BMJ Open

The impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-015888
Article Type:	Research
Date Submitted by the Author:	17-Jan-2017
Complete List of Authors:	Garcia-Alamino, Josep; University of Oxford, Nuffield Department of Primary Care Health Sciences Bankhead, Clare; University of Oxford, University of Oxford, Nuffield Department of Primary Care Health Sciences Heneghan, Carl; Oxford University, Primary Health Care Pidduck, Nicola; University of Oxford, Primary Health Care Perera, Rafael; University of Oxford, Primary Health Care
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Health policy, Public health, Research methods
Keywords:	STATISTICS & RESEARCH METHODS, Meta-analysis, Optimal information size, Trial sequential analysis, Heterogeneity, Effect size

Title: The impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses.

Authors: Josep M Garcia-Alamino[1], Clare Bankhead[1], Carl Heneghan[1], Nicola Pidduck[1], Rafael Perera[1](*).

1. Nuffield Department of Primary Care Health Sciences. ROQ. Woodstock Road. Oxford OX2 6GG. United Kingdom.

Financial support: non-financial support.

***Corresponding author:** Professor Rafael Perera. Nuffield Department of Primary Care Health Sciences. ROQ. Woodstock Road. Oxford OX2 6GG. United Kingdom. E-mail: rafael.perera@phc.ox.ac.uk Phone: +44 (0)1865 617300.

Word count: text (3,128) and abstract (250)

ABSTRACT

Objective: Published meta-analyses that include optimal information size (OIS) use a broad range of statistical assumptions to estimate the minimal information required to obtain reliable conclusions.

Methods: We carried out Medline and Cochrane electronic searches to retrieve meta-analyses published during 2010-2012, and analyzed several heterogeneity and effect size scenarios to evaluate the impact that these parameters have on the estimation of the OIS stratified by Cochrane/Non-Cochrane reviews. We compared estimates using dedicated software for the estimation of OIS (TSA software; TSA v0.9) and generic software for sample size estimation (Power and Sample size; Power and sample size calculations v3.1.2).

Results: We included a total of 137 out of 514 (26.6%) potential systematic reviews (one meta-analysis from each systematic review), 60.6% were Cochrane SR and 39.4% non-Cochrane. The meta-analyses included a total of 1,256 trials (individual studies), 1,291,364 patients and 65,087 events in the control group. The percentage of reviews meta-analyses that achieved the OIS were as follow: all cause mortality outcome for Cochrane 0% and 25% for non-Cochrane reviews (-0.074, 0.571), semi-objective outcome for Cochrane 16.6% and 45.8% for non-Cochrane reviews (-0.031, 0.534), subjective outcome for Cochrane 45.1% and 72.2% for non-Cochrane reviews (-0.024, 0.410).

Conclusions: These results demonstrate that the type of outcome is relevant for the estimation of the OIS, particularly to account for potential heterogeneity. More worrying is that less than fifty percent of recently published meta-analysis in high quality journals

achieves the OIS, and therefore conclusions based on these could still be subject to substantial change.

Keywords: meta-analysis, systematic review, optimal information size, trial sequential analysis, heterogeneity, effect size.

Strengths and limitations of this study

- This is an empirical review of meta-analyses that provide data about the statistical assumptions to estimate the minimal information required to obtain reliable conclusions.
- From our knowledge this is the first time that taking into account the type of outcome for the estimation of the OIS has been proposed.
- The type of outcome is relevant for the estimation of the OIS and impacts on the range of heterogeneity observed and was particularly high for subjective outcomes.
- Less than fifty percent of recently published meta-analysis in high quality journals achieves the OIS.

INTRODUCTION

The concept of optimum information size (OIS) was first proposed in 1998 by Pogue et al[1-2] and is defined as the minimum amount of information required in a meta-analysis for reliable conclusions to be drawn. For example, the required number of participants (information size) for a meta-analysis should match those required in an adequately powered single trial[3]. The estimation of the OIS helps to define whether firm evidence has been reached to draw robust conclusions[4]. Brok et al. demonstrated, in a subset of Cochrane reviews, that many meta-analyses have insufficient information size leading potentially to false positive results[3]. However, it is currently not known if the OIS is achieved in meta-analyses for different interventions and outcomes.

The calculation of the sample size and therefore also the OIS is influenced by several variables such as the control event rate (baseline risk), effect size [e.g. relative risk reduction (RRR)], power, and alpha value. Deciding on the values to use for the control event rate (baseline risk) or intervention effect that are required for OIS calculation can therefore be difficult.

Methods for the estimation of the baseline risk (control event rate) have been proposed in a structured way since 1996 or even before[5]. For example, based on single estimates (mean or median of all studies) or defining multiple risk groups (e.g. high, medium or low) for the control group. Also other methods taking baseline risk estimates from real-world populations (observational studies) have been recommended[6].

Trials used in meta-analyses include patients from several populations, different regimens of an intervention, as well as different methods in the study design. Therefore unexplained variation among trial results greater than the variation expected by chance often occurs;

typically referred to as statistical heterogeneity. Increased variation would decrease the precision of results, and authors like Wetterslev et al.[4] consider that the estimation of the optimal information size should include the sources of variation in a meta-analysis, including a measure of heterogeneity. The assessment of the between-study heterogeneity is an essential component of meta-analysis[7]. However, different statistical tests are used, such as Cochran's Q. Higgins and colleagues,[8] proposed the routine use of the I^2 (inconsistency factor), which is implemented in Cochrane reviews and widely used. Some studies, however, have showed several drawbacks with this statistic – issues mainly related with its precision[8-9]. Thorlund et al.[10] showed that I^2 fluctuated over time and experience considerable fluctuations when a meta-analysis includes less than 500 events and less than 15 trials. As a consequence, the risk of undetected heterogeneity is much higher when the number of meta-analyzed studies is small[11] and there is no consensus about which value of heterogeneity should be used to calculate the OIS.

Currently, there is no agreement about selecting an effect size for the estimation of the OIS. It has been observed that treatment effects differ within meta-analyses solely based on trial sample size[12], with stronger effect estimates seen in small to moderately sized trials than in the largest trials. Some authors follow the effect size conventions of small (0.10), medium (0.30) and large (0.50) effect. However this is not broadly accepted. Conversely there is consensus regarding the alpha (significance) or power value used and are conventionally either 0.05 or 0.01 for significance level and a power of 80% or 90%.

Therefore we set out to determine the impact that heterogeneity and effect size (RRR) have on the OIS estimation.

METHODS

Classic and recent studies demonstrated a difference in methodological quality and average number of included trials between Cochrane and non-Cochrane reviews[13-15].

We defined two sets of systematic reviews to evaluate: Cochrane and non-Cochrane.

We identified all Cochrane systematic reviews published during 2010-2012 through the Archie Database (http://archie.cochrane.org), which contains all Cochrane published reviews and allows electronic searching. We then randomly selected a total of 120 of these based on random numbers generated using Microsoft Excel.

To search for non-Cochrane reviews, we identified all systematic reviews with meta-analyses published in the top five general medical journals (N Engl J Med, Lancet, JAMA Intern Med, Ann Intern Med and BMJ) using the following search strategy in Medline (PubMed): "BMJ"[Journal] OR "Ann Intern Med"[Journal] OR "JAMA"[Journal] OR "Lancet"[Journal] OR "N Engl J Med"[Journal] AND (systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti]) restricted to reviews published during 2010-2012.

Inclusion criteria

From the selected Cochrane and non-Cochrane reviews, we included all meta-analyses reporting binary outcome data from two or more individual studies (clinical trials or randomized controlled trials).

Meta-analyses including observational studies, diagnostic interventions and network meta-analysis were excluded. Meta-analyses showing no-effect (pooled effect = 1) or meta-analyses with no events in all included trials were also excluded. For more details see the algorithm for the inclusion of meta-analyses (Figure 1).

Data extraction

We screened the titles and abstracts of the search results according to the inclusion criteria to identify relevant systematic reviews/meta-analyses. Full texts were obtained for those abstracts that met the inclusion criteria and assessed for eligibility. One reviewer JGA extracted the data and a second reviewer (RP or NP) checked the extracted data. We developed customized Excel spreadsheets for the data extraction process. From each included meta-analysis we extracted and calculated the following items: outcome type as defined by Turner[16] ['all cause mortality', 'semi-objective' (cause-specific mortality, major morbidity event) and 'subjective' (pain, mental health outcomes)], comparison, number of included patients, number of trials, number of events in each arm, control event rate, effect size and heterogeneity (Appendix, Table A.1).

Analyses

We extracted data from each trial and repeated the meta-analysis using random-effects models [DerSimonian and Laird (DL)] to account for potential heterogeneity of effects. Groups with zero events were adjusted with a constant continuity adjustment of 0.5 in each arm (default adjustment in Revman). The results obtained, pooled effect (RR) and I^2, were compared with the published results to detect any relevant disagreement and if required, the analyses were repeated to identify the source of the difference. Meta-analyses and calculation of the optimal information size (OIS) was done using *Trial Sequential Analysis* (TSA v0.9) software[17] freely downloadable at www.ctu.dk/tsa. The TSA software allows meta-analysis of dichotomous or continuous data under fixed or random-effects models and has the option to estimate an information size and the stopping boundary. This estimation of the OIS is based on the alpha spending method (Lan and De Mets or O'Brien Fleming).

To evaluate the impact of changes in heterogeneity and effect size (RRR), we estimated the OIS under different scenarios. For heterogeneity, we analyzed three values of heterogeneity: "heterogeneity = rep" as reported in the meta-analysis; "heterogeneity = 0"; and "heterogeneity = Q3" (upper quartile or 75th percentile), which was determined based on estimates published by Rhodes et al.[18]. The "heterogeneity = Q3" was used as high level of heterogeneity in comparison with "heterogeneity = 0". Consistent with Rhodes, the estimation of the OIS took into account the outcome type: 'all cause mortality', 'semi-objective' (cause-specific mortality, major morbidity event) and 'subjective' (pain, mental health outcomes) and, for simplicity, was based on assuming an average mean study size between 50 and 200 participants.

To evaluate the impact of effect size on the OIS, we used two different estimates of the effect size for the meta-analyses with mortality outcome: the Relative Risk Reduction (RRR) obtained in each meta-analysis as well as an a-priori conservative value of 5% for the RRR as reported by Djulbegovic et al.[19]. For the transformation of Relative Risk (RR) measure to RRR we used the following formula RRR =1-RR. If the RR was greater than 1 we used the RRR as a negative value. We did not determine an alternative estimate for the effect size for the other two outcomes (semi-objective and subjective) as the distribution of possible effects makes the choice of "average effect" difficult to justify. The baseline risk or Control Event Rate (CER) was taken to be the median of the proportion of events in the included trials in each meta-analysis, following the method proposed by Hayden et al.[20].

We compared the estimation of OIS obtained using the TSA v0.9 software against the OIS defined using generic sample size calculation software, *Power and sample size calculation*

v3.1.2 free software (http://biostat.mc.vanderbilt.edu). This software is designed for a single study, and therefore does not provide an option to incorporate the heterogeneity parameter in the calculation. We therefore regressed the (natural logarithm transformed) OIS found using each software package to determine potential deterministic association, which could facilitate the calculation of the OIS without specialist software.

We used descriptive statistics and plots to quantify differences in control event rate, effect size and heterogeneity between Cochrane and non-Cochrane reviews, stratified by type of outcome (six groups in total). We also determined the proportion of reviews that have achieved the OIS based on reported results and our two extreme scenarios "heterogeneity = 0" and "heterogeneity = Q3" comparing between Cochrane and non-Cochrane and again stratifying by type of outcome ('all cause mortality', 'semi-objective' and 'subjective' outcomes).

The descriptive analysis of the characteristics of included meta-analyses was carried out using SPSS v.22 software.

RESULTS

Search Results

Figure 2 provides the results of the literature search and the selection of the meta-analyses included in this sample. We excluded 11 Cochrane systematic reviews due to no events occurring in included trials or due to only one study included in the review. In the case of non-Cochrane systematic reviews, all included more than one study and several events by trial.

We included a total of 137 out of 514 (26%) potential systematic reviews (one meta-analysis from each systematic review) (Figure 2), 60.6% were Cochrane SR and 39.4%

non-Cochrane. The meta-analyses included a total of 1,256 trials (individual studies), 1,291,364 patients and 65,087 events in the control group.

Scenarios under different parameter estimates

Table 1 describes the main characteristics of the included meta-analyses.

Table 1. Descriptive results of included systematic reviews by type of outcome and intervention

	Cochrane (N=83)		Non-Cochrane (N=54)		All reviews (N=137)	
Type of outcome	% (n/N)	95% CI	% (n/N)	95% CI	% (n/N)	95% CI
All cause mortality	16.8% (14/83)	[8.76, 24.84]	22.2% (12/54)	[11.12, 33.28]	19% (26/137)	[12.43, 25.57]
Semi-objective	21.7% (18/83)	[12.8, 30.57]	44.4% (24/54)	[31.15, 57.65]	30.7% (42/137)	[22.98, 38.42]
Subjective	61.4% (51/83)	[50.93, 71.87]	33.3% (18/54)	[20.73, 45.87]	50.4% (69/137)	[42.03, 58.77]
Type of intervention						
Pharmacological	56.6% (47/83)	[45.94, 67.26]	62.9% (34/54)	[50.02, 75.78]	59.1% (81/137)	[50.87, 67.33]
Non-pharmacological	44.6% (37/83)	[33.91, 55.29]	35.2% (19/54)	[22.46, 47.94]	40.9% (56/137)	[32.7, 49.1]

Twenty six (19%) of the included meta-analyses used 'all cause mortality' as the outcome, forty two (31%) were based on a 'semi-objective' outcome and sixty nine (50%) on a 'subjective' outcome. The type of intervention was pharmacological in 59% of the meta-analyses.

The descriptive analysis of the different items (CER, RRR, I^2) and statistical assumptions used by individual meta-analyses showed a significant difference depending on the type of outcome (Table 2 and Figure 3).

