OPEN ACCESS Freely available online

Journal of Clinical Trials

Research Article

Risk of Bias in Clinical Trials Reported for Foods with Functional Claims in Japan: A Cross-Sectional Study on Research Quality

Hiroharu Kamioka[1*], Hideki Origasa[2], Jun Kitayuguchi[3], Takahiro Yoshizaki[4], Mikiko Shimada[5], Yasuyo Wada[6], Hiromi Takano-Ohmuro[7], Kiichiro Tsutani[8]

[1]Faculty of Regional Environment Science, Tokyo University of Agriculture, Setagaya-ku, Tokyo 1568502, Japan; [2]Institute of Statistical Mathematics, Tachikawa City, Tokyo 1908562, Japan; [3]Physical Education and Medicine Research Center Unnan, Unnan City, Shimane 6991105, Japan; [4]Department of Food and Life Sciences, Faculty of Food and Nutritional Sciences, Toyo University, Itakura Town, Gunma 3740193, Japan; [5]Department of Nutrition, Faculty of Health Care, Kiryu University, Kasakake-cho, Midori City, Gunma 3792329, Japan; [6]Department of Health Promotion, National Institute of Public Health, Wako City, Saitama 3510197, Japan; [7]Research Institute of Pharmaceutical Sciences, Musashino University, Nishitokyo-shi, Tokyo 2028585, Japan; [8]The Institute of Seizon and Life Sciences, Chuo-ku, Tokyo 1040061, Japan

ABSTRACT

Background: The Foods with Function Claims (FFC) notification system was introduced in Japan in April 2015. We hypothesized that there would be risk of bias (RoB) specific to health food interventions. The purpose of this cross-sectional study was to clarify RoB and related factors of clinical trials (CTs) reported as the scientific basis of efficacy in the FFC.

Methods: All 103 articles based on CTs published on the Consumer Affairs Agency website from 1 July 2018 to 30 June 2021 were reviewed. We evaluated 14 items, the highest RoB: 14 points (pts), as well as related items including first author characteristics, journal name, year published, journal impact factor, article language, and name of clinical trial registration.

Results: The RoB score was 5.7 ± 2.5 pts. In general, there was a remarkable lack of execution and/or description of the intention-to-treatment (ITT) analysis (81.6%), compliance (68.0%), and multiple outcome tests (67.0%). There was no significant difference (p=0.051) in RoB score between the published year categories of 2015-2017 (6.5 ± 2.4 pts) and 2018-2021 (5.5 ± 2.4 pts). There was also no significant difference (p=0.247) in RoB score between English (5.5 ± 2.6 pts) and Japanese (6.0 ± 2.3 pts) language publications, and no significant difference (p=0.740) between for-profit (5.7 ± 2.4 pts) and academia (6.0 ± 2.8 pts) in authors' organization. A significant correlation (p=0.099) between IF and RoB score was not observed with Spearman's rank correlation coefficient; r = -0.163.

Conclusions: Four common biases in most CTs reported in the FFC were randomization, deviations from intended interventions, measurement of outcome, and selective reporting. In particular, RoB including lack of ITT analysis, unknown compliance, and multiple outcome tests seriously damaged the study quality.

Review registration: The study was registered as UMIN 000046267 by the University Hospital Medical Information Network Clinical Trials Registry (UMIN-CTR)* in Japan (refer: https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000052795).

Keywords: Clinical trial; Risk of bias; Foods; Randomized controlled trial; Systematic review

LIST OF ABBREVIATIONS

CAA: Consumer Affairs Agency; CT: Clinical Trial; FFC: Foods with Function Claims; ICMJE: International Committee of Medical Journal Editors; IF: Impact Factor; ITT: Intension-To-Treatment; RCT: Randomized Controlled Trial; RoB: Risk of Bias; SR: Systematic Review

BACKGROUND

In Japan, a new type of foods with health claims, the Foods with Function Claims (FFC), was introduced in April 2015 [1,2]. The FFC allows manufacturers to submit labeling to the Secretary-General of the Consumer Affairs Agency (CAA) in Japan that indicates the food is expected to have a specific effect on health.

Correspondence to: Hiroharu Kamioka, Faculty of Regional Environment Science, Tokyo University of Agriculture, Setagaya-ku, Tokyo 1568502, Japan; E-mail: h1kamiok@nodai.ac.jp

Received: 01-Aug-2022, Manuscript No. JCTR-22-18603; **Editor assigned:** 10-Aug-2022, PreQC No. JCTR-22-18603 (PQ); **Reviewed:** 24-Aug-2022, QC No. JCTR-22-18603; **Revised:** 31-Aug-2022, Manuscript No. JCTR-22-18603 (R); **Published:** 07-Sep-2022, DOI:10.35248/2167-0870.22.12.503.

Citation: Kamioka H, Origasa H, Kitayuguchi J, Yoshizaki T, Shimada M, Wada Y, et al. (2022) Risk of Bias in Clinical Trials Reported for Foods with Functional Claims in Japan: A Cross-Sectional Study on Research Quality. J Clin Trials. 12:503.

The FFC is only a notification system in which food manufacturers must meet the following five specific criteria: i) it is for people not suffering from any disease (excluding minors, pregnant women and those planning a pregnancy], and lactating women); ii) all food products are subject to this system; iii) prior to market entry (before at least 60 days), food business operators are required to submit information, such as food safety and effectiveness and the system in place to collect information on adverse health effects, to the Secretary-General of CAA; iv) the government does not evaluate the safety and effectiveness of the submitted product, i.e., notification system; and v) all the submitted information is disclosed on the CAA website.

For a food product to claim effectiveness on its label, evidence for its proposed function claims must be substantiated by one of two standard scientific methods: clinical trials (CTs) such as randomized controlled trials (RCTs), or systematic reviews (SRs). Detailed guidelines about use of these two methods for the FFC have been published on the CAA website [3].

