Contemporary Clinical Trials Communications 1 (2015) 32–38

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: www.elsevier.com/locate/conctc

The relationship between external and internal validity of randomized controlled trials: A sample of hypertension trials from China

Xin Zhang [a], Yuxia Wu [b, c], Pengwei Ren [b], Xueting Liu [b], Deying Kang [b, *]

[a] Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu 610041, China
[b] Department of Evidence-based Medicine and Clinical Epidemiology, West China Hospital, Sichuan University, Chengdu 610041, China
[c] Department of Internal Medicine, Mianyang People's Hospital, Mianyang City, 621000, China

ARTICLE INFO

Article history:
Received 9 July 2015
Received in revised form
22 October 2015
Accepted 30 October 2015
Available online 19 November 2015

Keywords:
External validity
Internal validity
Hypertension
Randomized controlled trial (RCT)

ABSTRACT

Objective: To explore the relationship between the external validity and the internal validity of hypertension RCTs conducted in China.

Methods: Comprehensive literature searches were performed in Medline, Embase, Cochrane Central Register of Controlled Trials (CCTR), CBMdisc (Chinese biomedical literature database), CNKI (China National Knowledge Infrastructure/China Academic Journals Full-text Database) and VIP (Chinese scientific journals database) as well as advanced search strategies were used to locate hypertension RCTs. The risk of bias in RCTs was assessed by a modified scale, Jadad scale respectively, and then studies with 3 or more grading scores were included for the purpose of evaluating of external validity. A data extract form including 4 domains and 25 items was used to explore relationship of the external validity and the internal validity. Statistic analyses were performed by using SPSS software, version 21.0 (SPSS, Chicago, IL).

Results: 226 hypertension RCTs were included for final analysis. RCTs conducted in university affiliated hospitals ($P < 0.001$) or secondary/tertiary hospitals ($P < 0.001$) were scored at higher internal validity. Multi-center studies (median = 4.0, IQR = 2.0) were scored higher internal validity score than single-center studies (median = 3.0, IQR = 1.0) ($P < 0.001$). Funding-supported trials had better methodological quality ($P < 0.001$). In addition, the reporting of inclusion criteria also leads to better internal validity ($P = 0.004$). Multivariate regression indicated sample size, industry-funding, quality of life (QOL) taken as measure and the university affiliated hospital as trial setting had statistical significance ($P < 0.001$, $P < 0.001$, $P = 0.001$, $P = 0.006$ respectively).

Conclusion: Several components relate to the external validity of RCTs do associate with the internal validity, that do not stand in an easy relationship to each other. Regarding the poor reporting, other possible links between two variables need to trace in the future methodological researches.

1. Introduction

As the design and conduct has effectively eliminated the possibility of bias and confounding [1], randomized controlled trials (RCTs) having a favorable internal validity and being the gold standard for determining the effects of treatments, have been widely recognized in clinical researches [2–5]. Much of the methodological discussion around RCTs is framed in terms of the notions of internal and external validity. Both validities appeal to us all as obvious requisites for the worth of a RCT. Internal validity reflects the extent of confidence to RCTs' results, while the external validity needs to be emphasized too as it reflects the extent of RCT's conclusions to be generalized [6,7]. If a RCT is not externally valid, then its results cannot be said to hold outside of the research setting, and thus, even if internally valid, we cannot use its results to say anything relevant of the clinical setting; if RCTs were misused or the results from RCTs were irrelevant to the patients in a particular clinical setting [1,8,9], that may adversely affect to patients. Lack of external validity is frequently advocated as one of the obstacles to the translation of research evidence into clinical practice, which is why interventions found to be effective in clinical trials and recommended in guidelines are underused in clinical practice [1,10,11]. Although most of the current arguments and disputes around the

* Corresponding author.
E-mail address: deyingkang@126.com (D. Kang).

http://dx.doi.org/10.1016/j.conctc.2015.10.004

use of RCTs in clinical practice refer to either type of validity, it is surprising that not much has been researched systematically about the relationship between the internal and the external validity of RCTs. Hypertension has become a serious burden disease in China [12,13]; although a great number of clinical trials on hypertension have been conducted within China, few studies were successful in developing as evidence based information and disseminating to patients under specific circumstances [13]. Taking the example of hypertension, this study intends to explore the relationship between the external and internal validity of RCTs systematically.

2. Materials and methods

2.1. Search strategy and study selection

A systematic literature search was conducted to identify all relevant randomized controlled trials on hypertension using databases (incept-2010.6) including Medline (Ovid), Embase, CCTR (Cochrane Central Register of Controlled Trials, Ovid), CBMdisc (Chinese biomedical literature database), CNKI (China National Knowledge Infrastructure/China Academic Journals Full-text Database) and VIP (Chinese scientific journals database); articles with 'hypertension', 'randomized controlled trial', 'controlled clinical trial' and 'random allocation' as general keyword terms, free words or exploded MeSH terms were searched as English and corresponding Chinese search terms to identify studies from above databases. In addition, reference lists of included articles were screened for additional articles.

