Cukurova Medical Journal
ÇUKUROVA ÜNİVERSİTESİ TIP FAKÜLTESİ

Cukurova Med J 2019;44(4):1310-1316
DOI: 10.17826/cumj.512585

ARAŞTIRMA / RESEARCH

Evaluation of statistical significance of randomized controlled trials using diagnostic imaging techniques with fragility index

Tanısal görüntüleme teknikleri kullanılan randomize kontrollü çalışmaların istatistiksel anlamlılığının kırılganlık indeksi ile değerlendirilmesi

Didem Derici Yıldırım[1], Bahar Taşdelen[1]

[1]Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin, Turkey

Cukurova Medical Journal 2019;44(4):1310-1316.

Abstract

Purpose: The aim of our study was to evaluate the robustness and success of fragility index (FI) for randomized controlled trials (RCT) utilized diagnostic imaging techniques.

Materials and Methods: We performed a systematic survey of RCTs using the terms of "randomized controlled trial", "exact test" and "diagnostic imaging" in PubMed. Two researchers independently reviewed the abstracts and identified the studies according to selection criteria.

Results: The median FI was 4.0 [1.0-4.0] and the median sample size was 83.5[36.0-148.0]. Fifty percent of RCTs reported p value was between 0.001 and 0.05. After one more event added, 94.4 % of studies became non-significant with a p value between 0.05 and 0.10. There was no significant correlation between the FI's and the sample sizes and outcome events. (r=0.144, p=0.570; r=0.169, p=0.504) There was a statistically significant difference between reported p value groups in terms of FI.

Conclusion: FI should be given along with the p value for binary outcomes. Researchers should be careful with the interpretation of FI because it only shows the level of evidence and the power of the statistical significance.

Keywords: Randomized controlled trials, fragility index, diagnostic imaging.

Öz

Amaç: Bu çalışmanın amacı kırılganlık indeksinin (Kİ) başarısını, tanı amaçlı görüntüleme yöntemleri kullanılan randomize kontrollü çalışmalar (RKÇ) için değerlendirmektir.

Gereç ve Yöntem: Pubmed veri tabanında "Randomize Kontrollü Çalışma", "Kesin Test" ve "Tanısal Görüntüleme" terimlerini içeren bir sistematik araştırma yapılmıştır. İki araştırmacı birbirinden bağımsız olarak özetleri ve çalışmaları seçim kriterlerine göre incelemiştir.

Bulgular: Kİ medyan değeri 4.0 [1.0-4.0] ve örnek genişliğinin medyan değeri 83.5[36.0-148.0] olarak bulunmuştur. Dahil edilen çalışmaların %50'sinin p değeri 0.001 ile 0.05 arasında yer almaktadır. Bir kişi daha eklendiği durumda çalışmaların %94,4'ünün p değeri anlamsız hale gelmiştir. Kİ ile örnek genişliği ve ilgilenilen olayın gözlendiği kişi sayısı arasında anlamlı bir ilişki yoktur. (r=0.144, p=0.570; r=0.169, p=0.504) Rapor edilen p değeri grupları arasında Kİ değerleri bakımından istatistiksel olarak anlmalı bir farklılık vardır. (

Sonuç: İkili sonuç değişkenleri için Kİ p değeri ile birlikte verilmelidir. Araştırmacılar Kİ'yi yorumlarken dikkatli olmalıdır çünkü Kİ sadece çalışmanın kanıt değerini ve istatistiksel gücünü göstermektedir.

Anahtar kelimeler: Randomize kontrollü çalışmalar, kırılganlık indeksi, tanısal görüntüleme.

