

ORIGINAL ARTICLE

Many meta-analyses of rare events in the Cochrane Database of Systematic Reviews were underpowered

Pengli Jia^a, Lifeng Lin^b, Joey S.W. Kwong^c, Chang Xu^{d,*}

^aSchool of Management, Shanxi Medical University, Taiyuan, China

^bDepartment of Statistics, Florida State University, Tallahassee, FL, USA

^cGlobal Health Nursing, St. Luke's International University, Tokyo, Japan

^dDepartment of Population Medicine, College of Medicine, Qatar University, Doha, Qatar

Accepted 23 November 2020; Published online 30 November 2020

Abstract

Background and Objective: Meta-analysis is a statistical method with the ability to increase the power for statistical inference, while it may still face the problem of being underpowered. In this study, we investigated the power to detect certain true effects for published meta-analyses of rare events.

Methods: We extracted data from the Cochrane Database of Systematic Reviews for meta-analyses of rare events from January 2003 to May 2018. We retrospectively estimated the power to detect a 10–50% relative risk reduction (RRR) of eligible meta-analyses. The proportion of meta-analyses achieved a sufficient power (≥ 0.8) were estimated.

Results: We identified 4,177 meta-analyses. The median power to detect 10%, 30%, and 50% RRR were 0.06 (interquartile range [IQR]: 0.05 to 0.06), 0.08 (IQR: 0.06 to 0.15), and 0.17 (IQR: 0.10 to 0.42), respectively; the corresponding proportion of meta-analyses that reached sufficient power were 0.32%, 3.68%, and 11.81%. Meta-analyses incorporating data from more studies had higher probability to achieve a sufficient power (rate ratio = 2.49, 95% CI: 1.76, 3.52, $P < 0.001$).

Conclusion: Most of the meta-analyses of rare events in Cochrane systematic reviews were underpowered. Future meta-analysis of rare events should report the power of the results to support informative conclusions. © 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Keywords: Meta-analysis of rare events; Power analysis; Cochrane systematic reviews; Statistical inference; Decision-making; Relative risk reduction

1. Introduction

Meta-analysis is a crucial tool to synthesize findings from available studies of the same topic, and it has the

Funding: PJ was supported by the Science Research Start-up Fund for Doctor of Shanxi Province (SD1819). LF-K is funded by an Australian National Health and Medical Research Council Early Career Fellowship (APP1158469). The funding sources had no role in the design, data collection, analysis, interpretation of the data or decision to submit the results.

Conflicts of interests: None.

Data sharing: We have no copyright to share the data to public. For researchers wishing to obtain data for academic use, they are advised to contact the corresponding author.

Authors' contributions: X.C. conceived, designed the study, and developed the code; X.C. and P.J. drafted the manuscript; X.C. and L.L. acquired the data, analyzed the data, and interpreted the results; L.L. and J.K. contributed careful edits for the manuscript. All authors approved the final version.

* Corresponding author. Department of Population Medicine, College of Medicine, Qatar University, Al Jamiaa Street, P. O. Box 2713, Doha, Qatar. Tel.: +974 30070343; fax: +974 7843.

E-mail address: xuchang2016@runbox.com (C. Xu).

ability to increase power in testing whether the true effect actually exists [1–3]. This advantage is expected to be more apparent for homogeneous studies synthesized under a fixed-effect analytical model [4]. For heterogeneous studies that are synthesized using a random-effects model, due to additional between-study variance, the power is generally lower than fixed-effect model but still higher, on average, than a single study [5]. This is one of the reasons as to why evidence obtained from a meta-analysis is generally more conclusive than those from a single trial.

Power is defined as the probability to reject the null hypothesis when there is a true effect [6]. A study with insufficient power means there is an increased probability to produce a false-negative result, that is, *type II* error, and may mislead the healthcare decision. Power analysis is therefore an important measure to determine whether the results of a study are credible or not. Although meta-analysis has the potential to increase power, it faces the problem of being underpowered as well either due to the limited number of studies or the substantial variance across

What's new?

Key findings

- Most of the meta-analyses of rare events in Cochrane systematic reviews with the results were underpowered that are unable to support a conclusive decision.
- Meta-analyses with more studies included tend to have a higher probability to achieve a sufficient power to detect a certain true effect, but this did not guarantee a sufficient power; even for those with a larger number of studies, the majority still had a very low power.

What this study adds to what is known?

- Power analysis is an important measure to determine whether the results of a study are credible or not. Although meta-analysis has the potential to increase power, it faces the problem of being underpowered as well. This problem is even more common and severer for meta-analysis of rare events. In this study, we investigated the power to detect certain true effects for published meta-analyses of rare events. The elucidation of this article is expected to have implications for methodology guidelines, clinical practice, and health care policy.

What are the implications, and what should be changed?

- Some well-developed methods such as the beta-binomial model, the generalized linear mixed models, and the stratified exact logistic regression are especially useful to cooperate the information of studies with no events in both arms with other studies that may gain more power for meta-analysis of rare events and are recommended to be considered in practice.
- We advocate that meta-analysis of rare events should report the *post hoc* power of the results to help evidence users to form a better and informative decision.

studies [7]. For meta-analysis of rare events, this problem is even more common and severer because the low-event rate resulted in very large variances and thus lower precisions [7,8]. Hypothetically, meta-analysis of rare events may be at a high risk of overinflated *type II* error.

There has been a dramatic increase on the number of meta-analyses published over time; many of them were meta-analyses of rare events investigating mainly, but not limited to, safety end points. It is currently unclear as to

whether these meta-analyses have sufficient power to support their conclusions and how many of them may have produced false-negative results. In this study, we used a *post hoc* power estimation for meta-analyses of rare events in the Cochrane Database of Systematic Reviews (CDSR) to address the above questions.

2. Methods

2.1. Data source

The CDSR (ISSN: 1469-493X) is the official product of Cochrane and a listed database of the Cochrane Library. It enlists all published Cochrane protocols and reviews and is updated continuously [9]. A total of 6,781 systematic reviews were identified from CDSR in the period of January 2003 to May 2018. We excluded 1,295 reviews that did not contain useable data, 89 diagnostic test accuracy reviews, and 6 reviews were of incompatible file format. Finally, 5,391 Cochrane intervention reviews were further processed for data extraction (See details in [supplementary file](#)).

