Accuracy of the I2 point estimate for testing selection bias risk in meta-analyses – a simulation study (Protocol)

Steffen Mickenautsch (✉ neem@global.co.za )
University of the Western Cape  https://orcid.org/0000-0002-4994-2328

Veerasamy Yengopal
University of the Western Cape

Method Article

Keywords: meta-analysis, selection bias, bias testing, clinical trials, test accuracy

Posted Date: August 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3298532/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Accuracy of the $I^2$ point estimate for testing selection bias risk in meta-analyses – a simulation study (Protocol)

Steffen Mickenautsch,¹,²,³ Veerasamy Yengopal³

¹ Review Centre for Health Science Research, 84 Concorde Road East, Bedfordview/Johannesburg, 2008, South Africa.
²Department of Community Dentistry, School of Oral Health Sciences, Faculty of Health Sciences, University of the Witwatersrand, 7 York Rd, Parktown/Johannesburg, 2193, South Africa.
³Faculty of Dentistry, University of the Western Cape, Francie van Zijl Avenue, Tygerberg/Cape Town, 7505, South Africa.

*Corresponding author
Email: neem@global.co.za (SM)
Abstract

Objectives: To establish the test sensitivity and specificity of the $I^2$ point estimate for testing selection bias risk in meta-analyses under condition of large versus small trial sample size and large versus small trial number in meta-analysis and to test the null-hypothesis that the differences are not statistically significant.

Methods: Simulation trials will be generated in MS Excel, each consisting of three components, entered in form of three parallel columns: a sequence of subject ID (accession) numbers representing trial subjects; a random sequence of allocation to group A or B and a random sequence of a simulated baseline variable (‘age’) per subject, ranging between 50 to 55. These simulation trials will be included into five types of meta-analyses with large/small numbers of trials, as well as trials with large and small sample sizes. All meta-analyses will be tested using the $I^2$ point estimate. The numbers of true positive, false positive, false negative and true negative test results will be established. From these, the test sensitivity and specificity will be computed for each of the meta-analysis types and statistically compared.

Ethics and dissemination: Ethical approval is not required for simulation-based studies. The results will be disseminated as a prior preprint version and subsequent peer-reviewed publication.
Introduction

Meta-analyses form an important part in the methodology of systematic reviews, particularly that of randomized controlled trials (RCTs), with the purpose of statistically pooling the results of individual trials into one summary effect estimate. However, the internal validity of any meta-analysis result is directly dependant on that of its included trials. If the results of one or more single trial(s) are affected by selection bias then so will the result of the meta-analysis be, too.

Selection bias is present in an RCT when the random allocation of patients to the trial’s treatment groups was in some form subverted and patients with characteristics that are known to be conducive to a successful trial outcome are especially allocated to one treatment group, only [1]. In theory, such subversion causes an imbalance in the measurements of baseline variables (such as the mean age of the patients per group) in the trial. In contrast, in RCTs with true random allocation such baseline measurements should differ between the treatment groups, due to the play of chance, only. Consequently, a meta-analysis of such trials should show zero heterogeneity of the same baseline variables in-between trials and therefore reflect an $I^2$ value of zero percent [2]. The $I^2$ – statistic (measured in %) was developed for the purpose of estimating the proportion of variance in trial estimates that is due to heterogeneity. It ranges between 0 – 100% and is now common as an integrated part of meta-analysis [3].

In 2014, Clark et al reported that imbalances of baseline measurements in RCTs are reflected by an increased $I^2$ – point estimate when such measurements are statistically pooled with that of other trials [4]. On this basis, Hicks et al (2018) presented a simple test that would assist in identifying selection bias in meta-analyses. The test includes the extraction of measurements (mean value, standard deviation, sample size) for a baseline variable from all RCTs per treatment group that are included in a particular outcome meta-analysis and the calculation of the t-statistic for each trial. The baseline measurements are included into a fixed effect meta-analysis. If the result yields an $I^2$ – point estimate > 0%, the meta-analysis is repeated after step-wise removal of trials with the largest t-statistic until reaching $I^2 = 0%$. The outcome meta-analysis is then repeated without the removed
RCTs and the new pooled effect estimate noted. If the new pooled effect estimate differs from the previous one, then it is concluded that the outcome meta-analysis was affected by selection bias [2].

However, Rücker et al (2008) established in a simulation study that the $I^2$ – point estimate increases with the number of subjects included in the trials in a meta-analysis. The authors observed that by artificially inflating the sample size under a random effects model meta-analysis, the $I^2$ – point estimate tended to 100% and thus argued that $I^2$ is of limited use in assessing heterogeneity [5].

In addition, von Hippel (2005) noted that the $I^2$ – point estimate can be biased when a meta-analysis includes too few trials. For example, the $I^2$ – point estimate of a small meta-analysis with seven trials for which true heterogeneity is absent overestimated the heterogeneity by average of 12 percentage points and underestimated the heterogeneity by average of 28 percentage points when true heterogeneity was actually present [3].

Against this background [3, 5], it may be concluded that the $I^2$ – point estimate is not only imprecise due to its dependence on trial sample size but also easily affected by systematic error (bias) due to its dependence on the number of trials that are include in a meta-analysis. For that reason it may be assumed that such lack of precision and validity would negatively affect the accuracy of the $I^2$ – point estimate as test for selection bias in meta-analyses.