A problem with CTs notified through the FFC system is the adequacy of research reports, which was highlighted by the CAA and research groups at the beginning of the system [4,5]. The CAA examined 50 reported CTs and determined that many had inappropriate protocols and methods for evaluating the risk of bias, and also had conflicts of interest [4]. In addition, the report clarified specific points regarding research methodologies in food-based CTs. Tanemura et al. identified problems with the reporting quality and associated issues for 33 RCTs in the FFC; specifically, 29 check items in CONSORT 2010 met only 13.8 items (47.6%) on average [5].

The latest study performed in 2021 reported that compliance of CT protocols in the FFC system were suboptimal in transparency [6]. In addition to selective reporting, a new problem identified was that content of the intervention (test food) was intentionally concealed. To protect consumers, the report suggested that researchers should monitor and confirm that referenced RCTs are above a certain level of quality, journal editors and peer reviewers in this field should scrutinize differences between and ambiguities of the submitted manuscript and its protocol, and regulators should confirm the protocol before accepting it, even if it is included in a notification system.

Scrutiny of RoB is paramount in a SR, which ultimately assesses and integrates the effects of an intervention, and a CT with an unacceptably high RoB should be excluded from the analysis. There are many tools for assessing RoB in healthcare science [7-13]. The latest tool that provides rigorous coverage of a wide range of RCTs is “Revised Cochrane risk-of-bias tool for randomized trials: RoB 2” [7]. One of its major features is that it defines five domains: i) bias arising from the randomization process; ii) bias due to deviations from intended interventions; iii) bias due to missing outcomes data; iv) bias in measurement of the outcome; and v) bias in selection of the reported results.

Among the many tools mentioned above, we paid special attention to additional interventions, which, unlike medications, are unique to health food interventions. The most applicable intervention is a research method that is effective when a specific exercise (resistance training or endurance exercise) is used in combination with a supplement that is a functionally involved component. For example, there are studies that demonstrate some effectiveness with combined use of a test food and resistance training [14], walking [15], or a comprehensive exercise program [16]. Therefore, we decided to adopt a method to evaluate RoB through a revised Cochrane tool that van Tulder et al. (2009) described because it includes an item about "co-intervention" and is simply 11 items [13]. We hypothesized that there would be RoB specific to health food interventions.

The purpose of this cross-sectional study was to clarify RoB and related factors of CTs reported as the scientific basis of efficacy in the FFC system. Based on results, the following questions about issues necessary to improve CT quality were addressed: i) How much RoB do the CTs include? ii) What are the biased items and their characteristics? iii) Is RoB correlated with author characteristics (for-profit corporate authors, academia authors), the year in which the paper was published, the relationship between English and other languages, and the impact factor?

MATERIALS AND METHODS

Eligibility and exclusion criteria (Target Article)

All reported articles based on CTs published on the CAA website during the three years from 1 July 2018 to 30 June 2021 were reviewed. For articles that were duplicates, such as multiple notifications using the same article, only the first article was adopted. Articles about SRs and observational studies were excluded, as were articles in which clinical trial registration (CTR) was not specified. This was the same procedure followed for the previous study (i.e., compliance of CT protocols for FFC) our research group conducted [6]. Eligible articles were published in 27 journals, and most (62%) were published in 2018-2019 (Table 1).

Data items and evaluation of methodological quality (RoB score)

We evaluated a total of 14 items about RoB in the target articles (RoB score). These items included the following combined elements (Table 1): 11 items described by van Tulder et al. in their revised Cochrane’s criteria list [13], and three items for main outcome and multiple tests (outcomes and time points) that we added to the revised Cochrane’s list. In order to ensure that variation was not caused by systematic errors during study execution, five reviewer authors (HK, JK, TY, MS, and YW) independently assessed the quality of articles. Disagreements and uncertainties were resolved by discussion with other authors (HO and HT-O).

Each of the 14 items evaluated for an article was scored as: "1" for "not properly implemented, not described, or unclear"; "0" for "proper implementation/description" or "not applicable". For RoB score, the higher the total RoB score (i.e., 14 pts.), the higher the risk of bias in a study. Inter-rater reliability was calculated on a dichotomous scale using percentage agreement and Cohen’s kappa coefficient (k). In addition, to make it easier to understand RoB results, we created graphs for each of the five domains in Cochrane's RoB 2 [7].

The following data items were collected for sub-research questions: first author characteristics (for-profit researchers, academia researchers, or other), journal name, year published, journal impact factor (IF) in 2020, and language of the article. The IF was assessed according to the Clarivate Analytics’s gate (https://jcr.clarivate.com/). If a journal does not have the IF, the number "0" was substituted for it.

Table 1: Risk of Bias (RoB) of target articles.

#	RoB item	Number of articles with each RoB item n=103 n=103
1	Was the method of randomization adequate?	44 (42.7%)
2	Was the treatment allocation concealed?	23 (22.3%)
3	Were the groups similar at baseline regarding the most important prognostic indicators?	48 (46.6%)
4	Was the patient (participant) blinded to the intervention?	17 (16.5%)
5	Was the care provider (intervener) blinded to the intervention?	51 (49.5%)
6	Was the outcome assessor blinded to the intervention?	46 (44.7%)
7	Were co-interventions avoided or similar?	0 (0%)
8	Was the compliance acceptable in all groups?	70 (68.0%)
9	Was the drop-out rate described and acceptable?	50 (48.5%)
10	Did the analysis include an intention-to-treat analysis?	84 (81.6%)
11	Was the timing of the outcome assessment similar in all groups?	2 (1.9%)
12	Was the main outcome clear?	34 (33.0%)
13	Was the problem of multiple tests (outcomes) avoided?	69 (67.0%)
14	Was the problem of multiple tests (time points) avoided?	54 (52.4%)
-	RoB score (pts)*	5.7 ± 2.5 [0-11]

Value: n (%). Mean ± standard deviation [range].

* 0 pt indicates no bias in an article and 14 pts indicates all biases present in an article.

