Methodological tools and sensitivity analysis for assessing quality or risk of bias used in systematic reviews published in the high-impact anesthesiology journals

Marušić, Marija Franka

UNIVERSITY OF SPLIT
SCHOOL OF MEDICINE

Marija Franka Marušić

METHODOLOGICAL TOOLS AND SENSITIVITY ANALYSIS FOR ASSESSING QUALITY OR RISK OF BIAS USED IN SYSTEMATIC REVIEWS PUBLISHED IN THE HIGH-IMPACT ANESTHESIOLOGY JOURNALS

Diploma thesis

Academic year:

2019/2020

Mentor:

Livia Puljak, MD, PhD

Split, July 2020

TABLE OF CONTENTS

ACKNOWLEDGEMENT

This thesis contains text from a published manuscript:

Marusic MF, Fidahic M, Cepeha CM, Farcas LG, Tseke A, Puljak L. Methodological tools and sensitivity analysis for assessing quality or risk of bias used in systematic reviews published in the high-impact anesthesiology journals. BMC Medical Research Methodology. 2020;20:121. doi: 10.1186/s12874-020-00966-4. (2019 Journal Impact Factor: 3.031)

Link to the published article:

https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-00966-4

The text from the published study is unchanged, except for the Introduction, which has been modified and expanded to a format appropriate for a diploma thesis. Minor formatting has been used where applicable to satisfy the correct thesis format.

I wish to express gratitude to my mentor, Livia Puljak, for her guidance and everything she has taught me in the time we have worked together. During my studies, prof. Puljak was an important example and a role model to look up to.

I also thank my co-authors on the published article.

Endless thanks to my parents, for being there for me and always thinking of me, showing nothing but unconditional support for my aspirations.

A final thank you to Sandro and my friends and colleagues who have shared this journey with me.

LIST OF ABBREVIATIONS

AHRQ – Agency for Healthcare Research and Quality

CONSORT – Consolidated Standards of Reporting Trials

CRD – Centre for Reviews and Dissemination

GRADE – Grading of Recommendations Assessment, Development and Evaluation

JCR – Journal Citation Reports

MOOSE – Meta-Analysis of Observational Studies in Epidemiology

NOS – Newcastle-Ottawa Scale

QUADAS – Quality Assessment of Diagnostic Accuracy Studies

QUADAS-2 – Quality Assessment of Diagnostic Accuracy Studies 2

QUIPS – Quality in prognosis studies

QUOROM – Quality of Reporting of Meta-analyses

PEDro – Physiotherapy Evidence Database

PRISMA – Preferred reporting items for systematic review and meta-analysis

RCT – randomized controlled trial

RoB – risk of bias

SIGN – Scottish Intercollegiate Guidelines Network

SR – systematic review

STROBE – Strengthening of the Reporting of Observational Studies in Epidemiology

USPSTF – U.S. Preventive Services Task Force

1. INTRODUCTION

1.1. Evidence-based medicine (EBM)

Evidence-based medicine (EBM) began as a movement in the early 1990s, after early initiatives in the 1980s that highlighted the need for an empirical approach in medicine (1), with stricter guides and rules in making clinical decisions.

After first being mentioned in 1991 (2), one of the other most important early articles on EBM described it as [quote] "*the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients*" (3). According to Sacket et al, EBM incorporates three components: individual clinical expertise, patient values and expectations, and best external evidence (4).

The initial focus of the EBM movement was to educate clinicians on how to understand the data presented in clinical studies, how to evaluate the best possible evidence, and how to implement this knowledge into everyday clinical practice. Soon EBM began to evaluate the quality of evidence in general, emphasizing a need to continuously and critically appraise new evidence as it is published (5). Today, physicians are strongly encouraged to incorporate EBM into their daily practice, not only by reading scientific literature themselves, but by following clinical guidelines and protocols, which are formulated by a review of available evidence and a consensus of expert opinion (6).

1.2. Levels of evidence in medicine

In order to determine what the "best current evidence" is, it became necessary to determine a hierarchy of the quality of evidence. As evidence started to be the object of scrutiny, it became clear that not all evidence is the same. An early initial division of the quality of evidence into a pyramidal hierarchy focused mainly on the quality of study design, thereby placing randomized controlled trials (RCTs) at the top among the primary evidence, and with systematic reviews of RCTs on the very top of the pyramid (7). However, as time passed, this division proved to be inadequate and inflexible. In 2011, The Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach provided a system for rating the quality of evidence and strength of recommendations that is explicit, comprehensive, transparent, and pragmatic and is being increasingly adopted in large numbers by organizations worldwide. According to GRADE, RCTs can move further down the pyramid if they contain any kind of publication bias, which can lead to the overestimation of their effect (8).

The GRADE framework also recognized that the quality of evidence in medicine also depends on many different factors, including the extent of risk of bias in study implementation, imprecision, inconsistency, indirectness (the study being inapplicable or irrelevant to the patient case at hand), and the possibility of publication bias (8). Furthermore, this system has already been endorsed and supported by many different organizations and scientific societies such as the World Health Organization (WHO), the Cochrane (previously named the Cochrane Collaboration), and the American College of Physicians, among others (9).

One can conclude that the pyramid of evidence or the evidence hierarchy is constantly changing and improving. There are some disagreements on how the pyramid should look like, but it is still commonplace for systematic reviews and meta-analyses to be placed at the top, since they usually represent a pooling of the results of the highest quality studies (usually RCTs) (10). However, GRADE protects against a superficial assessment and unwarranted confidence in RCTs. The increasing application and use of GRADE have resulted in a significant improvement in the quality of systematic reviews. It has also allowed for observational studies to be a potential source of high-quality causal evidence, being relevant and appropriate in certain types of research questions, as well as in situations when no adequate RCTs are available (5).

1.3. Systematic reviews (SRs)

Systematic reviews (SRs) are studies that combine and appraise the available evidence to answer a specific research question. SRs vary in their methods and scope, but they most often follow a systematic methodology, including the following: pre-defined inclusion criteria, a suitable search strategy, quantitative analytical methods if applicable, and a systematic approach to minimizing biases and random errors, all of which is documented in a methods section (11).

