PROTOCOL

Apparently conclusive meta-analyses may be inconclusive - The reliability of positive, neutral, and negative results in apparently conclusive critical care meta-analyses

Authors:
F Keus [1], G Imberger [2], J Wetterslev [2], C Gluud [2], SL Klingenberg [2], JG Zijlstra [1], ICC van de Horst [1]

[1] Department of Critical Care, University of Groningen, University Medical Center Groningen, The Netherlands
and
[2] The Copenhagen Trial Unit (CTU), Centre of Clinical Intervention Research, Rigshospitalet, Dept.7812, Copenhagen University Hospital, DK-2100 Copenhagen, Denmark.

Version: 05-04-2013

INTRODUCTION

Systematic reviews with meta-analyses of randomised clinical trials are considered the highest level of evidence for intervention research.[1-3] However, systematic reviews might lose credibility when their meta-analyses are invalid.[4] Random errors ('the play of chance') are one major reason for misleading results in meta-analyses and the risk of random error may increase considerably due to early testing of sparse data [5] or due to multiple testing on accumulating evidence when new trials emerge.[6,7]

In a randomised clinical trial, a hypothesis is tested based on a required sample size estimated a priori. Logically, a meta-analysis should include an information size at least as large as the sample size of an adequately powered single trial to reduce the risk of random error.[8-11] However, in meta-analyses, multiple trials are included and heterogeneity, both clinical and statistical, may be present. Therefore, the information size in meta-analyses should be heterogeneity adjusted and more information is required when statistical heterogeneity increases.[10,11] Despite these prerequisites, medical communities have largely ignored the issues of information size and risks of random errors in meta-analyses.

The aim of a meta-analysis is to identify the benefit or harm of an intervention as early as possible. Thus, meta-analyses are subjected to repeated significance testing when updated [1,2], which is prone to exacerbate the risk of random error if a decision to use the intervention is made based on a conventional *p*-value criterion (typically a two-sided α=5%).[6,7] The situation is comparable with interim analyses of a single randomised clinical trial. In interim analyses of a randomised clinical trial, when assessing whether to stop the trial early, formal sequential monitoring boundaries are used to adjust the thresholds for the employed test statistic.[12] Similar utilisation of formal boundaries as guides for cumulative meta-analyses is desirable to distinguish real effects from random errors.[11]

Trial sequential analysis (TSA) is a methodology that combines an *a priori* estimated required information size for a meta-analysis with the adaptation of monitoring boundaries to evaluate the accumulating data (i.e., meta-analytic updates).[11,13] The information size calculation is similar to the sample size calculation in a single trial with the addition of a heterogeneity adjustment.[11,13] Once the information size is estimated, trial sequential monitoring boundaries can be adapted as new trials are published and meta-analyses are updated over time. In this

context, TSA may serve as a tool for quantifying the reliability of cumulative data in meta-analyses.[11,13]

In this study we will identify all systematic reviews or journal articles with meta-analyses related to critical care. We will apply TSA on these meta-analyses, i.e., we will calculate the heterogeneity-adjusted required information size and construct the trial sequential monitoring boundaries. We will use TSA to evaluate the risk of random error, and explore the extent to which apparently conclusive traditional meta-analyses remain conclusive when accounting for potentially exacerbated risk of random error due to sparse data and repetitive testing. We will also evaluate how many of the statistically insignificant meta-analyses actually have the power to exclude the anticipated intervention effect used for the information size estimation due to reaching the futility area in the TSA or due to having included a number of randomized patients which is greater than the required information size for that specific meta-analysis.

METHODS

Material

We will search the databases *'Cochrane Reviews'* and *'Other Reviews'* in *The Cochrane Library* [14], *PubMed/MEDLINE*, and *EMBASE,* for all meta-analyses related to critical care medicine. We give the search strategies in Appendix 1 with the time span of the searches until April 2013.