We previously used the same data set to examine the measurements of between-study heterogeneity, publication bias, and the information contained in studies with no events in both arms for meta-analysis [10–13]. Access to the Cochrane Library and CDSR was granted via the Florida State University (Tallahassee, FL, USA). We declare that these data were used only for academic research purposes.

2.2. Inclusion criteria

Our “population” of interest was generic *meta-analyses of rare events* from the CDSR. Exclusion criteria were as follows: 1) duplicates, defined as have exactly the same data to another meta-analysis (see details below); 2) total event count across studies in both arms were zero; 3) reviews that did not conduct pooling of effect measures; and 4) meta-analyses with only one study.

We extracted information from meta-analyses of rare events based on the maximum event rate of the included studies. The cutoff point of an event rate was set to 0.05 for defining rare events; this indicated a relatively small probability of observing an event [14]. A meta-analysis with both arms having a maximum event rate ≤ 0.05 across included studies was regarded as a meta-analysis of rare events.

2.3. Data cleaning and extraction

Data cleaning and extraction were conducted by a Stata (Version 14.0/SE, Stata, College Station, TX, USA) program developed by the lead author (X.C). The Stata program is presented in the supplementary file. The following meta-analysis information was collected: name

of the outcome measure (e.g., adverse events), data type (e.g., dichotomy), statistical method (e.g., Peto), effect measure (e.g., odds ratio and risk ratio), analysis model (e.g., fixed-effect), total event count in each arm, total sample size of each arm, maximum event rate in each arm, meta-analyzed effect estimate with its confidence interval (CI), weight, *P*-value and *z*-score of the statistical inference, between-study variance metrics, and number of studies included in each meta-analysis. Those meta-analyses with the same effect measure, total events, total sample, and effect estimate were regarded as duplicates and only one was used in our analysis.

2.4. Power estimation for meta-analysis

We have prespecified definitions of “two-stage” and “one-stage” meta-analysis. A “two-stage” meta-analysis estimates the effect of each study in the first stage and then combine these study-specific effects although the standard methods (e.g., inverse variance) in the second stage. The “one-stage” approach uses a generalized linear mixed model that treats all included studies as a whole and estimates the “average effect” directly, without the process of calculating the effect of each study [15,16]. There is a difference in the estimation of the power for “one-stage” meta-analysis and “two-stage” meta-analysis [17]; this study focused on the “two-stage” framework due to the nature of Cochrane systematic reviews.

Jackson et al. [18] have summarized three methods for the estimation of power of “two-stage” meta-analysis. The three methods include the *analytical approach* by Hedges and Pigott [6,19], the *Monte Carlo approach*, and the approach assuming that all studies are the same “size” [18]. We first discard the third approach as it assumes all studies provide the same amount of information, which is unrealistic in the present study with empirical data set. For the rest two approaches, the *analytical approach* uses the moment-based method for the estimation of between-study variance, whereas the *Monte C. approach* considers the uncertainty of the between-study variance by Monte Carlo simulations. For Cochrane reviews, where the empirical data we used, the official software (RevMan) uses the moment-based method for the between-study variance [18]. To keep the same between-study variance estimator to the original analysis, we used the *analytical approach* to estimate the *post hoc* power of eligible meta-analyses. Actually, the *analytical approach* is also the most commonly used power estimation method in practice. Under this method, the power could be estimated as

Here, θ is the true effect of the meta-analysis, $\hat{\tau}^2$ is the estimated between-study variance, and $\hat{\sigma}_i^2$ is the “averaged” within-study variance across *i* included studies. In addition, $\Phi(\cdot)$ is the standard normal cumulative distribution function, Z_α is the *z*-score at the significance level α (e.g., $Z_\alpha=1.96$ for $\alpha = 0.05$), and SE denotes the standard

Table 1. Baseline characteristics of eligible meta-analyses of rare events in CDSR

Baseline characteristics	All publication (N = 4,177)
Number of studies within meta-analysis (median, IQR)	2 (2 to 4)
1–5	3,647 (87.31%)
6–10	425 (10.17%)
11–30	97 (2.32%)
>30	8 (0.19%)
Effect measure	
Odds ratio	993 (23.77%)
Risk ratio	3,055 (73.14%)
Risk difference	129 (3.09%)
Statistical method	
Inverse variance	196 (4.69%)
Mantel–Haenszel	3,660 (87.62%)
Peto	321 (7.68%)
Analysis model	
Fixed-effect	2,768 (66.27%)
Random-effect	1,409 (33.73%)
Sample size of each meta-analysis	1,132 (489 to 3,226)
≤ 500	1,068 (25.57%)
501–1,000	869 (20.80%)
1,001–2,000	800 (19.15%)
>2,000	1,440 (34.47%)
<i>P</i> -value of the meta-analysis (median, IQR)	0.47 (0.21 to 0.73)
> 0.1	3,566 (85.37%)
0.05–0.1 (not include 0.05)	184 (4.41%)
0.01–0.05 (not include 0.01)	213 (5.10%)
0.001–0.01 (not include 0.001)	77 (1.84%)
<0.001	137 (3.28%)
Between-study heterogeneity (I ²) (median, IQR)	0 (0 to 0)
≤ 30%	3,651 (87.41%)
30.001%–60%	335 (8.02%)
60.001%–75%	120 (2.87%)
> 75%	71 (1.70%)

IQR, interquartile range.

error of estimated effect ($\hat{\theta}$), that is, $SE = \sqrt{1/\sum(\hat{\tau}^2 + \hat{\sigma}_i^2)}$. We regarded a power ≥ 0.8 is sufficient.