Table 2. Descriptive results for the statistical assumptions in the included meta-analyses

	CER	RRR*	Heterogeneity (I^2)	Included patients	OIS estimated†
All reviews n=137					
Mean (SD)	26.9 (26.1)	28.2 (31.5)	20.4 (26.1)	9426.0 (22753.9)	386441.1 (1645397.1)
Cochrane reviews n=83					
Mean (SD)	24.0 (27.9)	21.0 (36.6)	0.0 (25.5)	586 (6245.8)	2301.0 (1422086.6)
Non Cochrane reviews n=54					
Mean (SD)	10.0 (21.7)	20.0 (20.6)	14.5 (26.9)	6566.5 (32925.7)	7299.5 (1946750.2)
All cause Mortality n=26					
Mean (SD)	12.7 (16.5)	20.3 (22.1)	10.9 (17.6)	14314.6 (25880.1)	499090.7 (1966940.8)
Cochrane reviews n=14					
Mean (SD)	10.0 (11.2)	25.5 (27.3)	6.7 (13.7)	6902.5 (12971.1)	813301.7 (2678677.7)
Non Cochrane reviews n=12					
Mean (SD)	15.8 (21.4)	14.3 (12.6)	15.7 (20.8)	22962.1 (34232.8)	132511.2 (201708.8)
Semi-Objective n=42					
Mean (SD)	17.2 (20.9)	19.0 (18.7)	18.8 (25.9)	18683.5 (33125.7)	828450.9 (2444468.1)
Cochrane reviews n=18					
Mean (SD)	18.5 (10.5)	21.7 (24.0)	12.8 (22.7)	3384.8 (4762.7)	536616.8 (1803091.2)
Non Cochrane reviews n=24					
Mean (SD)	16.3 (20.0)	16.9 (13.7)	23.3 (27.5)	30157.0 (40233.9)	1047326.5 (2851698.9)
Subjective n=69					
Mean (SD)	38.2 (27.1)	36.8 (37.9)	24.9 (28.1)	1948.9 (2971.0)	74944.0 (406792.0)
Cochrane reviews n=51					

Mean (SD)	41.6 (27.9)	36.5 (41.6)	23.5 (27.6)	1173.0 (2244.3)	78688.6 (449189.7)
Non Cochrane reviews n=18					
Mean (SD)	28.5 (22.9)	37.6 (25.3)	28.9 (29.7)	4147.3 (3683.7)	64334.2 (261366.6)

*For the calculation of this descriptive variable all the values were considered as positive

† This estimation of the OIS was done under the conditions of the 'scenario 2' (heterogeneity = 0, alpha 5%).

The number of included patients was higher for non-Cochrane meta-analyses for all the outcomes analyzed (Figure 3). In the case of the CER, 'all cause mortality' had the lowest mean value (Figure 3). The distribution of the control event rate differed between the outcome types and showed an increase in the range of values as the type of outcome moved from 'all cause mortality' to a 'subjective' outcome; with 'all cause mortality' showing the narrowest range of values. Given that the majority of reviews/MA are of "subjective outcomes", it is unsurprising that the median CER of all systematic reviews is above the median for the reviews reporting on 'all cause mortality' or a 'semi-objective' outcome.

For the RRR the highest mean value was observed in the 'subjective' outcome (Figure 3). For heterogeneity (I^2) the pharmacological 'subjective' outcome had the highest mean value (Figure 3). The distribution of the relative risk reduction differs among the outcome types and shows a broad confidence interval for the three outcomes. Also for this statistical assumption the median of all systematic reviews is above the median of 'all cause mortality' and 'semi-objective' outcome.

The distribution of the heterogeneity differs depending on the outcome type, and has wide range of values particularly for the 'semi-objective' outcome. Again, the median of all systematic reviews is above the median of 'all cause mortality' and 'semi-objective' outcome (Figure 3).

Meta-analyses that have reached the OIS

Figure 4.a presents the OIS required for each SR considering the scenario when there is no heterogeneity (best-case scenario: heterogeneity = 0). The estimation of the OIS shows a wide range of values. 'All cause mortality' required the highest OIS for non-Cochrane and

Cochrane meta-analyses closely followed by the 'semi-objective' and further behind those MA reporting 'subjective' outcomes. Of note, the range for OIS in the MA reporting 'subjective outcome' is narrower when compared to the other two outcomes. In this case the median of the OIS is below the median of all outcomes. Figure 4.b shows the number of meta-analyses that have already achieved sample sizes equal or higher to this estimated OIS (those SR/MA above the diagonal line) while Figure 4.c shows that more non-Cochrane reviews have already achieved this estimate for the OIS compared to Cochrane reviews.

When estimating the OIS based on the heterogeneity reported in each review, the number of SR/MA that have already achieved the necessary sample was significantly reduced only for Cochrane SR with MA reporting subjective outcomes (Table 3).

Table 3. Percentage of Reviews that achieve the OIS by Heterogeneity (I^2) level assumed

	All cause Mortality			Semi-objective			Subjective		
% (n/N)	Coch	Non-Coch	95% CI Differnce	Coch	Non-Coch	95% CI Difference	Coch	Non-Coch	95% CI Difference
OIS Achieved I^2=reported	0% 0/14	25% 3/12	[-0.074, 0.571]	11.1% 2/18	37.5% 9/24	[-0.043, 0.499]	31.4% 16/51	72.2% 13/18	[0.111, 0.616]
OIS Achieved I^2=0	0% 0/14	25% 3/12	[-0.074, 0.571]	16.6% 3/18	45.8% 11/24	[-0.031, 0.534]	45.1% 23/51	72.2% 13/18	[-0.024, 0.490]
OIS Achieved I^2= Q3	0% 0/14	16.6% 2/12	[-0.134, 0.491]	11.1% 2/18	33.3% 8/24	[-0.079, 0.460]	29.4% 15/51	61.1% 11/18	[0.027, 0.553]

Coch = Cochrane meta-analyses, Non-Coch = Non-Cochrane meta-analyses

There were no statistically significant differences between the percentage of reviews achieving the OIS (except for 'subjective' outcome) with reported I^2 heterogeneity or the Q3.

When using a more stringent estimate for the effect size (5% RRR) for "all cause mortality" none of the identified SR/MA had achieved the necessary sample size to meet the OIS (Table 4).

Table 4. Percentage of Reviews that achieve the OIS by Effect size assumed (All cause Mortality only)

	Cochrane	Non-Cochrane	95% CI Difference
OIS Achieved, Effect reported % (n/N)	0% 0/14	25% 3/12	[-0.074, 0.576]
OIS Achieved, Effect = 5% % (n/N)	0% 0/14	0% 0/12	[-0.301, 0.267]

Box 1 presents five examples of meta-analyses with 'all cause mortality' as the main outcome to illustrate which ones achieve, or not the OIS.

In the figure 4 is showed the correlation between the power and sample size software and the TSA software without considering the heterogeneity.

Predicting the value of OIS with heterogeneity Q3 from a value of OIS without heterogeneity. From the value of the OIS without heterogeneity was predicted the OIS average value considering heterogeneity Q3 (Table 5).

Table 5. OIS estimated (n patients) incorporating the heterogeneity using two methods (TSA software vs Formula)

Review	**Outcome**	**$I^2 = 0$ n patients**	**I^2 = Heterogeneity Q3 n patients**	**Formula (y=0.9593x+0.8021) n patients**
#01	Mortality	3374	3627	5405
#02	Mortality	255866	284295	343725
#03	Semi-objective	630	863	1080
#04	Semi-objective	7642	8886	11843
#05	Subjective	935	1833	1578
#06	Subjective	17803	34907	26656

Box 1 . Example of five meta-analyses with the 'all cause mortality' as the main outcome that do, or do not, achieve the OIS

Meta-analysis that meet the OIS

Weng et al. 2010 (Annals)
This meta-analysis evaluated the use of a non-pharmacological intervention (Noninvasive ventilation) to treat patients with acute cardiogenic pulmonary edema including a total of 1,369 patients with a CER of 23%, RRR 27% and 0% heterogeneity. For this systematic review assuming a 0% heterogeneity the OIS estimated was 1296 patients.

Gastric team 2010 (JAMA)
This meta-analysis evaluated the use of adjuvant chemotherapy for resectable gastric cancer including a total of 3,781 patients with a CER 69%, RRR 9% and 24% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity the OIS estimated was 1,828 patients.

NSCLC meta-analysis 2010 (The Lancet)
This meta-analysis evaluated the use of Adjuvant chemotherapy in patients with operable non-small-cell lung cancer including a total of 8,447 patients with a CER 49%, RRR 11% and 1% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity the OIS estimated was 2,686 patients.

Meta-analysis with a large number of included patients that not meet the OIS

Adam et al. 2012 (Annals)
This meta-analysis evaluated the use of warfarin versus new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism including a total of 14,143 patients with a CER of 2%, RRR 12% and 0% heterogeneity reported. For this systematic review assuming a 0% heterogeneity the OIS estimated was 100562 patients.

Rizos et al. 2012 (JAMA)
This meta-analysis evaluated the administration of Omega-3 Fatty Acid Supplementation and Risk of Major Cardiovascular Disease Events including a total of 125,410 patients with a CER of 7%, RRR 4% and 1% heterogeneity reported. For this systematic review assuming a 0% heterogeneity the OIS estimated was 255, 912patients.

The regression line gives the average slope through a set of scattered points, the predicted value of Y is only the average for a given value of X.

DISCUSSION

Our results show that there is wide variability in the range of values for the parameters that impact on the OIS calculation: effect size (RRR), heterogeneity (I^2) and control event rate (CER) regardless of source (Cochrane or Non-Cochrane). Performing the analysis stratified by the different types of outcomes as proposed by Turner et al.[16] and Rhodes et al.[18] shows that the distribution of these values depend on the type of outcome being evaluated ('all cause mortality', 'semi-objective' or 'subjective'). From our knowledge this is the first time that taking into account the type of outcome for the estimation of the OIS has been proposed.

Our analysis also shows that the number of included patients in the meta-analyses using 'subjective' outcomes is lower than for other outcomes, particularly for Cochrane reviews. Possibly related to this, our analysis has also found that the CER for MA evaluating 'all cause mortality', particularly for Cochrane reviews, is lower than for other outcomes. This relationship could be explained as for a stable effect size (e.g. relative risk) the number of patients included in randomized controlled trials (RCTs) would need to be larger when the CER is smaller. This explanation is also consistent with our findings for the distribution of the effect sizes found which tend to show less variation between type of outcome. Finally, as reported elsewhere by Turner et al.[16] we also found that the type of outcome impacts on the range of heterogeneity observed and was particularly high for 'subjective' outcomes.

One possible explanation for this could be the higher number of small RCTs included in these MA.

The estimation of the OIS assuming different levels of heterogeneity and alpha values showed a strong correlation. The scenario of no heterogeneity and the scenario of high heterogeneity (heterogeneity = Q3) in the logarithmic scale showed perfect correlation. Therefore, our results show it is possible to predict the OIS, by using heterogeneity from the third interquartile range from the equation of the regression line. Furthermore, it is possible to estimate the value of OIS with high heterogeneity (Q3) without the use of specific software. The TSA software uses a linear formula when calculating the OIS[17].

However, there is currently no consensus on what heterogeneity assumptions to adopt. Even a paper published by Wetterslev et al.[21] proposes the use of an alternative index named the diversity (D^2) statistic as opposed to the I^2 factor. Our analysis suggests that the level of heterogeneity should depend on the type of outcome and the estimate of the OIS obtained following the statistical assumptions/estimates used in this analysis. Published meta-analyses with estimation of optimal information size often use one or more statistical assumptions, such as a RRR of 10% and 20% or the median RRR of trials with low risk of bias[22-24]. The analysis of all pooled reviews in this research showed that the 50th percentile (median) of the RRR is 20%. However, this distribution varied by the different outcome types. Therefore, in some cases, optimal information size is underestimated, whilst in others it is overestimated.

The OIS estimated is higher for mortality Non-Cochrane meta-analyses, which could be related to lower CER, RRR and heterogeneity in comparison with other type of outcomes and meta-analyses (Cochrane vs non-Cochrane). The obtained results show that globally

less than fifty percent of recent published meta-analysis in high quality journals achieved the OIS. This is important, as it shows that a significant number of meta-analysis, published in high quality journals, do not have appropriate statistical power to draw firm conclusions.

Limitations

Reviews conducted by the Cochrane collaboration are considered to be of higher quality[25-26] and of greater methodological rigor than meta-analyses published in paper-based journals. Besides Cochrane reviews, our study also included meta-analyses only from the top five medical journals and therefore our results may not be applicable to other meta-analyses published in other journals. Nevertheless this would bias our results towards better evidence being evaluated to what is currently being generated. Also, our results do not generalize to network meta-analyses which is an area of evidence synthesis that has grown rapidly since the mid 2000s[27]. A recently published study demonstrated that substantial variation exists in the network-based meta-analysis[28] and the statistical methodology to estimate the OIS in these meta-analyses is less developed than for traditional meta-analysis, hence our exclusion of these studies in our analyses.

CONCLUSIONS

In summary, heterogeneity and effect size impact on the estimation of the OIS. It is possible to estimate the OIS with the use of traditional software for sample size estimation using the regression equation obtained in this analysis. These results demonstrate that the type of outcome is relevant for the estimation of the OIS, particularly to account for potential heterogeneity. More worrying is that less than fifty percent of recently published

meta-analysis in high quality journals achieves the OIS, and therefore conclusions based on these could still be subject to substantial change.

Bibliography

1. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet. 1998; 351:47–52.

2. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials [Internet]. 1997; 18(6):580-93-6.

3. 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(8):763–9.

4. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol 2008; 61(1):64–75.

5. Sharp SJ, Thompson SG, Altman DG. The relation between treatment benefit and underlying risk in. 1996; 313 (September).

6. Barocas DA, Chen V, Cooperberg M, Goodman M, Graff JJ, Greenfield S, et al. Using a population-based observational cohort study to address difficult comparative effectiveness research questions: the CEASAR study. J Comp Eff Res 2013;2(4):445–60.

7. Patsopoulos N a, Evangelou E, Ioannidis JP a. Sensitivity of between-study heterogeneity in meta-analysis: proposed metrics and empirical evaluation. Int J Epidemiol 2008; 37(5):1148–57.

8. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327:557–60.

9. Malone DC. The Good, the Bad, and the Different: A Primer on Aspects of Heterogeneity of Treatment Effects. J Manag Care Pharm. 2014;20(6):555–63.

10. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, et al. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One 2012; 7(7):e39471.

11. Kontopantelis E, Springate D a, Reeves D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. PLoS One 2013 Jan;8(7):e69930.

12. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013;346:f2304.

13. Moja LP, Telaro E, D'Amico R, Moschetti I, Coe L, Liberati A. Assessment of methodological quality of primary studies by systematic reviews: results of the metaquality cross sectional study. BMJ 2005;330(7499):1053.

14. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, et al. AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 2009;62(10):1013–20.

15. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and Reporting Characteristics of Systematic Reviews of Biomedical Research: A Cross-Sectional Study. PLoS Med. 2016;13(5):1–30.

16. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of

heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012; 41(3):818–27.

17. Thorlund K, Engstrøm J, Wetterslev J. User manual for trial sequential analysis (TSA). Trial Unit Available Internet. 2011 [cited 2014 Jun 11].

18. Rhodes KM, Turner RM, Higgins JPT. Predictive distributions were developed for the extent of heterogeneity in meta-analyses of continuous outcome data. J Clin Epidemiol; 2014;68(1):52–60.

19. Djulbegovic B, Kumar A, Pp G, Perera R, Reljic T, Dent L, et al. New treatments compared to established treatments in randomized trials. Cochrane Database Syst Rev. 2012;(10).

20. Ja H, Tomlinson G, Ni A, Rm P, Cartwright J, Mw VT, et al. Approaches to Estimate and Present Baseline Risks Considerations for Cochrane Reviews Evidence-Informed Practice Integration of : 19th Cochrane Colloquium Madrid.

21. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol 2009; 9:86.

22. Imberger G, Orr a., Thorlund K, Wetterslev J, Myles P, Møller a. M, et al. Does anaesthesia with nitrous oxide affect mortality or cardiovascular morbidity? A systematic review with meta-analysis and trial sequential analysis. Br J Anaesth. 2014; 112(3):410–26.

23. Bangalore S, Kumar S, Wetterslev J, Messerli FH. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. BMJ. 2011;342:d2234.

24. Afshari A, Wetterslev J, Brok J, Møller A. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. BMJ. 2007;335(7632):1248–51.

25. Remschmidt C, Wichmann O HT. Methodological quality of systematic reviews on influenza vaccination. Vaccine. 2014;32(15):1678–84.

26. Windsor B, Popovich I, Jordan V, Showell M, Shea B, Farquhar C. Methodological quality of systematic reviews in subfertility: A comparison of Cochrane and non-Cochrane systematic reviews in assisted reproductive technologies. Hum Reprod. 2012; 27(12):3460–6.

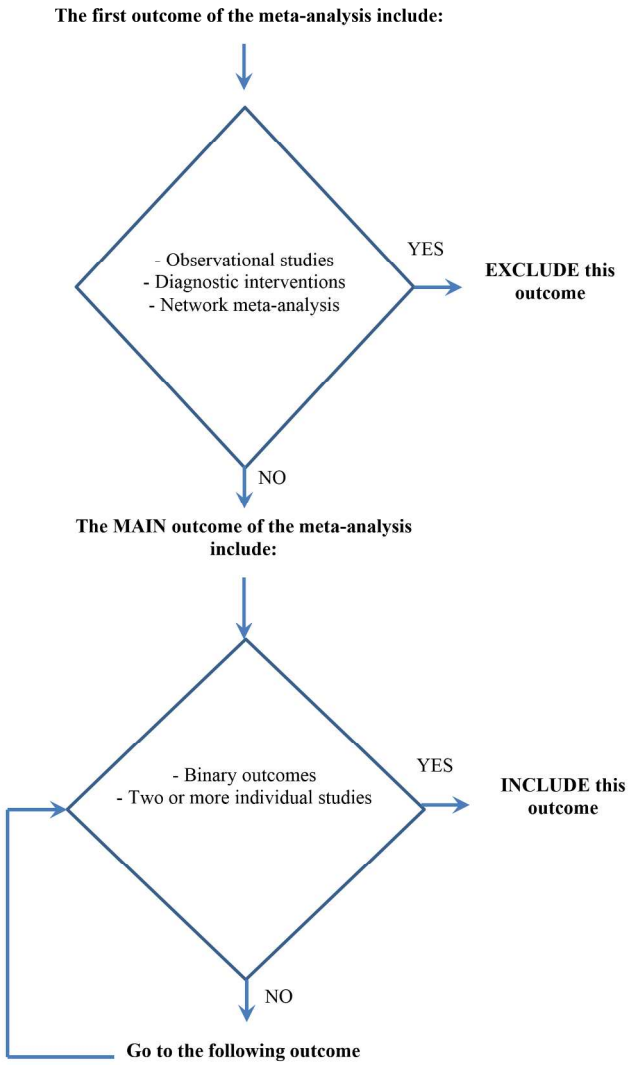
27. Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, et al. The quality of reporting methods and results in network meta-analyses: An overview of reviews and suggestions for improvement. PLoS One. 2014; 9(3).

28. Chambers JD, Naci H, Wouters OJ, Pyo J, Gunjal S, Kennedy IR, et al. An Assessment of the Methodological Quality of Published Network Meta-Analyses: A Systematic Review. PLoS One 2015;10(4):e0121715.

Competing interests: Non-conflict of interest.

Authors' Contributions: JGA, RP conceived and designed the study. JGA, NP extracted the data. JGA, RP analyzed the data. JGA, RP, CB, NP, CH interpreted the data. JGA, RP wrote the first draft. All authors revised the intellectual content and approved the version to be published.

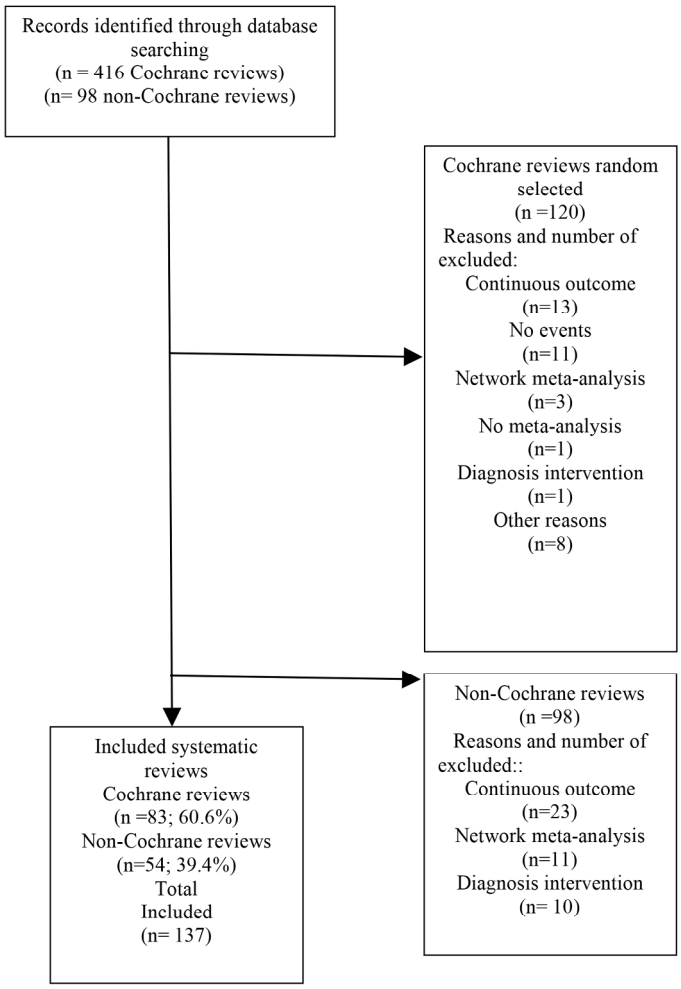
Figure 1. ALGORITHM FOR THE SELECTION OF THE OUTCOME



Algorithm for the selection of the outcome

181x261mm (300 x 300 DPI)

Figure 2. Flow-chart identification of Cochrane and non-Cochrane systematic reviews.



Flow-chart identification of Cochrane and non-Cochrane systematic reviews

164x226mm (300 x 300 DPI)

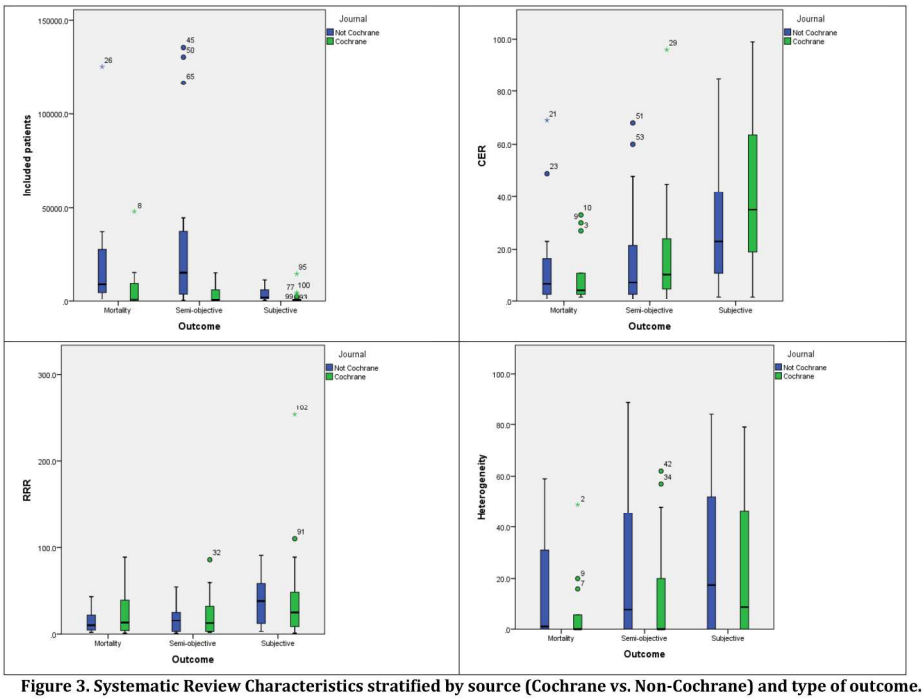


Figure 3. Systematic Review Characteristics stratified by source (Cochrane vs. Non-Cochrane) and type of outcome.

Systematic review characteristics stratified by source (Cochrane vs. Non-Cochrane) and type of outcome

246x185mm (300 x 300 DPI)

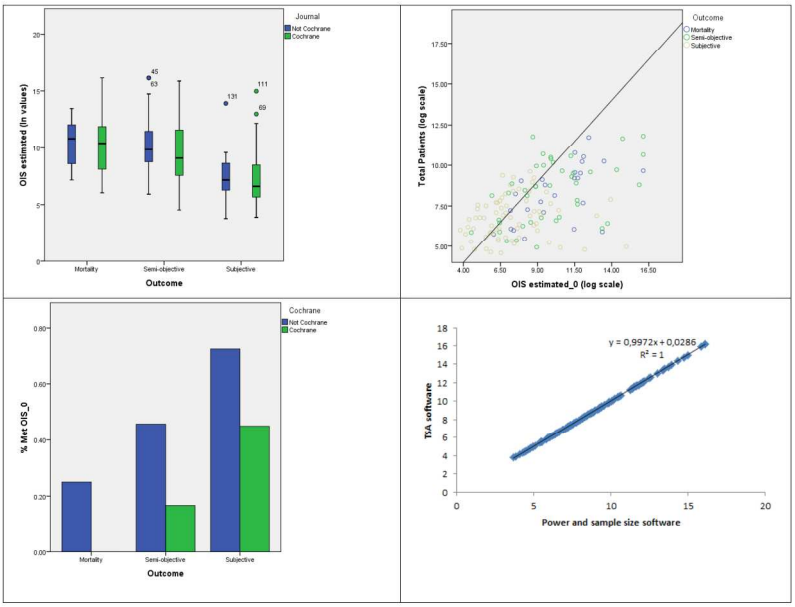


Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source, b) related to the total number of patients included in each SR, and c) percentage of SRs achieving the OIS by type of outcome and source d) OIS estimated comparing Power and sample size software versus TSA software (heterogeneity =0)

OIS estimated considering best-case scenario (heterogeneity=0) by a) type of outcome and source, b) related to the total number of patients included in each SR, and c) percentage of SRs achieving the OIS by type of outcome and source d) OIS estimated comparing power and sample size software versus TSA software (heterogeneity=0)

281x215mm (300 x 300 DPI)

BMJ Open

The impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-015888.R1
Article Type:	Research
Date Submitted by the Author:	23-May-2017
Complete List of Authors:	Garcia-Alamino, Josep; University of Oxford, Nuffield Department of Primary Care Health Sciences Bankhead, Clare; University of Oxford, University of Oxford, Nuffield Department of Primary Care Health Sciences Heneghan, Carl; Oxford University, Primary Health Care Pidduck, Nicola; University of Oxford, Primary Health Care Perera, Rafael; University of Oxford, Primary Health Care
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Health policy, Public health, Research methods
Keywords:	STATISTICS & RESEARCH METHODS, Meta-analysis, Optimal information size, Trial sequential analysis, Heterogeneity, Effect size

Title: The impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses.

Corresponding author: Professor Rafael Perera. Nuffield Department of Primary Care Health Sciences. ROQ. Woodstock Road. Oxford OX2 6GG. United Kingdom. E-mail: rafael.perera@phc.ox.ac.uk Phone: +44 (0)1865 617300.

Authors: Josep M Garcia-Alamino[1], Clare Bankhead[1], Carl Heneghan[1], Nicola Pidduck[1], Rafael Perera[1].

1. Nuffield Department of Primary Care Health Sciences. ROQ. Woodstock Road. Oxford OX2 6GG. United Kingdom.

Word count: text (3,392) and abstract (258)

ABSTRACT

Objective: To estimate the proportion of Systematic Reviews that meet the Optimal Information Size (OIS) and assess the impact heterogeneity and effect size have on the OIS estimate by type of outcome (e.g. mortality, semi-objective, or subjective).

Methods: We carried out searches of Medline and Cochrane to retrieve meta-analyses published in systematic reviews from 2010 to 2012. We estimated the OIS using *Trial Sequential Analysis* (TSA v0.9) software, and based on several heterogeneity and effect size scenarios, stratifying by type of outcome (mortality/semi-objective/subjective) and by Cochrane/non-Cochrane reviews.