We will select all meta-analyses that include two or more randomised trials reporting on a binary outcome. There will be no language restriction.[2] We will only include meta-analyses that include randomised clinical trials. If a meta-analysis includes observational studies next to randomised clinical trials, we will only use the data from the randomised clinical trials. For each included meta-analysis, we will include all the primary outcomes (as defined by that study) up to a maximum of three, but only if they are dichotomous. Continuous outcomes will be excluded. If it is unclear from the full text which outcome was considered by the authors to be the most important primary outcome measure, then we will select the first three reported outcome measures in the text. If the data required to conduct the TSA is not available in the full publication (including links to online material), we will exclude that meta-analysis.

Trial sequential analyses

TSA necessitates the pre-specification of a relevant (worthwhile) intervention effect (μ) and risk of type 1 (α) and type 2 (β) errors.[10,12] We will set a two-sided α = 5% and β = 20% (1- β = 80% power). The required information size will be calculated using the formula

$$2 \cdot (Z_{\alpha/2} + Z_{\beta})^2 \cdot 2 \cdot \nu/\mu^2$$

The intervention effect $\mu = P_C - P_E$ (where P_C being the proportion in the control group and P_E being the proportion in the intervention group) and its variance $\nu = P^* \cdot (1-P^*)$ where $P^* = (P_C + P_E)/2$, assuming equal group sizes.

Heterogeneity increases the uncertainty in meta-analyses.[15] Heterogeneity may be measured by diversity (D^2).[15,16] We will adjust the required information size according to the degree of diversity expressed by D^2 found in the conventional meta-analysis by multiplying the required information size (see above) by $1/(1 - D^2)$. If the actual diversity in the meta-analysis is 0 and the required information size has not been reached, we will perform a sensitivity analysis with a D^2 of 25% as our best guess of a heterogeneity adjustment when the meta-analysis eventually reaches its required information size.[16,17] This estimate of prospective heterogeneity corresponds to the heterogeneity adjustment suggested for multi-centre trials.[18]

We will use TSA software (www.ctu.dk/tsa) to conduct the TSA analyses. For each meta-analysis, we will calculate the diversity-adjusted required information size as described and apply the trial sequential monitoring boundaries.[11] The monitoring boundaries are based on the O'Brien–Fleming α-spending function that controls the overall type I error by spending it in an appropriate manner, as statistical tests are employed throughout the accumulation of trials.[12,19] We will chose the α-spending function that results in the well-known Lan–DeMets monitoring boundaries.[20] We will calculate the cumulative *z*-curve for each cumulative meta-analysis (i.e., the series of *z* statistics after each consecutive trial) and we will assess its crossing of monitoring boundaries with the DerSimonian-Laird random-effects model.[22] The monitoring boundaries should be crossed by the cumulative *z*-curve to obtain firm evidence for an intervention effect (Figure 1). In meta-analyses, *z*-values of ±1.96 correspond to the conventional $p = 0.05$ in a two-sided hypothesis test.

We will use relative risk as the effect measure for dichotomous data. When odds ratios, risk differences, or hazard ratios are used in the original publication, we will recalculate the reported pooled intervention effect measure into relative risks. We will perform a sensitivity analysis using odds ratios as the effect measure in case event proportions are below 5% (in the control group). In the case of zero events, we will make an empirical adjustment of 0.001 to the number of events in the control and intervention groups.[23]

The event proportion in the control arm will be estimated by meta-analyzing the control group event proportions of all included trials. For each meta-analysis, we will conduct the TSA analysis using the estimated intervention effect in the published meta-analysis. We will use an overall maximum type I error (α) of 0.05 and a maximum type II error (β) of 0.20 (1- β = 80% power).

We will also conduct a sensitivity analysis using a relative risk reduction (RRR) of 25% since early testing with sparse data may overestimate the intervention effect. On the other hand, the true intervention effect may eventually appear to be higher (e.g., around 25% RRR) than an initially underestimated intervention effect (e.g., 10%).

Apart from testing for a significant beneficial (or harmful) effect by applying trial sequential monitoring boundaries TSA may also assess when an intervention is unlikely to have some anticipated effect. If a meta-analysis has found that a given intervention has no significant effect, this finding may be due to lack of power or the intervention is likely to have no effect when the required information size has been reached.

