

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



Xin Zhang <sup>a</sup>, Yuxia Wu <sup>b, c</sup>, Pengwei Ren <sup>b</sup>, Xueting Liu <sup>b</sup>, Deying Kang <sup>b, \*</sup>

<sup>a</sup> Department of Integrated Traditional Chinese and Western Medicine, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>b</sup> Department of Evidence-based Medicine and Clinical Epidemiology, West China Hospital, Sichuan University, Chengdu 610041, China

<sup>c</sup> 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 (CTR), 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.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## 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](mailto:deyingkang@126.com) (D. Kang).

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  $\pm$  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).

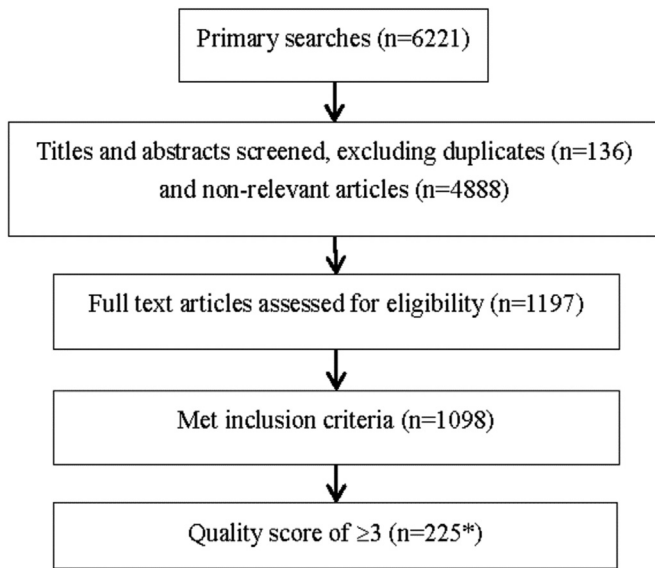
## 3. Results

### 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  $\geq 3$  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



\*Total 225 articles were remained (one report includes 2 RCTs).

Fig. 1. Flow of the RCTs selection.

scale were highly convergence with the results of Jadad score.

And then, the rater-agreement validation of full-sample RCTs was performed. Substantial agreements were also observed for all of the items. Specially, the rater agreements for each items varied from 88% to 100% with a total percentage of 76%, of that, 4 items reach to almost perfect agreements (>90.0%); however, the kappa values of inter-rater ranged widely from 0.63 to 0.90 with a total of 0.72, of that, 3 items had excellent agreements (above 80.0%) between raters (Table 1).

### 3.3. Internal validity assessment by applying the modified scale

226 RCTs with a score of 3 or more were included after applying the modified scale. The median grade of included RCTs was 3 with IQR of 1, the minimum was 3, and maximum was 9. Of those, 56.2% (n = 127) of RCTs were scored at 3, only 3.1% (n = 7) of RCTs were ≥7 (Table 2).

### 3.4. The relationship between the internal validity and external validity

#### 3.4.1. Characteristics of RCTs

3.4.1.1. *Research setting and research date.* The median scores of internal validity within RCTs conducted in two regions (south China or north China) were both at 3.0 with an IQR of 1.0 (P = 0.200). The scores among RCTs conducted in different level hospitals were calculated and significant difference was found (P < 0.001), with

**Table 2**  
Grading scores of internal validity for included RCTs.

Grading score	n	Percentage (%)	Cumulative percentage (%)
9	1	0.4	0.4
8	2	0.9	1.3
7	4	1.8	3.1
6	17	7.5	10.6
5	27	11.9	22.6
4	48	21.2	43.8
3	127	56.2	100.0

RCT, Randomized Controlled Trial.

averages of 3.00 (IQR = 0.00) and 4.00 (IQR = 2.00) respectively; 73.5% RCTs were conducted in university affiliated hospitals and were scored at higher points too (P < 0.001) (Table 3). Based on the RCTs carried out in the periods before 2000, 2001 to 2005 and 2006 to 2009 respectively, the difference of internal validity was not statistically significant (P = 0.272) (Table 3).

#### 3.4.1.2. Number of research centers and funding status.

Multi-center studies were graded higher score of internal validity than that of single-center studies with medians of 4.0 and 3.0 respectively (P < 0.001) (Table 3). In addition, RCTs with funding support also have higher score of internal validity (P < 0.001). We found that industry (manufacturer of the experimental drug) accounted 30.8% of funding source in China. RCTs either drug industry-funded or/and nonprofit-funded (e.g., from the government or nonprofit institutes) had higher score of internal validity than that of RCTs without funding, the median scores were 4.0 (IQR = 2.0), 4.0 (IQR = 1.0) and 6.0 (IQR = 2.0) accordingly (Table 3).

