

ORIGINAL ARTICLE

High quality of evidence is uncommon in Cochrane systematic reviews in Anaesthesia, Critical Care and Emergency Medicine

Aaron Conway, Zachary Conway, Kathleen Soalheira and Joanna Sutherland

BACKGROUND The association between the quality of evidence in systematic reviews and authors' conclusions regarding the effectiveness of interventions relevant to anaesthesia has not been examined.

OBJECTIVE The objectives of this study were: to determine the proportion of systematic reviews in which the authors made a conclusive statement about the effect of an intervention; to describe the quality of evidence derived from outcomes in reviews that used the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) working group system for grading the quality of evidence; and to identify review characteristics associated with conclusiveness.

DESIGN Cross-sectional analysis of Cochrane systematic reviews from the Anaesthesia, Critical Care and Emergency Review Group was undertaken.

DATA SOURCES The Cochrane webpage was used to identify reviews for inclusion (<http://ace.cochrane.org/>).

ELIGIBILITY CRITERIA New and updated versions of systematic reviews published up to 17 September 2015 were eligible. Protocols for systematic reviews were excluded.

RESULTS A total of 159 reviews were included. GRADE was used in 103 reviews (65%). Of these, high-level evidence for the primary outcome was identified in 11 reviews (10%). The main reasons that quality of evidence for the primary outcome was downgraded were risk of bias ($n = 44$; 43%) and imprecision ($n = 36$; 35%). Authors of 47% ($n = 75$) of the total number of reviews made conclusive statements about the effects of interventions. Independent predictors of conclusiveness in the subgroup of reviews with GRADE assessments were quality of evidence for the primary outcome (odds ratio 2.03; 95% confidence interval: [1.18 to 3.52]) and an increasing number of studies included in reviews (OR 1.05; 95% CI: [1.01 to 1.09]).

CONCLUSION It was common for conclusive statements to be made about the effects of interventions despite evidence for the primary outcome being rated less than high quality. Improving methodological quality of trials would have the greatest impact on improving the quality of evidence.

Published online 22 August 2017

Introduction

Clinicians rely on high-quality evidence to underpin valid clinical decision-making. One of the widely used sources of evidence is the Cochrane Database of Systematic Reviews. Yet a recent study found that high-quality evidence for medical and health-related interventions was uncommon among all the Cochrane systematic reviews published in 2014.¹ Although this is an important finding, it is unclear whether or not it was specific to a

single year or whether it was more generally applicable to all Cochrane reviews, and especially to the evidence base that has been evaluated by the Cochrane Anaesthesia, Critical Care and Emergency Review Group. Furthermore, the influence of the quality of evidence on the review authors' ability to draw conclusions regarding the effectiveness or lack of effectiveness of an intervention has not been determined. The Grades of

From the Institute of Health and Biomedical Innovation, Kelvin Grove; Queensland University Technology, Brisbane, Queensland (AC, ZC, KS) and Coffs Harbour Health Campus, Coffs Harbour; Rural Clinical School, University of New South Wales, Kensington, New South Wales, Australia (JS)

Correspondence to Aaron Conway, Institute of Health and Biomedical Innovation, Queensland University Technology, 60 Musk Ave, Kelvin Grove, QLD, 4059, Australia
Tel: +61 731386124; e-mail: aaron.conway@qut.edu.au

0265-0215 Copyright © 2017 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Society of Anaesthesiology.

DOI:10.1097/EJA.0000000000000691

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal

Recommendation, Assessment, Development and Evaluation (GRADE) tool can allow a systematic appraisal of the quality of evidence for an outcome included in a systematic review.² GRADE was adopted by Cochrane in 2008.³ The objectives of this study were: to determine the proportion of reviews in which the authors were able to make a conclusive statement about the effects of an intervention; to describe the quality of evidence derived from primary and secondary outcomes in reviews that used the GRADE system for grading the quality of evidence; and to identify review characteristics associated with conclusiveness.