In some situations, however, TSA may be able to conclude earlier before an appropriate information size (IS) has been reached that a treatment effect is unlikely to be as large as the anticipated intervention effect. Futility boundaries are a set of thresholds that reflect the uncertainty of obtaining a chance negative finding in relation to the strength of the available evidence (i.e., the accumulated number of patients). Above the thresholds, the test statistic may not have yielded statistical significance due to lack of power, but there is still a chance that a statistically significant effect will be found before the meta-analysis surpasses the required information size. Below the threshold, the test statistic is so low that the likelihood of a significant effect being found becomes negligible. In the latter case, further randomisation of patients is futile; the intervention does not possess the postulated intervention effect. The

methods for controlling for type II error are an extension of the Lan-DeMets methodology that allows for non-superiority and non-inferiority testing. That is, instead of constructing adjusted thresholds for statistical significance, the method constructs adjusted thresholds for non-superiority and non-inferiority (or no difference). Together, adjusted non-superiority and non-inferiority boundaries make up what is referred to as futility boundaries or inner wedge boundaries.

When the cumulative z-curve has not reached the futility monitoring boundaries then there is absence of evidence to support or refute a certain intervention effect. If the cumulative z-curve crosses into the futility area then there is evidence to refute a certain intervention effect. Once the futility boundary is crossed further trials on that intervention and outcome and that specific intervention effect are futile. However, one might claim a lesser intervention effect, if clinically relevant, and then more randomised trials may be needed.

All data will be analysed using the Copenhagen Trial Unit computer program, TSA version 0.9 beta (www.ctu.dk/tsa). The TSA v0.9 displays the relationship between the cumulative *z*-score, the information size, and the two-sided monitoring boundaries on a graph. We will present the conventional confidence intervals and the TSA adjusted 95% confidence intervals with the according p-values. The graphs may be shown in an additional file.

Assessment of bias risk

Assuming all included trials in a meta-analysis being of low risk of bias the result of the TSA can conclude that there is a significant difference, if the trial sequential monitoring boundary for benefit has been crossed. However, as we all know, any risk of bias would challenge a significant result for benefit and in this situation a thorough bias evaluation will be mandatory to declare that the meta-analysis shows firm evidence for a beneficial effect. For the meta-analyses that break the trial sequential monitoring boundary for benefit there is suggestion that the intervention seems to work without any risk for random errors provided that there is also no risk for systematic error (bias). Therefore, we will assess the risk of bias of the meta-analyses that break the trial sequential boundary for benefit. For meta-analyses that do not break the trial sequential monitoring boundaries for benefit there is insufficient evidence to propagate the implementation of that intervention and we will not assess the risk of bias for these meta-analyses.

In contrast, for the meta-analyses that break the trial sequential monitoring boundary for harm in case of two different interventions (intervention A versus intervention B) there is suggestion that the intervention A does not seem to work or that intervention B seems to work without any risk for random errors provided that there is also no risk for systematic error (bias). Therefore we will also assess the risk of bias for the meta-analyses that break the trial sequential monitoring boundary for harm.

The risk of bias of the included randomized trials will be assessed using the Cochrane's tool for bias assessment according to the domains of allocation sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessors, incomplete outcome data, selective outcome reporting, vested interest bias, and any other bias risk. We will consider trials classified as low risk of bias if all the domains are with low risk of bias. Trials with one or more of these mentioned risk of bias domains scored as unclear or high risk of bias will be considered high risk of bias trials.

We intend to perform subgroup analyses according to bias risk assessment of course only in meta-analyses that break the trial sequential monitoring boundary for benefit. Trials with low risk of bias will be compared to trials with high risk of bias (one or more of the domains of bias assessed as inadequate or unclear).

Overview of TSA analyses

Overall, a total of four TSA analyses will be conducted for each meta-analysis: one primary analysis and three sensitivity analyses (Table 1).