INTRODUCTION

Diagnostic imaging techniques like magnetic resonance imaging (MRI), ultrasound (US) or X-ray have been used by physicians to make diagnoses for nearly hundred years. For physicians, the integration of the best evidence while determining the most successful imaging method according to the diagnosis is very important. In recent years, the evidence based medicine (EBM) has become popular in making clinical decisions. EBM is the composite of clinical experiment of physicians, patient status and best evidence. Randomized controlled trials (RCTs) are the gold standard of evidence and recommended to

Yazışma Adresi/Address for Correspondence: Dr. Didem Derici Yıldırım, Mersin Üniversitesi Tıp Fakültesi, Biyoistatistik ve Tıbbi Bilişim Anabilim Dalı, Mersin, Turkey E-mail: didemderici@hotmail.com
Geliş tarihi/Received: 14.01.2019 Kabul tarihi/Accepted: 03.04.2019 Çevrimiçi yayın/Published online: 15.09.2019

evaluate the effectiveness of diagnostic imaging techniques because confounding factors and potential bias can be controlled and minimized with randomization[1,2,3].

Traditionally, the p value of 0.05 has been used to demonstrate the efficiency of a specific method, treatment or intervention statistically in RCTs. However, researches have been discussing the validity of this statement for a long time. There are some limitations in assessing treatment effects by a p value, especially related to studies with binary outcome. One limitation is that the p value of a study with large sample sizes and small effect sizes is similar to another with small sample sizes and large effect sizes. In addition, the p value changes according to the event rates in these trials. Especially, the Fisher exact test or the chi square tests are used in the studies which are evaluating the effectiveness of diagnostic techniques. Making a decision only according to the p value is sometimes misleading in these types of studies[4,5,6]. For example, a randomized controlled trial in which 23 patients with angioscopy for end luminal evaluation and 19 conventional (CON) *in situ* grafting were assessed in terms of wound morbidity, morbidity was seen in 6 of 23 patients of the ANG group and 12 of 19 patients of the CON group. According to the Fisher Exact test result, the p value was calculated as 0.043. But if only one more morbidity in the CON group were detected, the trial result would not remain statistically significant[7]. FI for the trial was 1 event, that means adding only one event to one group changed the result from significant to non-significant. Therefore, the quality of RCTs is controversial because just few events in one group alter the hypothesis, which is considered statistically significant.

To our knowledge, there are no studies present evaluating FI in RCTs using diagnostic imaging techniques although the Fisher Exact test is used commonly. Our aim in this study is to evaluate the robustness of randomized controlled trials utilized diagnostic imaging techniques.

MATERIALS AND METHODS

Selection of studies

In the first phase, PubMed was systematically searched for RCTs using the terms of "randomized controlled trial [Publication Type]", "exact test [Title/Abstract])" and "diagnostic imaging [MeSH Terms]". Diagnostic imaging techniques in MESH terms were included in the list which is given in Table 1. We filtered PubMed to detect all types of Clinical Trial publications which have free full texts and are concerned with human studies. In our search strategy, there was no year or journal restriction.

Table 1. Diagnostic imaging techniques and procedures

Diagnostic imaging
• Brain Mapping
• Cardiac-Gated Imaging Techniques
• Image interpretation, Computer-Assisted
• Imaging, Three Dimensional
• Magnetic Resonance Imaging
• Microscopy
• Molecular Imaging
• Photography
• Radiography
• Radionuclide Imaging
• Respiratory-Gated Imaging Techniques
• Spectroscopy, Near-Infrared
• Stroboscopy
• Subtraction Technique
• Terahertz Imaging
• Tomography
• Transillumination
• Ultrasonography
• Voltage-Sensitive Dye Imaging
• Whole Body Imaging

In the second phase, two researchers independently reviewed the abstracts and identified the studies according to selection criteria which were specified above. Only one result was taken into account from studies which reported more than one significant result. The flow diagram of the studies which were evaluated systematically is given in Figure 1. The characteristics of included studies are given in Table 2[7-24].

$$p = ((a+b)!(c+d)!(a+c)!(b+d)!) / a!\,b!\,c!\,d!\,N! \quad (1)$$

Statistical analysis

The Fisher Exact test is used for 2x2 contingency tables when 25% percent and more of the expected cells are smaller than 5. It gives a better approximation to the exact probability under small sample sizes than the Pearson Chi-Squared test. The Exact test probability is calculated by the formula given in Equation 1[25]. An example table required for

the calculation of Fisher Exact test is given in Table 3.