In practice, it is impossible to obtain the true effect (θ). In standard trial sequential analysis, the true effect is generally defined by the relative risk reduction (RRR), that is, $RRR = 1 - \text{relative risk}$, where RRR generally took values of 10–50% [20–23]. Therefore, in the present study, we used 5 RRRs, say, $RRR = 10\%$, 20% , 30% , 40% , and 50% , with the corresponding relative risk were 0.9, 0.8, 0.7, 0.6, and 0.5, as the potential true effect and estimate the *post hoc* power to detect such an effect for meta-analyses used relative risk (i.e., risk ratio and odds ratio) as effect measure. For meta-analyses used risk difference

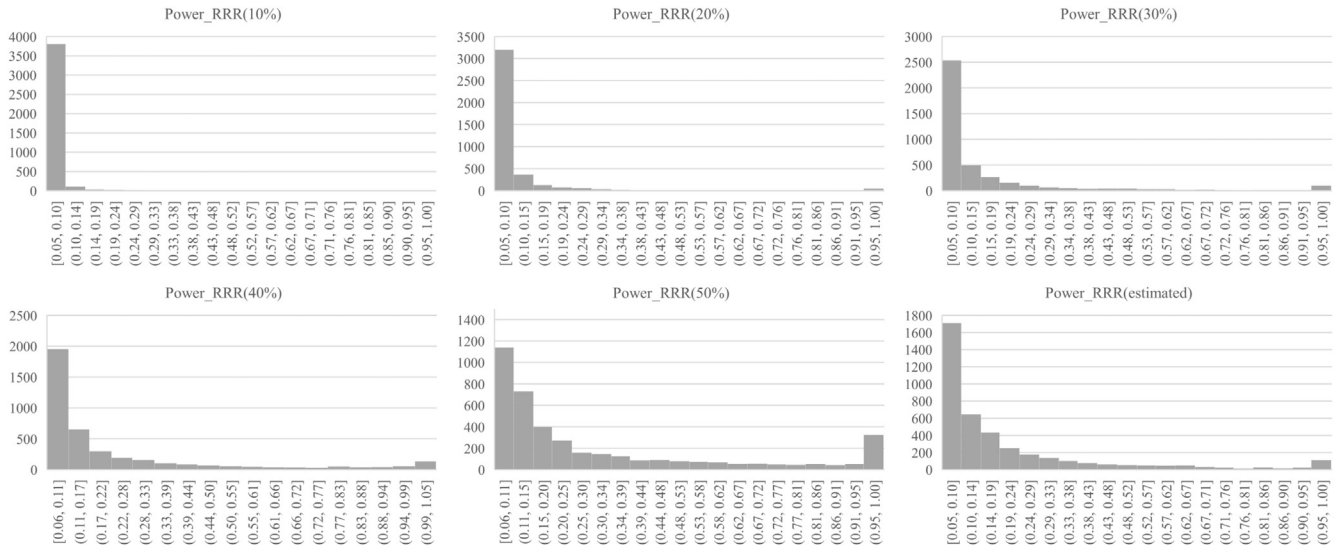


Fig. 1. The histogram distribution of power to detect a certain relative risk reduction (RRR) of 4,048 meta-analyses of rare events measured with relative risk.

as effect measure, the true effect was set as 0.001, 0.005, and 0.01 based on the distribution of empirical data. We also calculated a power to detect the estimated effect ($\hat{\theta}$) to as comparison. To be simple, we assume that risk ratio and odds ratio were approximately equal in meta-analysis of rare events [24].

2.5. Data analysis

Baseline characteristics were summarized based on descriptive statistics, including the proportion as well as the median value and the interquartile range (IQR). For the main outcomes, we estimated the proportion of how many meta-analyses had sufficient power and how many

were underpowered. The bar plot was used to visualize the distribution of the power in different scenarios.

A scatter plot was used to describe the relationship between effect size, *P*-value, and power. We further investigated whether a large number of studies within a meta-analysis had a higher probability to reach a sufficient power (≥ 0.8) based on the estimated effect. This was conducted by categorizing the number of studies into four groups: small (1–5), moderate (6–10), large (11–30), and very large (> 30) meta-analysis. The rate ratio (RR) was used as the effect measure.

All statistical analyses were achieved by the Stata (version 14.0/SE, Stata, College Station, TX, USA) software, with the prespecified significance level $\alpha = 0.05$.

Table 2. Post hoc power for meta-analyses of rare events measured with relative risk ($n = 4,048$)

Target effect	Power (95% CI)			Median (IQR)
	Power ≥ 0.8	$0.5 \leq \text{Power} < 0.8$	Power < 0.5	
For all included meta-analyses				
10% RRR	0.32%	0.62%	99.06%	0.06 (0.05 to 0.06)
20% RRR	1.73%	1.46%	96.81%	0.06 (0.06 to 0.09)
30% RRR	3.68%	3.48%	92.84%	0.08 (0.06 to 0.15)
40% RRR	7.16%	5.76%	87.08%	0.11 (0.08 to 0.26)
50% RRR	11.81%	9.66%	78.53%	0.17 (0.10 to 0.42)
Estimated RRR	4.50%	5.93%	89.58%	0.11 (0.06 to 0.24)
For meta-analyses with 6 or more studies				
10% RRR	0.27%	0.94%	98.79%	0.05 (0.05 to 0.07)
20% RRR	3.24%	1.75%	95.01%	0.08 (0.06 to 0.16)
30% RRR	5.94%	6.48%	87.58%	0.14 (0.08 to 0.34)
40% RRR	12.42%	10.80%	76.79%	0.23 (0.11 to 0.60)
50% RRR	20.92%	16.06%	63.02%	0.39 (0.16 to 0.85)
Estimated RRR	8.50%	7.42%	84.08%	0.14 (0.06 to 0.35)

RRR, relative risk reduction; IQR, interquartile range.

Table 3. Post hoc power for meta-analyses of rare events measured with risk difference ($n = 129$)

Target effect	Power ≥ 0.8	$0.5 \leq$ Power < 0.8	Power < 0.5	Median (IQR)
RD = 0.001	0.00%	0.78%	99.22%	0.05 (0.05 to 0.06)
RD = 0.005	2.33%	5.43%	92.25%	0.08 (0.06 to 0.18)
RD = 0.01	17.83%	8.53%	73.64%	0.18 (0.07 to 0.53)
Estimated RD	1.55%	2.33%	96.12%	0.06 (0.05 to 0.14)

3. Results

3.1. Baseline information

From the 5,391 data files, we identified a total of 118,741 meta-analyses. After data cleaning (supplementary file), there were 25,840 meta-analyses of binary rare outcomes. We further excluded those identified as duplicates, with no events in both arms across included studies, did not conduct pooling of effect measures or analyses of only one study. Eventually, we included 4,177 meta-analyses in the present study (Fig. S1, supplementary file).