Primary analysis

The primary analysis will be conducting the TSA using the estimated meta-analysed intervention effect for the RRR and the actual diversity (D^2) in the meta-analysis for the required information size adjustment with an overall type I error (α) of 0.05 and a type II error (β) of 0.20 (1- β = 80% power).

Sensitivity analyses

We will conduct a total of three sensitivity analyses for each meta-analysis:

1) RRR = 25% and D^2 as measured.
2) RRR as measured and D^2 = 25% if the measured D^2 < 25%.

3) RRR=25% and D^2 = 25% if the measured D^2 < 25%.

In case the event proportion in the meta-analysis is below 5%, we will additionally conduct a sensitivity analysis using odds ratios.

		Relative risk reduction (RRR)	
		as measured in the meta-analysis	***25%***
Heterogeneity (measured by diversity D^2)	***as measured in the meta-analysis***	Primary analysis	Sensitivity analysis
	25% ***(if D^2 measured < 25)***	Sensitivity analysis	Sensitivity analysis

Table 1: Overview of all TSA calculations for each meta-analysis

Data extraction

We will retrieve full copies of all the systematic reviews that contain meta-analyses selected based on the inclusion criteria available in the abstract.

For each included meta-analysis, we will extract the following data about the meta-analysis:

1. First author
2. Year of publication
3. Population included
4. Description of the intervention
5. Description of the control
6. Definition of the outcome
7. Type of meta-analysis used
8. Measure of heterogeneity used (if described)
9. Meta-analytic result

For each included meta-analysis, we will extract the following data about each included trial in the meta-analysis:

1. First author
2. Year of publication of each included trial

3. Results: dichotomous data – the proportion with the outcome in the intervention group and in the control group

Presentation of results

Eventually, we will produce a table listing the results of the original meta-analyses as well as our primary and sensitivity analyses (Table 2). Assessment of the reliability of conclusions will be facilitated by comparing the results in the columns of Table 2.

A	B	C	D	E	F	G	H	I
				Primary analysis	Sensitivity analyses			
Author, year	**Outcome**	**Original MA** ***published result***	**Original MA *** ***RCTs only (RR)***	**TSA MA** ***RRR: MA estimate*** ***D^2: MA estimate***	**TSA MA** ***RRR: MA estimate*** ***D^2: 25%***	**TSA MA** ***RRR: 25%*** ***D^2: 25%***	**TSA MA** ***RRR: 25%*** ***D^2: MA estimate***	**Futility boundary crossed? (Y/N)**
1								
2								
3								
etc								

Table 2: Overview of original meta-analysis results and recalculations using Trial Sequential Analysis techniques for each meta-analysis including primary and sensitivity analyses.
MA: meta-analysis; TSA Trial Sequential Analysis; RR: relative risk; RRR: Relative Risk Reduction; D^2: diversity (measure for heterogeneity); Y: yes; N: no.
* Column D will list the original meta-analysis result recalculated in relative risks if odds ratios or risk differences are used and excluding observational studies if these were included in the original meta-analytic pooled effect estimate.

Assessment of the reliability of conclusions

The reliability of results in apparently conclusive critical care meta-analyses will be expressed in proportions. We will assess the following proportions: true positive (benefit), potentially false positive, true neutral, potentially false neutral, true negative (harm), and potentially false negative (Figure 2).

True positive (benefit)

True positive meta-analyses are meta-analyses in which the cumulative z-curve crosses the TSA monitoring boundary for benefit (Table 2, significance for benefit in column E). This number will be expressed in relation to all conventional positive meta-analyses (the z-curve crosses the conventional z=1.96 boundary for benefit) as well as in relation to all meta-analyses included in this study.

Potentially false positive (benefit)

Potentially false positive meta-analyses are meta-analyses in which the cumulative z-curve crosses the conventional z=1.96 boundary for benefit, but does not cross the TSA monitoring boundary for benefit (Figure 1A; Table 2, significance for benefit in column C but no significance in column E). This number will be expressed in relation to all conventional positive meta-analyses (the z-curve crosses the conventional z=1.96 boundary for benefit) as well as in relation to all meta-analyses included in this study.