#### 3.4.2. Baseline characteristics of subjects

3.4.2.1. *Gender and age.* 200 (88.5%) RCTs reported the proportion of female patients, no significant difference was observed (P = 0.582) among RCTs with different grades of internal validity (Table 4). Similar result for age was observed too, patient ages were presented in 83.6% (n = 189) of RCTs and no statistical significance was found (P = 0.568) (Table 4).

3.4.2.2. *Sample source and sample size.* RCTs recruited patients in the out-patient department was more likely to have a higher internal validity, that is, the patients source was relevant to internal validity (P = 0.043) (Table 3).

Sample size within group 4 (median = 221, IQR = 235) was significantly larger than that of two groups (group 1, median = 94, IQR = 75, P < 0.001; group 2, median = 99.5, IQR = 115.5, P = 0.006). Sample sizes within groups 1, 2 and 3 did not differ significantly from each other (P > 0.0125).

3.4.2.3. *Inclusion/exclusion criteria and diagnostic criteria.* Internal validity in RCTs with setting inclusion criteria was higher significantly than that of without (P = 0.004); however, no

**Table 1**  
Assessment of the internal validity of selected 226 RCTs and agreements of inter-raters.

Item	Yes, n (%)	No, n (%)	Unclear, n (%)	Raters		P value
				Agreements	Kappa	
Randomization	48 (21.2)	3 (1.3)	175 (77.4)	98%	0.90	0.001
Allocation concealment	14 (6.2)	212 (93.8)	0 (0.0)	100%	–	–
Blinding	82 (36.3)	73 (32.3)	71 (31.4)	96%	0.88	0.001
Attrition	16 (7.1)	98 (43.4)	112 (49.6)	88%	0.63	0.001
Baseline condition	185 (81.9)	41 (18.1)	–	90%	0.80	0.001
Total	–	–	–	76%	0.72	0.001

**Table 3**  
Scores of methodological quality of included RCTs according to different strata.

	n (%)	Grading internal validity [median (IQR)]	P value
<b>External characteristics of RCTs</b>			
Research setting			
Region of trial setting, south China vs north China	105(53.3%)/92(46.7%)	3.0(1.0)/3.0(1.0)	0.200
University affiliated hospital, yes/no	166(73.5%)/60(16.5%)	4.00(2.00)/3.00(0.00)	<0.001
Hospital class primary vs secondary or tertiary	41(18.1%)/185(81.9%)	3.00(0.00)/4.00(2.00)	<0.001
Date of study 2000/2005/2006 vs 2009	45(19.9%)/92(40.7%)/89(39.4%)	3.00(1.00)/4.00(1.75)/3.00(1.00)	0.272
Number of centers involved, single-center vs multi-center study	172(76.1%)/54(23.9%)	3.0(1.0)/4.0(2.0)	<0.001
Funding source <sup>a</sup>			
Funding, yes/no	82(36.3%)/144(63.7%)	4.0(2.0)/3.0(1.0)	<0.001
Industry, yes/no	64(30.8%)/144(69.2%)	4.0(2.0)/3.0(1.0)	<0.001
Non-profit, yes/no	11(7.1%)/144(92.9%)	4.0(1.0)/3.0(1.0)	0.015
mixed, yes/no	7(4.6%)/144(95.4)	6.0(2.0)/3.0(1.0)	0.001
<b>Baseline characteristics of subjects</b>			
Source of patients, outpatient/inpatient/both	82(73.9%)/7(6.3%)/22(19.8%)	3.00(2.00)/3.00(0.00)/3.00(0.00)	0.043
Inclusion and exclusion criteria			
Inclusion criteria, yes/no	199(88.4%)/26(11.6%)	3.00(1.00)/3.00(0.00)	0.004
Exclusion criteria, yes/no	190(88.4%)/25(11.6%)	3.00(1.00)/7.00(6.00)	0.480
Diagnostic criteria			
Diagnostic criteria, yes/no	125(55.6%)/100(44.4%)	3.00(1.00)/3.00(1.00)	0.731
Diagnosis criteria, WHO/China	84(70.0%)/36(30.0%)	3.00(1.00)/3.00(2.00)	0.802
Complications			
Complications excluded, yes/no	185(81.9%)/41(18.1%)	3.00(1.00)/3.00(1.50)	0.633
CHD excluded, yes/no	131(58.0%)/95(42.0%)	3.00(1.00)/3.00(1.00)	0.066
Stroke excluded, yes/no	104(46.0%)/122(54.0%)	3.50(2.00)/3.00(1.00)	0.081
Renal insufficiency excluded, yes/no	167(73.9%)/59(26.1%)	3.00(1.00)/3.00(2.00)	0.865
Diabetes excluded, yes/no	92(40.7%)/134(59.3%)	3.00(1.00)/3.00(1.00)	0.445
Heart failure excluded, yes/no	128(56.6%)/98(43.4%)	3.00(2.00)/3.00(1.00)	0.060
<b>Interventions</b>			
Drugs			
alpha-blocker	25 (11.1%)	3.00 (1.00)	0.196
beta-blocker	19 (8.4%)	4.00 (1.00)	
ACEI	29 (12.8%)	3.00 (1.00)	
ARB	59 (26.1%)	3.00 (2.00)	
CCB	43 (19.0%)	4.00 (2.00)	
Diuretics	8 (3.5%)	4.00 (1.75)	
Drug combination or compound preparation	43 (19.0%)	3.00 (1.00)	
<b>Outcomes</b>			
Effectiveness outcomes			
Blood pressure value, yes/no	213(94.2%)/13(5.8%)	3.00(1.00)/3.00(4.00)	0.880
Effective rate, yes/no	175(77.4%)/51(22.6%)	3.00(2.00)/3.00(1.00)	0.058
Laboratory index, yes/no	55(24.3%)/171(75.7%)	3.00(1.00)/3.00(2.00)	0.078
Quality of life, yes/no	14(6.2%)/212(93.8%)	4.50(4.00)/3.00(1.00)	0.025

