A brief overview of systematic reviews and meta-analyses

Kamath Sriganesh[1], Harsha Shanthanna[2], Jason W Busse[2,3,4]

[1]Department of Neuroanaesthesia, National Institute of Mental Health and Neurosciences, Bengaluru, Karnataka, India, Departments of [2]Anesthesia and [3]Clinical Epidemiology and Biostatistics, McMaster University, [4]The Michael G. DeGroote Institute for Pain Research and Care, McMaster University, Hamilton, Ontario, Canada

Address for correspondence:
Dr. Kamath Sriganesh, Department of Neuroanaesthesia, National Institute of Mental Health and Neurosciences, Faculty Block, 3rd Floor, Hosur Road, Bengaluru - 560 029, Karnataka, India.
E-mail: drsri23@gmail.com

ABSTRACT

Systematic reviews (SRs) are performed to acquire all evidence to address a specific clinical question and involve a reproducible and thorough search of the literature and critical appraisal of eligible studies. When combined with a meta-analysis (quantitatively pooling of results of individual studies), a rigorously conducted SR provides the best available evidence for informing clinical practice. In this article, we provide a brief overview of SRs and meta-analyses for anaesthesiologists.

Key words: Evidence-based medicine, meta-analysis, systematic review

INTRODUCTION

The first tenet of evidence-based medicine (EBM) is that clinical decisions should be influenced by all relevant high-quality evidences, as opposed to select studies.[1] Systematic reviews (SRs) are aimed at acquiring all evidence to address a specific research question and involve a reproducible and thorough search of the literature and critical evaluation of eligible studies. This is different than narrative reviews, which allow authors to highlight the findings from select studies. The salient features and differences between systematic and narrative reviews are described in Table 1. There are several other types of reviews such as scoping reviews, rapid reviews, systematised reviews, overviews and critical reviews, which have their own strengths and limitations.[2] For the purposes of this overview, we will focus on SRs of therapeutic interventions.

DEFINITIONS OF SYSTEMATIC REVIEW AND META-ANALYSIS

A rigorously performed SR identifies all empirical evidence that meets pre-specified eligibility criteria to answer a specific clinical question using explicit, systematic methods to minimise bias and provides reliable findings to inform evidence-based clinical care.[3] A SR can be either qualitative, in which eligible studies are summarised, or quantitative (meta-analysis) when data from individual studies are statistically combined. Not all SRs may result in meta-analyses. Similarly, not all meta-analyses may have been preceded by a SR, though this element is essential to ensure that findings are not affected by selection bias.

PURPOSE OF SYSTEMATIC REVIEWS AND ITS POSITION IN THE HIERARCHY OF EVIDENCE

Large volumes of research are conducted and published every year, and at times with conflicting

Table 1: Differences between a narrative review and a systematic review			
Elements of review	**Narrative review**	**SR**	**Advantages of SR**
Scope of clinical or research question	Generally broad (overview of the topic)	Generally specific, with a research question in the PICO* format	Provides focused answer
Protocol	Absent	Often developed and can be registered	Avoids reporting bias
Source of evidence and search strategy	Not usually specified, selective search	Specifies the literature databases that were searched, and provides the search strategy	Addresses selection bias
Selection of articles	Not clearly stated, often selective	Uses a predefined criteria for inclusion and exclusion and includes all relevant evidence	Reproducible by anyone using similar methods
Evaluation of evidence	Subjective or absent	Critical appraisal of risk of bias of individual studies, and overall assessment of strength of evidence by outcome	Transparent reporting of study quality, and confidence of the evidence (e.g., very low, low, moderate, high) to inform clinical decision-making
Synthesis of evidence	Qualitative summary	Mostly quantitative (meta-analysis), when data can be pooled	Combines evidence and provides precision associated with treatment effects
Authorship	Usually by invitation to an individual with clinical expertise	A team with content and methodological expertise, at least 2 independent reviewers and librarian	Minimises selection bias, enhances quality of review and reports overall evidence