Methods

Design

A cross-sectional analysis of the quality of evidence in Cochrane systematic reviews from the Anaesthesia, Critical Care and Emergency Review Group was undertaken.

Inclusion criteria

New and updated versions of systematic reviews published by the Cochrane Anaesthesia, Critical Care and Emergency Review Group published up to 17 September 2015 were eligible for inclusion in this analysis. Protocols for systematic reviews were excluded.

Outcome definitions

For a review to be conclusive, the authors of the Cochrane systematic review had to have made a definitive statement about the effects of an intervention. As such, this assessment of conclusiveness should be considered subjective; it was based on the author's conclusions section of the abstract and the implications for practice sections of the main text of the review. An example of a statement that was interpreted to be 'conclusive' was 'bispectral index-guided anaesthesia can reduce the risk of intraoperative awareness in surgical patients at high risk for awareness in comparison to using clinical signs as a guide for anaesthetic depth'.⁴

Data sources

The Cochrane Anaesthesia, Critical Care and Emergency Review Group webpage was used to identify reviews for inclusion (<http://ace.cochrane.org/>). The full version of all identified reviews was then accessed for data collection.

Data collection

Data from reviews were extracted onto a standardised form by Z.C. after A.C. first subjected it to a pilot test of 10 reviews. Information on the following aspects of the included reviews was extracted:

Characteristics of the review: year of publication, status of review (new or update), number of studies included, type of studies included [coded as only randomised controlled trial (RCT) or other], number of participants included, discipline area (anaesthesia, critical care, emergency),

type of interventions evaluated (coded as pharmacological, nonpharmacological or medical device).

Quality of evidence: number of outcomes assigned a GRADE rating, GRADE rating of the first listed primary outcome, highest GRADE rating for secondary outcomes, number of outcomes for each GRADE category and reasons for downgrading primary outcome (risk of bias, imprecision, inconsistency, publication bias, indirectness).

Two authors (Z.C. and K.S.) accessed the full version of the reviews and independently evaluated those included for conclusiveness according to the predefined criteria. Disagreements were resolved by discussion with a third author (A.C.).

Statistical analysis

Descriptive statistics were used to calculate frequency and percentages for dichotomous data. Median and interquartile ranges (IQRs) were calculated for continuous data. Backward stepwise logistic regression was used for multivariate analyses. The final model was determined using a removal probability of 0.10. As this study was concerned with the interface between the authors' decision-making regarding conclusiveness about the evidence for an effect of interventions and the quality of evidence available to inform those decisions, we included only the following variables in the multivariate analysis: the number of studies included in each review; the total number of participants included in each review; the number of outcomes assigned a GRADE rating; the quality of evidence for the primary outcome and the highest quality of evidence for a secondary outcome. Reviews that did not assign a GRADE rating were not included in the multivariate analysis. The review category (anaesthesia, critical care, emergency) or intervention category (pharmacological, nonpharmacological, medical device) were not included. Quality of evidence was included in the analyses as a numerical variable. Accordingly, this assumes that the distances between the quality levels were equal and relevant (i.e. a one unit change from very low quality to low quality is equal to a change from moderate to high quality). Very low-quality evidence was coded as 1, low-quality evidence coded as 2, moderate-quality evidence coded as 3 and high-quality evidence coded as 4. The Hosmer–Lemeshow goodness of fit test was used to assess model fit. A *P* value of 0.05 was the threshold accepted for statistical significance.

