External Validity of Knee Osteoarthritis Clinical Trials: A Systematic Review

Yun Hyung KOOG,[1,2] Jin Su LEE,[2,3] Hyungsun WI[2]

[1]Department of Oriental Medicine, Medifarm Hospital, Suncheon, South Korea
[2]Honam Research Center, Medifarm Hospital, Suncheon, South Korea
[3]Department of Rehabilitation, Medifarm Hospital, Suncheon, South Korea

ABSTRACT

Objectives: This study aims to investigate the external validity of knee osteoarthritis trials through a systematic review of randomized, placebo-controlled, clinical trials.

Materials and methods: Randomized trials were identified by searches conducted in PubMed, SCOPUS, and the Cochrane Central Register of Controlled Trials. Then, the number of patients who were screened for eligibility, the number of patients who were eligible, and the number of patients who were randomized in each trial were identified.

Results: Overall, 345 reports presenting 352 trials were included in the analysis. Of the trials that reported quantitative recruitment data, the median proportion of screened patients who were eligible for participation was 71.9% (interquartile range: 52.7 to 86.5%) and the median proportion of eligible patients who were randomized in a trial was 92.9% (interquartile range: 82.5 to 100%). The median proportion of screened patients who were randomized in a trial was 67.9% (interquartile range: 48.9 to 92.9%), indicating that three patients were screened for every two patients randomized in trials. When this median value was considered as a reference point, trials conducted for 34% of individual treatments randomized lower proportions of screened patients.

Conclusion: Knee osteoarthritis trials were excellently generalizable. However, the generalizability should be considered in relation to all clinical information shown in trials, since a great number of treatments appear to be tested on highly selected patients.

Keywords: External validity; knee; osteoarthritis; systematic review.

Randomized trials are often used as key evidence for detecting the effectiveness of treatments in knee osteoarthritis (KOA).[1-3] However, for these trials to be clinically useful, trial participants should have similar characteristics with patients shown in clinical practices. To achieve this, trials need to be evaluated in terms of external validity, from which we can detect whether trial findings are generalizable to the population with KOA.

As external validity is an important issue for clinicians applying trial findings in a clinical setting, a number of studies have investigated the external validity of KOA trials.[4-7] According to these studies, majority of KOA trials reported information on external validity insufficiently.[4] Different criteria were employed even for items commonly observed in trials, thus making the generalization challenging.[5,6]

Previously, a possibility of delineating external validity was presented by calculating proportions of patients at each stage of the recruitment process (Figure 1).[8] Because flow diagrams depicting the recruitment process were emphasized in the Consolidated Standards of Reporting Trials (CONSORT) statement,[9] these scales are expected to be clearly calculated from KOA trials. Therefore, in this study, we aimed
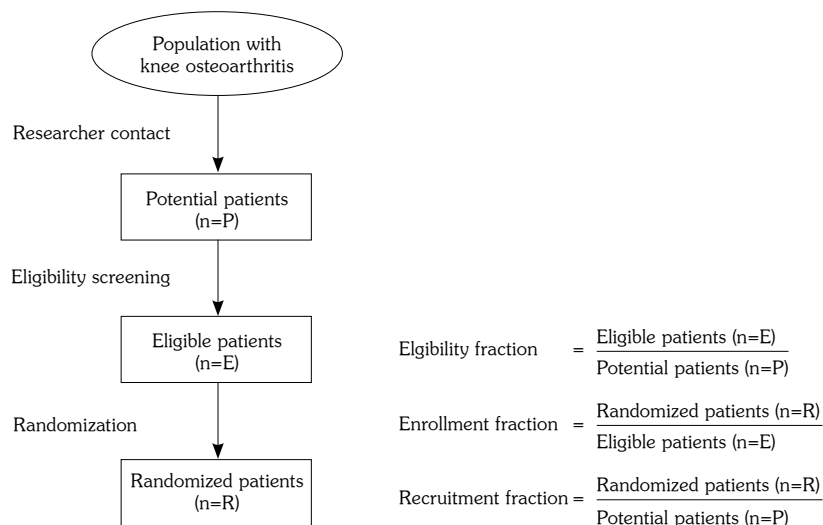
Figure 1. Trial recruitment process.

to investigate the external validity of KOA trials through a systematic review of randomized, placebo-controlled, clinical trials.