The terms "systematic review" and "meta-analysis" are often used interchangeably, although this is not correct. Systematic review is a study design that includes defining a research question, systematic search, appraising and synthesizing of evidence. Meta-analysis is a statistical procedure that combines numerical data collected from multiple individual studies. Systematic review may or may not include one or more meta-analyses, depending on the data that will be collected. If there is a high statistical or clinical heterogeneity in collected data, meta-analysis should not be conducted (6).

The importance of SRs is not only due to their placement at the top of the evidence pyramid, but also because they, inform health practice guidelines, thereby translating the best available evidence to practice (6). Cochrane is responsible for the biggest advances in systematic review methodology. Cochrane is an international organization that specializes in conducting systematic reviews of health interventions and diagnostic tests and publishes them in the Cochrane Library (12). Cochrane SRs are specific due to their advancing methodology that provides a clear and guided framework for SR authors and other experts involved in the process, thus creating a platform that pools together evidence to create a repository of reliable SRs that are continuously reassessed.

As more and more SRs are published, one can find conflicting or otherwise discordant results from different systematic reviews on the same research question or topic. Navigating through these discordant results can be demanding even for readers who have a good knowledge of evidence-based medicine and know in detail about the specific methods and/or research questions. Not only is confusion generated, but there is also the problem of redundant results of overlapping SRs – the same topics are often covered more times than is necessary (13).

There has been a threefold increase in the number of systematic reviews over the last decade, with quality of their conducting and reporting improving, yet overall remaining suboptimal in many regards. If the methods used to conduct a SR are flawed or if the reporting is incomplete, it is of limited use to decision makers and does not aid further research agendas, thus contributing to research waste (14).

Well-conducted SRs of randomized controlled trials should be considered as top-level research evidence for guideline development or other evidence appraisal. However, with current poor methodological practices, observed discrepancies in the results, and the incorporation observational studies as sources of evidence for SRs, there is a need for a strict and rigorous appraisal of study quality and methodology used in SRs in order to make sure that their place at the top of the evidence pyramid is justified (15).

1.4. Risk of bias

Bias is defined as a systematic error, or deviation from the truth, in results or inferences. Biases can vary in magnitude, from small to substantial, and can lead to an underestimation or overestimation of the true intervention effect. Bias should not be confused with imprecision. Bias refers to systematic error, an example being: if the same study would be replicated multiple

times, it would, on average, always reach the wrong answer. Imprecision refers to random error, meaning that the same study, replicated multiple times, would produce different estimates of effect. In imprecision, this is due to sampling variation even if the different replication attempts would give the right answer on average. The results of smaller studies are subject to greater sampling variation and hence are less precise (16).

There is empirical evidence that some features of the design, conduct and analysis of RCTs often lead to bias. However, it is usually impossible to know to what exact extent biases have affected the results of a particular study or analysis (17). Due to these reasons, one should consider whether a result is at risk of bias rather than claiming with certainty that bias is present. The most recent tools for assessing the internal validity of findings from quantitative studies in health now focus on risk of bias, whereas tools in the past targeted the broader notion of "methodological quality" (18).

In a Cochrane systematic review, this process of bias appraisal of the included studies is named *the assessment of risk of bias in included studies*. A tool has been developed and implemented for assessing risk of bias, the Cochrane Risk of Bias (RoB) tool. The first version of the Cochrane RoB tool was published in 2008, an update of it was published in 2011 (19). In 2019, a complete revision of the tool was published, called the Cochrane RoB 2 tool, intended as the recommended Cochrane tool for assessing RCTs included in a systematic review. It provides a framework for assessing the risk of bias in a single result (an estimate of the effect of an experimental intervention compared with a comparator intervention on a particular outcome) from any type of randomized trial (20).

The work covered in this diploma thesis refers to the 2011 Cochrane RoB tool, since it was the recommended and relevant tool during the time of data extraction. In the meantime, RoB 2 was published, but at the time of writing of this thesis, the new version of the tool was still not mandatory for use in Cochrane reviews. Therefore, an emphasis will be placed on the characteristics of the 2011 version of the RoB tool (19).

The 2011 RoB tool covers six domains of bias: selection bias, performance bias, detection bias, attrition bias, reporting bias, and a domain covering other types of bias. Within each domain, assessments are performed for one or more items, which may cover different aspects of the domain, or different outcomes. For each item in the tool, the assessment of risk of bias is in two parts. The first part of the assessment includes providing a free text description for each of the domains or a summary of the relevant trial characteristic on which judgments

on risk of bias have been based. This clear and structured approach aims to ensure transparency in how the judgments have been made (19).

The second part of the assessment involves assigning a high, low, or unclear risk of material bias for each item, based on the explanatory comment. In this case, material bias is defined as bias of enough magnitude to have a notable effect on the results or conclusions of the trial. Detailed criteria for making judgments about the risk of bias from each of the items in the tool are available in the Cochrane Handbook (16). If there is no clear and sufficient detail reported on what happened in the trial, the judgment will usually be "unclear risk of bias". A judgment of unclear risk should also be made if what happened in the RCT is known but the associated risk of bias cannot be determined, for example, if participants take additional drugs of unknown effectiveness because of them being aware of their intervention assignment. The key factor in this type of bias assessment is that it should be performed independently by at least two people, and a third person should resolve any potential discrepancies between the first two (19).

In this way, each study included in the SR is comprehensively analyzed in a structured way, divided by the different bias domains. The results of an assessment are usually can be presented in a table, in which judgments for each item in each trial are presented alongside the description of their justification. There is also the possibility of presenting the results in an illustration that is easier to interpret, since making a table for every study is difficult and hard to navigate. This sort of data is then included in the SR (19).

1.5 Sensitivity analysis

Conducting a systematic review involves a sequence of decisions. Even though many of these decisions are planned and objective, there are always some that are arbitrary or unclear. For instance, if the inclusion criteria involve a numerical value, the choice of value is usually arbitrary: for example, defining "groups of older people" may have lower cut-off limits of 60, 65, 70 or 75 years, or any value in between. Other decisions may be unclear because an included study fails to clearly provide all or some of the required information. Some decisions are unclear because the included studies themselves never obtained the information required: for example, the outcomes of patients who were lost during follow-up. Another example is that of decisions that are unclear because there is no consensus on the best statistical method to use in that particular problem (16).