Results

A total of 159 reviews were included. Table 1 presents the review characteristics. There were 83 reviews (52%) categorised as being centred on anaesthesia. About 64 reviews (40%) were categorised as critical care. A smaller number of reviews were for emergency medicine ($n = 12$; 8%). Pharmacological therapy was the most common type of intervention ($n = 98$; 62%). The median number of

Table 1 Review characteristics categorised by conclusiveness

Characteristic	All <i>n</i> = 159 <i>n</i> (%) or median (IQR)	Conclusive <i>n</i> = 75 <i>n</i> (%) or median (IQR)	Inconclusive <i>n</i> = 84 <i>n</i> (%) or median (IQR)
Specialty area			
Anaesthesia	83 (52)	45 (60)	38 (45)
Critical care	65 (41)	27 (36)	38 (45)
Emergency	11 (7)	3 (4)	8 (10)
Intervention category			
Pharmacological therapy	98 (62)	49 (66)	49 (58)
Nonpharmacological	15 (9)	7 (9)	8 (10)
Medical device	46 (29)	19 (25)	27 (32)
Number of included studies	10 (4, 21)	14.5 (6, 28)	6 (2, 15)
Only randomised trials included	135 (85)	66 (88)	69 (85)
Total number of included participants	1046 (402, 2267)	1368 (614, 2833)	759 (261, 1738)

studies included in the reviews was 10 (IQR 4 to 21). Only RCTs were included in the majority of reviews ($n = 135$; 85%). One of the reviews was an overview of Cochrane reviews and one was a diagnostic test accuracy review. The median number of participants in the reviews was 1046 (IQR 402 to 2267).

Quality of evidence

A total of 103 reviews (65%) used the GRADE system to evaluate the quality of evidence. A summary of the quality of evidence ratings for primary and secondary outcomes is provided in Tables 1–3. Out of these reviews, only a small number identified high-level evidence for the first listed primary outcome ($n = 11$; 10%). There was moderate-quality evidence for the first listed primary outcomes of 35 reviews (34%), low-quality evidence in 42 reviews (41%) and very low-quality evidence in 15 reviews (15%). In total, 16 reviews had a secondary

outcome with high-level evidence (18%). A larger proportion had moderate ($n = 27$; 30%) and low ($n = 33$; 37%) level evidence. Secondary outcomes were graded as very low for 13 reviews (15%). A summary of GRADE ratings for primary and secondary outcomes across specialty areas and intervention categories is presented in Table 4. The level of evidence for the first listed primary outcome was downgraded because of risk of bias in 44 reviews (43%), for imprecision in 36 reviews (35%), inconsistency in 18 reviews (18%), publication bias in 18 reviews (18%) and indirectness in seven reviews (7%). A summary of the reasons for downgrading GRADE ratings across specialty areas and intervention categories is presented in Table 5.

Conclusiveness of the evidence

In 75 reviews (47%), we judged that the review authors made a conclusive statement about the effects of an intervention. On univariate analysis, a higher number of studies were included in conclusive reviews (median 14.5 in conclusive and six in inconclusive reviews; $P < 0.001$) as was the total number of participants (median 1368 in conclusive and 759 in inconclusive reviews; $P = 0.002$). The quality of evidence for the primary (16% conclusive and 6% inconclusive) and secondary outcomes (24% conclusive and 10% inconclusive) was graded as higher in conclusive reviews. Results of the multivariate analysis are presented in Table 6. Quality of evidence for the primary outcome was an independent predictor of conclusiveness (OR 2.03; 95% CI: [1.18 to 3.52]). Authors of reviews were 5% more likely to have made a conclusive statement about the effects of an intervention with each additional included study (OR 1.05; 95% CI: [1.01 to 1.09]). Variables not significantly associated with conclusiveness in the multivariate model were the number of outcomes assigned a GRADE rating, the quality of