MATERIALS AND METHODS

The search strategy used has been described previously.[10,11] In brief, the first search was performed in PubMed, SCOPUS, and the Cochrane Central Register of Controlled Trials up to December 2011 using the terms knee arthritis, KOA, gonarthritis, and gonarthrosis with the limits set to trials. The second search was carried out in Cochrane Reviews using the same terms. Additional trials were sought by referencing the retrieved reviews. Finally, the search was expanded to all studies referenced in the trials which were found. Based on previous raw data set,[5,10-13] the first author selected randomized, placebo-controlled trials written in English.

We independently extracted data including participant flow (i.e. number of patients screened, number of patients eligible, and number of patients randomized) from each trial. We also extracted eligible patients' reasons for non-participation and corresponding values. We then calculated the eligibility, enrollment, and recruitment fractions (Figure 1).[8] The eligibility fraction was defined as a proportion of eligible patients in screened patients, the enrollment fraction as a proportion of randomized patients in eligible patients, and the recruitment fraction as a proportion of randomized patients in screened patients. Furthermore, we calculated the number needed to screen (NNS) to randomize one patient in a trial by dividing one by the recruitment fraction.[8] Finally, groups reporting or not reporting the recruitment data were compared using Chi-square test and Fisher's exact test.

Since NNS was considered as an essential scale in original study,[8] we investigated if the NNSs were affected by trial characteristics such as center type, publication year, treatment type, delivery route, and flare design. Group differences were identified using an analysis of variance. We also examined the NNS categorized by individual treatments. Because the tested treatments were diverse, we explained group differences in a descriptive manner using a median value as a reference point. When necessary, we contacted authors of included trials. Statistical significance was defined as $p<0.05$. STATA version 11.0 (StataCorp. LP, College Station, Texas, USA) was used for all analyses.

RESULTS

We identified 36,691 citations (PubMed, 1,646; SCOPUS, 32,246; Cochrane Registered Trials, 2,484; and extra source, 315), of which 354
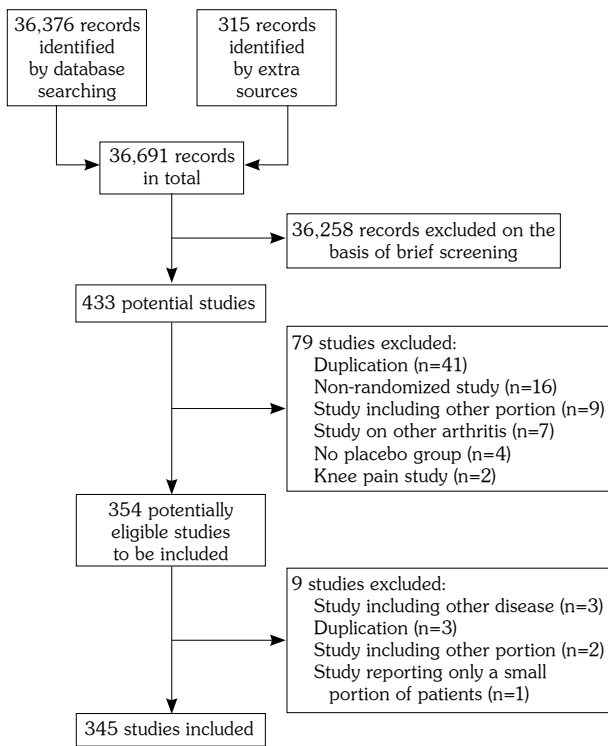
Figure 2. Study flow diagram.

reports were identified as potentially eligible for our analyses (Figure 2). Of these, a total of nine reports were excluded since some included other diseases (n=3), some were duplicates (n=3), some

reported data combined over the hip and knee joints (n=2), or included randomized patients of both genders but reported only female patients (n=1). The total number of reports analyzed was 345, which presented 352 trials.