The median number of studies included in each meta-analysis was 2 (IQR: 2 to 4). Among the included meta-analyses, 87.31% included 5 or less studies, and only 2.51% had included 11 or more studies. Moreover, 73.14% of the meta-analyses used the risk ratio as the effect measure, 23.77% used the odds ratio, and 3.09% used the risk difference. The Mantel–Haenszel method (87.62%) was most commonly used to synthesize the data, whereas

the Peto method was only used in 7.68% of the meta-analyses. For the choice of analytic model, 66.27% used the fixed-effect model and 33.73% used the random-effects model. The median sample size of each meta-analysis was 1,132 (IQR: 489 to 3226) and about a half of them had a sample size no more than 1,000 (46.37%).

For these meta-analyses, we documented 89.78% had statistically nonsignificant results (P -value > 0.05), and only 10.22% had a statistically significant result. Specifically, 213 meta-analyses (5.10%) had P -values ranging from 0.01 to 0.05, 77 (1.84%) had P -values ranging from 0.001 to 0.01, and 137 (3.28%) had P -values less than 0.001. The between-study variance of these meta-analyses were generally low, with the median value as 0 (IQR: 0 to 0); the majority (87.41%) of them had $I^2 \leq 30\%$. Table 1 presents the detailed baseline characteristics.

3.2. Power for meta-analyses of rare events

Fig. 1 presents the distribution of the *post hoc* power to detect a certain effect for the 4,048 meta-analyses of rare

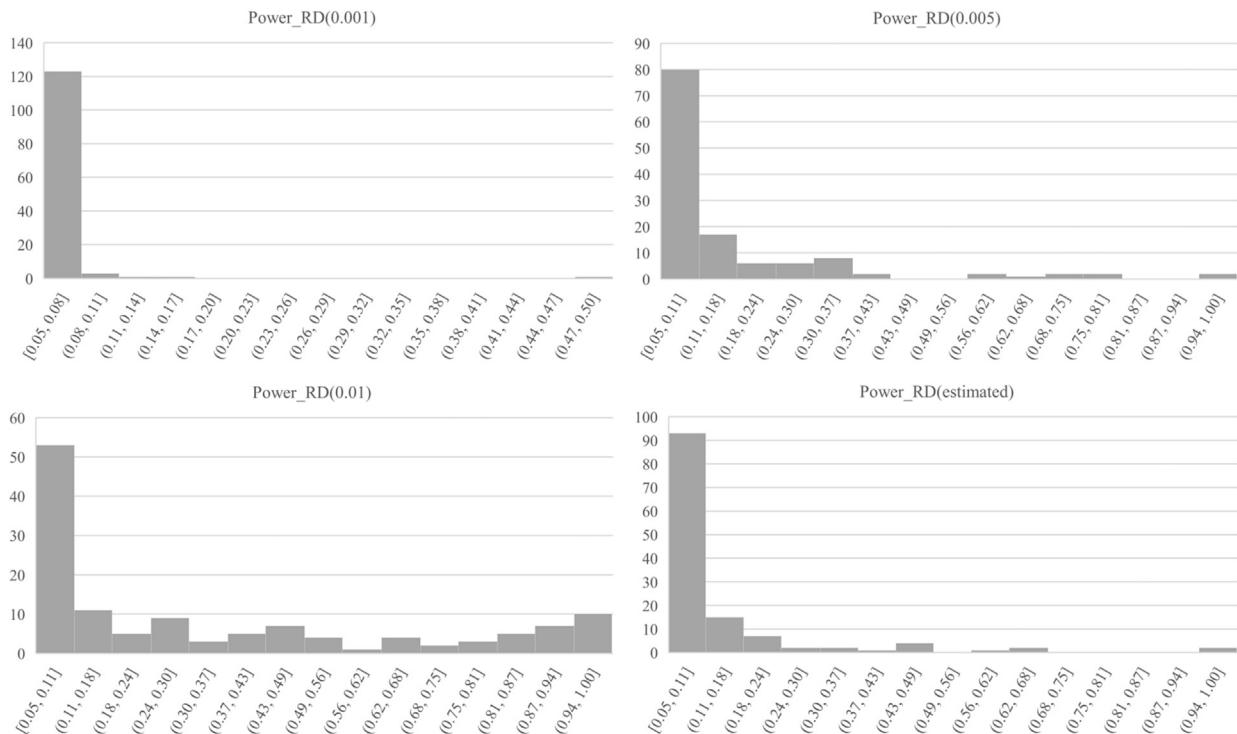


Fig. 2. The histogram distribution of power to detect a true risk difference (RD) of 129 meta-analyses of rare events measured with risk difference.