Table 2 Quality of evidence for primary outcome categorised by conclusiveness

Characteristic	All <i>n</i> = 103 <i>n</i> (%) or median (IQR)	Conclusive <i>n</i> = 51 <i>n</i> (%) or median (IQR)	Inconclusive <i>n</i> = 52 <i>n</i> (%) or median (IQR)
Number of outcomes assigned a GRADE rating	6 (4, 7)	6 (4, 7)	5 (3, 6)
High-quality evidence for primary outcome	11 (11)	8 (16)	3 (6)
Moderate-quality evidence for primary outcome	35 (34)	23 (45)	12 (23)
Low-quality evidence for primary outcome	42 (40)	17 (33)	25 (48)
Very low-quality evidence for primary outcome	15 (15)	3 (6)	12 (23)
Primary outcome downgraded because of risk of bias	44 (43)	18 (36)	26 (50)
Primary outcome downgraded because of imprecision	36 (35)	11 (22)	25 (48)
Primary outcome downgraded because of inconsistency	18 (18)	8 (16)	10 (19)
Primary outcome downgraded because of publication bias	18 (18)	11 (22)	7 (13)
Primary outcome downgraded because of indirectness	7 (7)	2 (4)	5 (10)

GRADE, Grades of Recommendation, Assessment, Development and Evaluation.

Table 3 Quality of evidence for secondary outcomes categorised by conclusiveness

Characteristic	All <i>n</i> = 89 <i>n</i> (%)	Conclusive <i>n</i> = 46 <i>n</i> (%)	Inconclusive <i>n</i> = 43 <i>n</i> (%)
High-quality evidence for any secondary outcome	16 (18)	11 (24)	5 (11)
Moderate-quality evidence for any secondary outcome	27 (30)	19 (41)	8 (19)
Low-quality evidence for any secondary outcome	33 (37)	14 (31)	19 (44)
Very low-quality evidence for any secondary outcome	13 (15)	2 (4)	11 (26)

Table 4 Summary of Grades of Recommendation, Assessment, Development and Evaluation ratings ($n = 103$)

Quality of evidence		Specialty area				Intervention category	
		Anaesthesia $n = 55$	Critical care $n = 42$	Emergency $n = 6$	Medication $n = 62$	Medical device $n = 30$	Nonpharmacological $n = 11$
First listed primary outcome	High	5 (9)	4 (9)	2 (33)	7 (11)	3 (10)	1 (9)
	Moderate	18 (33)	16 (38)	1 (17)	23 (37)	11 (36)	2 (20)
	Low	26 (47)	15 (36)	1 (17)	22 (36)	13 (43)	6 (54)
	Very low	6 (11)	7 (17)	2 (33)	10 (16)	3 (10)	2 (18)
Secondary outcome ^a		$n = 46$	$n = 39$	$n = 4$	$n = 51$	$n = 29$	$n = 9$
	High	6 (13)	10 (26)	0 (0)	7 (14)	8 (27)	1 (11)
	Moderate	14 (30)	12 (30)	1 (25)	20 (39)	6 (21)	1 (11)
	Low	22 (48)	10 (26)	1 (25)	13 (25)	13 (45)	7 (78)
	Very low	4 (9)	7 (18)	2 (50)	11 (22)	2 (7)	0 (0)

Data are given as n (%). ^aDenominator is different from primary outcome because not all reviews that used Grades of Recommendation, Assessment, Development and Evaluation included a secondary outcome.

Table 5 Reasons for downgrading Grades of Recommendation, Assessment, Development and Evaluation ratings for primary outcomes

Reasons	Specialty area					
	Anaesthesia $n = 50$	Critical care $n = 38$	Emergency $n = 6$	Medication $n = 55$	Medical device $n = 27$	Nonpharmacological $n = 10$
Risk of bias	23 (46)	18 (47)	3 (50)	23 (42)	14 (52)	7 (70)
Imprecision	16 (32)	17 (45)	3 (50)	17 (31)	14 (52)	5 (50)
Inconsistency	7 (14)	10 (26)	1 (17)	8 (15)	7 (26)	3 (30)
Publication bias	6 (12)	10 (26)	2 (33)	10 (18)	7 (26)	1 (10)
Indirectness	3 (6)	2 (5)	2 (33)	6 (11)	0 (0)	1 (10)

Data are given as n (%). Columns totals are the number of reviews that used Grades of Recommendation, Assessment, Development and Evaluation and downgraded the quality of evidence. Frequencies and percentages in each column are not cumulative because a primary outcome could have been downgraded for more than one reason.

