BJA

British Journal of Anaesthesia, 0 (0): 1–11 (2017)

doi: 10.1093/bja/aex251
Review Article

REVIEW ARTICLE

Heterogeneity of studies in anesthesiology systematic reviews: a meta-epidemiological review and proposal for evidence mapping

B. Umberham, R. Hedin, B. Detweiler, L. Kollmorgen, C. Hicks and M. Vassar*

Oklahoma State University Center for Health Sciences, 1111 W 17th St., Tulsa, OK 74107, USA

*Corresponding author. E-mail: matt.vassar@okstate.edu

Abstract

Heterogeneity among the primary studies included in a systematic review (SR) is one of the most challenging considerations for systematic reviewers. Current practices in anaesthesiology SRs have not been evaluated, but traditional methods may not provide sufficient information to evaluate the true nature of these differences. We address these issues by examining the practices for evaluating heterogeneity in anesthesiology reviews. Also, we propose a mapping method for presenting heterogeneous aspects of the primary studies in SRs.We evaluated heterogeneity practices reported in SRs published in highly ranked anesthesiology journals and Cochrane reviews. Elements extracted from the SRs included heterogeneity tests, models used, analyses conducted, plots used, and I^2 values. Additionally, we selected a SR to develop an evidence map in order to display clinical heterogeneity.

Our statistical analysis showed 150/207 SRs reporting a test for statistical heterogeneity. Plots were used in 138 reviews to display heterogeneity. Subgroup analyses were the most commonly reported analysis (54%). Meta-regression and sensitivity analyses were used sparingly (25%; 23% respectively). A random effects model was most commonly reported (33%). Heterogeneity statistics across meta-analyses suggested that, in our sample, the majority (55%) did not present sufficient heterogeneity to be of great concern. Cochrane reviews (n=58) were also analysed. Plots were used in 88% of Cochrane reviews. Subgroup analysis was used in 59% Cochrane reviews, while sensitivity analysis was used in 62%.

Many reviews did not provide sufficient detail regarding heterogeneity. We are calling for improvement to reporting practices.

Key words: heterogeneity; anaesthesia; review

Systematic reviews and meta-analyses are among the most popular methodologies in clinical research. These methodologies receive more citations, on average, than any other research design within the health sciences.[1] These reviews inform clinical decision-making and are considered level 1a evidence when developing anaesthesia treatment guidelines.[2] As these methodologies bring together evidence from individual studies that differ on a number of dimensions from patient characteristics to study designs, it is understandable that dealing with this diversity would be a significant methodological consideration. For example, a recent systematic review examining dexamethasone for peripheral nerve blocks noted that results should be interpreted with caution because of extreme differences among the included primary studies. General anaesthesia was combined with peripheral nerve blocks in one-quarter of the studies. Locations of the nerves, anaesthetic dose, and use of perineural

Editorial decision: May 31, 2017; **Accepted:** July 4, 2017

Editor's key points

- Statistical heterogeneity describes the variation of individual study results in a meta-analysis.
- Variation can occur in either the size or direction of effect.
- Inconsistent or marked variation of effects greatly limit external validity (generalizability) of results from a systematic review.
- Strong evidence exists when consistent, reproducible results are obtained from a variety of study settings.

adjuncts also varied across studies.[3] These differences were noted so that readers could consider the results and determine whether primary studies were suitable for synthesis.

In systematic reviews, the diversity among studies is referred to as heterogeneity. The Cochrane Collaboration delineates three types. *Clinical heterogeneity* refers to differences between patients, interventions, or outcomes. *Methodological heterogeneity* describes differences in study design and risk of bias. *Statistical heterogeneity* is represented by variability in the intervention effects being examined across studies and is a consequence of either clinical or methodological heterogeneity, or both.[4] Addressing heterogeneity is one of the most difficult aspects of many systematic reviews.[5] In addition, common reporting guidelines such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)[6] and the Meta-Analyses and Systematic Reviews of Observational Studies (MOOSE)[7] recommend that authors describe methods of combining study results and accounting for heterogeneity appropriately. Following these guidelines is now a requirement of many journals.