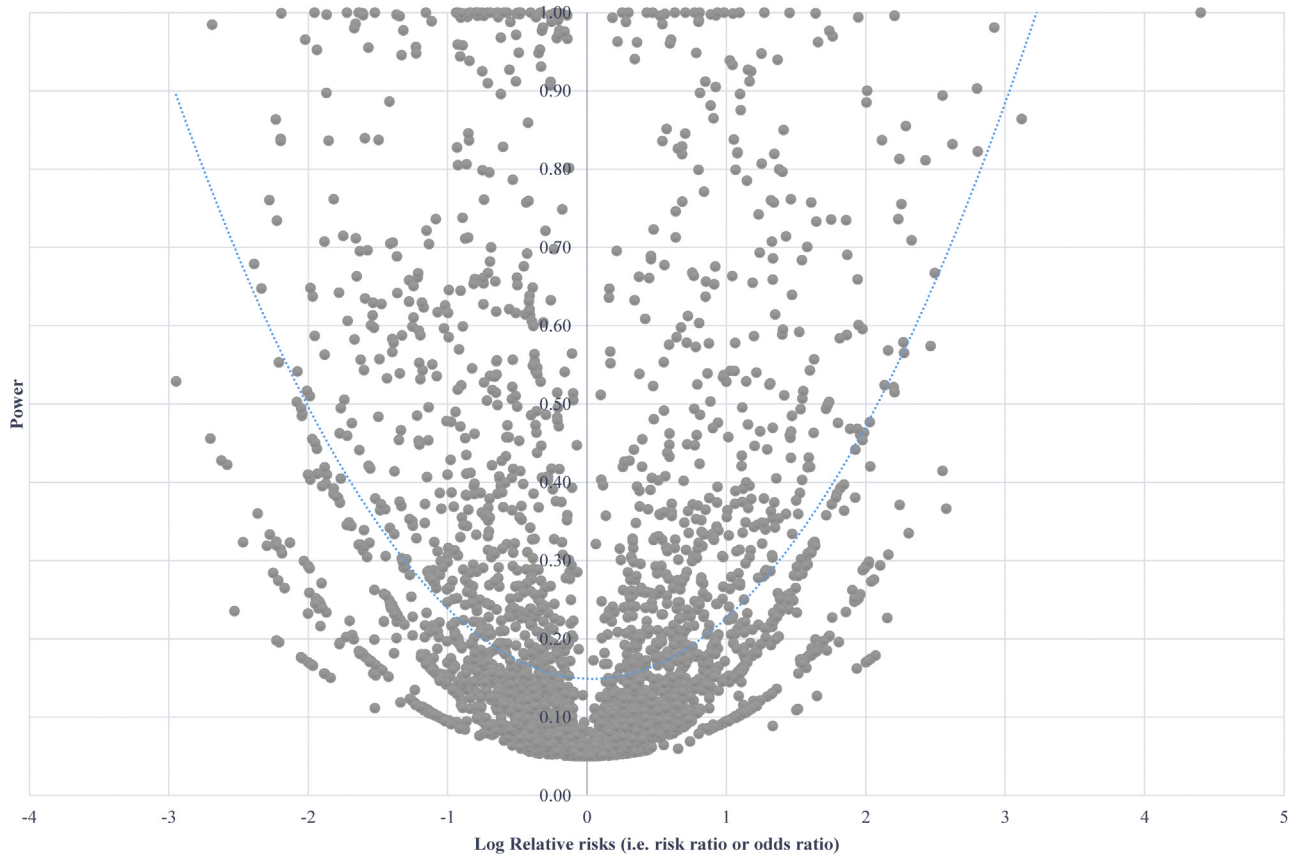


Fig. 3. The scatter plot of log relative risks and power.

events measured with relative risk. From the plot, we can see that most of the meta-analyses suffered from insufficient powers, and Table 2 presents the proportion in each power category. The median value of the power was 0.06 (0.05 to 0.06) to detect an RRR of 10%, and 0.32% with the results had a power ≥ 0.8 ; The median value of the power was 0.06 (0.06 to 0.09) to detect an RRR of 20%, with 1.73% of the results had a power ≥ 0.8 . For an RRR of 30%, median value of the power was 0.08 (0.06 to 0.15), with 3.68% of the results had a power ≥ 0.8 ; For an RRR of 40%, median value of the power was 0.11 (0.08 to 0.26), with 7.16% of the results had a power ≥ 0.8 ; For an RRR of 50%, 0.5, median value of the power was 0.17 (0.10 to 0.42), with 11.81% of the results had a power ≥ 0.8 . The median value of the power to detect an estimated effect was 0.11 (0.06 to 0.24), with 4.50% of the results had a power ≥ 0.8 , which was similar to the power to detect an RRR of 40%.

Fig. 2 presents the distribution of the *post hoc* power for the 129 meta-analyses of rare events measured with risk difference. Again, most of the meta-analyses were underpowered (Table 3). The median value of power to detect a risk difference of 0.001 was 0.05 (0.05 to 0.06), where 0.0% of the results had a power ≥ 0.8 ; The median value of power to detect a risk difference of 0.005 was 0.08

(0.06 to 0.18), with 2.33% of the results had a power ≥ 0.8 , and the median value of power to detect a risk difference of 0.01 was 0.18 (0.07 to 0.53), where 17.83% of the results had a power ≥ 0.8 , which was similar to the power with the true effect between 0.001 and 0.005.

3.3. Power for meta-analyses with more than 5 included studies

We further excluded those meta-analyses with 5 or less studies as a sensitivity analysis, the power was still low, however, there was a slight increase in the proportion of having powers ≥ 0.8 (Table 2). This was only used for meta-analyses measured with relative risk because for those measured with risk difference the sample was small. Again, the median value of the power to detect an RRR of 10% was 0.05 (0.05 to 0.07), and 0.27% with the results had a power ≥ 0.8 ; The median value of the power to detect an RRR of 20% was 0.08 (0.06 to 0.16), with 3.24% of the results had a power ≥ 0.8 ; The median value of the power to detect an RRR of 30% was 0.14 (0.08 to 0.34), with 5.94% of the results had a power ≥ 0.8 ; The median value of the power to detect an RRR of 40% was 0.23 (0.11 to 0.60), with 12.42% of the results had a power ≥ 0.8 , and the median value of the power to detect an RRR of 50% was 0.39