IQR, inter-quartile range.

<sup>a</sup> Industry, manufacturer of the experimental drug; nonprofit, such as the government; mixed, both industry and nonprofit sources.

significant differences were observed either in setting exclusion criteria or not ( $P = 0.480$ ) or in total number of exclusion criteria ( $P = 0.109$ ) (Table 4).

Regarding to diagnostic criteria, internal validity showed no significant difference in reporting diagnostic criteria or not ( $P = 0.731$ ); in addition, diagnostic criteria source was not related to internal validity ( $P = 0.802$ ) (Table 3).

**3.4.2.4. Duration and state of disease.** Duration of disease in four groups did not differ significantly from each other ( $P = 0.278$ ) (Table 4). Proportion of grade III hypertension was only reported in 4 RCTs, of those, 3 RCTs (proportion range: 0.163–0.317) scored 3 and 1 RCT with reporting of proportion of 0.068 was scored at 1.

**3.4.2.5. Complications.** 185 RCTs (81.9%) excluded patients with complications showed no significant difference in grading internal validity ( $P = 0.633$ ) (Table 3). Of those, 58.0% patients had coronary heart disease ( $P = 0.066$ ), 25% patients suffered from stroke ( $P = 0.081$ ), 46.0% patients with renal insufficiency ( $P = 0.865$ ), 40.7% patients with diabetes ( $P = 0.445$ ) and 56.6% patients with heart failure ( $P = 0.060$ ) removed, that is, significant differences of internal validity were not observed among RCTs excluding complications or not ( $P > 0.05$ ) (Table 3).

### 3.4.3. Drugs and course of treatment

Regarding to addressed drugs, 19.0% ( $n = 43$ ) RCTs were designated to test drug combination or compound preparation, while the proportion of RCTs designated to test single drug was 81.0% ( $n = 183$ ). Categories of drug therapy showed no significant difference in internal validity ( $P = 0.196$ ). Course of treatment was not relevant to internal validity ( $P = 0.070$ ) (Table 4).

### 3.4.4. Outcomes

**3.4.4.1. Safety measure.** RCTs having a higher internal validity were more likely to report lower adverse events rate ( $P = 0.047$ ) (Table 4). Of those, adverse events rate for group 4 (median = 0.10; IQR = 0.11) (Table 3) was significantly lower than that in group 2 (median = 0.15; IQR = 0.16,  $P = 0.011$ ). Adverse events rate in groups 1, 2 and 3 did not differ significantly from one another ( $P > 0.0125$ ) (Table 3).