A sensitivity analysis aims to prove that the findings obtained from a systematic review do not depend on these arbitrary decisions. It is performed by repeating the primary analysis (meta-analysis) by excluding some studies. In the primary analysis, all eligible studies are included in a meta-analysis; for sensitivity analysis meta-analysis will be repeated by including only studies that are definitely known to be eligible according to a certain criterion (e.g. as mentioned before, studies that do not provide all the necessary information) (16). For example, if the aim of a sensitivity analysis is to explore the effect of RoB, then the meta-analysis exploring this aspect may include only studies with low RoB, to see whether there will be a difference between meta-analysis that included all studies and meta-analysis that included only studies with low RoB. There are many aspects that can be studied within a sensitivity analysis, including characteristics of participants, interventions, comparators, outcomes, study design aspects, type of publication, etc. (16).

This essential part of SRs is often either not performed or reported. There is no set strategy for performing a sensitivity analysis. The underlying principle is to repeat the primary analysis with an altered dataset or statistical method and try to observe whether these changes have any effect on the combined outcome estimate. When altering the dataset, the choice of studies to add or remove is often up to the author's choice (21).

1.6 The importance of pain

Paying particular attention to the methodological/reporting quality of SRs in the high-impact journals published in the field of anesthesiology and pain is important because pain is the symptom that most commonly brings patients to see a physician. A pain-free life and access to pain treatment is considered a basic human right (22, 23). However, inadequate pain management is frequent, even in developed countries. This is caused both by insufficient attention devoted to pain measurement and treatment, as well as the fact that, for some painful conditions such as neuropathic pain, there are inadequate treatment options available (22). A pain-free state is very important for patients. Therefore, interventions for the treatment of pain are of a major public health importance. It has already been shown that methodological and reporting quality of SRs published in the highest-ranking journals in the field of pain needs to be improved (24), and therefore further methodological work in this field can help journal editors, reviewers, and authors to improve future studies.

Detweiler et al. have earlier analyzed a sample of studies from the field of anesthesiology and pain, and reported that, although 84% of those studies assessed quality/RoB,

many authors applied questionable methods (25). They reported that Jadad tool was used most commonly (25), but this tool is nowadays less used, in favor of the Cochrane RoB tool (19). It is unclear how the usage of different quality/RoB tools is changing over time, and how authors of SRs use sensitivity analysis when they want to check robustness of their result following quality/RoB indicators.

2. OBJECTIVES

The aim of this study was to assess quality/RoB assessment tools, the types of sensitivity analyses and quality/RoB thresholds for sensitivity analyses used within SRs published in the high-impact pain/anesthesiology journals between 2005 and 2018.

Our hypotheses were:

1. The majority of analyzed articles will report that they used quality/RoB assessment tools;
2. The minority of analyzed articles that used quality/RoB tools will use Cochrane RoB tool;
3. The majority of articles that have used sensitivity analyses for quality/RoB did not specify a threshold for quality/RoB.

3. MATERIALS AND METHODS

3.1 Data sources and study eligibility

We conducted a methodological study, i.e. a research-on-research study. We used an *a priori* defined research protocol; this protocol is available in Supplementary file 1. We analyzed systematic reviews and meta-analyses published in the 25% highest-ranking journals within the Journal Citation Reports (JCR) category "Anesthesiology". We limited our analysis to systematic reviews and meta-analyses published between January 2005 and June 2018. We did not include reviews published before 2005, since risk of bias assessment methodology is relatively recent, and the initial version of the Cochrane's risk of bias tool was published in 2008 (16).

We performed the search on July 3, 2018.

The following 7 journals were analyzed: Anaesthesia, Anesthesia and Analgesia, Anesthesiology, British Journal of Anaesthesia, Pain, Pain Physician, Regional Anesthesia & Pain Medicine. We did not use any language restrictions, as all the targeted journals publish articles in English.

Systematic reviews of both randomized and non-randomized studies were eligible. We excluded systematic reviews and meta-analyses of diagnostic accuracy or of individual patient data, as well as overviews of systematic reviews and guidelines. We also excluded systematic reviews published in a short form as a correspondence, and Cochrane reviews published as secondary articles in the analyzed journals. We did not include Cochrane reviews because for them use of Cochrane RoB tool is mandatory.

3.2 Definitions

For the purpose of this study, a systematic review was defined as an overview of scientific studies using explicit and systematic methods to locate, select, appraise, and synthesize relevant and reliable evidence. While meta-analysis is a statistical method used to pool results from more than one study, sometimes the terms "systematic review" and "meta-analysis" are used interchangeably, so we also included studies that were described by authors as a meta-analysis, if they fitted the definition of a systematic review.

While the Cochrane recommends using its RoB tool to assess the quality of individual studies included in their SRs, many systematic reviews use various quality assessment tools for

appraising studies. Sometimes authors use the terms "quality" and "bias" interchangeably. Therefore, in this study we analyzed any quality/RoB tool used by the SR authors, regardless of whether the authors called it a quality assessment tool, or a risk of bias assessment tool.

3.3 Search

We searched PubMed by using the advanced search with a journal name, a filter for systematic reviews and meta-analyses, and a filter for publication dates from January 2005 to June 2018. Search results were then exported and saved. The chosen publication dates and the included sample size were considered sufficient based on a previous similar publication (26).

3.4 Screening of records

A calibration exercise was performed on hundred first records to ensure compliance with eligibility criteria. Two authors independently performed each step in screening all the studies. The first step was the screening of titles and abstracts; the second step was the screening of full texts that were retained as eligible or potentially eligible in the first screening step. Disagreements about inclusion of full texts were resolved via discussion or discussion with the third author.

3.5 Data extraction

Two authors independently conducted data extraction, using a standardized data extraction form created for this study. Disagreements were resolved by a discussion with the third author.

Following the initial piloting on 10 reviews, two authors extracted data independently from each eligible study using the standardized extraction form. A third author compared two data sets and identified any possible discrepancies that were resolved by discussion with a third author and resulted in a final consensus.

The following data were extracted: i) the country of authors' affiliations (the whole count method was used, in which each country gets one mention when it appears in the address of an author, regardless of the number of times it was used for other authors), ii) the number of authors, iii) whether the involvement of a methodologist or statistician was mentioned in the Methods section, iv) whether a meta-analysis was performed, v) whether quality or RoB

assessment was performed, vi) the name of the specific quality/RoB tool (extracted verbatim, in the way the authors reported it), and vii) the name of the journal. We also recorded whether a threshold level of quality/RoB was set by the authors.