Table 6 Independent predictors of conclusiveness for reviews that used Grades of Recommendation, Assessment, Development and Evaluation

Variables	OR	Significance	95% CI for OR	
			Lower	Upper
Quality of evidence for the primary outcome	2.03	0.011	1.18	3.52
Total number of studies in the review	1.05	0.005	1.01	1.09
Number of outcomes assigned a GRADE rating	1.21	0.062	0.99	1.47

Variables removed from model: secondary outcome quality of evidence and total number of participants. CI, confidence interval; GRADE, Grades of Recommendation, Assessment, Development and Evaluation; OR, odds ratio.

evidence for secondary outcomes and total number of participants.

Discussion

We found that only one in 10 Cochrane reviews in Anaesthesia, Critical Care and Emergency Medicine that used the GRADE approach graded the quality evidence for a primary outcome as high. As such, clinicians do not have firm evidence to support the effectiveness of a large number of interventions across these medical specialty areas. An even lower proportion of Cochrane reviews with high-quality evidence was identified in the field of orthodontics (2%).⁵ Findings from previous analyses of Cochrane reviews in other medical specialties have varied, with some reporting a large proportion of conclusive reviews,^{6,7} whereas others were more likely to be inconclusive.^{8,9} The total number of studies and total number of included participants were associated with review conclusiveness.⁶⁻⁹ It is important to note that these

previous studies of the conclusiveness of Cochrane reviews did not evaluate reviewers' judgements about the quality of evidence. Examinations of non-Cochrane reviews have concentrated on the methodological quality of the reviews instead of an analysis of the conclusions.¹⁰ We were unable to locate a study that centred on the conclusiveness of non-Cochrane reviews in the literature.

As would reasonably be expected, the likelihood that a review was conclusive increased with the number of studies it included and its quality of evidence for the primary outcome. However, the small proportion of reviews in which there was high-quality evidence for the primary outcome (10%) and secondary outcomes (16%) is noteworthy and consistent with findings of a broader review of evidence in Cochrane reviews.¹ It indicates that many reviewers have drawn conclusions about the effects of interventions based on uncertain effect estimates that may change with results from further research.² It is possible that this could have negative

implications for clinical practice if further research indicates an intervention is not effective or if it is found to be harmful. Improving the quality of the design and reporting of RCTs in Anaesthesia, Critical Care and Emergency Medicine should be considered a high priority in addressing this problem; we have identified that the most common reason for downgrading the quality of evidence was risk of bias.

One benefit of a systematic review, which arises from accumulation of data from multiple studies, is increased statistical power to detect the effect of an intervention. The available evidence was still insufficient for a large number of primary outcomes examined by the Cochrane reviews included in our analysis. In line with previous studies of Cochrane reviews,¹¹ we identified that the quality of evidence for a large number of primary outcomes was downgraded because of imprecision. We did not investigate reasons for imprecision in our analysis. It may be that, for pragmatic reasons, surrogate outcomes were used for sample size estimation of RCTs for a specific intervention but it was then appropriately decided that clinical outcomes should be examined as the primary outcome of a Cochrane systematic review. For this reason, although we found that a large number of primary outcomes were downgraded because of imprecision, this should not necessarily be seen as a limitation.

Downgrading because of risk of publication bias was not common (18%) in the sample of Cochrane reviews that we analysed. This was consistent with results of a recent evaluation of publication bias reported in systematic reviews and meta-analyses published in anaesthesiology journals (16%; $n = 34$).¹² Of note, these authors identified that there was a greater likelihood of publication bias among reviews not performing these evaluations. We did not extract data about how publication bias was assessed in each review. However, assessment of reporting bias is a core component of Cochrane review methods and guidance on how to detect publication bias specifically is addressed in the handbook.³ Therefore, it could be assumed that the quality of evidence for primary outcomes that required downgrading for publication bias was accurate.

