			
Yes	13 (4.7)	8 (11.4)	0.03
No	264 (95.3)	62 (88.6)	
Conflicts of interest, n (%)			
Yes	177 (63.9)	40 (57.1)	0.30
No	100 (36.1)	30 (42.9)	
Parameters associated with study design			
Pre-specified margin in registration, n (%)			
Yes	11 (4.0)	2 (2.9)	0.66
No	266 (96.0)	68 (97.1)	
Type I error	0.05 (0.025–0.05)	0.05 (0.044–0.05)	0.10
Type II error	0.2 (0.10–0.20)	0.2 (0.11–0.20)	0.24
Non-inferiority margin	1.46 (1.23–2.02)	1.42 (1.23–1.88)	0.81
Margin justification, n (%)			
Yes	83 (30.0)	16 (22.9)	0.24
No	194 (70.0)	54 (77.1)	
Sample size	260 [140–820]	289 [118–737]	0.82

*, medians and quartiles were used for continuous variables; **, P values for chi-square tests except for follow-up time where a Mann-Whitney U test was used.

survey focusing on all types of surgical trials reported that general surgery accounted for the largest proportion (34.5%) of all published surgical trials (10). Our findings indicated that non-inferiority design might be more commonly adopted in trials of cardiovascular surgeries. In our study, 119 (34.3%) trials focused on comparisons across different types of coronary stents. Whether these trials adopted non-inferiority design in order to chase higher probability of achieving favorable outcomes, and what role funders played in selecting this type of study design remain unclear, and therefore are yet to be explored in-depth by future research.

The main limitation of our study is that we only enrolled published trials which were indexed in databases such as MEDLINE, Embase, and Cochrane Central which led to omission of unpublished data.

In summary, we systematically analyzed previously published non-inferiority trials in surgery and identified potential biases in such type of trials. Based on our findings, future trials should continue to improve transparent reporting of potential conflicts of interests especially the funding sources. In addition, trials are encouraged to be presented in conferences to increase visibility and to some extent prevent post-hoc manipulation of the study design. Last but not the least, trials should be registered with full details of study design in registries such as Clinicaltrials.gov, or publish these details in the protocol.

Footnote

Reporting Checklist: The authors have completed the PRISMA reporting checklist. Available at https://dx.doi.org/10.21037/atm-21-2626

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at https://dx.doi.org/10.21037/atm-21-2626). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

References

1. Mulla SM, Scott IA, Jackevicius CA, et al. How to use a noninferiority trial: users' guides to the medical literature. JAMA 2012;308:2605-11.
2. Merseburger AS, Herrmann TR, Shariat SF, et al. EAU guidelines on robotic and single-site surgery in urology. Eur Urol 2013;64:277-91.
3. Zerey M, Hawver LM, Awad Z, et al. SAGES evidence-based guidelines for the laparoscopic resection of curable colon and rectal cancer. Surg Endosc 2013;27:1-10.
4. Middleton LJ. Falling in the margin: Randomised controlled trials with a non-inferiority design. BJOG 2021. [Epub ahead of print]. doi: 10.1111/1471-0528.16702.
5. Burotto M, Prasad V, Fojo T. Non-inferiority trials: why oncologists must remain wary. Lancet Oncol 2015;16:364-6.
6. Gopal AD, Desai NR, Tse T, et al. Reporting of noninferiority trials in ClinicalTrials.gov and corresponding publications. JAMA 2015;313:1163-5.
7. Mauri L, D'Agostino RB Sr. Challenges in the Design and Interpretation of Noninferiority Trials. N Engl J Med 2017;377:1357-67.
8. Rehal S, Morris TP, Fielding K, et al. Non-inferiority trials: are they inferior? A systematic review of reporting in major medical journals. BMJ Open 2016;6:e012594.
9. Surgical Procedures, Operative. National Center for Biotechnology Information. Available online: https://www.ncbi.nlm.nih.gov/mesh/68013514. (Accessed in April 2020). 2020.
10. Yu J, Chen W, Chen S, et al. Design, Conduct, and Analysis of Surgical Randomized Controlled Trials: A Cross-sectional Survey. Ann Surg 2019;270:1065-9.
11. da Costa BR, Rutjes AW, Johnston BC, et al. Methods to convert continuous outcomes into odds ratios of treatment response and numbers needed to treat: meta-epidemiological study. Int J Epidemiol 2012;41:1445-59.

12. Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. Psychol Bull 1995;117:167-78.

13. Symons MJ, Moore DT. Hazard rate ratio and prospective epidemiological studies. J Clin Epidemiol 2002;55:893-9.

14. Parsyan A, Marini W, Fazelzad R, et al. Current Issues in Conduct and Reporting of Noninferiority Randomized Controlled Trials in Surgical Management of Cancer Patients. Ann Surg Oncol 2021;28:39-47.

15. Blencowe NS, Chana P, Whistance RN, et al. Outcome reporting in neoadjuvant surgical trials: a systematic review of the literature and proposals for new standards. J Natl Cancer Inst 2014;106:dju217.

16. Flacco ME, Manzoli L, Boccia S, et al. Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. J Clin Epidemiol 2015;68:811-20.

17. Zhou YL, Zhang YG, Zhang R, et al. Population diversity of cardiovascular outcome trials and real-world patients with diabetes in a Chinese tertiary hospital. Chin Med J (Engl) 2021;134:1317-23.