Table 1 describes the characteristics of the trials included in our analysis. A small number of trials (33%) reported information about the number of screened patients. Reporting the number of patients screened was significantly associated with the publication year and flare design. The number of patients who were eligible was reported in 82 (24%) trials. Reporting the number of patients eligible was significantly related to the center type and publication year. Overall, trials published after 2000 tended to report sufficient information about the recruitment process.

Table 2 presents data on the eligibility, enrollment, and recruitment fractions. With respect to the eligibility fraction, only 75 (21%) trials provided sufficient data for calculation. In these trials, the median proportion of potential patients who were eligible for trials was 71.9% (interquartile range: 52.7 to 86.5%). Regarding the enrollment fraction, 82 (23%) trials presented sufficient information. A median of 92.9% of eligible patients was randomized in these trials (interquartile range: 82.5 to 100%). Of these,

Table 1. Trial characteristics

	Total trials		Trials reporting number of patients screened			Trials reporting number of patients eligible		
	n	%	n	%	p	n	%	p
Total	352	100	115	33		82	24	
Center type								
Single-center	157	45	50	43	0.056	49	60	0.049
Multicenter	146	41	62	54		31	38	
Unclear	49	14	3	3	<0.000	2	2	<0.000
Publication year								
1950-1999	97	28	7	6	<0.000	3	4	<0.000
2000-2012	255	72	109	95		79	96	
Treatment type								
Pharmacological	259	74	86	75		56	68	
Non-pharmacological	92	26	29	25	=0.751	25	30	=0.107
Mixed	1	0	0	0		1	1	
Delivery route*								
Oral	157	46	52	47		36	45	
Invasive	99	29	30	27	=0.972	20	25	=0.322
Other	88	26	28	25		24	30	
Flare design								
Yes	32	9	17	15	=0.010	9	11	=0.498
No	320	91	98	85		73	89	

* Eight trials administering treatments via mixed routes were excluded.

Table 2. Eligibility, enrollment, and recruitment fractions

	Trials reporting data (n)	Median (%)	Interquartile range
Eligibility fraction	75	71.9	52.7-86.5
Enrollment fraction	82	92.9	82.5-100
Recruitment fraction	115	67.9	48.9-82.9

21 trials reported that 100% of eligible patients were randomized in a trial. In the remaining trials, reasons for why eligible patients were not randomized were reported as "refusing to participate" (38%), which was the most common, followed by "no interest" (17%), and "other" (16%). The overall recruitment fraction varied greatly across trials. In the 116 (33%) trials that reported adequate data for analysis, the median recruitment fraction was 67.9% (interquartile range: 48.9 to 82.9%). The median NNS was 1.5 (range: 1 to 10). To summarize, some trials randomized *every* potential patient screened for eligibility, whereas others screened as many as 10 potential patients for each patient finally randomized.

Table 3 shows the relationships between NNS and trial characteristics. While no significance was detected in the publication year, treatment type, delivery route, and flare

design, significance was detected for the center type (p=0.009). For example, pharmacological and non-pharmacological trials did not screen patients differently for randomization. In a post hoc analysis for center type, single center trials screened significantly more patients to randomize one patient than multicenter trials (p=0.003).

Figure 3 illustrates the recruitment fraction and NNS, categorized by 32 individual treatments. In most treatments, a diverse range of NNSs, centered at the median value, were found. For six (19%) treatments, including anthraquinone and antibiotics, trials randomized higher proportions of screened patients than the median trial. Meanwhile, for 11 (34%) treatments, including opioid and pulsed electromagnetic field, trials randomized lower proportions of screened patients than the median trial.