**3.4.4.2. Efficacy measure.** Internal validity within RCTs using blood pressure value ( $P = 0.880$ ), response rate ( $P = 0.058$ ), or laboratory index ( $P = 0.078$ ) as outcomes or not were different insignificantly; however, significant difference was observed in adopting quality of life as the outcome or not ( $P = 0.025$ ) (Table 3).

**Table 4**  
Characteristics of included RCTs according to different groups of internal validity score.

Grading internal validity	n (%)	Median (IQR)	Mean rank	P value
<b>Baseline characteristics of subjects</b>				
Proportion of female patients				
Group 1	111 (55.5%)	0.41 (0.11)	96.03	0.582
Group 2	43 (21.5%)	0.44 (0.09)	107.85	
Group 3	23 (11.5%)	0.42 (0.10)	99.13	
Group 4	23 (11.5%)	0.44 (0.16)	109.72	
Age				
Group 1	100 (52.9%)	53.00 (8.71)	98.79	0.568
Group 2	43 (22.8%)	53.50 (4.85)	96.19	
Group 3	24 (12.7%)	51.68 (4.57)	82.35	
Group 4	22 (11.6%)	52.21 (9.34)	89.27	
Number of exclusion criteria				
Group 1	120 (56.1%)	6.00 (4.00)	99.33	0.109
Group 2	47 (22.0%)	7.00 (6.00)	117.06	
Group 3	23 (10.7%)	6.00 (7.00)	108.59	
Group 4	24 (11.2%)	7.00 (7.00)	128.56	
Sample size				
Group 1	127 (56.2%)	94.00 (75.00)	105.92	0.002
Group 2	48 (21.2%)	99.50 (115.50)	115.90	
Group 3	27 (11.9%)	63.00 (176.00)	103.22	
Group 4	24 (10.6%)	221.00 (235.00)	160.40	
Duration of disease				
Group 1	46 (47.9%)	12.00 (97.50)	53.15	0.278
Group 2	24 (25.0%)	12.00 (45.00)	43.81	
Group 3	17 (65.4%)	12.00 (64.80)	46.74	
Group 4	9 (9.4%)	12.00 (30.00)	40.56	
<b>Interventions</b>				
Course of treatment				
Group 1	126 (56.3%)	8.00 (6.00)	111.19	0.070
Group 2	47 (21.0%)	8.00 (2.00)	99.91	
Group 3	27 (12.1%)	8.00 (0.00)	116.04	
Group 4	24 (10.7%)	8.00 (18.00)	140.06	
<b>Outcomes</b>				
Adverse events rate				
Group 1	120 (51.9%)	0.13 (0.12)	114.79	0.047
Group 2	53 (22.9%)	0.15 (0.16)	132.80	
Group 3	33 (14.3%)	0.11 (0.12)	115.09	
Group 4	25 (10.8%)	0.10 (0.11)	87.40	

IQR, inter-quartile range.

#### 3.4.5. Multiple linear regression

Multiple linear regressions were further used to explore possible dominants relate to internal validity (The grading score of internal validity taken as dependent variable), we found sample size, industry-funding, the reporting of quality of life and university affiliated hospital as the trial setting, were associated with internal validity ( $P < 0.001$ ,  $P < 0.001$ ,  $P = 0.001$ ,  $P = 0.006$  respectively), of those, the sample size rank as the first choice (0.253) based on the standardized beta coefficients (Table 5).

## 4. Discussion

More and more methodologists and practitioners of researches consider that internal validity and external validity stand in a

relationship best described as a “trade-off” [19–21] (the more we ensure that the treatment is isolated from potential confounders in order to make certain that the observed effect is attributable to the treatment, the more unlikely it is that the experimental results can be representative of phenomena of the outside world). Although this seems to be the standard view regarding the relationship between internal and external validity, it is not the only one. Another idea that internal validity is in a prerequisite to external validity was also found [22–25]. Although these two positions need not necessarily be contradictory, they do not stand in an easy association to each other. The existence of a relationship between internal and external validity constitutes a commonplace in the experimental and in the methodological literature around experimental medicine. In this study, we attempt to use a sample of hypertension RCTs conducted in China to explore the relationship between the external and internal validity systematically.