These meta-analyses may be considered potentially unreliable. For such meta-analyses, we will also calculate the additional number of participants that may be required to reach the required information size. We recognize that assessment of the number of additional randomized patients needed is part of a dynamic model which needs adaptation when further data accumulates.

True neutral

Truly neutral meta-analyses are meta-analyses in which the cumulative z-curve crosses the TSA monitoring boundary for futility (Table 2, column I). This number will be expressed in relation to all neutral meta-analyses (the z-curve crosses neither the conventional z=1.96 boundary for benefit or harm) as well as in relation to all meta-analyses included in this study.

Potentially false neutral

Potentially false neutral meta-analyses are meta-analyses in which the cumulative z-curve crosses neither the conventional z=1.96 boundary for benefit or harm nor the TSA monitoring boundary for futility (Table 2, column I). This number will be expressed in relation to all neutral meta-analyses (the z-curve crosses neither the conventional z=1.96 boundary for benefit or harm) as well as in relation to all meta-analyses included in this study.

True negative (harm)

True negative meta-analyses are meta-analyses in which the cumulative z-curve crosses the TSA monitoring boundary for harm (Table 2, significance for harm in column E). This number will be expressed in relation to all conventional negative meta-analyses (the z-curve crosses the conventional z=1.96 boundary for harm) as well as in relation to all meta-analyses included in this study.

Potentially false negative (harm)

Potentially false negative meta-analyses are meta-analyses in which the cumulative z-curve crosses the conventional z=1.96 boundary for harm, but does not cross the TSA monitoring boundary for harm (Table 2, significance for harm in column D but no significance for harm in column E). This number will be expressed in relation to all traditionally negative meta-analyses (the z-curve crosses the conventional z=1.96 boundary for harm) as well as in relation to all meta-analyses included in this study.

Figure 1:

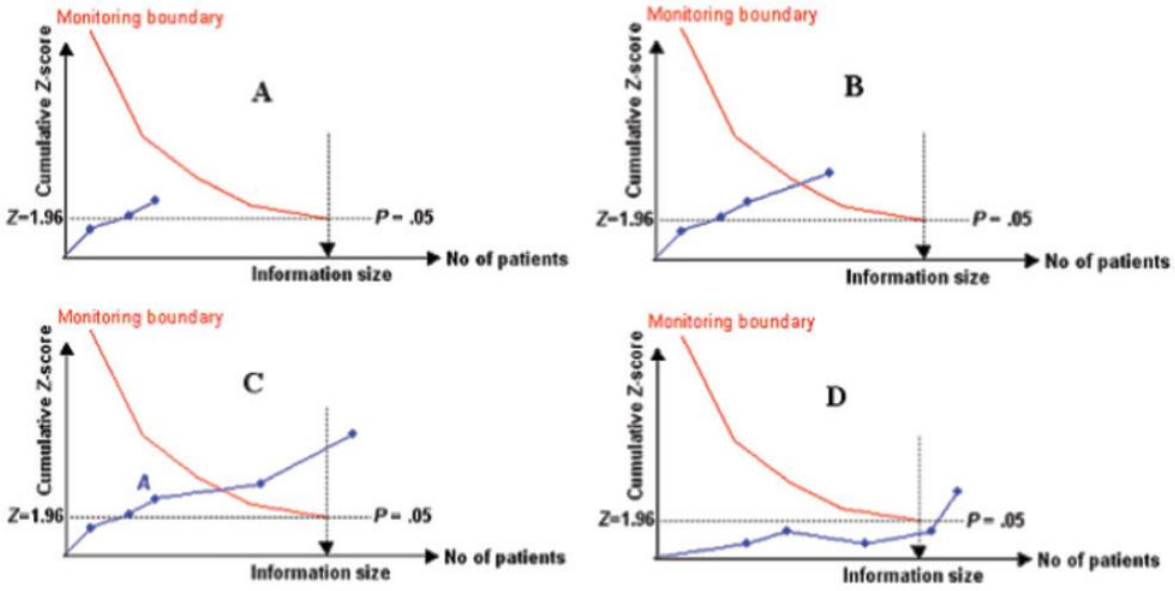

Figure 1

Four examples of TSA. The cumulative Z-curves (blue) were constructed with each cumulative Z-value calculated after including a new trial according to publication date. Crossing of the two-sided Z=1.96 provides a traditionally significant result. Crossing of the trial sequential monitoring boundaries (red) is needed to obtain reliable evidence.