Table 3. Two by two contingency table

	Event	No Event	Total
Diagnostic Method I	a	b	a+b
Diagnostic Method II	c	d	c+d
Total	a+c	b+d	N

The new approach, called the Fragility Index (FI) for 2x2 contingency tables, has been proposed as a measure of weakness for RCTs. For a trial which reported a statistical significant result, the FI is the minimum number of "non-events" that changed to "events" to get the non-significant p value[26]. The study results are more robust when the FI is larger.

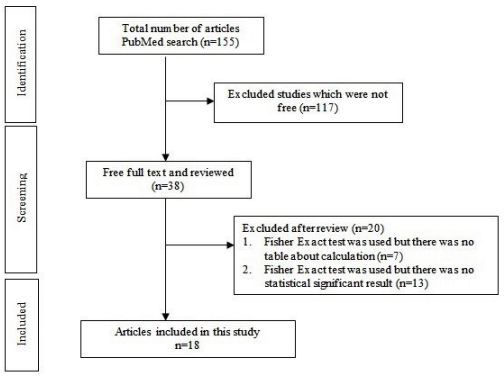

Figure 1. Flow diagram of studies reviewed systematically

In the calculation step of the FI, events are added iteratively to the smallest event group until the statistical test becomes non-significant. If a<c, the exact p value is calculated as

$$p = ((a+b+1)!(c+d)!(a+c+1)!(b+d)!)/(a+1)!\,b!\,c!\,d!\,(N+1)! \quad (2)$$

and, if the p value is greater than 0.05, the process is stopped. Otherwise, one more event is added until the non-significant p value is obtained. Then, the added event number is FI. Statistica 13.3.1 (TIBCO, CA, USA) was used for statistical analysis.

RESULTS

In the third phase, randomized controlled trials were examined in terms of sample sizes, the number of events and the reported p values. Besides the p value of the primary outcome for each study, the Fragility indices were obtained using the formulation reported by Walsh et al.[26,27]. We reviewed 38 studies, of which 18 studies were appropriate for the calculation of the FI. These studies met all inclusion criteria mentioned above. Trial statistics of included studies are given in Table 4.

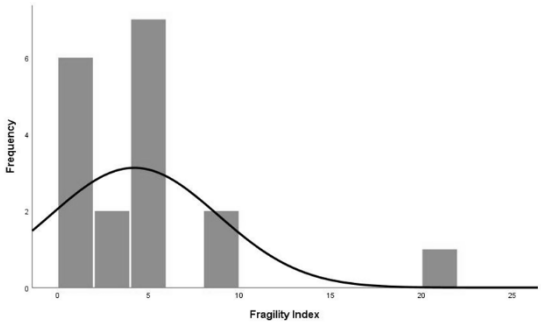

Figure 2. The distribution of Fragility Indices for diagnostic imaging studies obtained from systematic review

The median FI was 4.0 [1.0-4.0] and the median sample size was 83.5[36.0-148.0]. The median of the event numbers (both in treatment and control group) was 18 [13-32]. Fifty percent of RCTs reported p values between 0.001 and 0.05.

After one more event added, 94.4 % of the studies became non-significant with a p value between 0.05 and 0.10. There was no significant correlation between the FI and sample sizes and outcome events. (r=0.144, p=0.570; r=0.169, p=0.504) The histogram of the fragility index is given in Figure 2. There were 6 articles with the FI as 1. The publication years of the studies were between 1996 and 2017. Studies with zero fragility indices were excluded.