Results: We included 137 meta-analyses out of 218 (63%) potential systematic reviews (one meta-analysis from each systematic review). Of these reviews, 83 (61%) were Cochrane and 54 (39%) non-Cochrane. The Cochrane reviews included a mean of 6.5 (SD 6.1) studies and the non-Cochrane included a mean of 13.2 (SD 10.2) studies. The mean number of patients was 2619.1 (SD 6245.8 or median 586.0) for the Cochrane and 19888.5 (SD 32925.7 or median 6566.5) patients for the non-Cochrane reviews. The percentage of systematic reviews that achieved the OIS for all-cause mortality outcome were 0% Cochrane and 25% for non-Cochrane reviews; for semi-objective outcome 17% for Cochrane and 46% for non-Cochrane reviews and for subjective outcome 45% for Cochrane and 72% for non-Cochrane reviews.

Conclusions: The number of systematic reviews that meet an optimal information size is low and varies depending on the type of outcome and the type of publication. Less than half of primary outcomes synthesized in systematic reviews achieve the OIS, and therefore the conclusions are subject to substantial uncertainty.

Keywords: meta-analysis, systematic review, optimal information size, trial sequential analysis, heterogeneity, effect size.

Strengths and limitations of this study

- To our Knowledge this is the first analysis to estimate the optimal information size by type of outcome.
- This study includes only systematic reviews from the Cochrane library and the top five general medical journals therefore our results may not be generalizable to systematic reviews published in other journals.

INTRODUCTION

The concept of optimum information size (OIS) was first proposed in 1998 by Pogue et al. [1-2] as "the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached". This OIS estimate is based on standard sample size calculations. For example, the required number of participants (information size) for a meta-analysis should match those required in an adequately powered single trial [3]. Other measures of information size have been proposed [4-5], however the OIS involves a relatively simple calculation, which under some scenarios will underestimate the information required to define whether firm evidence has been reached to draw robust conclusions[6]. Brok et al. demonstrated, in a subset of Cochrane reviews, that many meta-analyses have false positive results due to insufficient information [3] and Turner et al. showed that most meta-analysis do not have sufficient power to identify even moderate effects [7-8].

Sample size calculation and the OIS is influenced by several variables such as the control event rate (baseline risk), effect size, the power and the alpha value. Deciding on which values to use can be difficult and is typically based on values observed or estimated from the meta-analysis, or one of the included studies. In addition, increased variation can also effect the estimate of the OIS, and there is currently no consensus about which value of heterogeneity should be used to calculate the OIS.

The OIS can help determine the stability of an effect and whether treatment effect estimates are likely to differ based on further information. However, they are difficult to define in

advance and there is no consensus regarding the alpha (significance) or power value used at the outset.

It is therefore not currently known if evidence accumulation and its associated OIS depend on the type of outcome studied and if this varies by publication type (Cochrane or non-Cochrane review. Therefore, we set out to quantify this by studying systematic reviews published in the Cochrane Library and the top five general medical journals and in the process describe the impact that observed variation in heterogeneity and effect size (RRR) have on the OIS estimation.

METHODS

We defined two sets of systematic reviews to evaluate: Cochrane and non-Cochrane. We identified all Cochrane systematic reviews published during 2010-2012 through the Archie Database (http://archie.cochrane.org), which contains all Cochrane published reviews and allows electronic searching. We randomly selected a total of 120 of these based on random numbers generated using Microsoft Excel for inclusion.

To search for non-Cochrane reviews, we identified all systematic reviews with meta-analyses published in the top five general medical journals (N Engl J Med, Lancet, JAMA Intern Med, Ann Intern Med and BMJ) using the following search strategy in Medline (PubMed): "BMJ"[Journal] OR "Ann Intern Med"[Journal] OR "JAMA"[Journal] OR "Lancet"[Journal] OR "N Engl J Med"[Journal] AND (systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti]) restricted to systematic reviews published during 2010-2012.

Inclusion/Exclusion criteria

From all the selected Cochrane and non-Cochrane reviews, we included one meta-analysis from each. Based on the order the outcomes were reported (e.g. outcome 1.1 for Cochrane SRs) we selected the first outcome presented in the meta-analysis that was based on: binary data from two or more individual studies (clinical trials or randomized controlled trials). If the first outcome did not meet this inclusion criteria we continued through the listed outcomes until one was identified or we had exhausted the list of outcomes reported. (Figure 1). Meta-analyses that included observational studies, of diagnostic interventions or that were based on network meta-analysis were excluded. Meta-analyses showing no-effect (pooled effect = 1), or meta-analyses with no events in all included trials were also excluded.

Data extraction

Full texts were obtained for those abstracts that met the inclusion criteria and assessed for eligibility. One reviewer JGA extracted the data and a second reviewer (RP or NP) checked the data. We developed customized Excel spreadsheets for the data extraction process. From each included meta-analysis we extracted and calculated the following items: outcome type as defined by Turner [9] [‘all cause mortality’, ‘semi-objective’ (cause-specific mortality, major morbidity event) and ‘subjective’ (pain, mental health outcomes)], comparison, number of included patients, number of trials, number of events in each arm, control event rate, effect size and heterogeneity.

Analyses

We extracted data from each trial and repeated the meta-analysis using random-effects models [DerSimonian and Laird (DL)] to account for potential heterogeneity of effects. Estimates for trials with only one group reporting zero events were adjusted with a constant continuity adjustment of 0.5 in each arm (default adjustment in Revman). The obtained estimates for the pooled effect (e.g. RR) and I^2 were compared with the published results to detect any relevant disagreement, and if required, the analyses were repeated to identify the source of the difference. Meta-analyses and calculation of the optimal information size (OIS) were done using *Trial Sequential Analysis* (TSA v0.9) software [10] freely downloadable at www.ctu.dk/tsa. The TSA software allows meta-analysis of dichotomous or continuous data under fixed or random-effects models and has the option to estimate an information size and the stopping boundary. This estimation of the OIS is based on the alpha spending method (O'Brien Fleming and Lan-DeMets).

To evaluate the impact of changes in heterogeneity and effect size (RRR), we estimated the OIS under different scenarios. For heterogeneity, we analyzed three values of heterogeneity: "heterogeneity = rep" as that reported in the meta-analysis using a random-effects model (or obtained from fitting a random effects model if a fixed effect model was used originally); "heterogeneity = 0"; and "heterogeneity = Q3" (upper quartile or 75th percentile), which was determined based on estimates of predictive distributions published by Rhodes et al. [11]. These two estimates of "heterogeneity = Q3" and "heterogeneity = 0" were chosen as extreme scenarios to evaluate the impact that this parameter has on the OIS. Consistent with Rhodes, the estimation of the OIS took into account the outcome type: 'all cause mortality', 'semi-objective' (cause-specific mortality, major morbidity

event) and 'subjective' (pain, mental health outcomes) and, for simplicity, was based on assuming an average mean study size between 50 and 200 participants.

To evaluate the impact of effect size on the OIS, we used two different estimates of the effect size for the meta-analyses with mortality outcome: the Relative Risk Reduction (RRR) obtained in each meta-analysis as well as an a-priori conservative value of 5% for the RRR as reported by Djulbegovic et al. [12]. For the transformation of Relative Risk (RR) measure to RRR we used the following formula RRR =1-RR. If the RR was greater than 1 we used the RRR as a negative value. We did not determine an alternative estimate for the effect size for the other two outcomes (semi-objective and subjective) as the distribution of possible effects makes the choice of "average effect" difficult to justify. We used only one value, per meta-analysis, for the baseline risk or Control Event Rate (CER). This was taken to be the median of the proportion of events in the included trials in each meta-analysis, following the method proposed by Hayden et al. [13].

We used descriptive statistics and plots to quantify differences in control event rate, effect size and heterogeneity between Cochrane and non-Cochrane reviews, stratified by type of outcome (six groups in total). We also determined the proportion of reviews that have achieved the OIS based on reported results and our two extreme scenarios "heterogeneity = 0" and "heterogeneity = Q3" comparing between Cochrane and non-Cochrane and again stratifying by type of outcome ('all-cause mortality', 'semi-objective' and 'subjective' outcomes).

The descriptive analysis of the characteristics of included meta-analyses was carried out using SPSS v.22 software.

RESULTS

Search Results

Figure 2 presents a flow chart of the results. We excluded 11 Cochrane systematic reviews due to no events reported in the included trials, or due to only one study being included in the review. We included a total of 137 meta-analyses out of 218 (63%) potential systematic reviews (Figure 2): 83 (61%) were Cochrane SR and 54 (39%) non-Cochrane.

The Cochrane reviews included a mean of 6.5 (SD 6.1) studies and the non-Cochrane included a mean of 13.2 (SD 10.2) studies. The number of patients was 2619.1 (SD 6245.8 or median 586.0) for the Cochrane and 19888.5 (SD 32925.7 or median 6566.5) patients for the non-Cochrane reviews.

Scenarios under different parameter estimates

Table 1 provides results on the types of outcomes and type of intervention studied for the included meta-analyses by publication type.

Table 1. Descriptive results of included meta-analyses by type of outcome and intervention

	Cochrane (N=83)	Non-Cochrane (N=54)	All reviews (N=137)
Type of outcome	% (n/N)	% (n/N)	% (n/N)
All cause mortality	16.8% (14/83)	22.2% (12/54)	19% (26/137)
Semi-objective	21.7% (18/83)	44.4% (24/54)	30.7% (42/137)
Subjective	61.4% (51/83)	33.3% (18/54)	50.4% (69/137)
Type of intervention			
Pharmacological	56.6% (47/83)	62.9% (34/54)	59.1% (81/137)
Non-pharmacological	44.6% (36/83)	35.2% (20/54)	40.9% (56/137)

Of the included meta-analyses 26 (19%) used 'all cause mortality' as an outcome; 42 (31%) were based on 'semi-objective' outcomes and 69 (50%) on a 'subjective' outcome. The type of intervention was pharmacological in 59% of the meta-analyses. There were significant differences in the type of outcome reported by publication type (χ^2; 2df = 11.15, p = 0.004) but not in the type of intervention reported (χ^2; 1df = 0.54, p = 0.46).

The descriptive analysis of the different parameter estimates (CER, RRR, I^2) used in the calculation of the OIS showed considerable variation depending on the type of outcome (Table 2 and Figure 3).

The number of included patients was higher in non-Cochrane reviews for all outcomes analyzed (Table 2). The CER for 'all cause mortality' had the lowest mean value and the distribution differed between outcome types. For RRR the highest mean value and heterogeneity was observed for 'subjective' outcomes (Table 2) (Figure 3).

Table 2. Descriptive results for the statistical assumptions in the included meta-analyses

	CER	RRR*	Heterogeneity (I^2)	Included patients	OIS estimated†
All reviews n=137					
Mean (SD)	26.9 (26.1)	28.2 (31.5)	20.4 (26.1)	9426.0 (22753.9)	386441.1 (1645397.1)
Cochrane reviews n=83					
Mean (SD)	24.0 (27.9)	21.0 (36.6)	0.0 (25.5)	586 (6245.8)	2301.0 (1422086.6)
Non-Cochrane reviews n=54					
Mean (SD)	10.0 (21.7)	20.0 (20.6)	14.5 (26.9)	6566.5 (32925.7)	7299.5 (1946750.2)
All cause Mortality n=26					
Mean (SD)	12.7 (16.5)	20.3 (22.1)	10.9 (17.6)	14314.6 (25880.1)	499090.7 (1966940.8)
Cochrane reviews n=14					
Mean (SD)	10.0 (11.2)	25.5 (27.3)	6.7 (13.7)	6902.5 (12971.1)	813301.7 (2678677.7)
Non-Cochrane reviews n=12					
Mean (SD)	15.8 (21.4)	14.3 (12.6)	15.7 (20.8)	22962.1 (34232.8)	132511.2 (201708.8)
Semi-Objective n=42					
Mean (SD)	17.2 (20.9)	19.0 (18.7)	18.8 (25.9)	18683.5 (33125.7)	828450.9 (2444468.1)
Cochrane reviews n=18					
Mean (SD)	18.5 (10.5)	21.7 (24.0)	12.8 (22.7)	3384.8 (4762.7)	536616.8 (1803091.2)
Non-Cochrane reviews n=24					
Mean (SD)	16.3 (20.0)	16.9 (13.7)	23.3 (27.5)	30157.0 (40233.9)	1047326.5 (2851698.9)
Subjective n=69					

Mean (SD)	38.2 (27.1)	36.8 (37.9)	24.9 (28.1)	1948.9 (2971.0)	74944.0 (406792.0)
Cochrane reviews n=51					
Mean (SD)	41.6 (27.9)	36.5 (41.6)	23.5 (27.6)	1173.0 (2244.3)	78688.6 (449189.7)
Non-Cochrane reviews n=18					
Mean (SD)	28.5 (22.9)	37.6 (25.3)	28.9 (29.7)	4147.3 (3683.7)	64334.2 (261366.6)

*For the calculation of this descriptive variable all the values were considered as positive

† This estimation of the OIS was done under the conditions of the 'scenario 2' (heterogeneity = 0, alpha 5%).

Meta-analyses that reached the OIS

Figure 4a presents the estimated OIS for each meta-analysis in the extreme scenario of no heterogeneity. All-cause mortality required the highest OIS for both types of reviews. But this was only marginally higher than 'semi-objective' outcomes. For 'subjective' outcomes OIS estimates are considerably smaller due to higher CERs and RRR. Figure 4b shows the number of meta-analyses that have achieved sample sizes equal or higher to the estimated OIS with more non-Cochrane reviews achieving this estimate (see Figure 4c).

Estimation of the OIS based on reported heterogeneity, shows that the necessary sample was only reduced for Cochrane SR reporting subjective outcomes (Table 3). Further increasing the level of heterogeneity (worst-case scenario: heterogeneity = Q3) did not substantially change the proportion of meta-analyses achieving the OIS.