Apart from analyzing the quality/RoB assessment tools, we also analyzed whether the authors used or planned to use a sensitivity analysis. We analyzed whether the study mentioned sensitivity analysis in the Methods section, regardless of whether it was actually conducted or not, because sensitivity analyses may be planned, but not conducted if they are not feasible subsequently. We also analyzed the frequency of use of sensitivity analyses, and which issues were explored in sensitivity analyses. If sensitivity analysis was done for quality/ RoB, we analyzed how did the authors define quality/RoB threshold (for example, authors may report "sensitivity analysis was conducted by excluding trials at high risk of bias", but if they do not define what did they consider a study at high risk of bias, a reader cannot know which quality threshold was used for such analysis). We did not have an a priori definition of what a sensitivity analysis is or should be; instead, we extracted all the information that study authors reported as a method of sensitivity analysis.

3.6 Data analysis

A descriptive statistical analysis was performed, including frequencies and percentages, using GraphPad Prism (GraphPad Software, La Jolla, CA, USA).

4. RESULTS

We retrieved 1413 results via a database search. After screening, we included 678 studies that were indexed as systematic reviews/meta-analyses. List of included studies is available in Supplementary file 2. The authors' affiliations originated from 48 countries, most commonly from the USA (N=230; 34%), Canada (N=124; 18%), UK (N=120; 18%), Australia (N=56; 8%) and Germany (N=56; 8%). The median number of authors was 5 (range: 1 to 16). In 35 (5.1%) articles it was stated that a methodologist/statistician was involved in the study. In our sample of 678 SRs, 382 (56%) included only RCTs, 181 (27%) included both RCT and non-randomized studies, 72 (11%) included non-randomized studies, while the remaining 43 (6%) did not report which types of studies were eligible.

4.1 Quality/risk of bias assessment tools

Authors reported that they assessed "quality" or "risk of bias" in 513 (76%) of the included studies. Some articles (N=75; 11%) reported using more than one tool for assessing quality/RoB (range: 2-4). The most commonly reported quality/RoB tools used were the Cochrane tool for RoB assessment (37%) and the Jadad tool (15%), either as a non-modified or a modified version (Table 1). Among studies that reported that only non-randomized studies were eligible, none of the studies reported using "Cochrane risk of bias tool" (with only that expression); two of those reviews reported using modified Cochrane risk of bias tool for observational studies, and one reported use of "ACROBAT-NRSI: Cochrane RoB tool for nonrandomized studies".

Some of the tools that authors reported were actually reporting checklists, or were intended for grading of overall evidence, such as QUROM, PRISMA and GRADE (Table 1). Since we analyzed articles published over the span of 14 years, we noticed a trend of a decrease in the use of the Jadad and Oxford scales, and increased use of the Cochrane tool for RoB assessment (Figure 1). In 44 (6.5%) articles, the authors reported that they analyzed the quality or RoB of their included studies, but they did not report the name of the tool they used, or provided a reference for the tool they used.

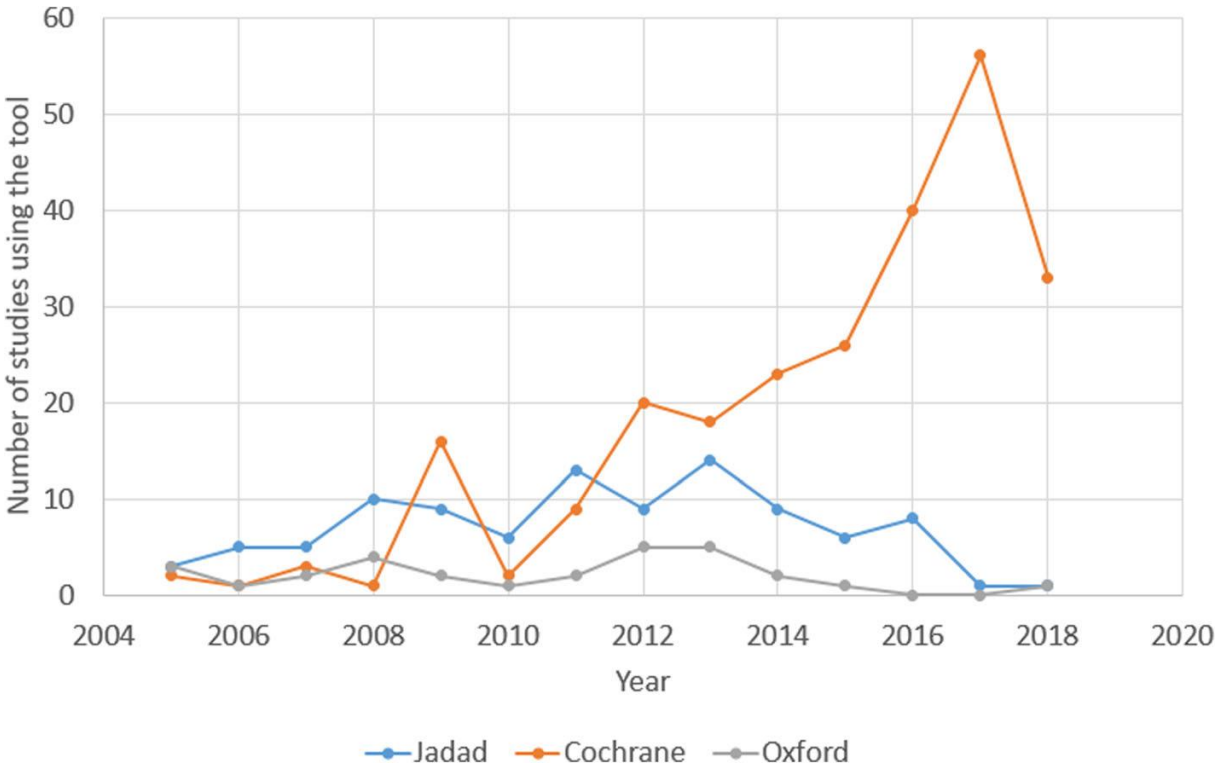

Figure 1. The time trend of using quality/risk of bias tools in the analyzed articles. The three most commonly used quality/risk of bias tools in articles analyzed within this study were Cochrane, Jadad, and Oxford tools. The figure indicates that the usage of Cochrane's tool is increasing, while the use of Jadad and Oxford tool is decreasing over time. Drop in the use of Cochrane's tool for RoB assessment in year 2018 is explained by our inclusion criteria – unlike other analyzed years, we included only articles published in the first half of 2018.