DISCUSSION

The statistical significance generally depends on the p value in medical research[6]. However, researches have been debated on this fact for years although the p value's application and interpretation is very easy. Previous studies reported that statistically significant results could be changed just with a single event. An approach named Fragility Index was introduced by Feinstein and then expanded by Walter et al. to solve this problem[28].

Table 2. The sample size, group size, the events number of each group, reported p and FI of included studies

#	Outcome	n	Treatment Group	Treatment Event	Control Group	Control Event	p	FI
7	Incidence of wound morbidity	42	19	12	23	6	0.043	1
8	Frequency of remnants of high-viscosity glass-ionomer sealants	201	157	2	44	12	<0.0001	20
9	Baseline low-luminance visual acuity (LLVA)	492	246	20	246	5	0.0010	4
10	The rate of positive surgical margins (PSMs)	239	125	20	114	31	0.035	1
11	Improvement in self-reported pain score	147	112	51	35	1	<0.0001	9
12	Early (<30 day) morbidity; incidence of device-related complications	41	22	13	19	19	0.0055	4
13	Efficacy of ultrasound (US)-guided three-in-one femoral nerve blocks	36	18	12	18	1	<0.0001	5
14	Change of Oswestry Disability Index (ODI) scores	87	42	7	45	26	<0.01	8
15	Pain intensity reduction	71	52	40	19	8	<0.001	2
16	The microtensile bond strength (μTBS)	80	40	14	40	2	0.001	4
17	The effectiveness of pancreatic duct (PD) stent placement	120	60	1	60	8	0.0322	1
18	The Mucus Shaver performance	24	12	1	12	10	<0.001	4
19	The effect of home-use bleaching agents	20	10	8	10	2	0.023	1
20	The Hobbs clinical score for treated limbs	124	70	13	54	19	0.038	1
21	The effect of bipolar radial head prosthesis replacement	22	8	1	14	13	0.0004	4
22	A modified way of performing Vacuum-Assisted Breast Biopsy.	18	9	4	9	9	0.029	1
23	Radiographic progression in patients with ankylosing spondylitis (AS).	214	111	15	103	4	0.034	2
24	The effect of intravenous N-acetylcysteineon renal function	100	46	14	54	4	0.004	4

*FI: Fragility Index

Table 4. Trial statistics of included studies

Sample Size*	83.5[36.0-148.0]
The number of event *	18.0[13.0-32.0]
Reported p valueϕ	n=18
• < 0.0001	• 3(16.7)
• 0.001-0.0001	• 6(33.3)
• > 0.001-0.05	• 9(50.0)
P value after one more eventϕ	n=18
• 0.05-0.10	• 17(94.4)
• >0.10	• 1(5.6)
Fragility Index *	4.0[1.0-4.0]

***** Data was presented as median[25.Percentile-75.Percentile],
ϕ Data was presented as count(percentage)

FI was calculated for different medical literatures as urologic, orthopedic surgery, giant cell arthritis, anesthesiology and etc.[29-32]. There was no research found, which evaluated the robustness of the results for randomized controlled trials using diagnostic imaging techniques and the Fisher exact test together

before. According to previous results, testing the hypothesis only according to the p value was insufficient. In Evidence Based Medicine, it was recommended that the p value should be evaluated with additional metrics like 95% confidence intervals. But there was no metric for categorical data analysis besides the chi square or the fisher exact test, therefore the FI should be used for this purpose.

The median FI for our study was 4. This meant that the result of the study became non-significant by adding only 4 patients. The previous studies reported that the median FI was calculated as 2 in spine surgery [27], 3 in urology literature [29], 4 for anesthesiology trials[31] and 4 for giant cell arteritis[32]. The results of our study were compatible with published results evaluating the FI in other fields. We expected to find significant differences between the p value groups in terms of the FI. There was a statistically significant difference between the three following groups ($p<0.001$). The median [25Q-75Q] FI was 9.0[5.0-20.0] for the <0.0001 group, 4.0[2.0-8.0] for the 0.0001-0.001 group and 1.0[1.0-4.0] for the 0.001-0.05 group, respectively. The difference originated in the third group (0.001-0.05 group). It was statistically different from the first (<0.0001) and the second (0.0001-0.001) group ($p=0.008$, $p=0.010$).