The aim of this simulation study is to establish the test’s sensitivity and specificity under condition of large versus small trial sample size and large versus small trial number in meta-analysis and to test the null-hypothesis that the differences are not statistically significant.

**Methodology**

*Meta-analyses generation of simulation trials*

Simulation trials will be generated in MS Excel, each consisting of three components, entered in form of three parallel columns: a sequence of subject ID (accession) numbers representing trial subjects; a random sequence of allocation to group A or B and a random sequence of a simulated baseline variable (‘age’) per subject, drawn from uniform distribution, ranging between 50 to 55. The
random allocation sequence will be generated by block-randomization with block size 4 using the ‘Sealed Envelope’ online tool [6]. The random sequence of a simulated baseline variable (‘age’) per subject will be generated using an online random number generator [1]. The comprehensive version of the generator will be used for randomly selecting the values of the baseline variable for each subject with the following settings: Lower limit = 50; Upper limit = 55; Allow duplication of results? = Yes; Sort the results? = Ascend; Type of result to generate = Integer.

These simulation trials will be included into the following types of meta-analyses:

(i) Large meta-analysis ($N_{\text{Trials}} = 15$) with large trials ($N_{\text{Subjects}} = 200$ per trial);
(ii) Large meta-analysis ($N_{\text{Trials}} = 15$) with small trials ($N_{\text{Subjects}} = 60$ per trial);
(iii) Small meta-analysis ($N_{\text{Trials}} = 5$) with large trials ($N_{\text{Subjects}} = 200$ per trial);
(iv) Small meta-analysis ($N_{\text{Trials}} = 5$) with small trials ($N_{\text{Subjects}} = 60$ per trial).

The number of meta-analyses per type will be determined by use of sample size calculation according to the method by Buderer et al. (1996) [8]. For each meta-analysis type, 50% of the meta-analyses will be generated as affected by selection bias and 50% as non-biased. The former will include two biased simulation trials. Trials will be biased by sorting the subjects according to their baseline variable (‘age’) in ascending order and assigning the first half (with the lower values) to group A and the other half to group B.

In addition, the meta-analysis type:

- Small meta-analysis ($N_{\text{Trials}} = 5$) with small trials ($N_{\text{Subjects}} = 60$ per trial)

will be repeated but by replacing all biased with non-biased trials that include 200 subjects, instead.

Sample size calculation

Sample size calculation was conducted using sample size calculator by Arifin (2023) [9] in line with the formula by Buderer et al. (1996) [8] with the following settings: expected sensitivity and specificity for both 80%; prevalence of disease (i.e. prevalence of biased meta-analyses) = 50%; expected precision = 10%; confidence level 100 $(1 - \alpha) = 95%$; expected drop out rate = 0%.
Accordingly, the calculation generated a required sample size of 123 meta-analyses for each meta-analysis type. In order to accommodate an even 50/50% split between biased and non-biased meta-analyses, the number was increased to 124. Therefore, a total of 496 meta-analyses, with 124 for each meta-analysis type, will be generated of which 62 meta-analyses will be repeated and 62 existing meta-analyses altered by replacing biased small trials with a mix of non-biased small and large trials.

Meta-analysis of baseline variable and bias test

According to the test method presented by Hicks et al. (2018) [2], all meta-analyses of the baseline variable will be conducted using inverse variance method with a fixed effect model using Review Manager (RevMan) 5.0.24 software. From each simulation trial the following continuous data of the baseline variable (‘age’) will be calculated and entered per meta-analysis per group A and B: mean value, standard deviation (SD) and sample size (N).

The $I^2$ – point estimate will be recorded for bias testing.

Test accuracy measure

In order to investigate test accuracy, the numbers of true positive (TP), false positive (FP), false negative (FN) and true negative (TN) test results will be established. These results are defined as follows:

- $TP = $ Baseline heterogeneity of biased meta-analysis $I^2 > 0$;
- $FP = $ Baseline heterogeneity of non-biased meta-analysis $I^2 > 0$;
- $TN = $ Baseline heterogeneity of non-biased meta-analysis $I^2 = 0$;
- $FN = $ Baseline heterogeneity of biased meta-analysis $I^2 = 0$. 
From the TP, FP, TN and FN values, the sensitivity and specificity with 95% Confidence interval (CI) will be computed for each of each meta-analysis type.

The test sensitivity is defined as the proportion of all meta-analyses with bias who yield a positive test result (calculated as TP / (TP + FN). The test specificity is defined as the proportion of all meta-analyses without bias who yield a negative test result (calculated as TN / (FP + TN) [10].

**Statistical comparisons**

The established test sensitivities and specificities will be statistically compared between

(i) Large and small meta-analyses and

(ii) Meta-analyses with large and small trials, as well as

(iii) Non-biased small meta-analyses with small trials versus non-biased small meta-analyses with a mix of small and large trials.

by means of the z-test for proportions. The null-hypotheses will be tested that the sensitivities and specificities from large versus small meta-analyses and meta-analyses with large versus small trials do not significantly differ (p > 0.05). Alpha will be set at 5%.

**Reporting**

The final report will be made available online as preprint in one of the major preprint repositories and submitted to a peer-reviewed journal.

**Financial Disclosure**

The authors received no specific funding for this work.

**Data availability**

All data will be made fully available without restriction.
Competing Interests

The authors declare that no competing interests exist

References