Downgrading the quality of evidence for indirectness and inconsistency was not as common in the sample of Cochrane reviews that we analysed. This may indicate that Cochrane reviews in these fields were highly targeted at interventions for particular patient groups.

A finding worthy of further discussion is that conclusive statements about the effects of interventions were made in similar proportions of reviews that did ($n = 51$; 49%) and did not ($n = 52$; 51%) use the GRADE approach. One interpretation is that the GRADE approach may not play a relevant role when reviewers are crafting their statements about the effects of interventions. Conversely, this could also be interpreted that further efforts are required

to assist review authors to apply GRADE assessments in their interpretations about the evidence. However, it should be noted that an evaluation of the impact of using GRADE or not on review authors' interpretation as to whether an intervention may be of value for clinical practice was not a specific aim of this study. A comparison of Cochrane reviews that use GRADE with non-Cochrane reviews that may or may not have used GRADE at the same time period on the same topic would be a more appropriate way to investigate the potential benefits of using GRADE in drawing conclusions about the effects of interventions.

Our finding that it was common for review authors to make a conclusive statement about the effects of an intervention when there was less than high-quality evidence for the primary outcome suggests that the conclusiveness of a systematic review may not be reducible to the quality of evidence for a single primary outcome. This conflicts with general recommendations in the Cochrane Handbook that conclusions about the effects of interventions should largely be based on the primary outcomes.¹³ The rationale behind basing conclusions about the effects of interventions mainly on results of primary outcomes in a systematic review is not as clear as it is for RCTs. In a RCT, the primary outcome is the outcome used to calculate the sample size.¹⁴ It is not typical for a sample size calculation to be conducted for the primary outcome of a systematic review.

A further relevant issue is the potential for subjectivity in GRADE assessments. It is recognised that a potential drawback of GRADE assessments is its complexity and consequent potential for inconsistency in judgements between review authors. There have been inconsistent findings in studies that examined the agreement in GRADE assessment between multiple reviewers.^{15,16} The impact of this subjectivity in GRADE assessments on review authors' interpretations of the evidence in the sample of reviews we examined is unknown. Preliminary results of research into the effectiveness of automating quality assessment for systematic reviews are promising and could represent a potential solution to this problem.¹⁷

Limitations

We did not register a protocol for this cross-sectional study of Cochrane reviews. It is also important to consider the implications of the subjectivity of our assessment of the conclusiveness of the reviews included in our study. Although two authors evaluated all reviews for conclusiveness and a third author resolved initial discrepancies, it is possible that other readers of the included reviews may interpret the review authors' statements differently. We included GRADE ratings in the multivariate analysis as a numerical variable. There is no supporting evidence available to confirm that the difference in quality of evidence for an outcome is equal between each quality level. As variables related to GRADE assessment were

included in the multivariate analysis of predictors of review conclusiveness, all reviews without GRADE assessments were excluded. Although this must be considered a limitation, we considered that the results are informative for the contemporary context, because all Cochrane reviews must now include a GRADE assessment of outcomes.¹³ Further, we only studied reviews published in the Cochrane database of systematic reviews so the results cannot be generalised to non-Cochrane systematic reviews. The variables we selected for inclusion in our multivariable logistic regression analysis may not have encompassed all the factors that influenced review authors' interpretations about the conclusiveness of the evidence for the use of an intervention. For example, statistical significance may be considered as a variable relevant in the review of authors' interpretations about the evidence of an effect of an intervention. However, Fleming *et al.*¹ identified that none of the authors of the Cochrane and non-Cochrane reviews made a favourable interpretation of the evidence in the absence of a statistically significant result. Therefore, adding this variable to our analysis would probably not increase our understanding of factors that contribute to review authors' interpretations of evidence.