*PICO – Population, intervention, control, outcome. SR – Systematic reviews

results.[4] Differences in findings lead to confusion among physicians, patients and policymakers. Resolving such situations necessitates a comprehensive and transparent synthesis of all available literature that evaluates the quality of evidence and when present, explores possible reasons for discrepancies. SRs are best suited to address such a need, especially when combined with meta-analysis. The second tenet of EBM is that there is a hierarchy of evidence, from research designs at higher risk of bias, to study designs at lower risk of bias.[5] Because of the advantages described in Table 1, well-conducted SRs of high quality randomised controlled trials (RCTs) are placed at the highest level in the hierarchy of evidence.[6]

COMPONENTS OF SYSTEMATIC REVIEW

Conduct of a SR starts with a clinical question and a protocol with clear identification of objectives, outcomes and a well-laid out plan for analysis and reporting of results. It is highly desirable to register protocols in a publicly available database, such as PROSPERO.[7] This allows for transparency and increases confidence among readers in the study results. The stages in the conduct of SRs are highlighted in the following section and a summary is provided in Table 2.

Formulating a clinical question
A specific and clearly thought out question is the first step towards completing a SR. The question is generally framed in the population, intervention, comparator and outcomes (PICO) format and

influences the scope of review.[8] A review with a narrow definition of PICO elements (e.g., effect of dalteparin on deep vein thrombosis (DVT) prophylaxis after hip replacement surgery) will apply to a much more specific patient population than a review with broader criteria (e.g., effect of low molecular weight heparin medications on DVT prophylaxis after joint replacement surgeries).

The decision on the type of study design(s) to be included, RCTs or observational studies or both, is made based on the review question. In SRs of therapeutic interventions, RCTs are desirable over observational studies. Observational study designs are considered when there are no RCTs, or the available RCTs are of poor quality or if the review question pertains to rare events. It is not uncommon for SRs to include only English-language articles, but this risks selection bias (English-language bias) and including non-English language articles is highly preferable.[9]

A detailed review protocol that identifies study objectives, outcomes, methods of study selection, data extraction, analysis and reporting should be prepared before the conduct of the review.[10] This promotes more rigorous SRs as researchers are required to carefully think through the details of their study before proceeding, and ensures that any deviation from the intended methodology is transparent to stakeholders. A research team consisting of one or more content experts, a medical librarian, a methodologist and a statistician, along with other

Table 2: Summary of practical steps in the conduct of a systematic review

Essential steps

Formulating an answerable research question

  Use the PICO format to frame the question which is neither too broad nor too specific (e.g., does inhalational anaesthesia result in higher post-operative cognitive dysfunction compared to intravenous anaesthesia in elderly patients undergoing urological surgery?)

  Identify team who will be part of this review (content expert, methodologist, experienced librarian, statistician, pairs of reviewers, review coordinator)

  Prepare review protocol detailing the planned process and register (PROSPERO) before the review begins

Developing a literature search strategy

  Involve an experienced librarian to select appropriate search words, perform database searches using relevant filters and manage search result

  Provide appendix detailing search strategy and results

Selection of relevant studies

  Define inclusion and exclusion criteria for studies based on your PICO format question

  Screen title and abstracts independently, guided by an instruction manual, after pre-testing among reviewers

  Obtain full-texts of all promising studies identified through title and abstract screening, and evaluate for inclusion independently as per predefined criteria

  Maintain record of exclusions with reasons for settling disagreements

Data extraction

  Design and pre-test extraction form among reviewers to evaluate ease of use and extraction of all relevant data

  Achieve consensus for conflicts

Appraising the quality of selected studies

  Use standard checklists/scale to evaluate quality of all included studies

Data synthesis and meta-analysis

  Tabulate results of individual studies, explain excluded studies

  Evaluate and plan suitability for meta-analysis or qualitative synthesis

  Examine for and attempt to explain heterogeneity by subgroup or sensitivity analyses

Reporting the findings

  Adhere to the PRISMA guidelines while reporting results

  Evaluate, grade and report strength of evidence for each reported outcome

  Provide flow diagram depicting the flow of studies in the review

PICO – Population, intervention, control, outcome; PRISMA – Preferred reporting items for systematic review and meta-analysis

stakeholders (e.g., patients) ensures the requisite skill set for design, conduct and reporting of the review.