Table 3. Relationship between number of patients needed to screen to randomize one patient and trial characteristics

	Trials reporting data	Number needed to screen to randomize one patient		
	n	Median	Interquartile range	*p*
Total	115	1.5	1.2-2.1	
Center type				
Single-center	50	1.7	1.3-2.7	
Multicenter	62	1.4	1.2-1.8	0.009*
Unclear	3	1.3	1.2-5.0	
Publication year				
1950-1999	7	1.3	1.2-1.3	0.192
2000-2012	108	1.5	1.2-2.1	
Treatment type				
Pharmacological	86	1.4	1.2-1.9	0.187
Non-pharmacological	29	1.7	1.4-3.0	
Mixed	0			
Delivery route†				
Oral	52	1.5	1.2-2.0	
Invasive	30	1.4	1.2-2.0	0.622
Other	28	1.7	1.2-2.3	
Flare design				
Yes	17	1.6	1.4-2.0	0.467
No	98	1.4	1.2-2.1	

* In a post hoc analysis, a significant difference was found between single center and multicenter trials (p=0.003);
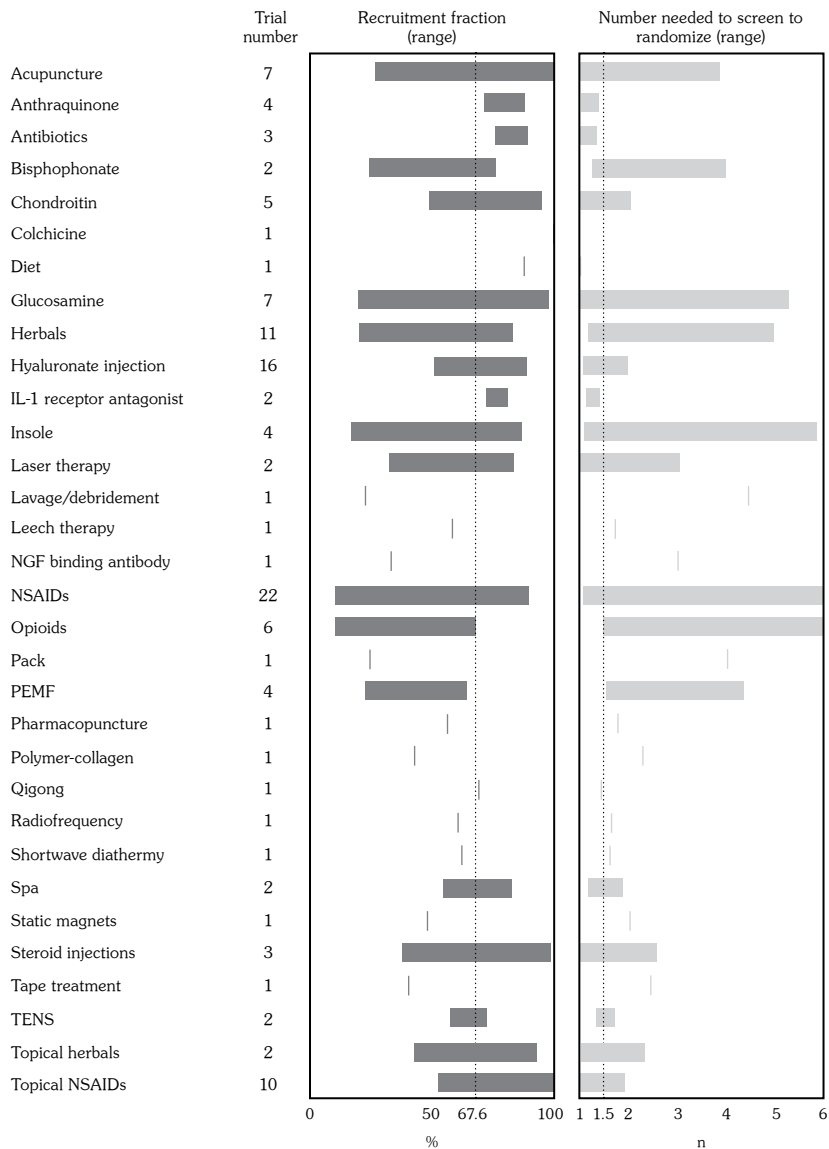† Five trials administering treatments via mixed routes were excluded.