There are several interesting findings in our study. Firstly, internal validity associated with trials setting, either university affiliated hospital or secondary/tertiary hospital, seem to have higher grading score of internal validity. That can be explained by the research capacity of personnel and institutes; trials conducted at above hospitals may suffer from less systematic error [14]. On the contrast, for clinicians in primary hospitals, high workloads and poor knowledge and research skills have been identified as main barriers to undertaking good research [26,27]. Secondly, industry-funded RCTs were graded higher score of internal validity too. This finding is consistent with that of a demonstrating that the trials supported by industry had better methodological quality [15]. Industry-funded trials having more financial resources were more likely to be designed as large scale, multicenter, international trials, the more likely it is that the trials results can be representative of phenomena of real world; besides, industry funding trials were more likely to be published in journals with a higher impact factor [16] and accordingly, were more likely to be of adequate methodological quality [15,17]. Domains relate to internal validity were further investigated in our study, but the associations between funding and each domains of internal validity were not found. Another study conducted by Mugambi et al. [18] identified there was no significant association between funding source and methodological quality of RCTs in terms of sequence generation, allocation concealment, blinding and selective reporting. However, there was a significant association between funding and methodological quality of RCTs in the domains of incomplete outcome data and free of other bias. Industry funded trials had a higher percentage of free of other bias as well as had less missing data than those of non-industry funded trials significantly.

Thirdly, similar to another report [17], our study revealed a significant association between the number of research centers and the internal validity of these trials. The benefits of multicenter trials include a larger number of participants (they differ along a wide range of factors, such in age, gender, behavior, height, intelligence, and so forth.) recruited, different geographic locations, the

**Table 5**  
Multiple linear regression for essential aspects of external validity to internal validity.

	Unstandardized coefficients		Standardized coefficients ( $\beta$ )	t	P value	95%CI for $\beta$
	$\beta$	SE				
Constant	2.698	0.176	—	15.338	<0.001	2.351–3.044
Sample size	0.308	0.075	0.253	4.135	<0.001	0.161–0.455
Industry-funding	0.577	0.154	0.229	3.751	<0.001	0.274–0.880
Quality of life	1.084	0.329	0.200	3.294	0.001	0.435–1.732
University affiliated hospital	0.455	0.164	0.172	2.770	0.006	0.131–0.778

t, t-value; CI, confidence interval; SE, standard error.

possibility of inclusion of a wider range of population groups, and the ability to compare results among centers, all of which increase the generalizability of those multicenter trials and meanwhile are more likely to reduce the risk of biases, increase statistical power and precision. As both internal and external validity benefit from multicenter trials, such kinds of trial seem to be a preferred choice in clinical researches.

Fourthly, internal validity benefits from the trial setting selected as well as the reporting of stringent eligibility criteria. However, in this situation, the internal validity and external validity stand in a “trade-off” relationship. The inclusion and exclusion criteria for a RCT are designed to identify a population of interest in whom an intervention has the greatest likelihood to produce a clinically important and statistically significant effect [28]. The advantages of stringent eligibility criteria are achieved at the risk of excluding patients who may be more likely to represent the population treated in clinical settings and who would better test an intervention’s effectiveness.

Fifthly, the RCTs with a larger sample size were more likely to be graded a higher score of internal validity. Trial large enough will have a high probability (power) of detecting as statistically significant a clinically important difference of a given size if such a difference exists [29], on contrast, reports of RCTs with small samples frequently include the erroneous conclusion that the intervention groups do not differ, when in fact too few patients were studied to make such a claim [30]. In addition, insufficient trial size may cause over homogenous patients to be enrolled (selection biases); simultaneously, since one of the main goals of dissertations that adopt RCT research design is to make generalizations from the sample being studied to the population the sample is drawn from, and in some cases, across populations, selection biases are arguably one of the most significant threats to external validity.

Sixthly, trials adopting quality of life as the patient preferred outcome have higher internal validity too. Quality of life (QoL) is the subjective perception of how an individual feels about their health status and/or the non-medical aspects of their lives [31,32]. As the need for a more holistic view of outcome during and following illness is required, the role of QoL measurement assumes increasing importance [33] and is particularly important for patients with chronic diseases (e.g., hypertension) [34–37]. Required as an end point of clinical trials [33], a measure of QoL instead of surrogate parameters (e.g. laboratory values) may improve the representativeness of a study population (improving external validity synchronously) [35]. However, the explanation by which the RCTs reporting QoL as an outcome may exert better internal validity has not been known.