Table 1. Tools reported by the systematic review/meta-analysis authors that were used for the assessment of quality or risk of bias of the included studies more than once (N=678)

Tool	N (%)
Cochrane tool for RoB assessment	251 (37)
Non-modified version	241 (36)
Modified version	10 (1.4)
Jadad tool	99 (15)
Non-modified version	92 (14)
Modified version	7 (1.0)
Newcastle-Ottawa scale or its adapted version	30 (4.4)
Oxford scale	29 (4.3)
Non-modified version	10 (1.5)
Modified version	19 (2.7)
Criteria of Agency for Healthcare Research and Quality (AHRQ)	24 (3.5)
Grading of Recommendations Assessment, Development and Evaluation (GRADE)	18 (2.7)
Quality of Reporting of Meta-analyses (QUOROM)	14 (2.0)
Preferred reporting items for systematic review and meta-analysis (PRISMA)	10 (1.5)
Quality Assessment of Diagnostic Accuracy Studies (QUADAS) or QUADAS-2	7 (1.0)
Criteria of the U.S. Preventive Services Task Force (USPSTF)	5 (0.7)
Consolidated Standards of Reporting Trials (CONSORT)	4 (0.6)
The Scottish Intercollegiate Guidelines Network (SIGN) checklist for RCTs	4 (0.6)
Quality in prognosis studies (QUIPS) tool	3 (0.4)
Downs and Black	3 (0.4)
Meta-Analysis of Observational Studies in Epidemiology (MOOSE) checklist	2 (0.3)
Physiotherapy Evidence Database (PEDro) evaluation scale	2 (0.3)
Strengthening of the Reporting of Observational Studies in Epidemiology (STROBE)	2 (0.3)
Centre for Reviews and Dissemination (CRD) recommendations checklist	2 (0.3)

Among 165 reviews that did not report use of quality/RoB tools, 56 (34%) reported that only RCTs were eligible, 47 (28%) included both RCTs and non-randomized studies, 35 (21%) included only non-randomized studies, while 27 (16%) reviews did not report which study designs were eligible.

4.2 Sensitivity analyses

The majority of included articles (N=451; 66%) reported at least one meta-analysis. In 219 (48%) of those 451 articles, the methods for sensitivity analysis were reported. There were 120 of the 219 (55%) studies that performed only one type of sensitivity analysis, while others performed from 2 to 9 various types of sensitivity analyses. Sensitivity analysis was most commonly conducted to explore various aspects of study quality/RoB (90/219; 41%), intervention variations, and various statistical aspects (Table 2).

Table 2. Variables analyzed in the sensitivity analyses, used more than 5 times in the analyzed sample (N=678)

Variables	N (%)
Various aspects of quality/risk of bias	90 (13)
Statistics and effect sizes (heterogeneity, effect sizes, imprecise effect estimates, intention-to-treat analysis, different methods for effect size calculations, different results, correlation coefficients, meta-regression, imputation of data, different analysis methods, event rate, standard deviation calculated from standard error)	58 (8.6)
Intervention variations	52 (7.7)
Impact of each individual study (sequential exclusion of single studies)	31 (4.6)
Patients' characteristics (such as smoking, gender, weight)	24 (3.5)
Type of outcomes (such as different pain scales)	23 (3.4)
Type of included studies (crossover studies, randomized controlled trials, non-randomized studies, non-blinded studies, data from retracted studies, mixed data, peer-reviewed manuscripts)	13 (1.9)
Trial size	7 (1.0)
Publication bias	6 (0.9)
Comparator	6 (0.9)
Covariates	5 (0.7)
Type of funding	5 (0.7)

4.3 Sensitivity analyses for study quality/risk of bias

Among 90 studies that conducted sensitivity analysis based on study quality/RoB, 47 (52%) clearly specified the threshold for defining different levels of study quality/RoB (Supplementary file 3). Those 47 studies provided clear descriptions of what they considered high or low quality studies, or the difference between a low, unclear, or high risk of bias.

However, thresholds for quality/RoB used in those articles were highly heterogeneous. The most common approach in those 47 studies was to use a certain number of points on the Jadad, Oxford or Newcastle-Ottawa scales to define what was considered high or low-quality study (N=19; 40%). The authors did not use consistent cut-off points for labelling high-quality studies (Table 3).

The next most common category used various numbers of individual pre-specified RoB domains (i.e. key domains) for assessing what was a high, unclear, or low RoB. There were 18 such studies and the most commonly used domain for contributing to the assessment of RoB was allocation concealment (used in 7 of 18 articles), followed by the 'blinding of outcome assessors' (N=4), the 'blinding of participants and personnel (N=3), the generation of a randomization sequence (N=3), and attrition bias (N=3). Even the definitions of acceptable attrition varied among those few studies, whereas one article indicated that they used the threshold of 10%, and another one used 20% (Supplementary file 3).

In 10 of 47 (21%) articles, any RoB domain could contribute equally to overall RoB assessment. For example, if any one domain was judged as having a high RoB, the whole study was considered to have a high RoB. Two of those 10 articles used numerical formulas for determining how many domains with high RoB need to be present to qualify the whole study as having a high RoB (e.g. "A decision to classify "overall bias" as low, unclear, or high was made by the reviewers using the following method: High: any trial with a high risk of bias listed on 3 or more domains.") (27).