It is known that the FI depends on the number of events. Studies with small number of events and sample sizes have a potential bias for false significant treatment effects. There were some studies that suggested at least 650 events to evaluate the treatment effect correctly in trials related to simulation results[33]. But many studies, utilizing diagnostic imaging techniques, used only qualified small sample sizes and small number of events. Therefore, an additional metric like the FI should be reported together with the results. On the other hand, a strong negative correlation between the FI and sample sizes was detected in some studies [27-31]. In our study, we could not detect such a correlation because most trials included in our review had small sample sizes. We thought that the ratio of sample sizes in the treatment and the control group affected the value of the FI because a high ratio of sample sizes (n_T/n_c) gives a high FI. In addition, relative risk, risk reduction, absolute risk and their confidence intervals can be calculated besides FI. Especially 95% confidence intervals of relative risk reduction denote the statistical significance[34]. This metrics may be helpful for more accurate decision.

The usage and interpretation of the FI is considered by researchers carefully. The FI should not be interpreted as a metric like p values. It should be always given with other metrics obtained from the study. A simulation study demonstrated that there was an inverse correlation between the FI and the p value[35]. When the p value is low, the FI is calculated higher but it does not provide significant clinical information about the diagnostic method or treatment. Therefore, the magnitude of the FI does not show that the study has a greater clinical effect. It only demonstrates the level of evidence and the power of the statistical significance.

Our study was limited by studies with small sample sizes. This limitation caused to underestimate the correlations between the FI and the trial characteristics. Therefore, our results, except the median of the FI and the p value group differences, were not consistent with earlier study results in medical literature. This study was limited because the FI could only be calculated for binary outcomes presented by two by two contingency tables and the exclusion of non-significant randomized controlled trials. Unfortunately, there was not a cut-off value for the evaluation of the FI.

Therefore, the FI should be given along with the p value for binary outcomes. There are some studies that plan the calculation of the FI for continuous outcomes. With this development, the usage of the FI increases because many RCTs investigate the differences and correlations between continuous or discrete outcomes. During this study, we have planned a simulation study which will show the effect of sample sizes for each group on the FI.

Yazar Katkıları: Çalışma konsepti/Tasarımı: DDY, BT; Veri toplama: -; Veri analizi ve yorumlama: DDY, BT; Yazı taslağı: DDY, BT; İçeriğin eleştirel incelenmesi: DDY, BT; Son onay ve sorumluluk: DDY, BT; Teknik ve malzeme desteği: -; Süpervizyon: DDY, BT; Fon sağlama (mevcut ise): yok.
Bilgilendirilmiş Onam: Katılımcılardan yazılı onam alınmıştır.
Hakem Değerlendirmesi: Dış bağımsız.
Çıkar Çatışması: Yazarlar çıkar çatışması beyan etmemişlerdir.
Finansal Destek: Yazarlar finansal destek beyan etmemişlerdir.
Teşekkür: Bu akademik çalışma dilsel olarak Mersin Üniversitesi Mersin Teknoloji Transfer Ofisi Akademik Yazı Merkezi tarafından desteklenmiştir.

Author Contributions: Concept/Design : DDY, BT; Data acquisition: -; Data analysis and interpretation: DDY, BT; Drafting manuscript: DDY, BT; Critical revision of manuscript: G DDY, BT; Final approval and accountability: DDY, BT; Technical or material support: -; Supervision: DDY, BT; Securing funding (if available): n/a.
Informed Consent: Written consent was obtained from the participants.
Peer-review: Externally peer-reviewed.
Conflict of Interest: Authors declared no conflict of interest.
Financial Disclosure: Authors declared no financial support
Acknowledgment: This academic work was linguistically supported by the Mersin Technology Transfer Office Academic Writing Center of Mersin University.