Table 3. Percentage of meta-analyses that achieve the OIS by Heterogeneity (I^2) level assumed

	All cause Mortality			Semi-objective			Subjective		
% (n/N)	Coch	Non-Coch	95% CI Differnce	Coch	Non-Coch	95% CI Difference	Coch	Non-Coch	95% CI Difference
OIS Achieved I^2=reported	0% 0/14	25% 3/12	[-0.074, 0.571]	11.1% 2/18	37.5% 9/24	[-0.043, 0.499]	31.4% 16/51	72.2% 13/18	[0.111, 0.616]
OIS Achieved I^2=0	0% 0/14	25% 3/12	[-0.074, 0.571]	16.6% 3/18	45.8% 11/24	[-0.031, 0.534]	45.1% 23/51	72.2% 13/18	[-0.024, 0.490]
OIS Achieved I^2= Q3	0% 0/14	16.6% 2/12	[-0.134, 0.491]	11.1% 2/18	33.3% 8/24	[-0.079, 0.460]	29.4% 15/51	61.1% 11/18	[0.027, 0.553]

Coch = Cochrane meta-analyses, Non-Coch = Non-Cochrane meta-analyses

When using a more stringent estimate for the effect size (5% RRR) for "all cause mortality" none of the identified meta-analyses had achieved the necessary sample size to meet the OIS (0/14 Cochrane and 0/12 non-Cochrane). Box 1 presents five examples of meta-analyses reporting 'all cause mortality' as an illustration of systematic reviews where the OIS has been reached, and where it has not.

Box 1. Example of five meta-analyses with the 'all cause mortality' as the main outcome that do, or do not, achieve the OIS.

Meta-analysis that meet the OIS

Weng et al. 2010 (Annals) [14]
This meta-analysis evaluated the use of a non-pharmacological intervention (Noninvasive ventilation) to treat patients with acute cardiogenic pulmonary edema including a total of 1,369 patients with a CER of 23%, RRR 27% and 0% heterogeneity. For this systematic review assuming a 0% heterogeneity the OIS estimated was 1,296 patients.

Gastric team 2010 (JAMA) [15]
This meta-analysis evaluated the use of adjuvant chemotherapy for resectable gastric cancer including a total of 3,781 patients with a CER 69%, RRR 9% and 24% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity the OIS estimated was 1,828 patients.

NSCLC meta-analysis 2010 (The Lancet) [16]
This meta-analysis evaluated the use of Adjuvant chemotherapy in patients with operable non-small-cell lung cancer including a total of 8,447 patients with a CER 49%, RRR 11% and 1% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity the OIS estimated was 2,686 patients.

Meta-analysis with a large number of included patients that not meet the OIS

Adam et al. 2012 (Annals) [17]
This meta-analysis evaluated the use of warfarin versus new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism including a total of 14,143 patients with a CER of 2%, RRR 12% and 0% heterogeneity reported. For this systematic review assuming a 0% heterogeneity the OIS estimated was 100,562 patients.

Rizos et al. 2012 (JAMA) [18]
This meta-analysis evaluated the administration of Omega-3 Fatty Acid Supplementation and Risk of Major Cardiovascular Disease Events including a total of 125,410 patients with a CER of 7%, RRR 4% and 1% heterogeneity reported. For this systematic review assuming a 0% heterogeneity the OIS estimated was 255,912patients.

DISCUSSION

Our results show that there is wide variability in the range of values that impact on the OIS calculation: effect size (RRR), heterogeneity (I^2) and control event rate (CER), regardless

of source (Cochrane or non-Cochrane). This variability is partially explained by the type of outcome ('all cause mortality', 'semi-objective' or 'subjective') evaluated.

OIS estimates could therefore be obtained from different types of outcomes, as previously proposed by Turner et al.[9] and Rhodes et al.[11] . To our knowledge this is the first time that accounting for the type of outcome in the estimation of the OIS has been proposed. We also found that the type of outcome impacts on the range of heterogeneity observed and was particularly high for 'subjective' outcomes. One possible explanation for this is the higher number of smaller RCTs. Nevertheless, these differences were more marked in Cochrane reviews while non-Cochrane reviews showed more similar levels of heterogeneity across all types of outcomes. The obtained results show that globally less than half of recent published meta-analysis in high quality journals achieved the OIS and therefore do not have appropriate statistical power to draw firm conclusions.

As expected, the estimation of the OIS assuming different levels of heterogeneity, and alpha values, showed a strong correlation. Although we used specialist software for the estimation of the OIS (TSA v0.9) it is possible to estimate this value using any software that allows sample size estimation if the heterogeneity level is assumed to be zero. Incorporation of heterogeneity can be done using a simple adjustment proposed by Wetterslev [5]. This author proposes the use of an alternative index named the diversity (D^2) statistic as opposed to the I^2 factor. However, there is currently no consensus on what measure of heterogeneity to adopt for the OIS [4,19].

Published meta-analyses that estimate optimal information size often use one or more statistical assumptions, such as a RRR of 10%, or the median RRR of trials with low risk of bias [20-22]. Our analysis shows that the median of the RRR is 20% for all pooled reviews. However, because the distribution of RRR varies by outcome type, in some cases optimal information size is underestimated, whilst in others it is overestimated.

Limitations

There are several proposed statistics to define a "desirable sample size in terms of numbers of participants across all studies" [4]. The OIS as described in this paper involves a relatively simple calculation, which if anything is likely to underestimate the information required to define whether firm evidence has been reached to draw robust conclusions [4,19] Therefore we used this definition of OIS as a measure to estimate what proportion of Systematic Reviews meet this minimum requirement.

We have focused exclusively on the calculation of a single threshold to define when/if a minimum level of evidence has been collected. However, retrospective analyses of meta-analytical results are more commonly used to inform prospective studies. For example, to determine the size of a new trial to answer definitively a question around efficacy. The use of Trial sequential methods has been proposed to identify early signals of effect with monitoring boundaries being defined by frequentist, semi-Bayesian and fully Bayesian methods [4,23,24] Although there is still considerable uncertainty about the estimates and the best method to use, empirical studies have provided examples to suggest these methods could help detect signals early (benefit, harm or futility) [8,25]. Of note, the identification

of the sample size required in a new study or studies will depend on the method used in the meta-analysis [19].

Reviews conducted by the Cochrane collaboration are considered to be higher quality [26-27] and of greater methodological rigor than meta-analyses published in paper-based journals. Our study only included meta-analyses from the top five medical journals and therefore our results may not be applicable to other meta-analyses published in other journals. Nevertheless, this would bias our results towards better evidence being evaluated to what is currently being generated. Also, our results do not generalize to network meta-analyses, which is an area of evidence synthesis that has grown rapidly [28]. A recently published study demonstrated that substantial variation exists in such network-based meta-analysis [29] and the statistical methodology to estimate the OIS in these meta-analyses is less developed than for traditional meta-analysis, hence our exclusion of these studies.

Implications for researchers and methodologists

This study has shown that the type of outcome when estimating the OIS can be used as a proxy for defining the basic parameters (CER, RRR, I^2) required to perform the calculation. Systematic reviewers can use these results to calculate an OIS value for their primary outcome independently of the confidence they have on the specific parameters obtained from their review. Therefore, we encourage reviewers to use the estimation of a sample size as a measure of the likely confidence in their results. Particularly as >50% of the primary outcomes in recent systematic reviews appear to fall below this minimum requirement, pointing out the need for further evidence to reduce uncertainty.

CONCLUSIONS

Heterogeneity and effect size impact on the estimation of the OIS. It is however possible to estimate the OIS using traditional sample size estimation software and if necessary adjust for heterogeneity. Our results demonstrate that the type of outcome is relevant to the estimation of the OIS, as well as the heterogeneity and the CER and RRR. Currently less than half of published meta-analysis in high quality journals have achieved the OIS, and therefore conclusions based on such results are subject to substantial uncertainty.

Competing interests: Non-conflict of interest.

Authors' Contributions: JGA, RP conceived and designed the study. JGA, NP extracted the data. JGA, RP analyzed the data. JGA, RP, CB, NP, CH interpreted the data. JGA, RP wrote the first draft. All authors revised the intellectual content and approved the version to be published.

Acknowledgements

We wish to thank Pol Escarpenter for his support to adapt the figures to the requested format of the journal.

REFERENCES

1. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet. 1998; 351:47–52.

2. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials [Internet]. 1997; 18(6):580-93-6.

3. 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(8):763–9.

4. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med. 2011; 30(9):903-21.

5. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol 2009; 9:86.

6. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol 2008; 61(1):64–75.

7. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2013;8(3):e59202. doi: 10.1371/journal.pone.0059202.

8. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open. 2016; 6(8):e011890. doi: 10.1136/bmjopen-2016-011890.

9. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012; 41(3):818–27.

10. Thorlund K, Engstrøm J, Wetterslev J. User manual for trial sequential analysis (TSA). Trial Unit Available Internet. 2011 [cited 2014 Jun 11].

11. Rhodes KM, Turner RM, Higgins JP. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. Res Synth Methods. 2016;7(4):346-370.

12. Djulbegovic B, Kumar A, Pp G, Perera R, Reljic T, Dent L, et al. New treatments compared to established treatments in randomized trials. Cochrane Database Syst Rev. 2012;(10).

13. Hayden JA, Tomlinson G, Ni A, Rm P, Cartwright J, Mw VT, et al. Approaches to Estimate and Present Baseline Risks Considerations for Cochrane Reviews Evidence-Informed Practice Integration of : 19th Cochrane Colloquium Madrid.

14. Weng CL(1), Zhao YT, Liu QH, Fu CJ, Sun F, Ma YL, Chen YW, He QY. Meta-analysis: Noninvasive ventilation in acute cardiogenic pulmonary edema. Ann Intern Med. 2010; 152(9):590-600.

15. GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group, Paoletti X, Oba K, Burzykowski T, Michiels S, Ohashi Y, Pignon JP, Rougier P, Sakamoto J, Sargent D, Sasako M, Van Cutsem E, Buyse M. Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. JAMA. 2010; 303(17):1729-37. doi: 10.1001/jama.2010.534.

16. NSCLC Meta-analyses Collaborative Group., Arriagada R, Auperin A, Burdett S, Higgins JP, Johnson DH, Le Chevalier T, Le Pechoux C, Parmar MK, Pignon JP, Souhami RL, Stephens RJ, Stewart LA, Tierney JF, Tribodet H, van Meerbeeck J. Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small-cell lung cancer: two meta-analyses of individual patient data. Lancet 2010; 375(9722):1267-77.

17. Adam SS, McDuffie JR, Ortel TL, Williams JW Jr. Comparative effectiveness of warfarin and new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism: a systematic review. Ann Intern Med. 2012; 157(11):796-807.

18. Rizos EC, Ntzani EE, Bika E, Kostapanos MS, Elisaf MS. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. JAMA. 2012; 308(10):1024-33.

19. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Res Synth Methods. 2014; 5(3):212-20.

20. Imberger G, Orr a., Thorlund K, Wetterslev J, Myles P, Møller a. M, et al. Does anaesthesia with nitrous oxide affect mortality or cardiovascular morbidity? A systematic review with meta-analysis and trial sequential analysis. Br J Anaesth. 2014; 112(3):410–26.

21. Bangalore S, Kumar S, Wetterslev J, Messerli FH. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. BMJ. 2011;342:d2234.

22. Afshari A, Wetterslev J, Brok J, Møller A. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. BMJ. 2007;335(7632):1248–51.

23. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017; 17(1):39.

24. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. Stat Med. 2016; 35(29):5356-5375.

25. AlBalawi Z, McAlister FA, Thorlund K, Wong M, Wetterslev J. Random error in cardiovascular meta-analyses: how common are false positive and false negative results?. Int J Cardiol. 2013; 168(2):1102-7.

26. Remschmidt C, Wichmann O HT. Methodological quality of systematic reviews on influenza vaccination. Vaccine. 2014;32(15):1678–84.

27. Windsor B, Popovich I, Jordan V, Showell M, Shea B, Farquhar C. Methodological quality of systematic reviews in subfertility: A comparison of Cochrane and non-Cochrane systematic reviews in assisted reproductive technologies. Hum Reprod. 2012; 27(12):3460–6.

28. Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, et al. The quality of reporting methods and results in network meta-analyses: An overview of reviews and suggestions for improvement. PLoS One. 2014; 9(3).

29. Chambers JD, Naci H, Wouters OJ, Pyo J, Gunjal S, Kennedy IR, et al. An Assessment of the Methodological Quality of Published Network Meta-Analyses: A Systematic Review. PLoS One 2015;10(4):e0121715.

Figures legends

-Figure 1. Algorithm for the selection of the meta-analysis (main comparison) in the systematic review.

-Figure 2. Flow-chart identification of Cochrane and non-Cochrane systematic reviews.

-Figure 3. Systematic Review Characteristics stratified by source (Cochrane vs. Non-Cochrane) and type of outcome.

-Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source; y-axis is on the log scale, b) related to the total number of patients included in each SR, and c) proportion of SRs achieving the OIS by type of outcome and source

Data sharing statement: All of the data used in this research is provided within this publication, its appendices, and the publications referenced in the online supplementary appendices.