Table 3. Specific quality threshold for sensitivity analysis used for different tools*

Study	Tool and threshold
Wong, 2013	Chorti et al criteria: The maximum score of the checklist is 26; 50% of maximum score is cut-off for high-quality study
Grant, 2016	Jadad: high risk of bias Jadad score <4
Johnson, 2007	Jadad: limiting the analysis to those studies with a Jadad score of at least 4
Raiman, 2016	Jadad: removing high bias studies (Jadad score <3)
Hamilton, 2011	Jadad: score 3 classified as a higher quality study
Hauser, 2011	Jadad: studies with a low (1 to 2) and moderate (3 to 5) Jadad score
Toner, 2017	Jadad: high-quality trials only (Jadad scale score, 4 to 5).
Aya 2013	Jadad: score >3 classified as a higher quality study
Morrison, 2013	Jadad: studies with low quality (Jadad score ≤3) vs studies with high quality (Jadad score >3)
Wang, 2009	Jadad: study quality (Jadad score ≥3 vs Jadad score ≤3)
Sanfilippo, 2017	Newcastle-Ottawa Scale tool: Low risk of bias score ranging between 6 and 9
Nagappa, 2017	Newcastle-Ottawa scale: good quality is score ≥8 of 9
Schnabel, 2011	Oxford scale: low quality study with 2 points
Schnabel, 2010	Oxford scale: the studies were rated as high (Oxford scale ≥3) or low (Oxford scale >3) quality studies.
Suppan, 2015	Oxford: lower quality studies (Oxford score <4)
Schnabel, 2012	Oxford: 'high quality': Oxford scale >3 versus 'low quality': Oxford scale 3 points
Schnabel, 2013a	Oxford: high-quality trials [modified Oxford scale >4] vs low-quality trials [modified Oxford scale ≤4
Schnabel, 2013b	Oxford: high-quality trials [modified Oxford scale >4] vs low-quality trials [modified Oxford scale ≤4
Mishriky, 2012	Oxford: restricting the analysis to studies with a modified Oxford score of 4 or higher

*References to all included studies are available in Supplementary file 1.

5. DISCUSSION

In a large sample of the systematic reviews and meta-analyses published from 2005-2018 in the highest-ranking pain/anesthesiology journals, the authors reported that they assessed quality/RoB in 76% of the articles. The most commonly used tools were the Cochrane RoB tool and Jadad tool, and some of the tools that the authors reported for assessing quality/RoB were not actually tools that are meant to be used for that purpose. A sensitivity analysis based on quality/RoB was performed in less than half of articles that reported using sensitivity analyses, and the thresholds for quality/RoB were highly inconsistent.

In 2016, Detweiler et al. published their report about the usage of RoB and methodological appraisal practices in SRs published in anesthesiology journals, in which they analyzed 207 SRs published from 2007 to 2015. In their analysis, the Jadad tool was the most commonly used for methodological assessment (25). On the contrary, in our analysis, which included SRs published from 2005 to 2018, with 678 analyzed articles, the Cochrane tool for RoB assessment was overall the most commonly used; our analysis shows that the usage of Cochrane tool for RoB assessment is increasing over time, and that popularity of the Jadad and Oxford scales is decreasing among SR authors. The Cochrane RoB tool 2.0 was announced recently, but none of the reviews included in our analysis have used it.

In most of the studies, a single quality/RoB assessment tool was used, but some studies used multiple tools. In recent years, the Cochrane Risk of Bias tool has become established in the assessment of RoB in randomized controlled trials (16). However, a significant variation can be observed for RoB assessment in non-randomized trials. It is especially important to assess RoB in observational studies because, unlike controlled experiments or well-planned, experimental randomized clinical trials, observational studies are subject to a number of potential problems that may bias their results (28). In a 2007 study, 86 tools comprising 53 checklists and 33 scales were found in the literature, following an electronic search performed in March 2005. The majority of those tools included RoB items related to study variables (86%), design-related bias (86%), and confounding (78%), although, for example, assessment of the conflict of interest was under-represented (4%). The number of items ranged from 3-36 (29). An analysis of SRs in the field of epidemiology of chronic disease indicated that only 55% of reviews addressed quality assessment of primary studies (30). An analysis of interventional SRs within the field of general health research and physical therapy showed that, in addition to the Cochrane RoB tool, 26 quality tools were identified, with an extensive item variation across tools (31).

Although it appears that the majority of the SRs in the highest-ranking pain journals do incorporate some kind of tool for appraising quality of evidence/risk of bias, about half of them then did not determine the level of quality of primary studies as a threshold for conducting numerical analyses and reaching conclusions. This may have directly influenced the conclusions that were derived from the evidence synthesis conducted within the SRs (32). To prevent biased conclusions based on studies with a flawed methodology, an acceptable threshold of study quality should be clearly specified, preferentially already in the initial SR protocol (33, 34).

In our study, less than half of the analyzed articles reported conducting a sensitivity analysis, and, most commonly, the sensitivity analysis was conducted to test the effect of quality/RoB on the results. Only half of the studies that used sensitivity analyses for quality/RoB have clearly specified a threshold for methodological quality, i.e. what was considered a high or low quality/RoB. Without a clear threshold for methodological quality, it is likely that different studies have different definitions of high and low quality, which may lead to different SR results and conclusions, which is not desirable and does not foster a reproducibility of the results and a consistency of assessment across different systematic reviews. This hypothesis is further confirmed by our findings that the studies where authors reported threshold for quality/RoB had a highly inconsistent approach, even when using the same tool.

Another issue is the diversity of the quality/RoB tools used for methodological quality assessment. These tools can be widely different and the levels of quality may not be comparable if different tools are used. For example, the Jadad scale has faced considerable criticism (16, 35). Furthermore, the Cochrane Handbook states for the Jadad scale that "the use of this scale is explicitly discouraged" because it suffers from the generic problems of scales, has a strong emphasis on reporting rather than conduct, and does not cover the allocation concealment aspect (36). Therefore, future SRs should avoid using the Jadad scale for assessing the methodological quality of included studies. As we can see from our results, the usage of both the Jadad and Oxford tools for methodological assessment is decreasing.

Some of the quality assessment tools reported in the SRs we analyzed are actually reporting guidelines/checklists for systematic reviews, such as QUOROM, PRISMA, or MOOSE. This indicates that not all authors of SRs are aware of the proper tools for the methodological assessment of SRs. Our study is, therefore, highlighting the possible lack of knowledge on research methodology among some review authors.

Although we have noted that half of the articles reported clear thresholds for sensitivity analysis related to a methodological assessment, even in those cases the authors rarely provided any more specific information about these thresholds, probably due to the -insufficient space and constraining word limits in journals. Namely, even if authors clearly describe that a study will be considered to have a high RoB based on the assessment of RoB in the 'random sequence generation' domain, it is still possible that the authors will erroneously assess RoB judgments. Our recent analyses of RoB assessments made by authors of Cochrane reviews showed that many Cochrane reviews have inadequate and inconsistent RoB judgments (37–41).