Measures and statistical analysis

Since the IF and RoB score were not normally distributed, intergroup comparisons were performed with the nonparametric Mann-Whitney test and shown in a boxplot. However, for results presented in the text, the numerical values for comparison between groups are shown by mean ± standard deviation for easy understanding. For the relationship between IF and RoB score, Spearman's rank correlation coefficient was used. In addition to assessment of the above-mentioned 14 items, statistical associations in each study were analyzed using RoB score as the dependent variable and first author characteristics, year of publication (before 2017 or after 2018), language characteristics (English or Japanese), and IF as explanatory variables.

IBM SPSS Statistics 25.0 (IBM Corporation, Armonk, NY, USA) was used for statistical analyses. A P-value less than 0.05 was considered statistically significant.

Protocol registration

The study methodology (protocol) was established on 2 December 2021. The study was registered as UMIN 000046267 by the University Hospital Medical Information Network Clinical Trials Registry (UMIN-CTR)* in Japan (refer: https://upload.umin.ac.jp/cgi-open-bin/ctr/ctr_view.cgi?recptno=R000052795). However, UMIN-CTR could not register the contents of all protocols in the input settings, so the complete protocol was stored in an online cloud, which can be viewed from this link: https://drive.google.com/file/d/1c1M02bZrztnA7AR8myjgxN3rWtuEtzmV/view?usp=sharing.

*UMIN-CTR is the largest CTR in Japan and joined the WHO registry network in October 2008.

RESULTS

Included studies and characteristics

Preliminary research identified 177 applicable publications, of which 103 met the eligibility criteria before final confirmation (Figure 1 and Supplementary: Table S1). Eligible articles were published in 27 journals, and most (62%) were published in 2018-2019 (Table 2).

The languages of eligible publications were English (55%) and Japanese (45%). According to the affiliation classification of the first author, for-profit was 81% and academia was 19%. Seventy-four percent of journals had no IF.

Feature of RoB on CTs

Table 1 shows the RoB assessment on target articles. The RoB score was 5.7 ± 2.5 pts (range: 0-11). In general, there was a remarkable lack of execution and/or description in the intention-to-treatment (ITT) analysis, compliance, and multiple tests (outcomes and time points). There were four items with a very poor (>50%) description in the articles: "Did the analysis include an ITT analysis?" (81.6%); "Was the compliance acceptable in all groups?" (68.0%); and "Was the problem of multiple tests (outcomes and time points) avoided?" (67.0% and 52.4%, respectively). Another six items had a poor (30.0%-49.5%) description. Interrater reliability metrics for the quality assessment indicated substantial agreement (92.1%, k= 0.835) for all 1,442 items (14 items multiplied by 103 CTs).

Figure 2 shows the characteristics of flaw based on four bias domains (a fifth domain, "Bias due to missing outcomes data", was not planned in this study) with reference to RoB item 2 about treatment allocation. The four domains tended to have the same degree of bias. Bias due to deviation from intended interventions, including lack of ITT analysis (81.6%) and unknown compliance (68.0%), seemed to have a particularly serious impact on the deterioration of study quality.

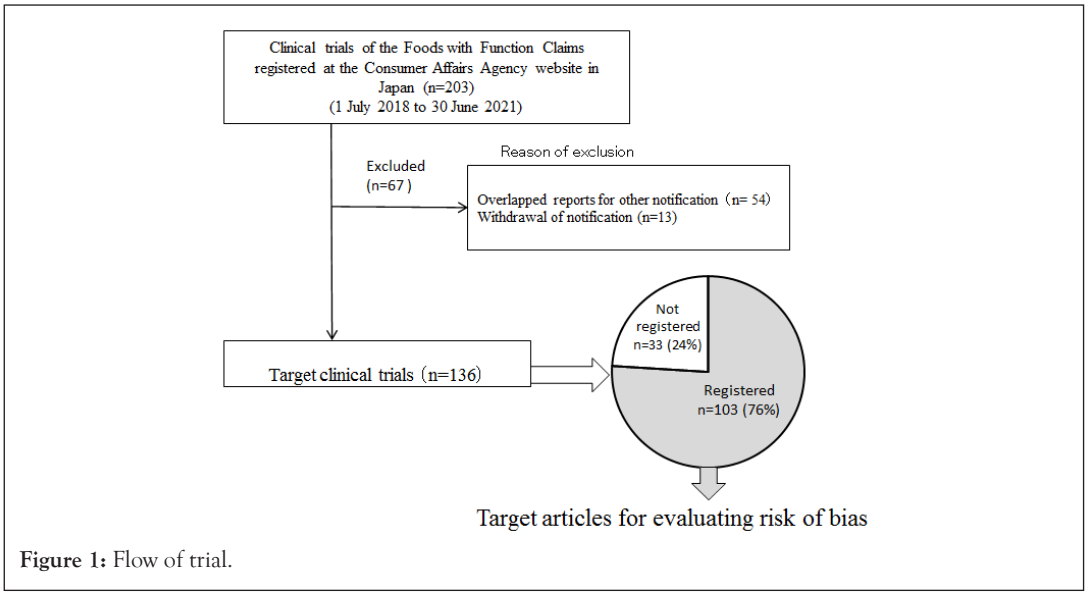

Figure 1: Flow of trial.

Table 2: Published journal's characteristics.

Journal name	Publications
薬理と治療/Japanese Pharmacological and Therapeutics	57 (55%)
診療と新薬/Medical Consultation and New Remedies	9 (9%)
応用薬理/Pharmacometrics	4 (4%)
Functional Foods in Health and Disease	4 (4%)
Nutrients	4 (4%)
Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy	2 (2%)
Frontiers in Neuroscience	2(2%)
Integrative Molecular Medicine	2 (2%)
American Journal of Geriatric Psychiatry Applied and Environmental Microbiology 機能性食品と薬理栄養/Associate Journal of Japanese Society for Medical Use of Functional Foods Benefical Microbes Biological and Pharmaceutical Bulletin Bioscience, Biotechnology, and Biochemistry Bioscience of Microbiota, Food and Health Complementary Therapies in Medicine International Journal of Food Sciences and Nutrition Journal of Clinical Biochemistry and Nutrition Journal of Dairy Science 日本栄養·食糧学会誌/Journal of Japanese Society of Nutrition and Food Science Journal of Traditional and Complementary Medicine Neurogastroenterology and Motility Nutrition Journal 調理食品と技術/Prepared Foods and Technology Science Reports Skin Pharmacology and Physiology	Common to all of the following journals: 1 (1%)
Published year	
2014-2015	1 (1%)
2016-2017	25 (24%)
2018-2019	64 (62%)
2020-2021	13 (13%)
Language	
English	56 (55%)
Japanese	47 (45%)
Category of first author's organaization	
For-profit	83 (81%)
Academia	20 (19%)
Journal's impact factor in 2020	
None (0)	76 (74%)
1.999>	4 (4%)
2.000-3.999	11 (11%)
>4.000	12 (12%)