Searching the literature

This process is ideally conducted by a research librarian. If unavailable, a librarian's assistance should be sought to design and complete the literature search. Common medical databases for searching include Medline, Embase and the Cochrane Central Register of Controlled Trials (CENTRAL); however, based on the research question, additional databases may be appropriate. The search strategy for a SR is based on the research question (PICO) and may vary slightly based on database-specific indexing practices; ideally, all dataset-specific search strategies will be included as appendices when SRs are published.[10]

Literature search strategies will typically use a combination of medical subject heading (MeSH) terms and free-text keywords. For example, if we are interested in the concept of anaesthesia through the breathing of gas, the MeSH term for this concept is 'Anaesthesia, Inhalation,' which will capture publications that have had this subject heading assigned to them by a database indexer. In addition, we might search in the title and abstract field of citations for a specific free-text term. For example, a free-text phrase we might use is 'an?esthe* adj3 (inhal* or insufflation).' The question mark will capture British or American spelling variations while the asterix will capture all variations of the root word (inhaled, inhalation, inhaling, etc.). The adjacency operator (adj3) requires that the two words appear within three words of one another so that an article containing a phrase such as 'anaesthesia delivered by inhalation' will be captured. By combining the results of the two sets of searches using the "OR" operator, a larger set of potentially relevant records is retrieved. A balance between sensitivity (capturing all possible evidence pertaining to review question) and specificity (retrieving more relevant studies) is necessary to achieve a comprehensive and practical search. After removal of duplicates, a final list of citations (titles and abstracts) from all reviewed databases is compiled.

Selection of relevant studies

Based on the selection criteria, the articles obtained through literature search are taken through a two-stage process of screening and selection. The first stage is carried out on titles and abstracts in which a more sensitive strategy is adopted to clearly exclude unrelated articles. For the second stage, all potentially relevant studies are acquired in full-text to complete final adjudication of eligibility. Both these stages should ideally be performed independently and in duplicate by teams of reviewers to provide assurances regarding reliability. Discrepancies in selection are resolved by arbitration. Consensus exercises between reviewers at the start of each stage help promote a high rate of agreement. The kappa statistic can be calculated to inform agreement between reviewers, and is interpreted as follows: Almost perfect agreement (0.81–0.99), substantial agreement (0.61–0.80),

moderate agreement (0.41–0.60), fair agreement (0.21–0.40), slight agreement (0.01–0.20) and chance agreement (0).[11] At the end of this process, a final list of articles to be included within the SR is obtained.

Data extraction
This stage involves extraction of individual study data necessary for the analysis and reporting of the review. Extractable data will typically include study design characteristics, risk of bias (RoB) items, participant demographic information and follow-up and study outcomes. Similar to screening, this is performed independently by paired reviewers to avoid errors in the acquisition of data. Clear organisation and recording of the extracted data are necessary to facilitate a review. Missing or unclear data items should be sought from the authors of studies and if unavailable, should be clearly documented. The use of web-based SR software such as DistillerSR™ (https://distillercer.com/), particularly for large reviews, helps promote faster, more organised and efficient study selection and data extraction.