Conclusion

High quality of evidence, according to the GRADE approach, was uncommon in the sample of Cochrane systematic reviews in Anaesthesia, Critical Care and Emergency Medicine that we analysed. We identified that authors of many of the systematic reviews made conclusive statements about the effects of interventions based on very low, low and moderate-quality evidence. In the subgroup of conclusive reviews, only 16% had high-quality evidence available to support the primary outcome. These are important findings considering that there could be negative implications for patient outcomes from concluding that an intervention is superior, inferior or equivalent to an alternative based on evidence that is not high quality. Improving methodological quality of trials in these medical disciplines would have the greatest impact on improving the quality of evidence.

Acknowledgements relating to this article

Assistance with the analyses: none.

Financial support and sponsorship: none.

Conflicts of interest: none.

Presentation: none.

References

- Fleming PS, Koletsi D, Ioannidis JP, *et al.* High quality of the evidence for medical and other health-related interventions was uncommon in Cochrane systematic reviews. *J Clin Epidemiol* 2016; **78**:34–42.
- Guyatt GH, Oxman AD, Vist GE, *et al.*, GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008; **336**:924–926.
- Higgins JP, Green S, The Cochrane Collaboration. *Cochrane handbook for systematic reviews of interventions*. Hoboken: Wiley-Blackwell; 2008.
- Punjasawadwong Y, Phongchiewboon A, Bunchungmongkol N. Bispectral index for improving anaesthetic delivery and postoperative recovery. *Cochrane Database Syst Rev* 2014;CD003843.
- Pandis N, Fleming PS, Worthington H, *et al.* The quality of the evidence according to GRADE is predominantly low or very low in oral health systematic reviews. *PLoS One* 2015; **10**:e0131644.
- Cohen S, Mandel D, Mimouni F, *et al.* Conclusiveness of the Cochrane reviews in nutrition: a systematic analysis. *Eur J Clin Nutr* 2014; **68**: 143–145.
- Yin S, Chuai Y, Wang A, *et al.* Conclusiveness of the Cochrane reviews in gynaecological cancer: a systematic analysis. *J Int Med Res* 2015; **43**:311–315.
- Zhang X, Wu Z, Zhao H, *et al.* Conclusiveness of the Cochrane reviews in palliative and supportive care for cancer. *Am J Hosp Palliat Care* 2017; **34**:53–56.
- Mandel D, Littner Y, Mimouni FB, *et al.* Conclusiveness of the Cochrane neonatal reviews: a systematic analysis. *Acta Paediatr* 2006; **95**: 1209–1212.
- Conway A, Inglis SC, Chang AM, *et al.* Not all systematic reviews are systematic: a meta-review of the quality of systematic reviews for noninvasive remote monitoring in heart failure. *J Telemed Telecare* 2013; **19**:326–337.
- 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**:e59202.
- Hedin RJ, Umberham BA, Detweiler BN, *et al.* Publication bias and nonreporting found in majority of systematic reviews and meta-analyses in anesthesiology journals. *Anesth Analg* 2016; **123**:1018–1025.
- Higgins J and Green S. *Cochrane Handbook for Systematic Reviews of Interventions* version 5.1 [Updated March 2011]. The Cochrane Collaboration: Available from www.cochrane-handbook.org. 2011.
- Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010; **8**:18.
- Hartling L, Fernandes RM, Seida J, *et al.* From the trenches: a cross-sectional study applying the GRADE tool in systematic reviews of healthcare interventions. *PLoS One* 2012; **7**:e34697.
- Mustafa RA, Santesso N, Brozek J, *et al.* The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013; **66**:736–742.
- Llewellyn A, Whittington C, Stewart G, *et al.* The use of Bayesian networks to assess the quality of evidence from research synthesis: 2. Inter-rater reliability and comparison with standard GRADE assessment. *PLoS One* 2015; **10**:e0123511.