Value: n (%)

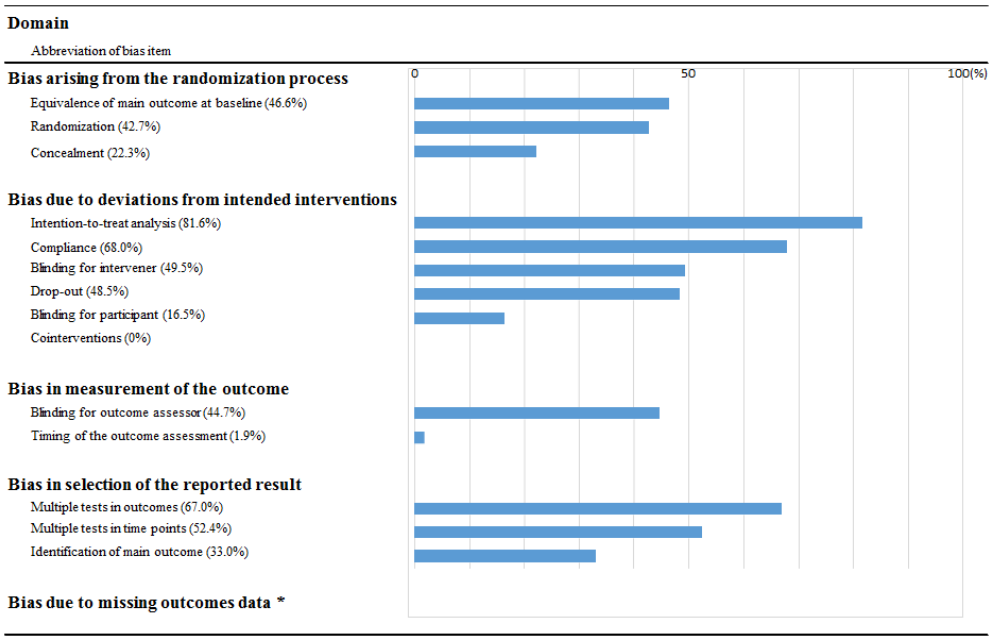

Figure 2: Characteristics on flaw based on five bias domains.
Note: *One domain, "Bias due to missing outcomes data", was not planned in this study

Elements correlated with RoB

There was no significant difference (p=0.051) in RoB score between the published year categories of 2015-2017 (6.5 ± 2.4 pts) and 2018-2021 (5.5 ± 2.4 pts) (Figure 3). There was also no significant difference (p=0.247) in RoB score between English (5.5 ± 2.6 pts) and Japanese (6.0 ± 2.3 pts) language publications (Figure 4), and no significant difference (p=0.740) between for-profit (5.7 ± 2.4 pts) and academia (6.0 ± 2.8 pts) in authors' organization (Figure 5). A significant correlation (p=0.099) between IF and RoB score was not observed with Spearman's rank correlation coefficient; r=-0.163 (Figure 6).

DISCUSSION

To the best of our knowledge, this was the first study to clarify RoB and related factors of CTs reported as the scientific basis of efficacy in the FFC system; unfortunately, we identified substantial RoB in the CTs included in this study. As in many other healthcare fields, SRs about nutrition are performed as the method of choice to synthesize data from CTs. Consequently, a SR of suboptimal quality CTs cannot describe with certainty evidence about the functionality of a food.

Feature of RoB on CTs

Bias is defined as systematic error of study results and is caused by incorrect research methodology [17]. As a characteristic of bias, it occurred in all four domains in common, especially "bias due to deviations from intended interventions" were serious. Since a CT of health food has a relatively short intervention period (8-12 weeks in most cases [4]) in general, it is desirable to analyze both an ITT population and a full-set population. The overall degree of reviewer agreement in this study was fair, but there were many articles for which the evaluations were confusing. A cross-sectional study reported that RoB assessments of RCTs included in more than one Cochrane Review differed substantially, and most disagreements were related to a difference in interpretation of an incomplete or unclear description in the study report [18]. As such, CTs on nutrition may also require a checklist to reduce food science-specific RoB.

Study reporting guidelines include the CONSORT 2010 statement [19,20] and the CONSORT 2010 statement: crossover extension [21], both of which describe the method for RCT reporting, and the SPIRIT 2013 statement [22], which summarizes the reports of CT protocols. However, these guidelines are unlikely to have direct implications for RoB. Researchers need to have a careful understanding of RoB in advance of conducting a study, because these guidelines were originally created with the aim of improving "quality of the report" and not "quality of the research".

There are two studies that highlight issues with the Cochrane Review, and both are thought to be very rigorous RoB assessments. After analyzing 10,103 trials, authors of one of these studies often identified RoB related to random sequence generation that were not in line with instructions given in the Cochrane Handbook [23]. In the other study, RoB assessments for blinding of participants and personnel (performance bias) were also frequently not in line with those in the Handbook recommendations [24]. These findings suggest that at the same time as following instructions in the Handbook to increase the degree of agreement between reviewers in SRs, proper and non-misleading planning, implementation, and reporting of individual CTs are essential.