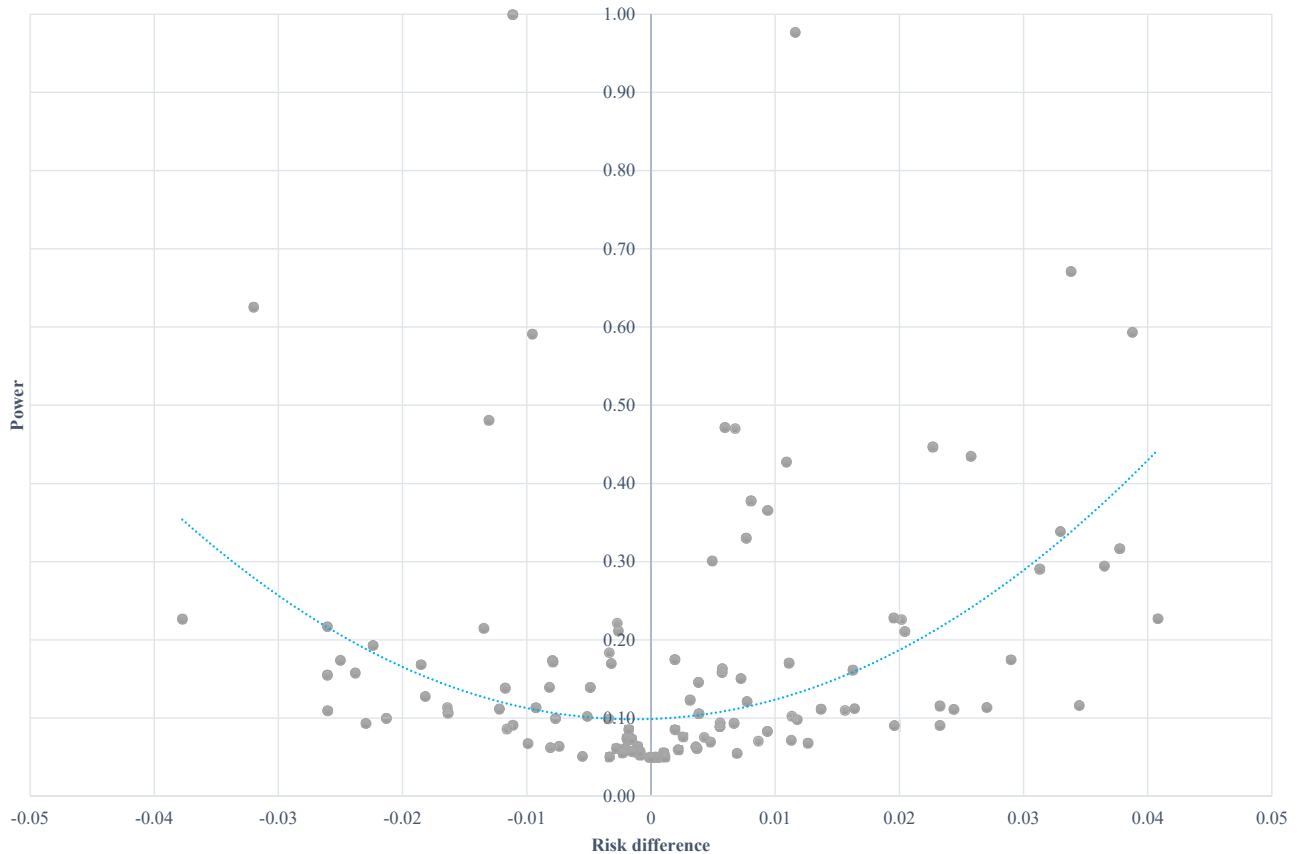


Fig. 4. The scatter plot of risk difference and power.

(0.16 to 0.85), with 20.92% of the results had a power ≥ 0.8 . For the estimated effect, the median value of the power to detect was 0.14 (0.06 to 0.35), and 8.50% with the results had a power ≥ 0.8 , which was similar to the power when the true effect was 0.7.

3.4. Relationship between effect size, *P*-value, number of studies, and power

We used the power in terms of estimated effect to reflect the relationship between the magnitude of effect size and power (Figs. 3 and 4). Generally, a larger effect size tends to have a higher power. Fig. 5 presents the relationship between *P*-value and power. As expected, there was a reverse relationship between them for which as the *P*-value increases, power decreases; as *P*-value decreases, power increases.

By dividing the included meta-analyses into four categories in accordance with the number of included studies, the median powers of each category were: 0.11 (IQR: 0.06 to 0.23), 0.14 (IQR: 0.07 to 0.33), 0.13 (IQR: 0.06 to 0.43), and 0.06 (IQR: 0.05 to 0.11) for small, moderate, large, and very large meta-analyses. For those meta-analyses that included more studies, there was a significant higher probability of achieving a sufficient power

(moderate vs. small: RR = 2.49, 95% CI: 1.76, 3.52, $P < 0.001$; large vs. small: RR = 4.02, 95% CI: 2.41, 6.71, $P < 0.001$; very large vs. small: RR = 3.48, 95% CI: 0.55, 21.94, $P = 0.184$); The *post hoc* powers of these three comparisons were as follows: 0.9990, 0.9996, and 0.2642. There was a very low power for the comparison of very large vs. small meta-analyses because in the category of very large meta-analyses, there were only 8 meta-analyses in total, making the sample size extremely low. The results suggested that for the group of very large vs. small meta-analyses, the power was insufficient to support the inference [25].

4. Discussion

In this study, we investigated the *post hoc* power of meta-analyses of rare events in Cochrane systematic reviews. Our results suggested that the majority of these meta-analyses were underpowered: only 11.81% of these meta-analyses reached sufficient power. Meta-analyses including more studies tended to have a higher probability for achieving sufficient power, but this did not guarantee that they have a sufficient power: even for those with a larger number of studies, the majority still had a very low