There are several limitations in our study. Of those, the poor reporting in most included RCTs make it hard to analyze the relationship between the external and internal validity thoroughly [38]; information related to external validity was not reported or was reported insufficiently too, there is marked room for improving quality of the reporting in RCTs, especially at the respects related to external validity. In addition, because the methodological quality of an RCT was assessed based on its published report, we cannot be certain whether our findings represent incomplete reporting or inadequate performance of these measures. Plausibly, investigators may not have reported important quality measures despite their adequate performance, causing an underestimation of the study internal validity.

Finally, including studies confined in China and in a certain field may impact the external validity of our research. Future research may concentrate on the other therapeutic areas or other counties using the approach reported in this research to explore the relationship between the external and internal validity of RCTs; the results of these two research studies could then be compared.

## 5. Conclusion

This study has identified the relationship between the internal validity and several domains of the external validity of RCTs in China, that do not stand in an easy relationship to each other. Taking factors that can influence the representativeness of a study population to makeup of external validity to explore the relationship between the internal and external validity is somewhat feasible; other possible links between two validities needed to demonstrate in the future methodological researches.

## Author contributions

Conceived and designed the experiments: DK. Performed the experiments: XZ YW. Analyzed the data: XZ YW RP. Wrote the manuscript: XZ DK RP XL.

## References

- [1] P.M. Rothwell, External validity of randomised controlled trials: “to whom do the results of this trial apply?”, *Lancet* 365 (2005) 82–93.
- [2] L.M. Marcia, A brief history of the randomized controlled trial – from oranges and lemons to the gold standard, *Hematol. Oncol. Clin. N. Am.* 14 (2000) 745–760.
- [3] D. Moher, B. Pham, A. Jones, D.J. Cook, A.R. Jadad, M. Moher, et al., Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 352 (1998) 609–613.
- [4] K.F. Schulz, I. Chalmers, R.J. Hayes, D.G. Altman, Empirical evidence of bias: dimensions of methodological quality associated with estimates of treatment effects in controlled trials, *JAMA* 273 (1995) 408–412.
- [5] G. Antes, The evidence base of clinical practice guidelines, health technology assessments and patient information as a basis for clinical decision-making, *Z. Arztl. Fortbild. Qual.* 98 (3) (2004) 180–184.
- [6] R.E. Glasgow, L.W. Green, L.M. Klesges, D.B. Abrams, E.B. Fisher, M.G. Goldstein, et al., External validity: we need to do more, *Ann. Behav. Med.* 31 (2006) 105–108.
- [7] O.M. Dekkers, E. von Elm, A. Algra, J.A. Romijn, J.P. Vandenbroucke, How to assess the external validity of therapeutic trials: a conceptual approach, *Int. J. Epidemiol.* 39 (1) (2010) 89–94.
- [8] H. Burchett, M. Umoquit, M. Dobrow, How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks, *J. Health Serv. Res. Policy* 16 (2011) 238–244.
- [9] B. Swinburn, T. Gill, S. Kumanyika, Obesity prevention: a proposed framework for translating evidence into action, *Obes. Rev.* 6 (2005) 23–33.
- [10] N. Ahmad, I. Boutron, D. Moher, I. Pitrou, C. Roy, P. Ravaud, Neglected external validity in reports of randomized trials: the example of hip and knee osteoarthritis, *Arthritis Care Res.* 61 (2009) 361–369.
- [11] C.J. Metge, What comes after producing the evidence? The importance of external validity to translating science to practice, *Clin. Ther.* 33 (2011) 578–580.
- [12] Y.X. Wu, D.Y. Kang, Q. Hong, J.L. Wang, External validity and its evaluation in clinical trials, *Chin. J. Epidemiol.* 32 (5) (2011) 514–518 (in Chinese).
- [13] National Center for Cardiovascular Disease Control, Report on Cardiovascular Disease in China (2008–2009), Encyclopedia of China Publishing House, Beijing, 2009, pp. 12–32.
- [14] B. Green, J. Segrott, J. Hewitt, Developing nursing and midwifery research capacity in a university department: a case study, *J. Adv. Nurs.* 56 (3) (2006) 302–313.
- [15] Y. Lu, Q. Yao, J. Gu, C. Shen, Methodological reporting of randomized clinical trials in respiratory research in 2010, *Respir. Care* 58 (9) (2013) 1546–1551.
- [16] N.A. Khan, J.I. Lombeida, M. Singh, H.J. Spencer, K.D. Torralba, Association of industry funding with the outcome and quality of randomized controlled trials of drug therapy for rheumatoid arthritis, *Arthritis Rheum.* 64 (7) (2012) 2059–2067.
- [17] V. Bridoux, G. Moutel, H. Roman, B. Kianifard, F. Michot, C. Herve, J.J. Tuech, Methodological and ethical quality of randomized controlled clinical trials in gastrointestinal surgery, *J. Gastrointest. Surg.* 16 (9) (2012) 1758–1767.
- [18] M.N. Mugambi, A. Musekiwa, M. Lombard, T. Young, R. Blaauw, Association between funding source, methodological quality and research outcomes in randomized controlled trials of synbiotics, probiotics and prebiotics added to infant formula: a systematic review, *BMC Med. Res. Methodol.* 13 (1) (2013) 137.
- [19] S.S. Brehm, S.M. Kassin, S. Fein, *Social Psychology*, fourth ed., Mifflin and Company, Boston: Houghton, 1999.
- [20] E.R. Smith, D.M. Mackie, *Social Psychology*, Psychology Press, Philadelphia, 1999.
- [21] F. Guala, On the scope of experiments in economics: comments on Siakantaris, *Camb. J. Econ.* 26 (2002) 261–267.