Elements correlated with RoB

For the sub-research question assessed in this study, "Is RoB correlated with author characteristics (for-profit corporate authors, academia authors), the year in which the paper was published, the relationship between English and other languages, and the IF?", no significant relationship was observed for any of the variables. In the case of industry funding, there is evidence that favorable results are more likely to be reported than unfavorable results [25]. Almost all FFC notifications were from industries, and 80% of authors of the target articles were affiliated with those industries. However, this

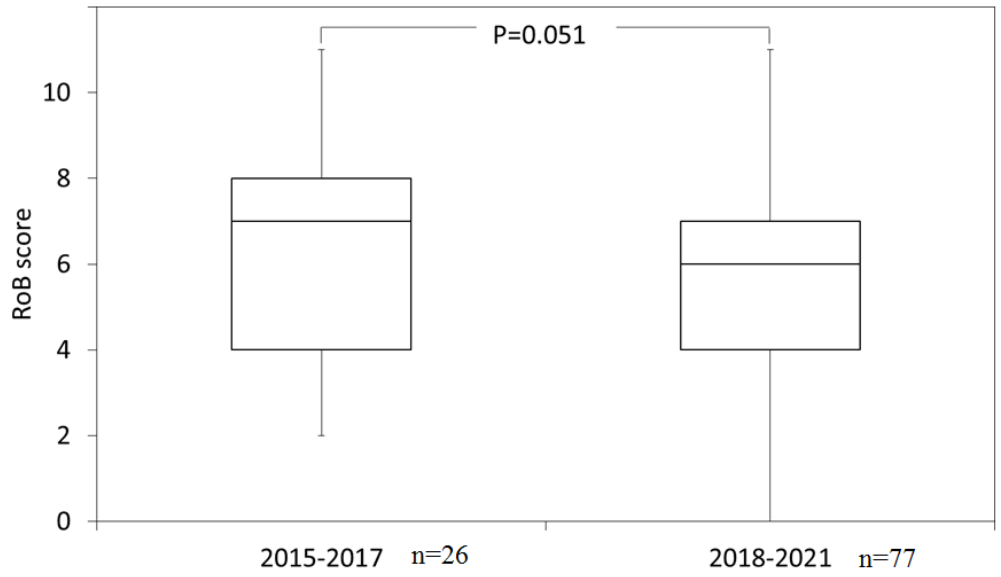

Figure 3: RoB score during the published year period.

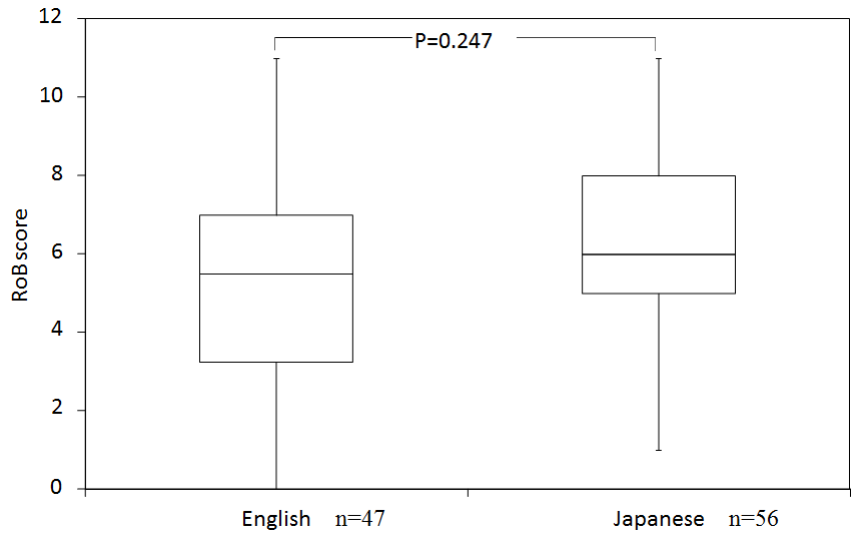

Figure 4: RoB score between Japanese and English publications.

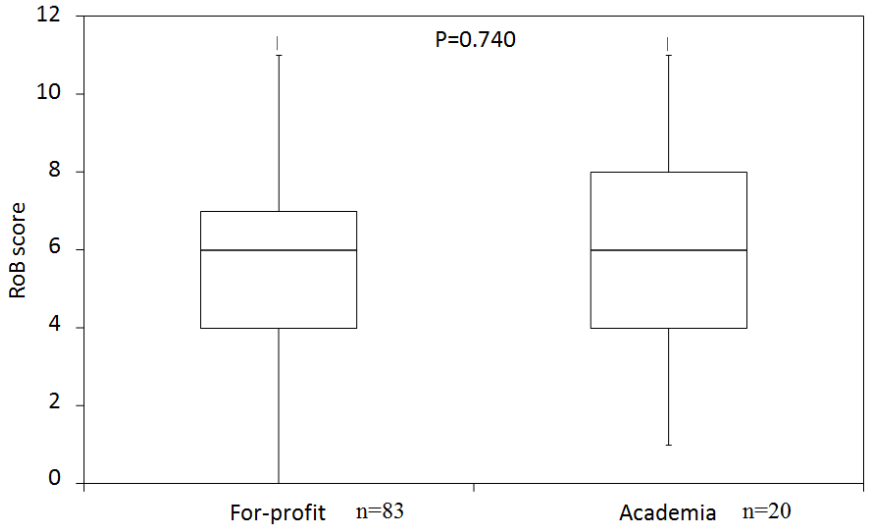

Figure 5: RoB score for category of first author's organization.