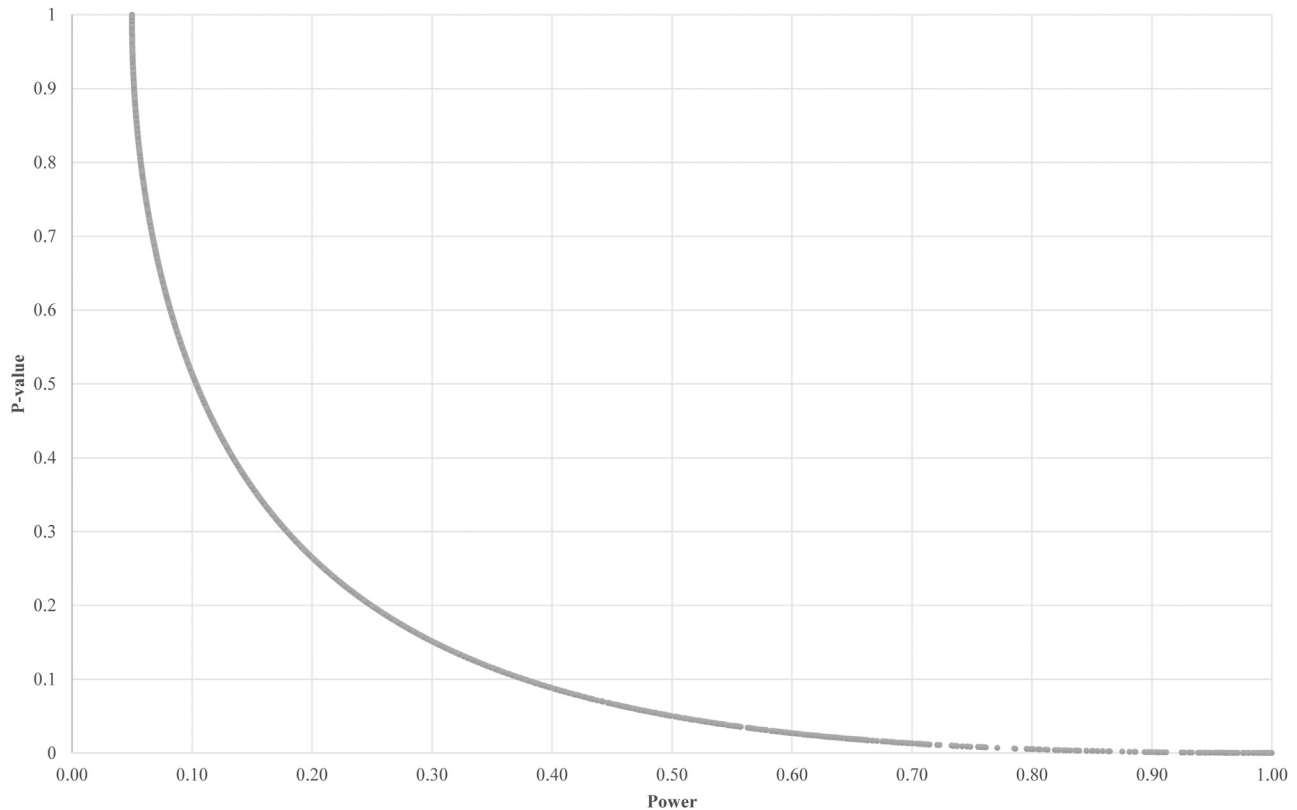


Fig. 5. The scatter plot of P -value and power.

power. Our findings imply that the results of these meta-analyses were mostly inconclusive and should be treated with caution.

The results of the present study concur with those from a study by Turner et al. [22]. The authors estimated the power of binary meta-analyses in the CDSR and documented a median power of 0.11 (0.06 to 0.21) for an RRR of 30% for safety outcomes (adverse events) [22]. In their study, 11% of the meta-analyses of adverse events were identified to possess a power of ≥ 0.5 . In our present study, 7.16% of the included meta-analyses had a power of ≥ 0.5 with regards to an RRR of 30%. The median power was also slightly smaller in our study. The potential difference was mainly due to the different inclusion criteria because the present study was not restricted to adverse events.

Similarly, Jackson et al. [18] compared the power of random-effects meta-analyses using data from the CDSR. They found that when there were 5 or more included studies in a meta-analysis, in more than 79.3% of the situation, a random-effects meta-analysis would have a greater power than the average power by individual studies. The findings suggested that a meta-analysis with 5 or more studies has a better guarantee for the conclusions. In our study, we focused on meta-analysis of rare events, and the results suggested that even for meta-analyses with 6 or more studies, at least 80% having insufficient powers.

Therefore, for meta-analysis of rare events, a much larger numbers of studies were needed to ensure a sufficient power for the conclusion. However, this seems unrealistic because in practice, only a small part of meta-analyses contained a large number of studies. As Jackson et al. pointed out that it is often difficult for researchers to predict how many eligible studies will be included and can contribute data for synthesis before the meta-analysis being conducted [18]. Nevertheless, Cochrane has proposed a valid solution that it is mandatory for all systematic reviews to be regularly updated. This is expected to improve the power and credibility for the results of their systematic reviews.

The amount of between-study variance has been widely used to determine which analytic model (fixed effect vs. random effects) to be used, for which a small variance indicates the fixed-effect model could be used while a large variance indicates the random-effects model is more suitable. Our study findings suggested that most meta-analyses of rare events have a small between-study variance. This is partly due to the wide confidence interval of each included study, and it is expected that the between-study variance obtained using classical methods (e.g., moment-based or likelihood-based [26,27]) is likely to be underestimated. Despite of this, the fixed-effect model might be more suitable for meta-analysis of rare events due to the following reasons. First, using the fixed-effect

model could be more effective to increase the power than using the random-effect model. Second, as most rare events are safety outcomes, the random-effect model may generate a conservative conclusion that may “cover-up” the potential increased risk of adverse events. Third, some methods for dealing with zero-events in meta-analysis, such as Mantel–Haenszel and Peto, were primarily driven by the homogeneous assumption, and thus, a fixed-effect model is preferred. Fourth, the random-effect model can be used as sensitivity analysis in addition to the main results estimated by the fixed-effect model. It should be highlighted that whether the fixed-effect model or the random-effect model was used and the *post hoc* power estimation is needed for meta-analysis of rare events. The TSA software (<http://www.ctu.dk/tsa>) developed by Copenhagen Trial Unit provides a user-friendly way to estimate the power.

In addition to the number of included studies, sample size, and between-study variance, the selection of synthesis methods is expected to have some influence on the power. For meta-analysis of rare events, due to the low events rate, single studies may fail to observe any events in both arms, whereas the standard “two-stage” method routinely discards these studies from the meta-analysis. This would lead to substantial loss of power. Some well-developed methods such as the beta-binomial model [28–30], the generalized linear mixed models [13,16], and the generalized estimating equations [31] are especially useful to cooperate the information of studies with no events in both arms with other studies. It is expected that these methods may gain more power for meta-analysis of rare events and should be considered for use in future practice.

We should highlight the importance of power analysis for meta-analysis. There are two important roles for a power analysis in meta-analysis: first, systematic review authors can use the estimated power as an indicator of precision as per the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) framework [32], where a lower power of a given analysis will result in wider CIs (i.e., imprecision), which leads to a rating down of evidence certainty [33]; for example, when the results were underpowered (e.g., ≤ 0.8), they may consider to rate down for certainty. Second, researchers can use the estimated power as an indicator for whether a meta-analysis needs an update. Both of these two roles of power analysis are expected to be useful for better health care decision-making. Therefore, we suggested further systematic review authors should routinely use a *post hoc* power analysis for their meta-analyses.