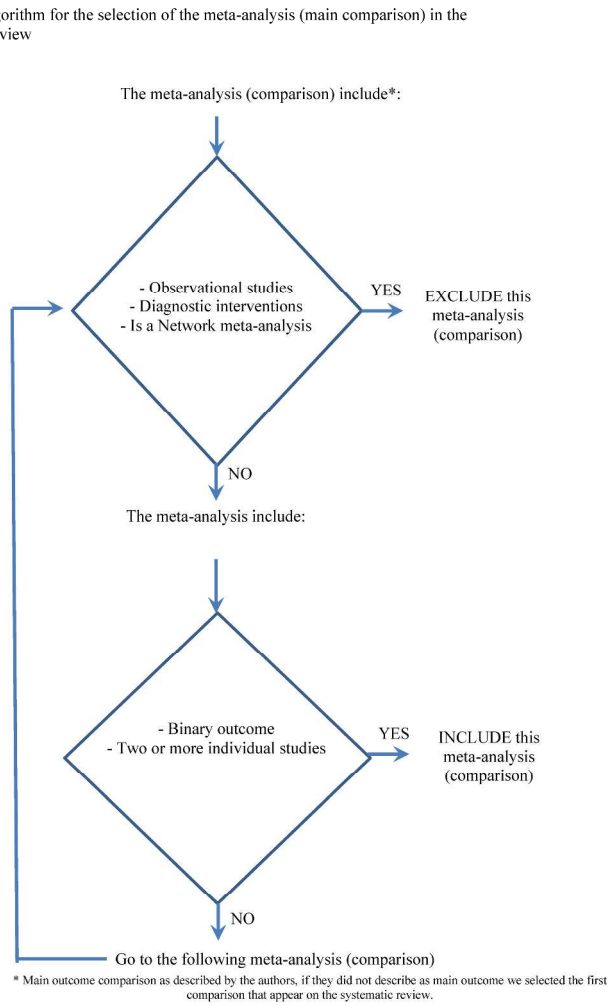


Figure 1. Algorithm for the selection of the meta-analysis (main comparison) in the systematic review

297x420mm (300 x 300 DPI)

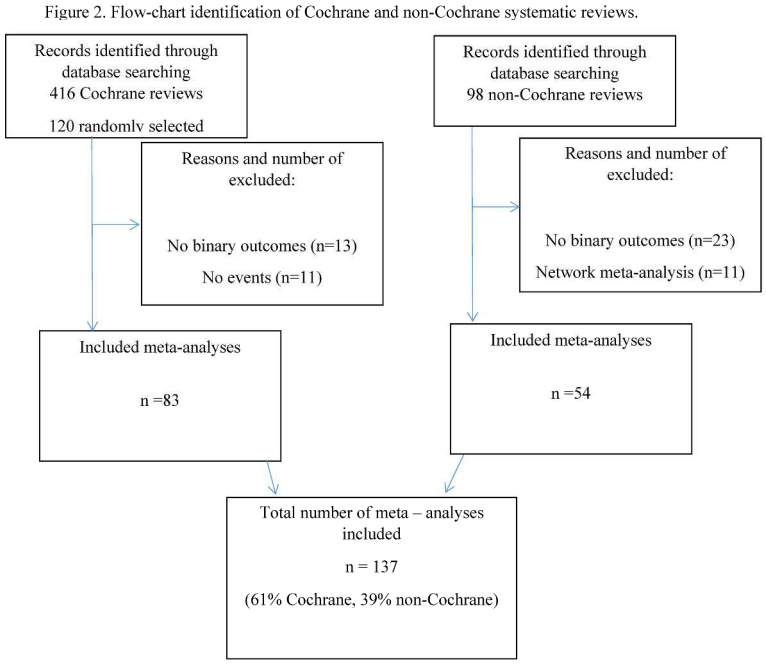


Figure 2. Flow-chart identification of Cochrane and non-Cochrane systematic reviews

297x420mm (300 x 300 DPI)

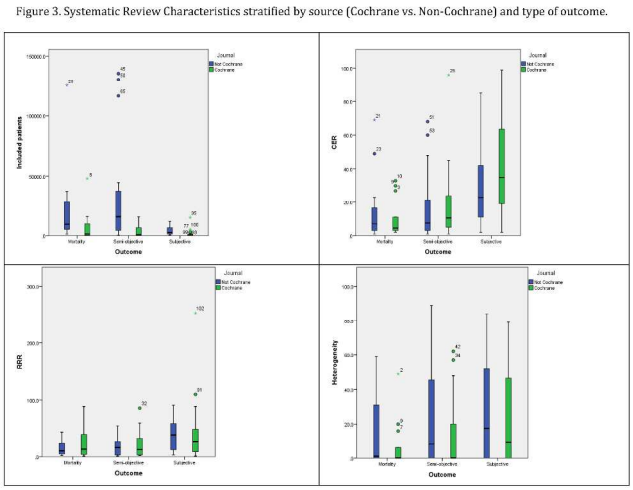


Figure 3. Systematic Review Characteristics stratified by source (Cochrane vs. Non-Cochrane) and type of outcome

297x420mm (300 x 300 DPI)

Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source; y-axis is on the log scale, b) related to the total number of patients included in each SR, and c) proportion of SRs achieving the OIS by type of outcome and source.

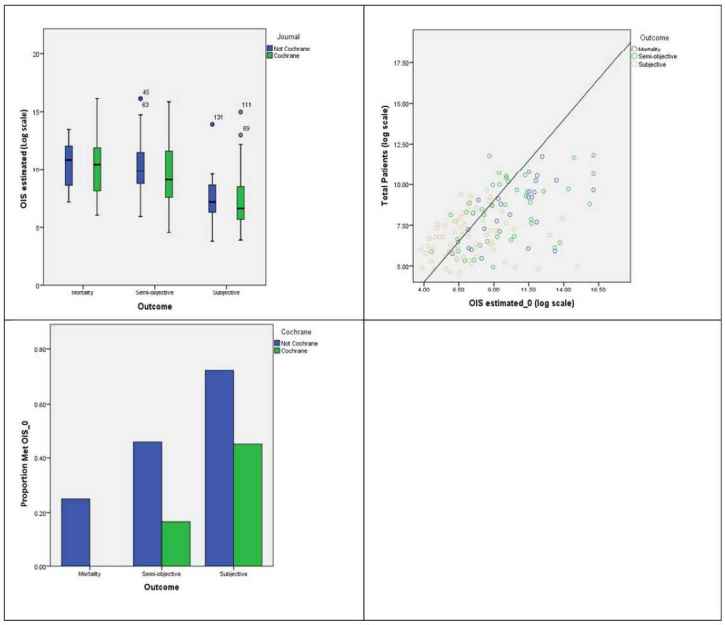


Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source; y-axis is on the log scale, b) related to the total number of patients included in each SR, and c) proportion of SRs achieving the OIS by type of outcome and source

297x420mm (300 x 300 DPI)

BMJ Open

The impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses.

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2017-015888.R2
Article Type:	Research
Date Submitted by the Author:	11-Jul-2017
Complete List of Authors:	Garcia-Alamino, Josep; University of Oxford, Nuffield Department of Primary Care Health Sciences Bankhead, Clare; University of Oxford, University of Oxford, Nuffield Department of Primary Care Health Sciences Heneghan, Carl; Oxford University, Primary Health Care Pidduck, Nicola; University of Oxford, Primary Health Care Perera, Rafael; University of Oxford, Primary Health Care
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Epidemiology, Health policy, Public health, Research methods
Keywords:	STATISTICS & RESEARCH METHODS, Meta-analysis, Optimal information size, Trial sequential analysis, Heterogeneity, Effect size

Title: The impact of heterogeneity and effect size on the estimation of the optimal information size: analysis of recently published meta-analyses.

Corresponding author: Professor Rafael Perera. Nuffield Department of Primary Care Health Sciences. ROQ. Woodstock Road. Oxford OX2 6GG. United Kingdom. E-mail: rafael.perera@phc.ox.ac.uk Phone: +44 (0)1865 617300.

Authors: Josep M Garcia-Alamino[1], Clare Bankhead[1], Carl Heneghan[1], Nicola Pidduck[1], Rafael Perera[1].

1. Nuffield Department of Primary Care Health Sciences. ROQ. Woodstock Road. Oxford OX2 6GG. United Kingdom.

Word count: text (3,392) and abstract (258)

ABSTRACT

Objective: To estimate the proportion of Systematic Reviews that meet the Optimal Information Size (OIS) and assess the impact heterogeneity and effect size have on the OIS estimate by type of outcome (e.g. mortality, semi-objective, or subjective).

Methods: We carried out searches of Medline and Cochrane to retrieve meta-analyses published in systematic reviews from 2010 to 2012. We estimated the OIS using *Trial Sequential Analysis* (TSA v0.9) software, and based on several heterogeneity and effect size scenarios, stratifying by type of outcome (mortality/semi-objective/subjective) and by Cochrane/non-Cochrane reviews.

Results: We included 137 meta-analyses out of 218 (63%) potential systematic reviews (one meta-analysis from each systematic review). Of these reviews, 83 (61%) were Cochrane and 54 (39%) non-Cochrane. The Cochrane reviews included a mean of 6.5 (SD 6.1) studies and the non-Cochrane included a mean of 13.2 (SD 10.2) studies. The mean number of patients was 2619.1 (SD 6245.8 or median 586.0) for the Cochrane and 19888.5 (SD 32925.7 or median 6566.5) patients for the non-Cochrane reviews. The percentage of systematic reviews that achieved the OIS for all-cause mortality outcome were 0% Cochrane and 25% for non-Cochrane reviews; for semi-objective outcome 17% for Cochrane and 46% for non-Cochrane reviews and for subjective outcome 45% for Cochrane and 72% for non-Cochrane reviews.

Conclusions: The number of systematic reviews that meet an optimal information size is low and varies depending on the type of outcome and the type of publication. Less than half of primary outcomes synthesized in systematic reviews achieve the OIS, and therefore the conclusions are subject to substantial uncertainty.

Keywords: meta-analysis, systematic review, optimal information size, trial sequential analysis, heterogeneity, effect size.

Strengths and limitations of this study

- To our Knowledge this is the first analysis to estimate the optimal information size by type of outcome.
- This study includes only systematic reviews from the Cochrane library and the top five general medical journals therefore our results may not be generalizable to systematic reviews published in other journals.

INTRODUCTION

The concept of optimum information size (OIS) was first proposed in 1998 by Pogue et al. [1-2] as "the minimum amount of information required in the collective literature for reliable conclusions about an intervention to be reached". This OIS estimate is based on standard sample size calculations. For example, the required number of participants (information size) for a meta-analysis should match those required in an adequately powered single trial [3]. Other measures of information size have been proposed [4-5], however the OIS involves a relatively simple calculation, which under some scenarios will underestimate the information required to define whether firm evidence has been reached to draw robust conclusions[6]. Brok et al. demonstrated, in a subset of Cochrane reviews, that many meta-analyses have false positive results due to insufficient information [3] and Turner et al. showed that most meta-analysis do not have sufficient power to identify even moderate effects [7-8].

Sample size calculation and the OIS is influenced by several variables such as the control event rate (baseline risk), effect size, the power and the alpha value. Deciding on which values to use can be difficult and is typically based on values observed or estimated from the meta-analysis, or one of the included studies. In addition, increased variation can also effect the estimate of the OIS, and there is currently no consensus about which value of heterogeneity should be used to calculate the OIS.

The OIS can help determine the stability of an effect and whether treatment effect estimates are likely to differ based on further information. However, they are difficult to define in

advance and there is no consensus regarding the alpha (significance) or power value used at the outset.

It is therefore not currently known if evidence accumulation and its associated OIS depend on the type of outcome studied and if this varies by publication type (Cochrane or non-Cochrane review. Therefore, we set out to quantify this by studying systematic reviews published in the Cochrane Library and the top five general medical journals and in the process describe the impact that observed variation in heterogeneity and effect size (RRR) have on the OIS estimation.

METHODS

We defined two sets of systematic reviews to evaluate: Cochrane and non-Cochrane. We identified all Cochrane systematic reviews published during 2010-2012 through the Archie Database (http://archie.cochrane.org), which contains all Cochrane published reviews and allows electronic searching. We randomly selected a total of 120 of these based on random numbers generated using Microsoft Excel for inclusion.

To search for non-Cochrane reviews, we identified all systematic reviews with meta-analyses published in the top five general medical journals (N Engl J Med, Lancet, JAMA Intern Med, Ann Intern Med and BMJ) using the following search strategy in Medline (PubMed): "BMJ"[Journal] OR "Ann Intern Med"[Journal] OR "JAMA"[Journal] OR "Lancet"[Journal] OR "N Engl J Med"[Journal] AND (systematic review [ti] OR meta-analysis [pt] OR meta-analysis [ti]) restricted to systematic reviews published during 2010-2012.

Inclusion/Exclusion criteria

From all the selected Cochrane and non-Cochrane reviews, we included one meta-analysis from each. Based on the order the outcomes were reported (e.g. outcome 1.1 for Cochrane SRs) we selected the first outcome presented in the meta-analysis that was based on: binary data from two or more individual studies (clinical trials or randomized controlled trials). If the first outcome did not meet this inclusion criteria we continued through the listed outcomes until one was identified or we had exhausted the list of outcomes reported. (Figure 1). Meta-analyses that included observational studies, of diagnostic interventions or that were based on network meta-analysis were excluded. Meta-analyses showing no-effect (pooled effect = 1), or meta-analyses with no events in all included trials were also excluded.

Data extraction

Full texts were obtained for those abstracts that met the inclusion criteria and assessed for eligibility. One reviewer JGA extracted the data and a second reviewer (RP or NP) checked the data. We developed customized Excel spreadsheets for the data extraction process. From each included meta-analysis we extracted and calculated the following items: outcome type as defined by Turner [9] [‘all cause mortality’, ‘semi-objective’ (cause-specific mortality, major morbidity event) and ‘subjective’ (pain, mental health outcomes)], comparison, number of included patients, number of trials, number of events in each arm, control event rate, effect size and heterogeneity.

Analyses

We extracted data from each trial and repeated the meta-analysis using random-effects models [DerSimonian and Laird (DL)] to account for potential heterogeneity of effects. Estimates for trials with only one group reporting zero events were adjusted with a constant continuity adjustment of 0.5 in each arm (default adjustment in Revman). The obtained estimates for the pooled effect (e.g. RR) and I^2 were compared with the published results to detect any relevant disagreement, and if required, the analyses were repeated to identify the source of the difference. Meta-analyses and calculation of the optimal information size (OIS) were done using *Trial Sequential Analysis* (TSA v0.9) software [10] freely downloadable at www.ctu.dk/tsa. The TSA software allows meta-analysis of dichotomous or continuous data under fixed or random-effects models and has the option to estimate an information size and the stopping boundary. This estimation of the OIS is based on the alpha spending method (O'Brien Fleming and Lan-DeMets).

To evaluate the impact of changes in heterogeneity and effect size (RRR), we estimated the OIS under different scenarios. For heterogeneity, we analyzed three values of heterogeneity: "heterogeneity = rep" as that reported in the meta-analysis using a random-effects model (or obtained from fitting a random effects model if a fixed effect model was used originally); "heterogeneity = 0"; and "heterogeneity = Q3" (upper quartile or 75th percentile), which was determined based on estimates of predictive distributions published by Rhodes et al. [11]. These two estimates of "heterogeneity = Q3" and "heterogeneity = 0" were chosen as extreme scenarios to evaluate the impact that this parameter has on the OIS. Consistent with Rhodes, the estimation of the OIS took into account the outcome type: 'all cause mortality', 'semi-objective' (cause-specific mortality, major morbidity

event) and 'subjective' (pain, mental health outcomes) and, for simplicity, was based on assuming an average mean study size between 50 and 200 participants.