(A) Inconclusive evidence: Number of participants does not reach the information size and the cumulative Z-curve does not cross the monitoring boundary.

(B) Evidence for at least 25% relative risk reduction: Number of participants does not reach the information size, but the cumulative Z-curve does cross the monitoring boundary.

(C) Evidence for at least 25% relative risk reduction: Number of participants does reach the information size and the cumulative Z-curve does cross the monitoring boundary.

(D) Evidence of less than 25% relative risk reduction: The cumulative Z-curve does not cross the monitoring boundary before reaching the information size

Figure 2:

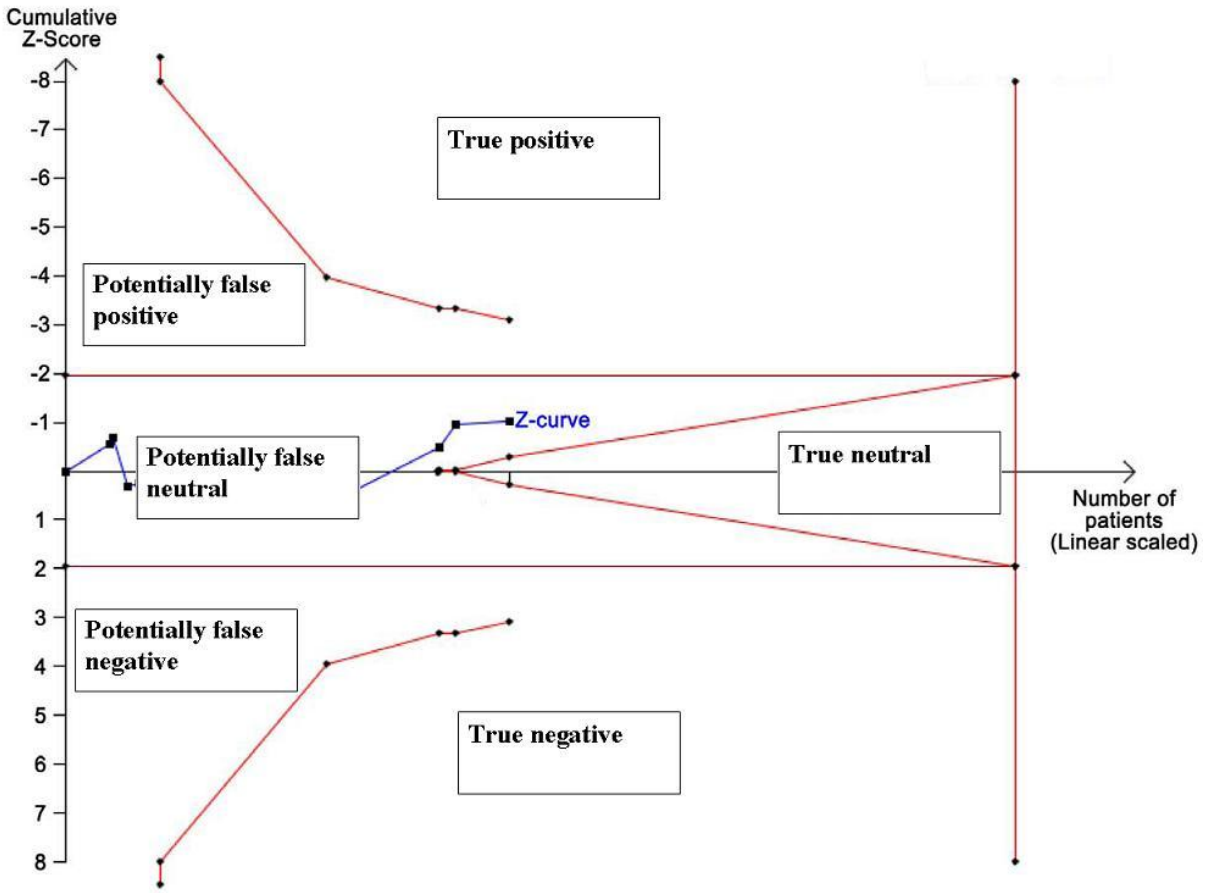

Figure 2

Illustration of possible conclusions based on the cumulative z-curve having crossed or not the conventional z = +/- 1.96 and the TSA monitoring boundaries for benefit, harm and futility.

Appendix 1: HAS TO BE CHECKED BY SARAH

Search Strategy to Identify Meta-analyses pertinent to Critical Care Medicine.

1. PubMed MEDLINE Search

- (*critical* OR intensive* OR intensive care OR intensive care units OR "intensive therapy" OR critically ill OR critical illness OR critical care)*
  OR
- (*cardiotonic agents OR sympathomimetic OR vasoconstrictor agents OR artificial respiration OR mechanical ventilation OR resuscitation OR shock OR multiple organ failure*)
  AND
- *(((meta-analysis [pt] OR meta-analysis [tw] OR metanalysis [tw]) OR ((review [pt] OR guideline [pt] OR consensus [ti] OR guideline* [ti] OR literature [ti] OR overview [ti] OR review [ti]) AND ((Cochrane [tw] OR Medline [tw] OR CINAHL [tw] OR (National [tw] AND Library [tw])) OR (handsearch* [tw] OR search* [tw] OR searching [tw]) AND (hand [tw] OR manual [tw] OR electronic [tw] OR bibliographi* [tw] OR database* OR (Cochrane [tw] OR Medline [tw] OR CINAHL [tw] OR (National [tw] AND Library [tw])))))) OR ((synthesis [ti] OR overview [ti] OR review [ti] OR survey [ti]) AND (systematic [ti] OR critical [ti] OR methodologic [ti] OR quantitative [ti] OR qualitative [ti] OR literature [ti] OR evidence [ti] OR evidence-based [ti]))) BUTNOT (case* [ti] OR report [ti] OR editorial [pt] OR comment [pt] OR letter [pt])[1]*

2. OVID EMBASE and Cochrane Library searches.

An initial search for articles pertinent to critical care was run in all three databases using the strategy:

1. intensive care.mp. or exp Intensive Care/
2. critical care.mp. or exp Critical Care/
3. critical illness.mp. or exp Critical Illness/
4. 1 or 2 or 3

AND a sensitive filter to identify meta-analyses[2]

1. meta-analysis.pt.
2. meta-anal:.tw.
3. metaanal:.tw.
4. quantitativ: review:.tw.
5. quantitativ: overview:.tw.
6. systematic: review:.tw.
7. systematic: overview:.tw.
8. methodologic: review:.tw.
9. methodologic: overview:.tw.
10. review.pt.
11. medline:.tw.
12. 10 and 11
13. 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 12

This search was supplemented by a search of EMBASE using the terms

1. exp TRAUMATIC SHOCK/ or exp HYPOVOLEMIC SHOCK/ or exp BURN SHOCK/ or shock.mp. or exp ANAPHYLACTIC SHOCK/ or exp SHOCK/ or exp

HEMORRHAGIC SHOCK/ or exp SHOCK LUNG/ or exp SEPTIC SHOCK/ or exp CARDIOGENIC SHOCK/

2. resuscitation.mp. or exp RESUSCITATION/

3. multiple organ failure.mp. or exp Multiple Organ Failure/

4. exp Noradrenalin/ or exp Dobutamine/ or exp Inotropic Agent/ or inotrope.mp. or exp Adrenalin/ or exp Dopamine/

5. mechanical ventilation.mp. or exp Artificial Ventilation/

Again combined with the filter to identify meta-analyses[2]