Current practices focus on statistical heterogeneity by reporting the outcomes of a statistical test and using the magnitude of the test statistic (or its significance level) as an indication of the degree of heterogeneity present among the included primary studies. These results form the basis for deciding next steps. While these practices are commonplace, they do not provide sufficient information to evaluate potential sources of heterogeneity or point to the clinical implications of the observed heterogeneity.[8] Identifying these sources may greatly enrich the readers and reviewers understanding of the SR and how findings may be interpreted.[9] In this study, we examine heterogeneity assessment practices among meta-analyses and systematic reviews in anesthesiology research. We focus on particular methods used to detect heterogeneity and examine the ways in which heterogeneity results inform decision-making. Second, we examine the extent of heterogeneity among the systematic reviews in our sample. Last, we present evidence mapping in the form of a case study using a recently published systematic review, to demonstrate how a more thorough evaluation of clinical and methodological heterogeneity moves beyond decisions based on statistical heterogeneity tests. This process allows for researchers to make informed decisions regarding the distinguishing clinical or methodological features of primary studies and for readers to form conclusions regarding the nature of heterogeneity of studies included in a meta-analysis.

Methods: statistical heterogeneity

Search criteria

PubMed searches were conducted on May 18 and May 26, 2015, using the following search string: ((((“Anesthesiology”[Journal] OR “Anesthesia and analgesia”[Journal]) OR “British journal of anaesthesia”[Journal]) OR “Anaesthesia”[Journal]) OR “Regional anesthesia and pain medicine”[Journal]) AND ((meta-analysis[Title/Abstract] OR meta-analysis[Publication Type]) OR systematic review[Title/Abstract]) AND ((“2007/01/01”[PDAT]: “2015/12/31”[PDAT]) AND “humans”[MeSH Terms]). This search strategy was adapted from a previously published approach that is sensitive to identifying systematic reviews and meta-analyses.[10] Journals were selected based on the 2014 h5-index of Google Scholar Metrics: Anesthesiology subcategory. The h5-index is a journal level metric based on the Hirsch index for authors that uses an algorithm to determine journal influence based on the number of citations received from articles published in the journal. Second, we searched for SRs produced by the Cochrane Anaesthesia, Critical, and Emergency Care Group (ACE) using the Cochrane Database of Systematic Reviews. This search occurred on July 15, 2016, and was achieved by limiting the Cochrane Database of Systematic Reviews to SRs produced by the Cochrane ACE Group between 2007 and 2015.

Screening and data extraction

Covidence (covidence.org), a systematic review platform, was used initially to screen titles and abstracts. To qualify as a systematic review, studies had to summarize evidence across multiple studies and provide information on the search strategy, such as search terms, databases, or inclusion/exclusion criteria.[11] Meta-analyses were considered studies that applied a quantitative synthesis of results across multiple studies.[12] Two authors (M.V. & R.H.) independently screened all articles based on title and abstract, after which a follow up meeting was held to discuss differences in screening. Any disagreements were settled by consensus. Full-text articles were next obtained via EndNote for extracting relevant study features.

The authors designed an extraction manual to ensure accuracy in extraction. Before releasing the extraction manual to individual authors, the manual was piloted based on a subset of systematic reviews. The following elements were included in the manual: a) the statistical test used to evaluate heterogeneity; b) sample size; c) *a priori* threshold for statistical significance; d) type of model (random, fixed, mixed, or both); e) whether reviewers selected a random effects model based on significance of the heterogeneity test; f) whether reviewers used a random effects model without explanation; g) what type of plot was used to evaluate heterogeneity, if any; h) whether the plot was published as a figure in the manuscript; i) whether follow-up analysis was conducted, and if so, the type of analysis (subgroup, meta-regression, and/or sensitivity analysis); j) what type of heterogeneity was analysed if subgroup analysis was performed (clinical or methodological); j) whether heterogeneity was mentioned in writing only; k) whether reviewers concluded there was too much heterogeneity to perform a meta-analysis; l) whether a confidence interval was reported with the heterogeneity statistic; m) whether the reviewers included an evidence map to explore clinical heterogeneity; n) whether reviewers included pre-specified subgroups; o) test of interaction, if any; p) prediction interval, if any; and q) the type of studies included in the SR (e.g. randomized trials, non-randomized trials, cohort studies).

Next, a training session was conducted to familiarize authors with the manual and extraction process. A subset of systematic reviews was used for training purposes and discussed as a group. Two authors (B.A.U. & R.H.) were next assigned three new systematic reviews for independent extraction. These data were analyzed