Appraisal of study reporting quality
The reporting quality of a study considers the extent to which aspects of a study's design and conduct are likely to protect against systematic error (deviation from the truth). Systematic errors broadly fall into three domains of selection, measurement and confounding biases. Several tools are available for assessment of the methodological quality of included studies such as the Newcastle-Ottawa scale for non-randomised studies,[12] and the Cochrane RoB tool for RCTs.[13] Quality assessment is required for interpreting results and for grading the strength of evidence. An example of assessing and representing the RoB in individual studies and across studies is available in the cited reference.[14]
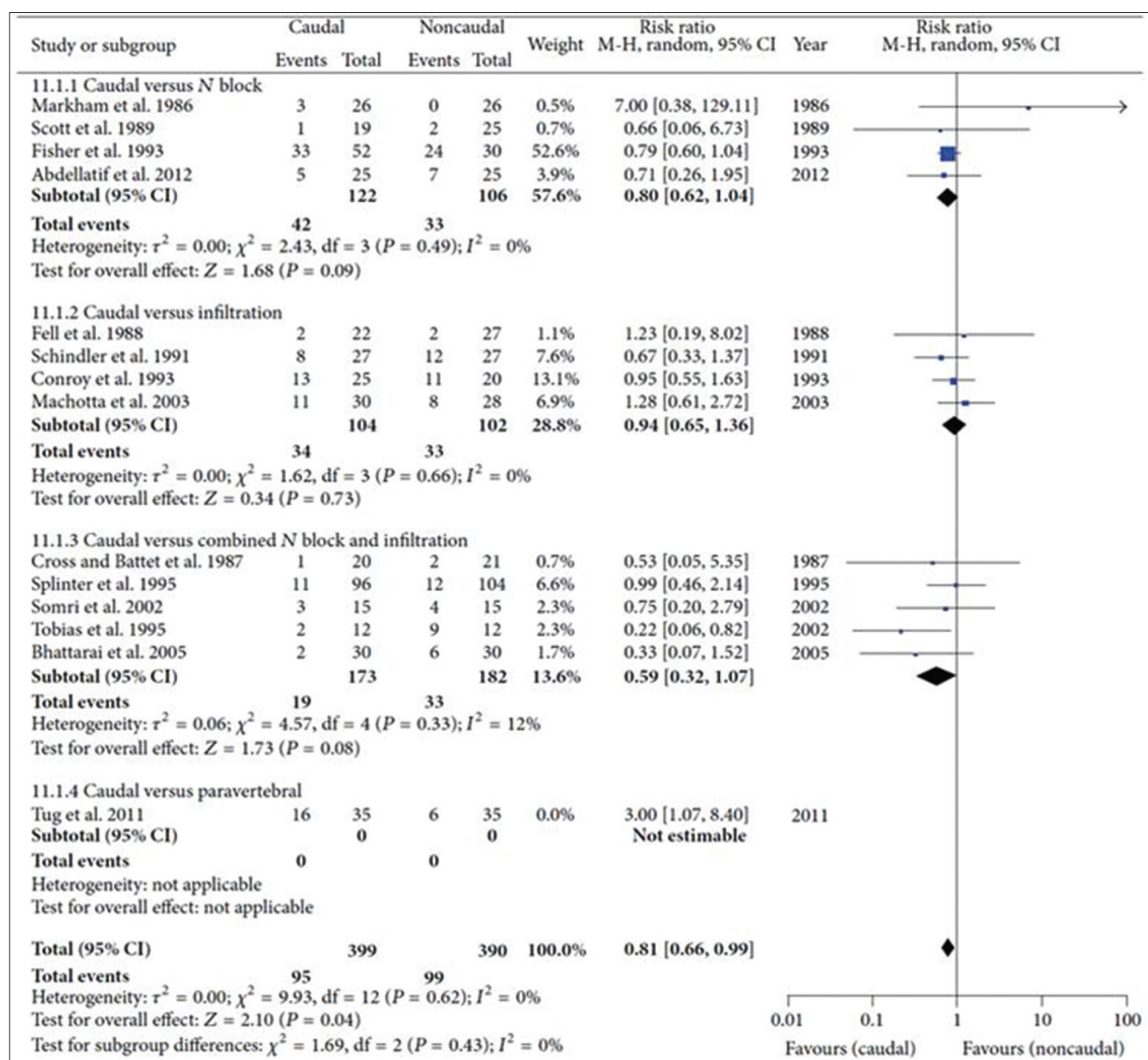
Data synthesis (meta-analysis)
It is desirable to explicitly state how data will be synthesised a priori to avoid data mining, and to reduce unsubstantiated claims during the synthesis.[15] Different study designs and outcome measures should be summarised and synthesised separately.