Figure 3. Recruitment fraction and number needed to screen to randomize one patient. For simplicity, only single treatments are presented. IL-1, interleukin-1; NGF: Nerve growth factor; NSAID: Non-steroidal antiinflammatory drug; PEMF: Pulsed electromagnetic field; TENS: Transcutaneous electrical nerve stimulation.

DISCUSSION

A number of previous studies have argued that the recruitment process necessary for calculating three fractions was poorly described in many randomized trials.[8,14-18] While some stages were well reported, others were depicted incompletely. In cases of trials published in major journals, 85-90% reported the enrollment fraction, whereas only 40-60% reported the eligibility fraction.[8,14]

Nevertheless, these values were definitely higher than our finding of 21-33%.

It can be argued that analyzing trials published only after 2000 may increase these proportions since description of the recruitment process was improving in 2000s.[16] Indeed, previous studies analyzed trials published after 2000,[8,14-18] whereas our study included all trials published after 1955. Our examination of KOA trials published after

Table 4. Eligibility, enrollment, and recruitment fractions categorized by publication year

	Total trials		Median	Interquartile range
	n	%	%	
Publication year 1950-1999	97	100		
Eligibility fraction	3	3	87.5	72.4-99.2
Enrollment fraction	3	3	87.0	77.2-96.1
Recruitment fraction	7	7	76.7	63.0-84.1
Publication year 2000-2012	255	100		
Eligibility fraction	72	28	71.5	52.2-84.7
Enrollment fraction	79	31	93.1	83.2-100
Recruitment fraction	108	42	66.2	48.6-82.2

2000 revealed that a majority of the trials actually reported the number of patients screened or eligible compared with those published before 2000 (Table 4). Nonetheless, the proportions of trials reporting three fractions were still low.

Considering the CONSORT statement, this poor reporting of the recruitment process in the KOA trials is surprising. Ever since the CONSORT statement was first published, it has strongly recommended that the possibility of whether trial findings are applicable in clinical practices should be discussed in trial reporting.[19-21] To better provide information on external validity, the CONSORT statement also emphasized depicting all stages of flow diagrams including enrollment in 2001.[20] However, such effort appears to be unsuccessful in the KOA area since a majority of KOA trials omitted data necessary for calculating the fractions.

The omission may be associated with the fact that the recruitment process regarding external validity was not actually considered in the CONSORT statement. In fact, the CONSORT guidelines encouraged clinicians to focus on characteristics of trial participants or results of previous works.[22] For example, one behavioral program was considered as a reliable treatment in different settings because a subsequent large-scale trial[23] successfully replicated a previous work.[24] For this reason, it was argued that the CONSORT statement lacked the reporting of external validity.[25] Alternative frameworks were proposed, such as qualitative studies or quality assessment checklists.[25,26]

In the meantime, one method that aims to gauge external validity quantitatively was proposed.[8] Specifically, an outcome of NNS was magnified. Since it has been shown that trial participants are

different from eligible non-participants,[27,28] trials showing lower NNSs may be more generalizable. According to many studies, the median NNSs were 1.8 in trials of major journals,[8] 2-5.6 in cancer trials,[16,18] and 2.4 in primary care trials.[17] Clearly, these results were greater than our finding of 1.5, indicating that these trials screened more patients than KOA trials to randomize one patient.