The final search was of the Cochrane Library using the search strategy

1. resuscitation.mp. [mp=title, short title, abstract, full text, keywords, caption text]

2. mechanical ventilation.mp. [mp=title, short title, abstract, full text, keywords, caption text]

3. artificial respiration.mp. [mp=title, short title, abstract, full text, keywords, caption text]

4. inotrope.mp. [mp=title, short title, abstract, full text, keywords, caption text]

5. shock.mp. [mp=title, short title, abstract, full text, keywords, caption text]

6. multiple organ failure.mp. [mp=title, short title, abstract, full text, keywords, caption text]

7. 1 or 2 or 3 or 4 or 5 or 6

8. limit 7 to systematic reviews

Searches were limited to articles published in English and dealing with human subjects published between January 1, 1994 and December 31, 2003 (including any updates).

References for search strategy

1. Shojania KG, Bero LA: Taking advantage of the explosion of systematic reviews: an efficient MEDLINE search strategy. Eff Clin Pract 2001, 4(4):157-162.
2. Hunt DL, McKibbon KA: Locating and appraising systematic reviews. Ann Intern Med 1997, 126(7):532-538.
3. Delaney A, Bagshaw SM, Ferland A, Manns B, Laupland KB, Doig CJ. A systematic evaluation of the quality of meta-analyses in the critical care literature. Crit Care. 2005 Oct 5;9(5):R575-82. Epub 2005 Sep 9. Review.

REFERENCES

1. Young C, Horton R. Putting clinical trials into context. Lancet 2005;366:107–8.
2. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from www.cochrane-handbook.org
3. Dickersin K, Rennie D. Registering clinical trials. JAMA 2003;290:516–23.
4. Humaidan P, Polyzos NP. (Meta)analyze this: Systematic reviews might lose credibility. Nat Med 2012;18:1321.
5. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. PLoS One. 2011;6(10):e25491. doi: 10.1371/journal.pone.0025491. Epub 2011 Oct 18.
6. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. J Clin Epidemiol 1995;48:45–57.
7. Berkey CS, Mosteller F, Lau J, Antman EM. Uncertainty of the time of first significance in random effects cumulative meta-analysis. Control Clin Trials 1996;17:357–71.
8. Pogue J, Yusuf S. Cumulating evidence from randomized trials: utilizing sequential monitoring boundaries for cumulative meta-analysis. Control Clin Trials 1997;18:580–93.
9. Pogue J, Yusuf S. Overcoming the limitations of current meta-analysis of randomised controlled trials. Lancet 1998;351:47–52.
10. Devereaux PJ, Beattie WS, Choi PT, et al. How strong is the evidence for the use of perioperative beta-blockers in non-cardiac surgery? Systematic review and meta-analysis of randomised controlled trials. Br Med J 2005;331:313-21.
11. 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;61:64–75.
12. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659–63.
13. Thorlund K, Devereaux PJ, Guyatt G, et al. Can trial sequential monitoring boundaries reduce spurious inferences from meta-analyses? Int J Epidemiol 2009; 38:276–86.
14. The Cochrane Library, Issue 1, 2013. Chichester: Wiley.
15. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002;21:1539–58.
16. 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.
17. Thorlund K, Imberger G, Johnston BC, Walsh M, Awad T, Thabane L, Gluud C, Devereaux PJ,Wetterslev J. Evolution of heterogeneity (I2) estimates and their 95% confidence intervals in large meta-analyses. PLoS One. 2012;7(7):e39471. doi: 10.1371/journal.pone.0039471. Epub 2012 Jul 25.
18. Fedorov V, Jones B. The design of multicentre trials. Stat Methods Med Res 2005;14:205–48.
19. Lan KKG, DeMets DL. Discrete sequential boundaries for clinical trials. Biometrika 1983;70:659–63.
20. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics 1979;35:549–56.

21. DeMets DL. Methods for combining randomized clinical trials: strengths and limitations. Stat Med 1987;6:341–50.
22. DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986;7:177–88.
23. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004;23:1351-75.