Meta-analysis is a statistical technique used for combining the effect on a common outcome domain (e.g., pain relief) across individual studies. By combining study results, the precision associated with a treatment effect becomes narrower.[3] Findings from meta-analyses are graphically depicted in the form a forest plot [Figure 1],[14] with individual studies and their treatment effect represented horizontally. The vertical line indicates the line of null effect, and effects either to the left or right of the line indicate the benefit of the intervention or control therapy. Individual study effect sizes are shown as a square along with the associated 95% confidence interval, and the pooled estimate is shown as a diamond.

For continuous outcomes that are reported using the same instrument across trials, the pooled estimate can be reported in natural units as a weighted mean difference. The more common situation is that trials eligible for meta-analysis will report outcomes tapping into the same domain (e.g., pain) using different instruments. In such cases, the effect estimates require standardisation before pooling. A common approach is to convert different instruments into standard deviation units, and report the pooled effect as a standardised mean difference (SMD) (a value of 0.2 represents a small difference, 0.5 a medium difference and 0.8 a large difference in treatment effect). However, the SMD is difficult for patients and clinicians to interpret and is vulnerable to baseline heterogeneity of trial patients: greater heterogeneity among pain scores at baseline will result in a smaller SMD versus studies that enrol patients that provide more homogeneous scores, even when the true underlying effect is the same. We have previously published guidance regarding approaches that can be used to pool different instruments in natural units, minimally important difference units, a ratio of means, or the odds or risk of achieving a patient-important level of improvement.[16] For binary outcomes, pooled treatment effects are typically expressed as an odds ratio or risk ratio.

Pooled treatment effects may demonstrate heterogeneity suggesting that an overall pooled effect may be misleading as treatment effects differ between studies. The statistical evidence of heterogeneity can be estimated using Cochrane's Q-test (heterogeneity detected if $P \leq 0.10$), and the I^2 statistic (a percentage measure of inconsistency) - higher I^2 indicates more heterogeneity, hence higher inconsistency.[17] A rough guide for interpreting heterogeneity using the I^2 statistic is as follows: 0%–40%, insignificant; 30%–60%, moderate; 50%–90%, significant and 75%–100%, considerable. If heterogeneity is high, it would not necessarily mean abandoning meta-analyses, but caution should be exercised during interpretation of results. Strategies to explore