To evaluate the impact of effect size on the OIS, we used two different estimates of the effect size for the meta-analyses with mortality outcome: the Relative Risk Reduction (RRR) obtained in each meta-analysis as well as an a-priori conservative value of 5% for the RRR as reported by Djulbegovic et al. [12]. For the transformation of Relative Risk (RR) measure to RRR we used the following formula RRR =1-RR. If the RR was greater than 1 we used the RRR as a negative value. We did not determine an alternative estimate for the effect size for the other two outcomes (semi-objective and subjective) as the distribution of possible effects makes the choice of "average effect" difficult to justify. We used only one value, per meta-analysis, for the baseline risk or Control Event Rate (CER). This was taken to be the median of the proportion of events in the included trials in each meta-analysis, following the method proposed by Hayden et al. [13].

We used descriptive statistics and plots to quantify differences in control event rate, effect size and heterogeneity between Cochrane and non-Cochrane reviews, stratified by type of outcome (six groups in total). We also determined the proportion of reviews that have achieved the OIS based on reported results and our two extreme scenarios "heterogeneity = 0" and "heterogeneity = Q3" comparing between Cochrane and non-Cochrane and again stratifying by type of outcome ('all-cause mortality', 'semi-objective' and 'subjective' outcomes).

The descriptive analysis of the characteristics of included meta-analyses was carried out using SPSS v.22 software.

RESULTS

Search Results

Figure 2 presents a flow chart of the results. We excluded 11 Cochrane systematic reviews due to no events reported in the included trials, or due to only one study being included in the review. We included a total of 137 meta-analyses out of 218 (63%) potential systematic reviews (Figure 2): 83 (61%) were Cochrane SR and 54 (39%) non-Cochrane.

The Cochrane reviews included a mean of 6.5 (SD 6.1) studies and the non-Cochrane included a mean of 13.2 (SD 10.2) studies. The number of patients was 2619.1 (SD 6245.8 or median 586.0) for the Cochrane and 19888.5 (SD 32925.7 or median 6566.5) patients for the non-Cochrane reviews.

Scenarios under different parameter estimates

Table 1 provides results on the types of outcomes and type of intervention studied for the included meta-analyses by publication type.

Table 1. Descriptive results of included meta-analyses by type of outcome and intervention

	Cochrane (N=83)	Non-Cochrane (N=54)	All reviews (N=137)
Type of outcome	% (n/N)	% (n/N)	% (n/N)
All cause mortality	16.8% (14/83)	22.2% (12/54)	19% (26/137)
Semi-objective	21.7% (18/83)	44.4% (24/54)	30.7% (42/137)
Subjective	61.4% (51/83)	33.3% (18/54)	50.4% (69/137)
Type of intervention			
Pharmacological	56.6% (47/83)	62.9% (34/54)	59.1% (81/137)
Non-pharmacological	44.6% (36/83)	35.2% (20/54)	40.9% (56/137)

Of the included meta-analyses 26 (19%) used 'all cause mortality' as an outcome; 42 (31%) were based on 'semi-objective' outcomes and 69 (50%) on a 'subjective' outcome. The type of intervention was pharmacological in 59% of the meta-analyses. There were significant differences in the type of outcome reported by publication type (χ^2; 2df = 11.15, p = 0.004) but not in the type of intervention reported (χ^2; 1df = 0.54, p = 0.46).

The descriptive analysis of the different parameter estimates (CER, RRR, I^2) used in the calculation of the OIS showed considerable variation depending on the type of outcome (Table 2 and Figure 3).

The number of included patients was higher in non-Cochrane reviews for all outcomes analyzed (Table 2). The CER for 'all cause mortality' had the lowest mean value and the distribution differed between outcome types. For RRR the highest mean value and heterogeneity was observed for 'subjective' outcomes (Table 2) (Figure 3).

Table 2. Descriptive results for the statistical assumptions in the included meta-analyses

	CER	RRR*	Heterogeneity (I^2)	Included patients	OIS estimated†
All reviews n=137					
Mean (SD)	26.9 (26.1)	28.2 (31.5)	20.4 (26.1)	9426.0 (22753.9)	386441.1 (1645397.1)
Cochrane reviews n=83					
Mean (SD)	24.0 (27.9)	21.0 (36.6)	0.0 (25.5)	586 (6245.8)	2301.0 (1422086.6)
Non-Cochrane reviews n=54					
Mean (SD)	10.0 (21.7)	20.0 (20.6)	14.5 (26.9)	6566.5 (32925.7)	7299.5 (1946750.2)
All cause Mortality n=26					
Mean (SD)	12.7 (16.5)	20.3 (22.1)	10.9 (17.6)	14314.6 (25880.1)	499090.7 (1966940.8)
Cochrane reviews n=14					
Mean (SD)	10.0 (11.2)	25.5 (27.3)	6.7 (13.7)	6902.5 (12971.1)	813301.7 (2678677.7)
Non-Cochrane reviews n=12					
Mean (SD)	15.8 (21.4)	14.3 (12.6)	15.7 (20.8)	22962.1 (34232.8)	132511.2 (201708.8)
Semi-Objective n=42					
Mean (SD)	17.2 (20.9)	19.0 (18.7)	18.8 (25.9)	18683.5 (33125.7)	828450.9 (2444468.1)
Cochrane reviews n=18					
Mean (SD)	18.5 (10.5)	21.7 (24.0)	12.8 (22.7)	3384.8 (4762.7)	536616.8 (1803091.2)
Non-Cochrane reviews n=24					
Mean (SD)	16.3 (20.0)	16.9 (13.7)	23.3 (27.5)	30157.0 (40233.9)	1047326.5 (2851698.9)
Subjective n=69					

Mean (SD)	38.2 (27.1)	36.8 (37.9)	24.9 (28.1)	1948.9 (2971.0)	74944.0 (406792.0)
Cochrane reviews n=51					
Mean (SD)	41.6 (27.9)	36.5 (41.6)	23.5 (27.6)	1173.0 (2244.3)	78688.6 (449189.7)
Non-Cochrane reviews n=18					
Mean (SD)	28.5 (22.9)	37.6 (25.3)	28.9 (29.7)	4147.3 (3683.7)	64334.2 (261366.6)

*For the calculation of this descriptive variable all the values were considered as positive

† This estimation of the OIS was done under the conditions of the 'scenario 2' (heterogeneity = 0, alpha 5%).

Meta-analyses that reached the OIS

Figure 4a presents the estimated OIS for each meta-analysis in the extreme scenario of no heterogeneity. All-cause mortality required the highest OIS for both types of reviews. But this was only marginally higher than 'semi-objective' outcomes. For 'subjective' outcomes OIS estimates are considerably smaller due to higher CERs and RRR. Figure 4b shows the number of meta-analyses that have achieved sample sizes equal or higher to the estimated OIS with more non-Cochrane reviews achieving this estimate (see Figure 4c).

Estimation of the OIS based on reported heterogeneity, shows that the necessary sample was only reduced for Cochrane SR reporting subjective outcomes (Table 3). Further increasing the level of heterogeneity (worst-case scenario: heterogeneity = Q3) did not substantially change the proportion of meta-analyses achieving the OIS.

Table 3. Percentage of meta-analyses that achieve the OIS by Heterogeneity (I^2) level assumed

	All cause Mortality			Semi-objective			Subjective		
% (n/N)	Coch	Non-Coch	95% CI Differnce	Coch	Non-Coch	95% CI Difference	Coch	Non-Coch	95% CI Difference
OIS Achieved I^2=reported	0% 0/14	25% 3/12	[-0.074, 0.571]	11.1% 2/18	37.5% 9/24	[-0.043, 0.499]	31.4% 16/51	72.2% 13/18	[0.111, 0.616]
OIS Achieved I^2=0	0% 0/14	25% 3/12	[-0.074, 0.571]	16.6% 3/18	45.8% 11/24	[-0.031, 0.534]	45.1% 23/51	72.2% 13/18	[-0.024, 0.490]
OIS Achieved I^2= Q3	0% 0/14	16.6% 2/12	[-0.134, 0.491]	11.1% 2/18	33.3% 8/24	[-0.079, 0.460]	29.4% 15/51	61.1% 11/18	[0.027, 0.553]

Coch = Cochrane meta-analyses, Non-Coch = Non-Cochrane meta-analyses

When using a more stringent estimate for the effect size (5% RRR) for "all cause mortality" none of the identified meta-analyses had achieved the necessary sample size to meet the OIS (0/14 Cochrane and 0/12 non-Cochrane). Box 1 presents five examples of meta-analyses reporting 'all cause mortality' as an illustration of systematic reviews where the OIS has been reached, and where it has not.

Box 1. Example of five meta-analyses with the 'all cause mortality' as the main outcome that do, or do not, achieve the OIS.

Meta-analysis that meet the OIS

Weng et al. 2010 (Annals) [14]
This meta-analysis evaluated the use of a non-pharmacological intervention (Noninvasive ventilation) to treat patients with acute cardiogenic pulmonary edema including a total of 1,369 patients with a CER of 23%, RRR 27% and 0% heterogeneity. For this systematic review assuming a 0% heterogeneity the OIS estimated was 1,296 patients.

Gastric team 2010 (JAMA) [15]
This meta-analysis evaluated the use of adjuvant chemotherapy for resectable gastric cancer including a total of 3,781 patients with a CER 69%, RRR 9% and 24% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity the OIS estimated was 1,828 patients.

NSCLC meta-analysis 2010 (The Lancet) [16]
This meta-analysis evaluated the use of Adjuvant chemotherapy in patients with operable non-small-cell lung cancer including a total of 8,447 patients with a CER 49%, RRR 11% and 1% heterogeneity reported by the meta-analysis. For this systematic review assuming a 0% heterogeneity the OIS estimated was 2,686 patients.

Meta-analysis with a large number of included patients that not meet the OIS

Adam et al. 2012 (Annals) [17]
This meta-analysis evaluated the use of warfarin versus new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism including a total of 14,143 patients with a CER of 2%, RRR 12% and 0% heterogeneity reported. For this systematic review assuming a 0% heterogeneity the OIS estimated was 100,562 patients.

Rizos et al. 2012 (JAMA) [18]
This meta-analysis evaluated the administration of Omega-3 Fatty Acid Supplementation and Risk of Major Cardiovascular Disease Events including a total of 125,410 patients with a CER of 7%, RRR 4% and 1% heterogeneity reported. For this systematic review assuming a 0% heterogeneity the OIS estimated was 255,912patients.

DISCUSSION

Our results show that there is wide variability in the range of values that impact on the OIS calculation: effect size (RRR), heterogeneity (I^2) and control event rate (CER), regardless

of source (Cochrane or non-Cochrane). This variability is partially explained by the type of outcome ('all cause mortality', 'semi-objective' or 'subjective') evaluated.

OIS estimates could therefore be obtained from different types of outcomes, as previously proposed by Turner et al.[9] and Rhodes et al.[11] . To our knowledge this is the first time that accounting for the type of outcome in the estimation of the OIS has been proposed. We also found that the type of outcome impacts on the range of heterogeneity observed and was particularly high for 'subjective' outcomes. One possible explanation for this is the higher number of smaller RCTs. Nevertheless, these differences were more marked in Cochrane reviews while non-Cochrane reviews showed more similar levels of heterogeneity across all types of outcomes. The obtained results show that globally less than half of recent published meta-analysis in high quality journals achieved the OIS and therefore do not have appropriate statistical power to draw firm conclusions.

As expected, the estimation of the OIS assuming different levels of heterogeneity, and alpha values, showed a strong correlation. Although we used specialist software for the estimation of the OIS (TSA v0.9) it is possible to estimate this value using any software that allows sample size estimation if the heterogeneity level is assumed to be zero. Incorporation of heterogeneity can be done using a simple adjustment proposed by Wetterslev [5]. This author proposes the use of an alternative index named the diversity (D^2) statistic as opposed to the I^2 factor. However, there is currently no consensus on what measure of heterogeneity to adopt for the OIS [4,19].

Published meta-analyses that estimate optimal information size often use one or more statistical assumptions, such as a RRR of 10%, or the median RRR of trials with low risk of bias [20-22]. Our analysis shows that the median of the RRR is 20% for all pooled reviews. However, because the distribution of RRR varies by outcome type, in some cases optimal information size is underestimated, whilst in others it is overestimated.

Limitations

There are several proposed statistics to define a "desirable sample size in terms of numbers of participants across all studies" [4]. The OIS as described in this paper involves a relatively simple calculation, which if anything is likely to underestimate the information required to define whether firm evidence has been reached to draw robust conclusions [4,19] Therefore we used this definition of OIS as a measure to estimate what proportion of Systematic Reviews meet this minimum requirement.

We have focused exclusively on the calculation of a single threshold to define when/if a minimum level of evidence has been collected. However, retrospective analyses of meta-analytical results are more commonly used to inform prospective studies. For example, to determine the size of a new trial to answer definitively a question around efficacy. The use of Trial sequential methods has been proposed to identify early signals of effect with monitoring boundaries being defined by frequentist, semi-Bayesian and fully Bayesian methods [4,23,24] Although there is still considerable uncertainty about the estimates and the best method to use, empirical studies have provided examples to suggest these methods could help detect signals early (benefit, harm or futility) [8,25]. Of note, the identification

of the sample size required in a new study or studies will depend on the method used in the meta-analysis [19].

Reviews conducted by the Cochrane collaboration are considered to be higher quality [26-27] and of greater methodological rigor than meta-analyses published in paper-based journals. Our study only included meta-analyses from the top five medical journals and therefore our results may not be applicable to other meta-analyses published in other journals. Nevertheless, this would bias our results towards better evidence being evaluated to what is currently being generated. Also, our results do not generalize to network meta-analyses, which is an area of evidence synthesis that has grown rapidly [28]. A recently published study demonstrated that substantial variation exists in such network-based meta-analysis [29] and the statistical methodology to estimate the OIS in these meta-analyses is less developed than for traditional meta-analysis, hence our exclusion of these studies.

Implications for researchers and methodologists

This study has shown that the type of outcome when estimating the OIS can be used as a proxy for defining the basic parameters (CER, RRR, I^2) required to perform the calculation. Systematic reviewers can use these results to calculate an OIS value for their primary outcome independently of the confidence they have on the specific parameters obtained from their review. Therefore, we encourage reviewers to use the estimation of a sample size as a measure of the likely confidence in their results. Particularly as >50% of the primary outcomes in recent systematic reviews appear to fall below this minimum requirement, pointing out the need for further evidence to reduce uncertainty.