Meanwhile, a previous study showed that the median NNS was 1.3 in rheumatoid arthritis trials,[15] which is consistent with our finding. This study suggested that rheumatoid arthritis trials were excellently generalizable since three patients were screened for every two enrolled in the trials.[15] Because KOA trials also screened three patients for every two enrolled, it can be argued in general that KOA trials were of good quality for external validity.

It may also be argued that not all KOA trials were generalizable, since 34% of treatments demonstrated their efficacy on highly selected patients compared with the median trial. However, it should be noted that NNS reflects whole clinical situations in a trial (e.g. center, care provider, eligibility criteria, etc.). Even trials showing high NNSs may be conducted so well that the trial results can be applied to a broader spectrum of patients. In this respect, the NNS should be considered as an indicator that determines the degree of challenge in enrolling patients who are screened. Therefore, clinicians who use such treatments should consider recruitment data in relation to whole information shown in the trials.

We also found that NNS was affected by the center type. Researchers in a single center may be better informed about the aim of trials, thus apply strict eligibility criteria to patients and may

need more patients for screening. Consequently, this may result in more pronounced effects of test treatments since highly selected patients may show better outcomes. In fact, single-center trials showed superior treatment effects than multicenter trials.[29-31] Therefore, further studies are required to demonstrate whether treatments on KOA show superior efficacy in single-center trials.

In conclusion, only 21-33% of 352 KOA trials provided sufficient data on the recruitment process necessary for calculating three fractions. This low reporting rate was contrary to the CONSORT statement recommending detailed reporting of the recruitment process. On closer analysis using available recruitment data, KOA trials were, in general, excellently generalizable. However, 34% of treatments were tested on highly selected patients compared with the median trial. Therefore, clinicians who wish to use such treatments should consider all clinical information and further trials should document the recruitment data in detail to help clinicians determine the applicability of trial findings in clinical practice.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

REFERENCES

1. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, part I: critical appraisal of existing treatment guidelines and systematic review of current research evidence. Osteoarthritis Cartilage 2007;15:981-1000.
2. Zhang W, Moskowitz RW, Nuki G, Abramson S, Altman RD, Arden N, et al. OARSI recommendations for the management of hip and knee osteoarthritis, Part II: OARSI evidence-based, expert consensus guidelines. Osteoarthritis Cartilage 2008;16:137-62.
3. Zhang W, Nuki G, Moskowitz RW, Abramson S, Altman RD, Arden NK, et al. OARSI recommendations for the management of hip and knee osteoarthritis: part III: Changes in evidence following systematic cumulative update of research published through January 2009. Osteoarthritis Cartilage 2010;18:476-99.
4. Ahmad N, Boutron I, Moher D, Pitrou I, Roy C, Ravaud P. Neglected external validity in reports of randomized trials: the example of hip and knee osteoarthritis. Arthritis Rheum 2009;61:361-9.
5. Koog YH, Wi H, Jung WY. Eligibility criteria in knee osteoarthritis clinical trials: systematic review. Clin Rheumatol 2013;32:1569-74.
6. Liberopoulos G, Trikalinos NA, Ioannidis JP. The elderly were under-represented in osteoarthritis clinical trials. J Clin Epidemiol 2009;62:1218-23.
7. Purepong N, Jitvimonrat A, Sitthipornvorakul E, Eksakulkla S, Janwantanakul P. External validity in randomised controlled trials of acupuncture for osteoarthritis knee pain. Acupunct Med 2012;30:187-94.
8. Gross CP, Mallory R, Heiat A, Krumholz HM. Reporting the recruitment process in clinical trials: who are these patients and how did they get there? Ann Intern Med 2002;137:10-6.
9. Egger M, Jüni P, Bartlett C; CONSORT Group (Consolidated Standards of Reporting of Trials). Value of flow diagrams in reports of randomized controlled trials. JAMA 2001;285:1996-9.
10. Koog YH, Gil M, We SR, Wi H, Min BI. Barriers to participant retention in knee osteoarthritis clinical trials: a systematic review. Semin Arthritis Rheum 2013;42:346-54.
11. Ryang We S, Koog YH, Jeong KI, Wi H. Effects of pulsed electromagnetic field on knee osteoarthritis: a systematic review. Rheumatology (Oxford) 2013;52:815-24.
12. Koog YH. Caution should be observed against the last observation carried forward analysis in opioid trials. Turk J Rheumatol 2013;28:282-3.
13. We SR, Jeong EO, Koog YH, Min BI. Effects of nutraceuticals on knee osteoarthritis: systematic review. Afr J Biotechnol 2012;11:2814-21.
14. Toerien M, Brookes ST, Metcalfe C, de Salis I, Tomlin Z, Peters TJ, et al. A review of reporting of participant recruitment and retention in RCTs in six major journals. Trials 2009;10:52.
15. Simsek I, Yazici Y. Incomplete reporting of recruitment information in clinical trials of biologic agents for the treatment of rheumatoid arthritis: a review. Arthritis Care Res (Hoboken). 2012;64:1611-6.
16. Treweek S, Loudon K. Incomplete reporting of recruitment information in breast cancer trials published between 2003 and 2008. J Clin Epidemiol 2011;64:1216-22.
17. Jones R, Jones RO, McCowan C, Montgomery AA, Fahey T. The external validity of published randomized controlled trials in primary care. BMC Fam Pract 2009;10:5.
18. Wright JR, Bouma S, Dayes I, Sussman J, Simunovic MR, Levine MN, et al. The importance of reporting patient recruitment details in phase III trials. J Clin