Study or subgroup	Caudal		Noncaudal		Weight	Risk ratio M-H, random, 95% CI	Year	Risk ratio M-H, random, 95% CI
	Events	Total	Events	Total				
11.1.1 Caudal versus N block								
Markham et al. 1986	3	26	0	26	0.5%	7.00 [0.38, 129.11]	1986	
Scott et al. 1989	1	19	2	25	0.7%	0.66 [0.06, 6.73]	1989	
Fisher et al. 1993	33	52	24	30	52.6%	0.79 [0.60, 1.04]	1993	
Abdellatif et al. 2012	5	25	7	25	3.9%	0.71 [0.26, 1.95]	2012	
Subtotal (95% CI)		**122**		**106**	**57.6%**	**0.80 [0.62, 1.04]**		
Total events	42		33					
Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 2.43$, df = 3 ($P = 0.49$); $I^2 = 0\%$								
Test for overall effect: $Z = 1.68$ ($P = 0.09$)								
11.1.2 Caudal versus infiltration								
Fell et al. 1988	2	22	2	27	1.1%	1.23 [0.19, 8.02]	1988	
Schindler et al. 1991	8	27	12	27	7.6%	0.67 [0.33, 1.37]	1991	
Conroy et al. 1993	13	25	11	20	13.1%	0.95 [0.55, 1.63]	1993	
Machotta et al. 2003	11	30	8	28	6.9%	1.28 [0.61, 2.72]	2003	
Subtotal (95% CI)		**104**		**102**	**28.8%**	**0.94 [0.65, 1.36]**		
Total events	34		33					
Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 1.62$, df = 3 ($P = 0.66$); $I^2 = 0\%$								
Test for overall effect: $Z = 0.34$ ($P = 0.73$)								
11.1.3 Caudal versus combined N block and infiltration								
Cross and Battet et al. 1987	1	20	2	21	0.7%	0.53 [0.05, 5.35]	1987	
Splinter et al. 1995	11	96	12	104	6.6%	0.99 [0.46, 2.14]	1995	
Somri et al. 2002	3	15	4	15	2.3%	0.75 [0.20, 2.79]	2002	
Tobias et al. 1995	2	12	9	12	2.3%	0.22 [0.06, 0.82]	2002	
Bhattarai et al. 2005	2	30	6	30	1.7%	0.33 [0.07, 1.52]	2005	
Subtotal (95% CI)		**173**		**182**	**13.6%**	**0.59 [0.32, 1.07]**		
Total events	19		33					
Heterogeneity: $\tau^2 = 0.06$; $\chi^2 = 4.57$, df = 4 ($P = 0.33$); $I^2 = 12\%$								
Test for overall effect: $Z = 1.73$ ($P = 0.08$)								
11.1.4 Caudal versus paravertebral								
Tug et al. 2011	16	35	6	35	0.0%	3.00 [1.07, 8.40]	2011	
Subtotal (95% CI)		**0**		**0**		**Not estimable**		
Total events	0		0					
Heterogeneity: not applicable								
Test for overall effect: not applicable								
Total (95% CI)		**399**		**390**	**100.0%**	**0.81 [0.66, 0.99]**		
Total events	95		99					
Heterogeneity: $\tau^2 = 0.00$; $\chi^2 = 9.93$, df = 12 ($P = 0.62$); $I^2 = 0\%$								
Test for overall effect: $Z = 2.10$ ($P = 0.04$)								
Test for subgroup differences: $\chi^2 = 1.69$, df = 2 ($P = 0.43$); $I^2 = 0\%$								

0.01   0.1   1   10   100

Favours (caudal)   Favours (noncaudal)

Figure 1: Forest plot depicting the use of rescue analgesia in the early post-operative period in caudal and non-caudal regional techniques in children undergoing inguinal surgeries. Harsha Shanthanna, Balpreet Singh, and Gordon Guyatt, "A Systematic Review and Meta-Analysis of Caudal Block as Compared to Noncaudal Regional Techniques for Inguinal Surgeries in Children," BioMed Research International, vol. 2014, Article ID 890626, 17 pages, 2014. doi:10.1155/2014/890626

heterogeneity include subgroup analysis (exploring outcomes for predefined subgroups, based on PICO, or RoB, etc.) or meta-regression (predicting outcome according to one or more continuous explanatory variables).[18]

The strength of a SR is the avoidance of selection bias, strengths of meta-analysis include an increase in power that improves statistical precision of treatment effect estimates, and the ability to explore for subgroup effects (although between-study subgroup effects are less credible than within-study subgroup effects). The main limitation of SRs and meta-analysis are related to the quality of included studies – aggregating poor quality studies will lead to a poor quality review.

Reporting the findings
The reporting of SRs should be guided by the preferred reporting items for SR and meta-analysis (PRISMA) for RCTs[19] and Meta-analysis of observational studies in epidemiology (MOOSE) for observational data.[20] The grading of recommendations assessment, development and evaluation (GRADE) approach should be used to evaluate and report the quality of

evidence for each outcome reported in a SR (http://www.gradeworkinggroup.org). Resources such as the Cochrane Database of SRs[21] provide access to high-quality SRs on treatment interventions.

SUMMARY

There are 3 tenets of EBM: (1) all relevant evidence should be considered to inform clinical decision-making, (2) there is a hierarchy of evidence that is based on the ability of different study designs to address RoB and (3) evidence alone is never enough (e.g., patient values and preferences). Rigorously conducted SRs and meta-analyses present an invaluable strategy to address the 1st two of these tenants, and are essential to the practice of EBM.