To the best of our knowledge, this is the first study to investigate the power of meta-analyses of rare events. The large sample size of the present study is expected to have good representativity. However, several limitations should be highlighted. First, we only considered Cochrane systematic reviews where most of the meta-analyses included a small number of studies, which may attribute to the low power. For meta-analyses published in other academic

journals, the number of included studies may differ and is expected to be greater; thus, the results may be only representative for Cochrane systematic reviews. Second, the definition of rare events is somewhat arbitrary. However, there is currently no consensus on defining rare events. A different definition of meta-analysis of rare events may have some influence on the results, and we believe the results will be similar or even worse because the cutoff point of 0.05 for the event rate is conservative compared with other choices such as 0.01 or 0.001 that were used by some previous studies [7,16]. Based on these limitations, a separate investigation on the power of meta-analyses of non-Cochrane reviews may prove useful to further verify our findings.

5. Conclusions

Our study findings indicated that most of the meta-analyses of rare events in Cochrane systematic reviews with the results were underpowered, where the results should be treated with caution. Our findings highlighted the importance of updating a meta-analysis regularly to provide a truly conclusive and credible evidence base. Considering the substantial impact of the power on the conclusions, we advocate that meta-analysis of rare events should report the *post hoc* power of the results to help evidence users to form a better and informative decision.

Acknowledgments

We would like to thank Qatar National Library for funding the Open Access payment.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2020.11.017>.

References

- [1] Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al, editors. Cochrane Handbook for Systematic Reviews of Interventions. 2nd ed. Chichester, UK: John Wiley & Sons; 2019.
- [2] Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to Meta-Analysis. Chichester, UK: Wiley; 2009.
- [3] Storz-Pfennig P. Potentially unnecessary and wasteful clinical trial research detected in cumulative meta-epidemiological and trial sequential analysis. *J Clin Epidemiol* 2017;82:61–70.
- [4] Cohn LD, Becker BJ. How meta-analysis increases statistical power. *Psychol Methods* 2003;8(3):243–53.
- [5] Borneman M. Using meta-analysis to increase power in Differential prediction analyses. *Ind Organizational Psychol* 2010;3(2):224–7.
- [6] Hedges LV, Pigott TD. The power of statistical tests in meta-analysis. *Psychol Methods* 2001;6(3):203–17.
- [7] Bradburn MJ, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Stat Med* 2007;26:53–77.

- [8] Kuss O. Statistical methods for meta-analyses including information from studies without any events—add nothing to nothing and succeed nevertheless. *Stat Med* 2015;34:1097–116.
- [9] The Cochrane library. About the Cochrane Database of Systematic Reviews. Available at <https://www.cochranelibrary.com/cdsr/about-cdsr>. Accessed October 27, 2020.
- [10] Ma X, Lin L, Qu Z, Zhu MT, Chu HT. Performance of between-study heterogeneity measurements in the Cochrane Library. *Epidemiology* 2018;29:821–4.
- [11] Lin L, Chu H, Murad MH, Hong C, Qu Z, Cole SR, et al. Empirical comparison of publication bias tests in meta-analysis. *J Gen Intern Med* 2018;33:1260–7.
- [12] Furuya-Kanamori L, Xu C, Lin L, Doan T, Chu H, Thalib L, et al. P value-driven methods were underpowered to detect publication bias: analysis of Cochrane review meta-analyses. *J Clin Epidemiol* 2020; 118:86–92.
- [13] Xu C, Li L, Lin L, Chu H, Thabane L, Zou K, et al. Exclusion of studies with no events in both arms in meta-analysis impacted the conclusions. *J Clin Epidemiol* 2020;123:91–9.
- [14] Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016; 31(4):337–50.
- [15] Jackson D, Turner R. Power analysis for random-effects meta-analysis. *Res Synth Methods* 2017;8(3):290–302.
- [16] Jackson D, Law M, Stijnen T, Viechtbauer W, White IR. A comparison of seven random-effects models for meta-analyses that estimate the summary odds ratio. *Stat Med* 2018;37:1059–85.
- [17] Kontopantelis E. A comparison of one-stage vs two-stage individual patient data meta-analysis methods: a simulation study. *Res Synth Methods* 2018;9(3):417–30.
- [18] Kontopantelis E, Springate DA, Parisi R, Reeves D. Simulation-based power calculations for mixed effects modeling: ipdpower in Stata. *J Stat Softw* 2016;74(12):1–25.
- [19] Hedges LV, Pigott TD. The power of statistical tests for moderators in meta-analysis. *Psychol Methods* 2004;9(4):426–45.
- [20] Wetterslev J, Jakobsen JC, Gluud C. Trial sequential analysis in systematic reviews with meta-analysis. *BMC Med Res Methodol* 2017; 17:39.
- [21] Brok J, Thorlund K, Gluud C, Wetterslev J. Trial sequential analysis reveals insufficient information size and potentially false positive results in many meta-analyses. *J Clin Epidemiol* 2008;61:763–9.
- [22] Turner R, Bird SM, Higgins JPT. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. *PLoS One* 2013;8:e59202.
- [23] Roshanov PS, Dennis BB, Pasic N, Garg AX, Walsh M. When is a meta-analysis conclusive? A guide to trial sequential analysis with an example of remote ischemic preconditioning for renoprotection in patients undergoing cardiac surgery. *Nephrol Dial Transpl* 2017; 32(suppl_2):ii23–30.
- [24] Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. *JAMA*; 280:1690–1691.
- [25] Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
- [26] Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials I: the inverse variance heterogeneity model. *Contemp Clin Trials* 2015;45:130–8.
- [27] Doi SA, Barendregt JJ, Khan S, Thalib L, Williams GM. Advances in the meta-analysis of heterogeneous clinical trials II: the quality effects model. *Contemp Clin Trials* 2015;45:123–9.
- [28] Bakbergenuly I, Kulinskaya E. Beta-binomial model for meta-analysis of odds ratios. *Stat Med* 2017;36:1715–34.
- [29] Kuss O, Hoyer A, Solms A. Meta-analysis for diagnostic accuracy studies: a new statistical model using beta-binomial distributions and bivariate copulas. *Stat Med* 2014;33:17–30.
- [30] Chu H, Nie L, Chen Y, Huang Y, Sun W. Bivariate random effects models for meta-analysis of comparative studies with binary outcomes: methods for the absolute risk difference and relative risk. *Stat Methods Med Res* 2012;21(6):621–33.
- [31] Zeger SL, Liang KY, Albert PS. Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1988;44:1049–60.
- [32] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
- [33] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64:1283–93.