- [22] S.R. Thye, Reliability in experimental sociology, *Soc. Forces* 78 (2000) 1277–1309.
- [23] J.W. Lucas, Theory-testing, generalization and the problem of external validity, *Sociol. Theory* 21 (2003) 236–253.
- [24] F. Guala, Experimental localism and external validity, *Philos. Sci.* 70 (2003) 1195–1205.
- [25] R.B. Hogarth, The challenge of representativeness design in psychology and economics, *J. Econ. Methodol.* 12 (2005) 253–263.
- [26] A.M. Hutchinson, L. Johnston, Bridging the divide: a survey of nurses' opinions regarding barriers to, and facilitators of, research utilization in the practice setting, *J. Clin. Nurs.* 13 (2004) 304–315.
- [27] K. Parahoo, Barriers to, and facilitators of, research utilization amongst nurses in Northern Ireland, *J. Adv. Nurs.* 31 (1) (2000) 89–98.
- [28] S.B. Hulley, T.B. Newman, S.R. Cummings, Choosing the study subjects: specification, sampling, and recruitment, in: S.B. Hulley, S.R. Cummings, W.S. Browner, et al. (Eds.), *Designing Clinical Research*, second ed., Lippincott Williams & Wilkins, Philadelphia, Pa, 2001, pp. 25–34.
- [29] D. Moher, S. Hopewell, K.F. Schulz, V. Montori, P.C. Gøtzsche, P.J. Devereaux, D. Elbourne, M. Egger, D.G. Altman, CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials, *Int. J. Surg.* 10 (1) (2012) 28–55.
- [30] D.G. Altman, J.M. Bland, Absence of evidence is not evidence of absence, *BMJ* 311 (1995) 485.
- [31] T.M. Gill, A.R. Feinstein, A critical appraisal of the quality of life measurements, *JAMA* 272 (1994) 619–626.
- [32] M. Bullinger, U. Ravens-Sieberer, Health related quality of life assessment in children: a review of the literature, *Rev. Eur. Psychol. App* 45 (1995) 245–254.
- [33] M. Barnes Peter, E.M. Jenney Meriel, Measuring quality of life, *Curr. Paediatr.* 12 (6) (2002) 476–480.
- [34] Julia Addington-Hall, Lalit Kalra, Who should measure quality of life? *BMJ* 322 (7299) (2001) 1417–1420.
- [35] G. Bornhöft, S. Maxion-Bergemann, U. Wolf, G.S. Kienle, A. Michalsen, H.C. Vollmar, S. Gilbertson, P.F. Matthiessen, Checklist for the qualitative evaluation of clinical studies with particular focus on external validity and model validity, *BMC Med. Res. Methodol.* 6 (2006) 56.
- [36] Medical Research Council Working Party, MRC trial of treatment of mild hypertension: principle results, *BMJ* 2 (1985) 97–104.
- [37] Medical Research Council Working Party, MRC trial of treatment of hypertension in older adults: principle results, *BMJ* 304 (1992) 405–412.
- [38] A.R. Jadad, R.A. Moore, D. Carroll, C. Jenkinson, D.J. Reynolds, D.J. Gavaghan, H.J. McQuay, Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin. Trials* 17 (1) (1996) 1–12.