REFERENCES

1. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine-new approaches and challenges. Acta Inform Med. 2008;16:219-25.
2. Bowers A, Meyer C, Tritz D, Cook C, Fuller K, Smith C et al. Assessing quality of randomized trials supporting guidelines for laparoscopic and endoscopic surgery. J Surg Res. 2018;224:233-39.
3. Lalumera E, Fanti S. Randomized controlled trials for diagnostic imaging: conceptual and pratical problems. Topoi. 2017;1-6.
4. Chakkera HA, Schold JD, Kaplan B. P value: significance is not all black and white. Transplantation. 2016;100:1607-9.
5. Pocock SJ, Stone GW. The primary outcome is positive is that good enough? N Engl J Med. 2016;375:971–9.
6. Shen Y, Cheng X, Zhang W. The fragility of randomized controlled trials in intracranial hemorrhage. Neurosurg Rev. 2019;42:9-14.
7. Wilson YG, Davies AH, Currie IC, McGrath C, Morgan M, Baird RN et al. Angioscopically-assisted in situ saphenous vein bypass for infrainguinal revascularisation. Eur J Vasc Surg. 1996;12:223-9.
8. Hu X, Zhang W, Fan M, Mulder J, Frencken JE. Frequency of remnants of sealants left behind in pits and fissures of occlusal surfaces after 2 and 3 years. Clin Oral Investig. 2017;21:143-9.
9. Frenkel RE, Shapiro H, Stoilov I. Predicting vision gains with anti-VEGF therapy in neovascular age-related macular degeneration patients by using low-luminance vision. Br J Ophthalmol. 2016;100:1052-7.
10. Rud E, Baco E, Klotz D, Rennesund K, Svindland A, Berge V et al. Does preoperative magnetic resonance imaging reduce the rate of positive surgical margins at radical prostatectomy in a randomised clinical trial? Eur Urol. 2015;68:487-96.
11. Hurwitz MD, Ghanouni P, Kanaev SV, Iozeffi D, Gianfelice D, Fennessy FM et al. Magnetic resonance-guided focused ultrasound for patients with painful bone metastases: phase III trial results. J Natl Cancer Inst. 2014;106:1-9.
12. Neethling WM, Strange G, Firth L, Smith FE. Evaluation of a tissue-engineered bovine pericardial patch in paediatric patients with congenital cardiac anomalies: initial experience with the ADAPT-treated CardioCel(R) patch. Interact Cardiovasc Thorac Surg. 2013;17:698-702.
13. Beaudoin FL, Haran JP, Liebmann O. A comparison of ultrasound-guided three-in-one femoral nerve block versus parentenal opioids alone for analgesia in emergency department patients with hip fractures: a randomized controlled trial. Acad Emerg Med. 2013;20:584-91.
14. Gerdesmeyer L, Wagenpfeil S, Birkenmaier C. Percutaneous epidural lysis of adhesions in chronic lumbar radicular pain: a randomized, double-blind, placebo-controlled trial. Pain Physician. 2013;16:185-96.
15. Jin GY, Lynch D, Chawla A. Interstitial lung abnormalities in a CT lung cancer screening population: prevalence and progression rate. Radiology. 2013;268:563-71.
16. Conde A, Mainieri V, Mota EG, Oshima HM. Influence of ultrasound and diamond burs treatments on microtensile bond strength. Indian J Dent Res. 2012;23:373-7.
17. Kawaguchi Y, Ogawa M, Omata F, Hiroyuki I, Shimosegawa T, Mine T. Randomized controlled trial of pancreatic stenting to prevent pancreatitis after endoscopic retrograde cholangiopan creatography. World J Gastroenterol. 2012;18:1635-41.
18. Berra L, Coppadoro A, Bittner EA, Kolobow T, Laquerriere P, Pohlmann JR et al. A clinical assessment of the Mucus Shaver: a device to keep the endotracheal tube free from secretions. Crit Care Med. 2012;40:119-24.
19. Sasaki RT, Arcanjo AJ, Florio FM, Basting RT. Micromorphology and microhardness of enamel after treatment with home-use bleaching agents containing 10% carbamide peroxide and 7.5% hydrogen peroxide. J Appl Oral Sci. 2009;17:611-6.
20. Carandina S, Mari C, De Palma M,Marcellino MG, Cisno C, Legnaro A et al. Varicose vein stripping vs haemodynamic correction (CHIVA): a long term randomised trial. Eur J Vasc Endovasc Surg. 2008;35:230-7.
21. Ruan HJ, Fan CY, Liu JJ, Big-fang Z. A comparative study of internal fixation and prosthesis replacement for radial head fractures of Mason type III. Int Orthop. 2009;33:249-53.
22. Zografos GC, Zagouri F, Sergentanis TN,Koulocheri D, Nonni A, Oikonomou V et al. Is zero underestimation feasible? Extended vacuum-assisted breast biopsy in solid lesions - a blind study. World J Surg Oncol. 2007;5:53.
23. Wanders A, Heijde DV, Landewe R,Behier JM, Calin A, Oliveri I et al. Nonsteroidal antiinflammatory drugs reduce radiographic progression in patients with ankylosing spondylitis: a randomized clinical trial. Arthritis Rheum. 2005;52:1756-65.
24. Rashid ST, Salman M, Myint F, Baker DM, Agarwal S, Sweny P et al. Prevention of contrast-induced nephropathy in vascular patients undergoing angiography: a randomized controlled trial of intravenous N-acetylcysteine. J Vasc Surg. 2004;40:1136-41.
25. Kim HY. Statistical notes for clinical researchers: Chi-squared test and Fisher's Exact test. Restor Dent Endod. 2017;42:152-5.
26. Walsh M, Srinathan SK, McAuley DF, Mrkobrada M, Levine O, Ribic C et al. The statistical significance of randomized controlled trial results is frequently fragile: a case for a fragility index. J Clin Epidemiol. 2014;67:622-8.