Our analysis of high-impact anesthesiology journals indicates a considerable inconsistency in the methods used for sensitivity analyses based on quality/RoB. Authors make an assessment of the overall risk of bias on the level of the whole study using different approaches, which may yield widely different conclusions. For example, it is not the same if the authors consider all RoB domains as equally contributing to the overall RoB of a study, or if they define certain key domains.

One solution for improving SRs in terms of their methodological assessments is to provide more detailed journal instructions for authors, where editors can indicate that all SRs need to conduct a methodological quality assessment of included studies and recommend adequate tools. Furthermore, editors and peer-reviewers analyzing submitted SRs should pay attention to adequate quality assessment and whether SRs with an included numerical analysis have conducted sensitivity analyses to account for the effect of study quality/RoB. Editors and peer-reviewers can request clear reporting of the methods that the authors have used. Editors are commonly perceived as gatekeepers protecting from the acceptance of low-quality manuscripts. Most authors will try to comply with editorial suggestions (42).

A limitation of our study is its reliance on reported data. The study authors were not contacted for clarifications regarding analyzed variables. Additionally, our study may be limited by publication bias, i.e. the fact that some results tend to be published in higher ranking journals, independent of the quality of research, just because of the direction of results. By analyzing the highest-ranking 25% of the journals in the chosen field, we may have introduced reporting bias ourselves. Furthermore, we limited our search to studies published from 2005 onwards, because methods for assessing RoB were developed relatively recently. We have searched for studies published in the targeted journals only via PubMed; it is possible that some relevant articles were missed due to erroneous indexing, and that we could have found additional relevant studies by employing additional search sources, such as hand-searching on

journal sites, or using another database. We did not include a librarian in designing our search strategy because the search for the targeted articles was simple, using the built-in filters.

Future studies should explore possible interventions for improving systematic review methodology in terms of its analyzing quality, including a sensitivity analysis for study quality, and clearly specifying a threshold. This methodological consistency will ensure a better comparability of the study results.

6. CONCLUSIONS

1. Our study indicates that a quarter of the SRs published in the highest-ranking pain journals published from 2005 to 2018 do not incorporate a methodological assessment of their included primary studies;
2. Among those with meta-analyses, a minority of the SRs had a sensitivity analysis for study quality/RoB performed, and, in only half of those, the methodological quality threshold criteria were clearly defined;
3. Without a consistent quality assessment and clear definitions of quality, untrustworthy evidence is piling up, in whose conclusions one cannot trust, much less safely implement it into clinical practice;
4. Systematic reviews need to appraise their included studies and plan sensitivity analyses because an inclusion of trials with a high RoB has the potential to meaningfully alter the conclusions;
5. The editors and peer-reviewers should act as gatekeepers protecting against the acceptance of systematic reviews that do not account for the quality of their included studies, and do not report their methods adequately, as well as help the authors to become aware of this crucial aspect of systematic review methodology.

7. REFERENCES

1. Eddy DM, Billings J. The quality of the medical evidence: Implications of care. Health Aff. 1988;7:19-32.
2. Guyatt G. Evidence-based medicine. ACP Journal Club Archives. 1991;114:2-4.
3. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: What it is and what it isn't. It's about integrating individual clinical expertise and the best external evidence. BMJ. 1996;312:71-2.
4. Wieten S. Expertise in evidence-based medicine: A tale of three models. Philos Ethics, Humanit Med. 2018;13:1-7.
5. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet. 2017;390:415-23.
6. Marusic M. Principles of Research in Medicine. 2nd ed. Zagreb: Medicinska naklada; 2016.
7. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. Chest. 1989;95(2 Suppl):2S-4S.
8. Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence - Publication bias. J Clin Epidemiol. 2011;64;1277-82.
9. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: An emerging consensus on rating quality of evidence and strength of recommendations. BMJ. 2008;336:924-6.
10. Murad MH, Asi N, Alsawas M, Alahdab F. New evidence pyramid. Evid Based Med. 2016;21:125-7.
11. Sackett DL, Rosenberg WMC. The need for evidence-based medicine. J R Soc Med. 1995;88:620-4.
12. Cochrane.org [Internet]. About us. 2020 [cited 2020 Jul 2]. Available from: https://www.cochrane.org/about-us.
13. Ioannidis JPA. The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. Milbank Q. 2016;94:485-514.
14. Page MJ, Shamseer L, Altman DG, Tetzlaff J, Sampson M, Tricco AC, et al. Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. PLoS Med. 2016;13:e1002028.
15. Paul M, Leibovici L. Systematic review or meta-analysis? Their place in the evidence hierarchy. Clin Microb Infect. 2014;20:97-100.

16. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011 [cited 2020 Jul 2]. Available from: www.handbook.cochrane.org.
17. Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med. 2012;157:429-38.
18. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA, editors. Cochrane handbook for systematic reviews of interventions Version 6.0 [updated July 2019]. Cochrane, 2019 [cited 2020 Jul 2]. Available from www.training.cochrane.org/handbook.
19. Higgins JP, Altman DG, Gotzsche PC, Juni P, Moher D, Oxman AD, et al. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. BMJ. 2011;343:d5928.
20. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:I4898.
21. Bown MJ, Sutton AJ. Quality control in systematic reviews and meta-analyses. Eur J Vasc Endovasc Surg. 2010;40:669-77.
22. Lohman D, Schleifer R, Amon JJ. Access to pain treatment as a human right. BMC Med. 2010;8:8.
23. Dosenovic S, Jelicic Kadic A, Miljanovic M, Biocic M, Boric K, Cavar M, et al. Interventions for neuropathic pain: an overview of systematic reviews. Anesth Analg. 2017;125:643-52.
24. Riado Minguez D, Kowalski M, Vallve Odena M, Longin Pontzen D, Jelicic Kadic A, Jeric M, et al. Methodological and reporting quality of systematic reviews published in the highest ranking journals in the field of pain. Anesth Analg. 2017;125:1348-54.
25. Detweiler BN, Kollmorgen LE, Umberham BA, Hedin RJ, Vassar BM. Risk of bias and methodological appraisal practices in systematic reviews published in anaesthetic journals: A meta-epidemiological study. Anaesthesia. 2016;71:955-68.
26. Cartes-Velásquez R, Manterola Delgado C, Aravena Torres P MCJ. Methodological quality of therapy research published in ISI dental journals: preliminary results. J Int Dent Med Res. 2015;8:46-50.
27. Meng H, Johnston B, Englesakis M, Moulin DE, Bhatia A. Selective cannabinoids for chronic neuropathic pain: a systematic review and meta-analysis. Anesth Analg. 2017;125:1638-52.