CONCLUSIONS

Heterogeneity and effect size impact on the estimation of the OIS. It is however possible to estimate the OIS using traditional sample size estimation software and if necessary adjust for heterogeneity. Our results demonstrate that the type of outcome is relevant to the estimation of the OIS, as well as the heterogeneity and the CER and RRR. Currently less than half of published meta-analysis in high quality journals have achieved the OIS, and therefore conclusions based on such results are subject to substantial uncertainty.

Competing interests: Non-conflict of interest.

Authors' Contributions: JGA, RP conceived and designed the study. JGA, NP extracted the data. JGA, RP analyzed the data. JGA, RP, CB, NP, CH interpreted the data. JGA, RP wrote the first draft. All authors revised the intellectual content and approved the version to be published.

Acknowledgements

We wish to thank Pol Escarpenter for his support to adapt the figures to the requested format of the journal.

REFERENCES

1. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet. 1998; 351:47–52.

2. Pogue JM, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials [Internet]. 1997; 18(6):580-93-6.

3. 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(8):763–9.

4. Higgins JP, Whitehead A, Simmonds M. Sequential methods for random-effects meta-analysis. Stat Med. 2011; 30(9):903-21.

5. Wetterslev J, Thorlund K, Brok J, Gluud C. Estimating required information size by quantifying diversity in random-effects model meta-analyses. BMC Med Res Methodol 2009; 9:86.

6. Wetterslev J, Thorlund K, Brok J, Gluud C. Trial sequential analysis may establish when firm evidence is reached in cumulative meta-analysis. J Clin Epidemiol 2008; 61(1):64–75.

7. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLoS One. 2013;8(3):e59202. doi: 10.1371/journal.pone.0059202.

8. Imberger G, Thorlund K, Gluud C, Wetterslev J. False-positive findings in Cochrane meta-analyses with and without application of trial sequential analysis: an empirical review. BMJ Open. 2016; 6(8):e011890. doi: 10.1136/bmjopen-2016-011890.

9. Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012; 41(3):818–27.

10. Thorlund K, Engstrøm J, Wetterslev J. User manual for trial sequential analysis (TSA). Trial Unit Available Internet. 2011 [cited 2014 Jun 11].

11. Rhodes KM, Turner RM, Higgins JP. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. Res Synth Methods. 2016;7(4):346-370.

12. Djulbegovic B, Kumar A, Pp G, Perera R, Reljic T, Dent L, et al. New treatments compared to established treatments in randomized trials. Cochrane Database Syst Rev. 2012;(10).

13. Hayden JA, Tomlinson G, Ni A, Rm P, Cartwright J, Mw VT, et al. Approaches to Estimate and Present Baseline Risks Considerations for Cochrane Reviews Evidence-Informed Practice Integration of : 19th Cochrane Colloquium Madrid.

14. Weng CL(1), Zhao YT, Liu QH, Fu CJ, Sun F, Ma YL, Chen YW, He QY. Meta-analysis: Noninvasive ventilation in acute cardiogenic pulmonary edema. Ann Intern Med. 2010; 152(9):590-600.

15. GASTRIC (Global Advanced/Adjuvant Stomach Tumor Research International Collaboration) Group, Paoletti X, Oba K, Burzykowski T, Michiels S, Ohashi Y, Pignon JP, Rougier P, Sakamoto J, Sargent D, Sasako M, Van Cutsem E, Buyse M. Benefit of adjuvant chemotherapy for resectable gastric cancer: a meta-analysis. JAMA. 2010; 303(17):1729-37. doi: 10.1001/jama.2010.534.

16. NSCLC Meta-analyses Collaborative Group., Arriagada R, Auperin A, Burdett S, Higgins JP, Johnson DH, Le Chevalier T, Le Pechoux C, Parmar MK, Pignon JP, Souhami RL, Stephens RJ, Stewart LA, Tierney JF, Tribodet H, van Meerbeeck J. Adjuvant chemotherapy, with or without postoperative radiotherapy, in operable non-small-cell lung cancer: two meta-analyses of individual patient data. Lancet 2010; 375(9722):1267-77.

17. Adam SS, McDuffie JR, Ortel TL, Williams JW Jr. Comparative effectiveness of warfarin and new oral anticoagulants for the management of atrial fibrillation and venous thromboembolism: a systematic review. Ann Intern Med. 2012; 157(11):796-807.

18. Rizos EC, Ntzani EE, Bika E, Kostapanos MS, Elisaf MS. Association between omega-3 fatty acid supplementation and risk of major cardiovascular disease events: a systematic review and meta-analysis. JAMA. 2012; 308(10):1024-33.

19. Kulinskaya E, Wood J. Trial sequential methods for meta-analysis. Res Synth Methods. 2014; 5(3):212-20.

20. Imberger G, Orr a., Thorlund K, Wetterslev J, Myles P, Møller a. M, et al. Does anaesthesia with nitrous oxide affect mortality or cardiovascular morbidity? A systematic review with meta-analysis and trial sequential analysis. Br J Anaesth. 2014; 112(3):410–26.

21. Bangalore S, Kumar S, Wetterslev J, Messerli FH. Angiotensin receptor blockers and risk of myocardial infarction: meta-analyses and trial sequential analyses of 147 020 patients from randomised trials. BMJ. 2011;342:d2234.

22. Afshari A, Wetterslev J, Brok J, Møller A. Antithrombin III in critically ill patients: systematic review with meta-analysis and trial sequential analysis. BMJ. 2007;335(7632):1248–51.

23. Wetterslev J, Jakobsen JC, Gluud C. Trial Sequential Analysis in systematic reviews with meta-analysis. BMC Med Res Methodol. 2017; 17(1):39.

24. Spence GT, Steinsaltz D, Fanshawe TR. A Bayesian approach to sequential meta-analysis. Stat Med. 2016; 35(29):5356-5375.

25. AlBalawi Z, McAlister FA, Thorlund K, Wong M, Wetterslev J. Random error in cardiovascular meta-analyses: how common are false positive and false negative results?. Int J Cardiol. 2013; 168(2):1102-7.

26. Remschmidt C, Wichmann O HT. Methodological quality of systematic reviews on influenza vaccination. Vaccine. 2014;32(15):1678–84.

27. Windsor B, Popovich I, Jordan V, Showell M, Shea B, Farquhar C. Methodological quality of systematic reviews in subfertility: A comparison of Cochrane and non-Cochrane systematic reviews in assisted reproductive technologies. Hum Reprod. 2012; 27(12):3460–6.

28. Hutton B, Salanti G, Chaimani A, Caldwell DM, Schmid C, Thorlund K, et al. The quality of reporting methods and results in network meta-analyses: An overview of reviews and suggestions for improvement. PLoS One. 2014; 9(3).

29. Chambers JD, Naci H, Wouters OJ, Pyo J, Gunjal S, Kennedy IR, et al. An Assessment of the Methodological Quality of Published Network Meta-Analyses: A Systematic Review. PLoS One 2015;10(4):e0121715.

Figures legends

-Figure 1. Algorithm for the selection of the meta-analysis (main comparison) in the systematic review.

-Figure 2. Flow-chart identification of Cochrane and non-Cochrane systematic reviews.

-Figure 3. Systematic Review Characteristics stratified by source (Cochrane vs. Non-Cochrane) and type of outcome.

-Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source; y-axis is on the log scale, b) related to the total number of patients included in each SR, and c) proportion of SRs achieving the OIS by type of outcome and source

Data sharing statement: All of the data used in this research is provided within this publication, its appendices, and the publications referenced in the online supplementary appendices.

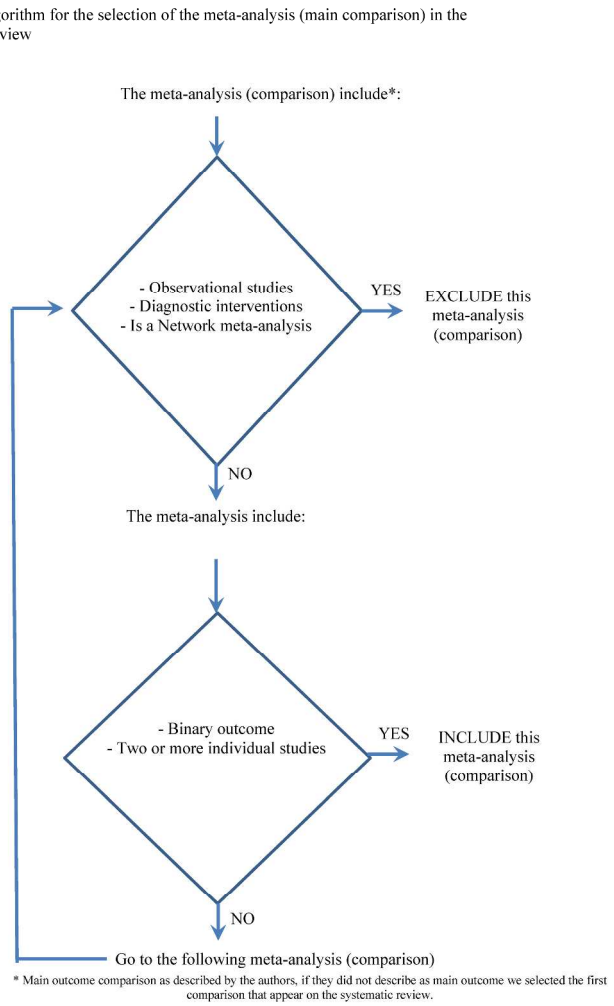


Figure 1. Algorithm for the selection of the meta-analysis (main comparison) in the systematic review

297x420mm (300 x 300 DPI)

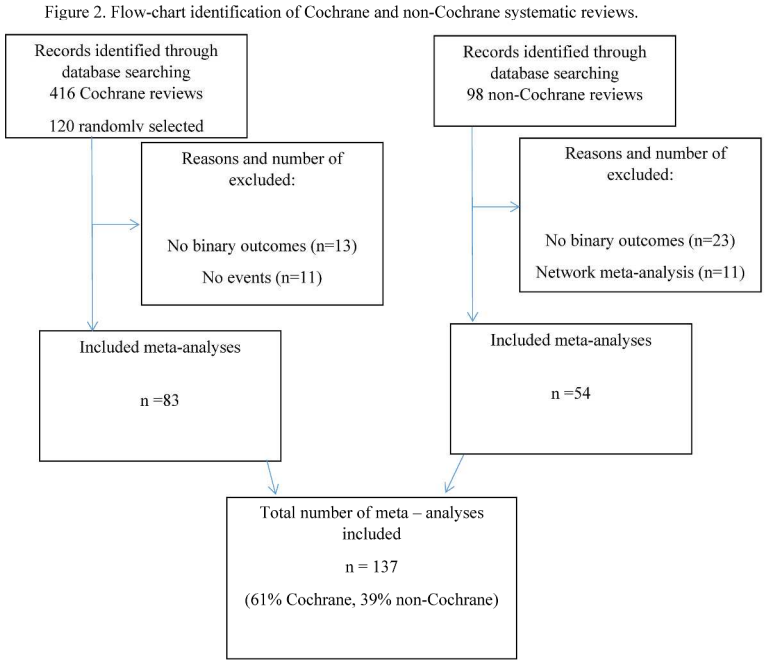


Figure 2. Flow-chart identification of Cochrane and non-Cochrane systematic reviews

297x420mm (300 x 300 DPI)

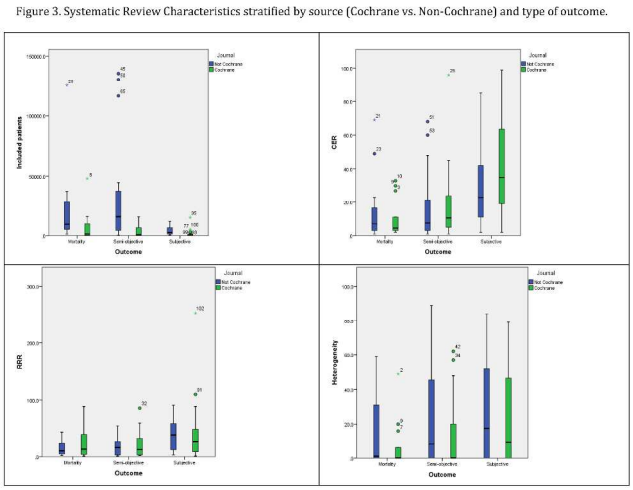


Figure 3. Systematic Review Characteristics stratified by source (Cochrane vs. Non-Cochrane) and type of outcome

297x420mm (300 x 300 DPI)

Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source; y-axis is on the log scale, b) related to the total number of patients included in each SR, and c) proportion of SRs achieving the OIS by type of outcome and source.

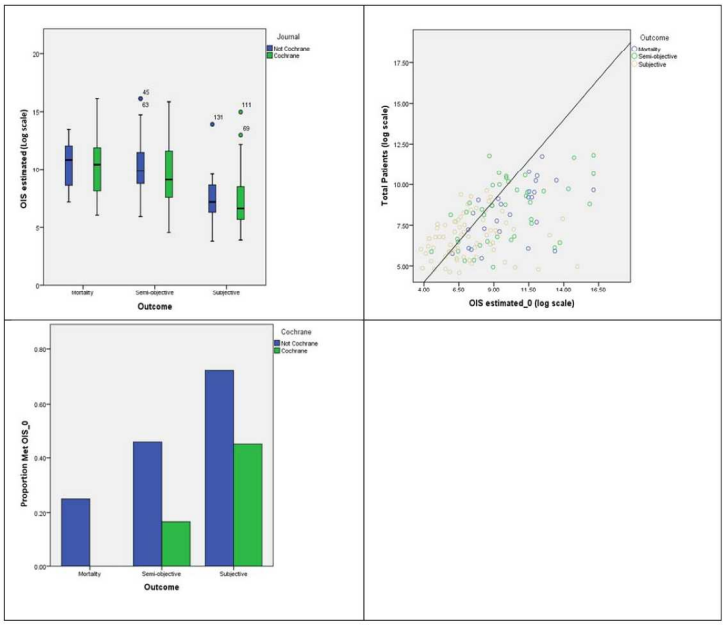


Figure 4. OIS estimated considering best-case scenario (heterogeneity = 0) by a) type of outcome and source; y-axis is on the log scale, b) related to the total number of patients included in each SR, and c) proportion of SRs achieving the OIS by type of outcome and source

297x420mm (300 x 300 DPI)