Financial support and sponsorship
Nil.

Conflicts of interest
There are no conflicts of interest.

REFERENCES

1. Guyatt G, Rennie D, Meade MO, Cook DJ. Users' Guides to the Medical Literature: Essentials of Evidence-based Clinical Practice. Evidence Based Medicine. McGraw-Hill Professional, New York; 2002. p. 1-359.
2. Grant MJ, Booth A. A typology of reviews: An analysis of 14 review types and associated methodologies. Health Info Libr J 2009;26:91-108.
3. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, *et al.* The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. J Clin Epidemiol 2009;62:e1-34.
4. Glasziou P, Burls A, Gilbert R. Evidence based medicine and the medical curriculum. BMJ 2008;337:a1253. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18815165. [Last cited on 2016 Jul 19].
5. Sackett DL, Strauss SE, Richardson WS, Rosenberg W, Haynes RB. Evidence-Based Medicine: How to Practice and Teach EBM. 2nd ed. Edinburgh: Churchill Livingstone; 2000.
6. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes BD. Oxford Centre for Evidence-Based Medicine Levels of Evidence. Available from: http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/. [Last cited on 2016 Jul 16].
7. Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. Lancet 2011;377:108-9.
8. Riva JJ, Malik KM, Burnie SJ, Endicott AR, Busse JW. What is your research question? An introduction to the PICOT format for clinicians. J Can Chiropr Assoc 2012;56:167-71.
9. Busse JW, Bruno P, Malik K, Connell G, Torrance D, Ngo T, *et al.* An efficient strategy allowed English-speaking reviewers to identify foreign-language articles eligible for a systematic review. J Clin Epidemiol 2014;67:547-53.
10. Shanthanna H, Busse JW, Thabane L, Paul J, Couban R, Choudhary H, *et al.* Local anesthetic injections with or without steroid for chronic non-cancer pain: A protocol for a systematic review and meta-analysis of randomized controlled trials. Syst Rev 2016; 5:18 doi: 10.1186/s13643-016-0190-z.
11. Viera AJ, Garrett JM. Understanding interobserver agreement: The kappa statistic. Fam Med 2005;37:360-3.
12. Wells G, Shea B, O'Connell D, Peterson J, Welch V, Losos M, *et al.* The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-analyses. Available from http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed on 15-07-2016.
13. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, *et al.* The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ 2011;343:d5928.
14. Shanthanna H, Singh B, Guyatt G. A systematic review and meta-analysis of caudal block as compared to noncaudal regional techniques for inguinal surgeries in children. Biomed Res Int 2014;2014:890626.
15. Ryan R; Cochrane Consumers and Communication Review Group: Data Synthesis and Analysis. Available from: http://www.cccrg.cochrane.org/sites/cccrg.cochrane.org/files/uploads/AnalysisRestyled. [Last cited on 2016 Jul 16].
16. Busse JW, Bartlett SJ, Dougados M, Johnston BC, Guyatt GH, Kirwan JR, *et al.* Optimal strategies for reporting pain in clinical trials and systematic reviews: Recommendations from an OMERACT 12 Workshop. J Rheumatol 2015. pii: jrheum. 141440.
17. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557-60.
18. Deeks JJ, Higgins JPT, Altman DG. on Behalf of the CSMG. Analysing data and undertaking meta-analyses. Cochrane Handbook for Systematic Reviews of Interventions. Ch. 9. 2011. p. 9.1-9.9. Available from: http://www.handbook.cochrane.org/chapter_9/9_analysing_data_and_undertaking_meta_analyses.htm. [Last accessed on 2016 Jul 15].
19. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009;339:b2535.
20. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, *et al.* Meta-analysis of observational studies in epidemiology: A proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. JAMA 2000;283:2008-12.
21. Cochrane Database of Systematic Reviews. Available from: http://www.cochranelibrary.com/cochrane-database-of-systematic-reviews/index.html. [Last cited on 2016 Jul 15].