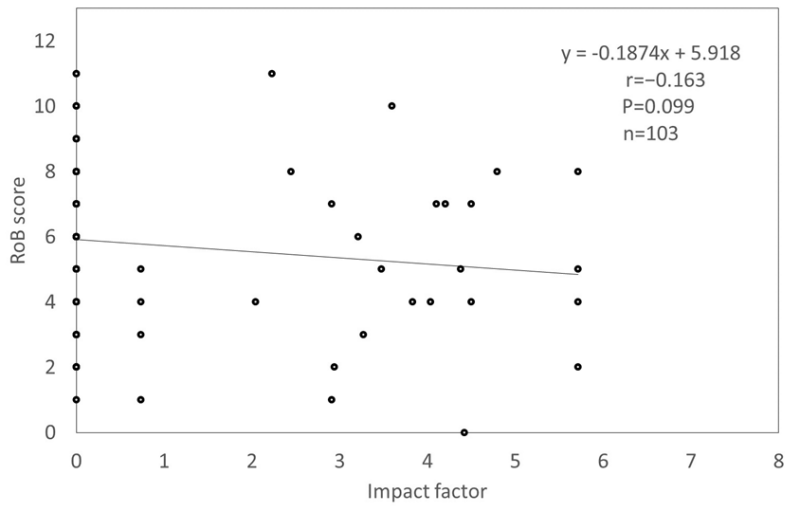

Figure 6: Correlation between journal's impact factor and RoB score.

study did not find such a specific feature from its RoB assessment.

Impact on SRs

In the FFC system, about 90% of notifications are from SRs that were substantiated by standard scientific methods [1-2]. The present study showed that the quality of CTs was very poor. Because SRs must have little RoB in the reviewed CTs, our findings bring into question the reliability of quality of SRs in the FFC system. In fact, the previous studies in 2017 [1] and 2019 [2] evaluated the quality of methodologies and reporting of SRs based on the FFC using the AMSTAR checklist [26] and noted very poor description and/or implementation of study selection, data extraction, search strategy, evaluation methodology for risk of bias, and assessment of publication bias. When determining credibility of study results by meta-analysis, it is very important to know whether only low RoB CTs were included or high RoB CTs were excluded. For example, a previous study that evaluated 59 SRs reported that only 50% of the SRs performed sensitivity analyses for low RoB CTs [27].

The latest cross-sectional studies that evaluated consistency between the description of published CTs and their protocols concluded that the registered protocols were suboptimal in transparency [6]. In particular, in addition to selective reporting, a new problem identified was that content of the intervention (test food) was intentionally concealed. This problem was also previously reported in a review of funding for pharmaceutical industry studies [25]. Therefore, because the FFC is a notifications system by the food industry and related businesses, readers especially need to pay attention to positive results of SRs in the FFC system. In addition, they should verify how the researcher evaluated RoB in the SR, i.e., whether a sensitivity analysis was performed or high RoB CTs were excluded prior to conducting a meta-analysis. Tools have been developed to assess the quality of SRs (AMSTAR 2 [28] and ROBIS [29]), but they require a rigorous RoB assessment of the extracted CTs. A primary study that can withstand such rigor, i.e., a CT with low RoB, is needed in the FFC system.

Conducting RoB assessments to validate SR findings imposes high demands on reviewers' expertise as well as on resources such as time and cost. With the rapid progress of computer science, text mining technology can now be applied in the field of healthcare science [30-32], and this has led to development of electronic applications that promote automatic reviewing [33] of articles. For example, RobotReviewer was developed to support RoB assessment of RCTs for the Cochrane domain [34]. A prospective, randomized user study reported that semi-automation was quicker than manual assessment (mean, 755 s vs. 824 s; relative time 0.75, 95% confidence interval 0.62-0.92) in RoB evaluations [35]. A recent study reported that RobotReviewer yielded a moderate degree of agreement with human reviewers for randomization and allocation concealment, and an adequate sensitivity for detecting low risk of selection bias [36]. However, it emphasized that human reviewers should supervise the semi-automated assessment process. In the RoB evaluation for this study, there were vague descriptions that confused the reviewers' assessments, which confirmed that final review by humans is indispensable.

Future research challenges to improve the quality of CT on the FFC

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) working group developed and updated a transparent approach for grading the certainty of evidence and strength of recommendations based on a body of evidence [37]. The GRADE method has also been adopted in the nutrition and food-related fields [38-40]. A methodological review of nutrition SRs [41] reported that, out of 800 SRs, 55 used GRADE; certainty of evidence was downgraded mostly for RoB (37.8%) and imprecision (33%) in SRs of RCTs, and for imprecision (32.7%), RoB (29.4%), and inconsistency (29%) in SRs of non-RCTs. Even in the FFC system, it may be necessary to adopt the GRADE theory in the near future, given the huge number of product notifications (4,982 as of 1 February 2022) since the FFC was launched in 2015.

In the planning stage for conducting a CT, researchers need to carefully review some type of RoB checklist, as well as reporting guidelines (i.e., SPIRIT 2013 [22], CONSORT 2010 statement [19-20], CONSORT 2010 statement: crossover extension [21]), and take steps to avoid bias. Although there are many evaluation checklists for RoB, the RoB 2 for RCTs and non-RCTs [7], and the ROBINS-1 [11] and Newcastle-Ottawa Scale [42] for non-RCTs are becoming more commonly used. A cross-sectional study to assess interrater reliability and usability of RoB 2 concluded that it is a detailed and comprehensive tool but difficult and demanding, and the difficulties were due to poor knowledge of subject matter

of primary studies, new technology, different approaches for some domains compared with the previous tool, and ways of formulating signaling questions [43]. Another recent study found that traditional double blinding of participants and clinicians should not be regarded as a gold standard to strive for, and should be used only if the negative effects are considered carefully and are outweighed by the potential benefit [44]. Furthermore, GRADE does not recommend a specific tool to assess RoB because the tools have different advantages and disadvantages that influence their utilization [45]. As long as RoB is assessed across studies, any validated and appropriate tool could be used, so we do not recommend a particular assessment tool.

Researchers planning to perform a SR should carefully examine the quality of each reviewed CT and should not include studies with high RoB. First, researchers should establish a complete research plan based on PRISMA 2020 and PRISMA-P, which are checklists for reporting methods. Next, since the SR itself will be evaluated later by AMSTAR 2 or another method, each CT that meets the eligibility criteria in the SR should be strictly evaluated for quality.