Oncol 2006;24:843-5.

19. Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, et al. Improving the quality of reporting of randomized controlled trials. The CONSORT statement. JAMA 1996;276:637-9.

20. Moher D, Schulz KF, Altman DG. The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. Lancet 2001;357:1191-4.

21. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. Int J Surg 2011;9:672-7.

22. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. J Clin Epidemiol 2010;63:1-37.

23. Garber J, Clarke GN, Weersing VR, Beardslee WR, Brent DA, Gladstone TR, et al. Prevention of depression in at-risk adolescents: a randomized controlled trial. JAMA 2009;301:2215-24

24. Clarke GN, Hornbrook M, Lynch F, Polen M, Gale J, Beardslee W, et al. A randomized trial of a group cognitive intervention for preventing depression in adolescent offspring of depressed parents. Arch Gen Psychiatry 2001;58:1127-34.

25. Bonell C, Oakley A, Hargreaves J, Strange V, Rees R. Assessment of generalisability in trials of health interventions: suggested framework and systematic review. BMJ 2006;333:346-9.

26. Bornhöft G, Maxion-Bergemann S, Wolf U, Kienle GS, Michalsen A, Vollmar HC, et al. Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity. BMC Med Res Methodol 2006;6:56.

27. Britton A, McKee M, Black N, McPherson K, Sanderson C, Bain C. Choosing between randomised and non-randomised studies: a systematic review. Health Technol Assess 1998;2:1-124.

28. Koog YH, Min BI. Does random participant assignment cause fewer benefits in research participants? Systematic review of partially randomised acupuncture trials. J Altern Complement 2009;15:1107-13.

29. Dechartres A, Boutron I, Trinquart L, Charles P, Ravaud P. Single-center trials show larger treatment effects than multicenter trials: evidence from a meta-epidemiologic study. Ann Intern Med 2011;155:39-51.

30. Bafeta A, Dechartres A, Trinquart L, Yavchitz A, Boutron I, Ravaud P. Impact of single centre status on estimates of intervention effects in trials with continuous outcomes: meta-epidemiological study. BMJ 2012;344:813.

31. Unverzagt S, Prondzinsky R, Peinemann F. Single-center trials tend to provide larger treatment effects than multicenter trials: a systematic review. J Clin Epidemiol. 2013;66:1271-80.