Titles and abstracts of all citations were independently evaluated by two reviewers (WYX and KD). The full texts of the potentially relevant articles were obtained and independently evaluated by the same two authors. Disagreement was resolved by consensus. Studies were included if (1) drug therapy for primary hypertension, covering the six kinds of anti-hypertension drugs in which recommended by WHO were included (ACEI, Angiotensin-Converting Enzyme Inhibitor; ARB, Angiotensin Receptor Blocker; CCB, Calcium Channel Blocker; alpha-blocker; beta-blocker; Diuretics); (2) studies grading score equal or greater than 3. Studies were excluded if (1) recruited patients with secondary hypertension; (2) that published as abstracts only; (3) reported partial data from multi-center research.

2.2. Internal validity assessment

The scale for assessing internal validity of RCTs were modified from two RCTs-based tools, the Jadad scale [13] and the evaluation criteria of risk of bias in Cochrane Review's Handbook [14]. The scale developing for RCTs include five items: randomization (0–2 points), allocation concealment (0–2 points), blinding (0–2 points), attrition (0–2 points) and baseline condition (0–1 points); the maximum score for a perfect RCT is 9. To study the relationship between internal validity and external validity, all included RCTs were divided into four groups (3-score group, 4-score group, 5-score group and 6–9 scores group).

Meanwhile, 50 RCTs were selected randomly using a computer-generated list to validate inter-rater agreement of applying the modified scale. The agreement for each item and the whole scale was explained by percentage of actual agreement as well as Kappa coefficient. We adopted the Kappa values of <0 rates as less than chance agreement, 0.01–0.20 as slight agreement, 0.21–0.40 as fair agreement, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial agreement, and 0.81–0.99 as almost perfect agreement [15]. In addition, Jadad scale [16] was taken as reference standard to validate criterion validity of this modified scale. Two authors (ZX, WYX) conducted a critical appraisal of the internal validity of all studies by using the modified scale; any disagreement between reviewers was submitted to the third author (KD) and resolved by consensus.

2.3. Data abstraction for evaluating external validity

From each publication, information was extracted regarding characteristics of included RCTs, such as subjects recruitment, baseline characteristics of subjects, interventions, outcomes and any further information about external validity by a pre-developed form [17,18]. The data extract form for evaluating external validity includes 4 domains and 25 items totally, the checklist has been developed by listing the most commonly used assessment criteria for clinical studies [1,35]. Of this, the domain of "source" has 5 items: region of trial setting, research setting, research date, number of centers involved, funding source; domain of "subjects recruitment" includes 7 items: location, setting, method, duration of recruitment, number of eligible patients, number of patients not meeting inclusion criteria, number of patients who refusing participation; domain of "baseline characteristics of subjects" has 8 items: sample size, source of patients, age, gender, diagnosis criteria, duration of disease, state of disease, complications; the last 4 domain relates to patient reported outcomes, includes "effectiveness outcomes" and "adverse events" respectively. A meeting followed in which the ratings were reviewed and any disagreements were resolved by discussion and consensus with the third author (KD). Two reviewer (ZX, WY) independently completed all the data extractions.

2.4. Statistical analysis

A description of the data included rate and proportion used for dichotomous data, and medians (inter-quartile range, IQR) or mean ± SD (standard deviation) for continuous data. Possible differences between groups were calculated with Mann–Whitney test or Kruskal–Wallis test for continuous variables. Correlation coefficients were taken to validate criterion validity of the modified scale for internal validity. The statistical significance level was set at 0.05 and all tests were two-sided. Bonferroni correction was used of multiple comparisons if possible; in that case, the statistical significance level was re-settled accordingly. Multiple linear regressions were used to test the relationship of internal and external validity in terms of characteristics of RCTs, baseline characteristics of subjects, interventions and outcomes, the grading score of internal validity was taken as dependent variable. Data analysis was done using SPSS software, version 21.0 (SPSS, Chicago, IL).

3. Results

3.1. Flow of included studies

1197 RCTs were identified from the searches (excluding 136 duplicates and 4888 non-relevant articles), additional 99 RCTs were excluded based on the inclusion criteria; after that, the evaluation of internal validity was performed by applying the modified scale, 226 RCTs with internal validity scores of ≥3 remained for final analysis (Fig. 1).

3.2. Validation of the modified scale for grading internal validity

In order to evaluate the criterion validity of the scale, we select 50 RCTs randomly using a computer-generated list to validate inter-rater agreement. Total mean score was converted into the percentage of the maximum score for the modified scale, the ICC against Jadad score was 0.84, that is, the results of the modified