LIMITATIONS

There were several limitations to the present study. First, we only focused on CTs based on notification to the FFC in Japan (single country), so our findings may not necessarily be generalized to all CTs of healthy foods. In fact, about half of the articles in our study were written in Japanese. Second, the 103 articles included in our study was a relatively medium sample size and was limited to three years of CTs, so we cannot be certain they were representative of all FFC notifications to the CAA. Third, although the RoB assessment method in our study covered major biases, it did not cover all biases; for example, it did not include the bias domain of missing outcome data. In addition, since each evaluation item simply gives a total score as 1 point, it does not reflect the severity (weight) of a certain bias on research results. Furthermore, we did not comprehensively interpret high-, middle-, and low-risk by each domain. Fourth, since this study was not designed to evaluate the reliability and utility of various checklists in nutrition-related CTs, we cannot recommend a checklist for researchers to use. According to a study that analyzed the PROSPERO protocol, as of 2018 [46] the Cochrane RoB tool (2011 version) [17] had become the standard for SRs of RCTs. Despite the existence of dozens of tools for assessing non-RCTs, relatively few were commonly used, with the Newcastle-Ottawa Scale [42] and ROBINS-1 [11] being the most frequent. Finally, although there could be many other potential elements related to RoB, we only assessed four aspects: first author characteristics, published year, languages, and the IF.

CONCLUSION

There were four common biases in most CTs that were reported as the scientific basis of efficacy in the FFC system: randomization, deviations from intended interventions, measurement of outcome, and selective reporting. In particular, RoB including lack of ITT analysis, unknown compliance, and multiple outcome tests seemed to seriously damage study quality. In the planning stage of study conduct, researchers in the nutrition field need to carefully review some type of RoB checklist as well as reporting guidelines, and take all steps to avoid bias.

Supplementary Information

The following are available online at.

Table S1: Characteristics of clinical trials of the Food with Function Claims

Table S2: RoB score, language, and category of first author's organization

And other data and materials were was stored in an online cloud, which can be viewed from this link: https://drive.google.com/file/d/14CqNt3c53gv7H3ftlsS3sHO1oS6rmINX/view?usp=sharing

COMPETING INTEREST

HK supervised systematic reviews for five corporations (FANCL CORPORATION, Morishita Jintan Co., Ltd., KAGOME Co., Ltd., MARUZEN PHARMACEUTICAL CO., LTD. and SUNTORY WELLNESS LIMITED) and was compensated for that work. The other authors have no conflicts to declare.

FUNDING

This study was supported by a Grant-in-Aid for Scientific Research (C) Number 21K11604 from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan, 2021.

AUTHORS CONTRIBUTIONS

HK conceived and designed the study and made a brief summary list of included studies and data extraction. HK, JK, TY, MS, and YW independently assessed the quality of articles. Disagreements and uncertainties were resolved by discussion with other authors (e.g., HO and HT-O). JK performed the statistical analysis. HO confirmed the appropriateness of the analysis method as a biostatistics expert. HK produced the draft article. KT was the guarantor.

ACKNOWLEDGMENTS

We would like to express our appreciation to Ms. Michiko Ishida (paper work) and Ms. Satoko Sayama (collecting related papers).

REFERENCES

1. Kamioka H, Tsutani K, Origasa H, Yoshizaki T, Kitayuguchi J, Shimada M, et al. Quality of systematic reviews of the Foods with Function Claims registered at the Consumer Affairs Agency web site in Japan: A prospective systematic review. Nutr Res. 2017;40:21-31.
2. Kamioka H, Tsutani K, Origasa H, Yoshizaki T, Kitayuguchi J, Shimada M, et al. Quality of systematic reviews of the Foods with Function Claims in Japan: Comparative before-and after-evaluation of verification reports by the Consumer Affairs Agency. Nutrients. 2019;11(7):1583.
3. Consumer Affairs Agency. Government of Japan: Guideline.2021.
4. Consumer Affairs Agency. Government of Japan: Verification of scientific evidence on “Foods with Function Claims”: Assessment of the submitted clinical trials.2017.
5. Tanemura N, Hamadate N, Urushihara H. Evaluation of randomized controlled trials of foods with functional claims request: The learning outcomes from studies in Japan. J Funct Foods. 2018;42:248-53.
6. Kamioka H, Origasa H, Kitayuguchi J, Tsutani K. Compliance of clinical trial protocols for Foods with Function Claims (FFC) in Japan: Consistency between clinical trial registrations and published reports. Nutrients. 2021;14(1):81.
7. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: A revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366.
8. Higgins JP, Savović J, Page MJ, Elbers RG, Sterne JA. Assessing risk of bias in a randomized trial. Cochrane Handbook for systematic reviews of interventions. 2019:205-28.
9. Healthcare improvement Scotland SIGN: Methodology checklist 2; Randomised controlled trials: 2021.
10. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Education for information. 2018;34(4):285-91. [Google scholar]
11. Sterne JA, Hernán MA, Reeves BC, Savovi J, Berkman ND, Viswanathan M, et al. ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. BMJ. 2016;355:i4919.
12. Jadad AR, Moore RA, Carroll D, Jenkinson C, Reynolds DJ, Gavaghan DJ, et al. Assessing the quality of reports of randomized clinical trials: Is blinding necessary?. Control Clin Trials. 1996;17(1):1-2.