27. Evaniew N, Files C, Smith C, Bhandari M, Ghert M, Walsh M et al. The fragility of statistically significant findings from randomized trials in spine surgery: a systematic survey. Spine J. 2015;15:2188-97.
28. Pocock SJ. Current issues in the design and interpretation of clinical trials. Stat Med. 1985;290:39-42.
29. Narayan VM, Gandhi S, Chrouser K, Evaniew N, Dahm P. The fragility of statistically significant findings from randomized controlled trials in the urologic literature. BJU Int. 2018; 122:160-6.
30. Checketts JX, Scott JT, Meyer C, Horn J, Jones J, Vassar M. The robustness of trials that guide evidence-based orthopaedic surgery. J Bone Joint Surg Am. 2018;100:e85.
31. Mazzinari G, Ball L, SerpaNeto A,Errando CL, Dondorp AM, Bos LD et al. The fragility of statistical significant findings in randomised controlled anaesthesiology trials: systematic review of the medical literature. Br J Anaesth. 2018;120:935-41.
32. Berti A, Cornec D, Medina Inojosa JR, Matteson EL, Murad MH. Treatments for giant cell arteritis: Meta-analysis and assessment of estimates reliability using the fragility index. Semin Arthritis Rheum. 2018;48:77-82.
33. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J et al. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis a simulation study. PLoS One. 2011;6:e25491.
34. Pocock SJ, McMurry JJV, Collier TJ. Making sense of statistics in clinical trial reports: part 1 of a 4-part series on statistics for clinical trials. J Am Coll Cardiol. 2015;66:2536-49.
35. Carter RE, McKie PM, Storlie CB. The Fragility Index: a P-value in sheep's clothing? Eur Heart J. 2017;38:346-8.