13. Van Tulder M, Furlan A, Bombardier C, Bouter L, Editorial Board of the Cochrane Collaboration Back Review Group. Updated method guidelines for systematic reviews in the Cochrane Collaboration Back Review Group. Spine. 2003;28(12):1290-9.
14. Nagai N, Yagyu S, Hata A, Nirengi S, Kotani K, Moritani T, et al. Maslinic acid derived from olive fruit in combination with resistance training improves muscle mass and mobility functions in the elderly. J Clin Biochem Nutr. 2019;64(3):224-30.
15. Ota N, Soga S, Hase T, Shimotoyodome A. Daily consumption of milk fat globule membrane plus habitual exercise improves physical performance in healthy middle-aged adults. SpringerPlus. 2015;4(1):120.
16. Ueda K, Sasai H, Tsujimoto T, Sanbongi C, Ikegami S, Kobayashi H, et al. Randomized trial of amino acid mixture combined with physical activity promotion for abdominal fat reduction in overweight adults. Diabetes Metab Syndr Obes. 2018;11:23.
17. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
18. Bertizzolo L, Bossuyt P, Atal I, Ravaud P, Dechartres A. Disagreements in risk of bias assessment for randomised controlled trials included in more than one Cochrane systematic reviews: A research on research study using cross-sectional design. BMJ Open. 2019;9(4):e028382.
19. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c332.
20. 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. BMJ. 2012;10(1):28-55.
21. Dwan K, Li T, Altman DG, Elbourne D. CONSORT 2010 statement: Extension to randomised crossover trials. BMJ. 2019;366:l4378.
22. Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gøtzsche PC, Krleža-Jerić K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. Ann Intern Med 2013;158(3):200-7.
23. Barcot O, Boric M, Poklepovic Pericic T, Cavar M, Dosenovic S, Vuka I, et al. Risk of bias judgments for random sequence generation in Ccochrane systematic reviews were frequently not in line with Cochrane Handbook. BMC Med Res Methodol. 2019;19(1):170.
24. Barcot O, Boric M, Dosenovic S, Pericic TP, Cavar M, Puljak L. Risk of bias assessments for blinding of participants and personnel in Cochrane Reviews were frequently inadequate. J Clin Epidemiol. 2019;113:104-13.
25. Melander H, Ahlqvist-Rastad J, Meijer G, Beermann B. Evidence b(i) ased medicine—selective reporting from studies sponsored by pharmaceutical industry: Review of studies in new drug applications. BMJ. 2003;326(7400):1171-3.
26. Shea BJ, Grimshaw JM, Wells GA, Boers M, Andersson N, Hamel C, et al. Development of AMSTAR: A measurement tool to assess the methodological quality of systematic reviews. BMC Med Res Methodol. 2007;7(1):1-7.
27. Katikireddi SV, Egan M, Petticrew M. How do systematic reviews incorporate risk of bias assessments into the synthesis of evidence? A methodological study. J Epidemiol Community Health. 2015;69(2):189-95.
28. Shea BJ, Reeves BC, Wells G, Thuku M, Hamel C, Moran J, et al. AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. BMJ. 2017;358.
29. Whiting P, Savović J, Higgins JP, Caldwell DM, Reeves BC, Shea B, et al. ROBIS: A new tool to assess risk of bias in systematic reviews was developed. J Clin Epidemiol. 2016;69:225-34.
30. O'Mara-Eves A, Thomas J, McNaught J, Miwa M, Ananiadou S. Using text mining for study identification in systematic reviews: A systematic review of current approaches. Syst Rev. 2015;4(1):5.
31. Marshall IJ, Noel-Storr A, Kuiper J, Thomas J, Wallace BC. Machine learning for identifying randomized controlled trials: an evaluation and practitioner's guide. Res Synth Methods. 2018;9(4):602-14.
32. Marshall IJ, Wallace BC. Toward systematic review automation: A practical guide to using machine learning tools in research synthesis. Syst Rev. 2019;8(1):163.
33. O'Connor AM, Tsafnat G, Gilbert SB, Thayer KA, Shemilt I, Thomas J, et al. Still moving toward automation of the systematic review process: A summary of discussions at the third meeting of the International Collaboration for Automation of Systematic Reviews (ICASR). Syst Rev. 2019;8(1):57.
34. Gates A, Vandermeer B, Hartling L. Technology-assisted risk of bias assessment in systematic reviews: A prospective cross-sectional evaluation of the RobotReviewer machine learning tool. J Clin Epidemiol. 2018;96:54-62.
35. Soboczenski F, Trikalinos TA, Kuiper J, Bias RG, Wallace BC, Marshall IJ. Machine learning to help researchers evaluate biases in clinical trials: A prospective, randomized user study. BMC Med Inform Decis Mak. 2019;19(1):96.
36. Hirt J, Meichlinger J, Schumacher P, Mueller G. Agreement in risk of bias assessment between robot reviewer and human reviewers: An evaluation study on randomised controlled trials in nursing-related cochrane reviews. J Nurs Scholarsh. 2021;53(2):246-54.
37. Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol. 2017;87:4-13.
38. Santesso N, Akl EA, Bianchi M, Mente A, Mustafa R, Heels-Ansdell D, et al. Effects of higher-versus lower-protein diets on health outcomes: A systematic review and meta-analysis. Eur J Clin Nutr. 2012;66(7):780-8.
39. De Souza RJ, Mente A, Maroleanu A, Cozma AI, Ha V, Kishibe T. Intake of saturated and trans unsaturated fatty acids and risk of all-cause mortality, cardiovascular disease, and type 2 diabetes: Systematic review and meta-analysis of observational studies. BMJ. 2015;351:h3978.
40. Reynolds A, Mann J, Cummings J, Winter N, Mete E, Te Morenga L. Carbohydrate quality and human health: A series of systematic reviews and meta-analyses. Lancet. 2019;393(10170):434-45.
41. Werner SS, Binder N, Toews I, Schunemann HJ, Meerpohl JJ, Schwingshackl L. Use of the GRADE approach for rating the certainty of evidence in evidence syntheses published in high impact factor nutrition journals: A methodological survey. J Clin Epidemiol. 2021;135:54-69.
42. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. 2022.
43. Minozzi S, Cinquini M, Gianola S, Gonzalez-Lorenzo M, Banzi R. The revised Cochrane risk of bias tool for randomized trials (RoB 2) showed low interrater reliability and challenges in its application. J Clin Epidemiol. 2020;126:37-44.
44. Anand R, Norrie J, Bradley JM, McAuley DF, Clarke M. Fool's gold? Why blinded trials are not always best. BMJ. 2020;368:l6228.
45. Schwingshackl L, Schünemann HJ, Meerpohl JJ. Improving the trustworthiness of findings from nutrition evidence syntheses: Assessing risk of bias and rating the certainty of evidence. Eur J Nutrition. 2021;60(6):2893-903.
46. Farrah K, Young K, Tunis MC, Zhao L. Risk of bias tools in systematic reviews of health interventions: An analysis of PROSPERO-registered protocols. Syst Rev. 2019;8(1):280.