28. Hammer GP, Du Prel JB, Blettner M. Avoiding bias in observational studies: Part 8 in a series of articles on evaluation of scientific publications. Dtsch Arztebl. 2009;106:664-8
29. Sanderson S, Tatt ID, Higgins JPT. Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. Int J Epidemiol. 2007;36:666-76.
30. Shamliyan T, Kane RL, Jansen S. Systematic reviews synthesized evidence without consistent quality assessment of primary studies examining epidemiology of chronic diseases. J Clin Epidemiol. 2012;65:610-8.
31. Armijo-Olivo S, Fuentes J, Ospina M, Saltaji H, Hartling L. Inconsistency in the items included in tools used in general health research and physical therapy to evaluate the methodological quality of randomized controlled trials: A descriptive analysis. BMC Med Res Methodol. 2013;13:116.
32. Seehra J, Pandis N, Koletsi D, Fleming PS. Use of quality assessment tools in systematic reviews was varied and inconsistent. J Clin Epidemiol. 2016;69:179-84 e175
33. Moher D, Liberati A, Tetzlaff J, Altman DG, Altman D, Antes G, et al. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. PLoS Medicine. 2009;6:e1000097.
34. Booth A, Clarke M, Dooley G, Ghersi D, Moher D, Petticrew M, et al. The nuts and bolts of PROSPERO: An international prospective register of systematic reviews. Syst Rev. 2012;1:2.
35. Clark HD, Wells GA, Huët C, McAlister FA, Salmi LR, Fergusson D, et al. Assessing the quality of randomized trials: Reliability of the Jadad scale. Control Clin Trials. 1999;20:448-52
36. Higgins JPT, Altman DG, Sterne JAC. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011 [cited 2020 Jul 2]. Available from www.handbook.cochrane.org.
37. Barcot O, Boric M, Dosenovic S, Poklepovic Pericic T, Cavar M, Puljak L. Risk of bias assessments for blinding of participants and personnel in Cochrane reviews were frequently inadequate. J Clin Epidemiol. 2019;113:104-13.
38. Propadalo I, Tranfic M, Vuka I, Barcot O, Pericic TP, Puljak L. In Cochrane reviews, risk of bias assessments for allocation concealment were frequently not in line with Cochrane's Handbook guidance. J Clin Epidemiol. 2019;106:10-7.

39. Saric F, Barcot O, Puljak L. Risk of bias assessments for selective reporting were inadequate in the majority of Cochrane reviews. J Clin Epidemiol. 2019;112:53-8.
40. Babic A, Pijuk A, Brazdilova L, Georgieva Y, Raposo Pereira MA, Poklepovic Pericic T, et al. The judgement of biases included in the category "other bias" in Cochrane systematic reviews of interventions: a systematic survey. BMC Med Res Methodol. 2019;19:77.
41. Babic A, Tokalic R, Amilcar Silva Cunha J, Novak I, Suto J, Vidak M, et al. Assessments of attrition bias in Cochrane systematic reviews are highly inconsistent and thus hindering trial comparability. BMC Med Res Methodol. 2019;19:76.
42. Tsang EWK. Ensuring manuscript quality and preserving authorial voice: The balancing act of editors. Management and Organization Review. 2014;10:191-7.

8. SUMMARY

Objectives: A crucial element in the systematic review (SR) methodology is the appraisal of included primary studies, using tools for assessment of methodological quality or risk of bias (RoB). SR authors can conduct sensitivity analyses to explore whether their results are sensitive to exclusion of low quality studies or a high RoB. However, it is unknown which tools do SR authors use for assessing quality/RoB, and how they set threshold for quality/RoB in sensitivity analyses. The aim of this study was to assess quality/RoB assessment tools, the types of sensitivity analyses and quality/RoB thresholds for sensitivity analyses used within SRs published in high-impact pain/anesthesiology journals.

Materials and methods: This was a methodological study. We analyzed SRs published from January 2005 to June 2018 in the 25% highest-ranking journals within the Journal Citation Reports (JCR) "Anesthesiology" category. We retrieved the SRs from PubMed. Two authors independently screened records, full texts, and extracted data on quality/RoB tools and sensitivity analyses. We extracted data about quality/RoB tools, types of sensitivity analyses and the thresholds for quality/RoB used in them.

Results: Out of 678 analyzed SRs, 513 (76%) reported the use of quality/RoB assessments. The most commonly reported tools for assessing quality/RoB in the studies were the Cochrane tool for risk of bias assessment (N=251; 37%) and Jadad scale (N=99; 15%). Meta-analysis was conducted in 451 (66%) of SRs and sensitivity analysis in 219/451 (49%). Most commonly, sensitivity analysis was conducted to explore the influence of study quality/RoB (90/219; 41%) on the results. Quality/RoB thresholds used for sensitivity analysis for those studies were clearly reported in 47 (52%) articles that used them. The quality/RoB thresholds used for sensitivity analyses were highly heterogeneous and inconsistent, even when the same tool was used.

Conclusions: A quarter of SRs reported using quality/RoB assessments, and some that did cited tools that are not meant for assessing quality/RoB. Authors who use quality/RoB to explore the robustness of their results in meta-analyses use highly heterogeneous quality/RoB thresholds in sensitivity analyses. Better methodological consistency for quality/RoB sensitivity analyses is needed.

9. CROATIAN SUMMARY

Naslov: Metodološki alati i analiza osjetljivosti za procjenu kvalitete ili rizika od pristranosti korišteni u sustavnim pregledima objavljenim u anesteziološkim časopisima visokog odjeka

Ciljevi: Procjena primarnih istraživanja koristeći alate za procjenu metodološke kvalitete ili rizika od pristranosti (RP) ključan je element u metodologiji sustavnih pregleda. Autori sustavnih pregleda mogu provesti analize osjetljivosti kako bi istražili jesu li im rezultati osjetljivi na isključivanje studija niže kvalitete ili onih s visokim RP. Međutim, nije poznato koje alate autori sustavnih pregleda koriste za procjenu kvalitete ili RP, te kako postavljaju prag kvalitete ili RP pri analizama osjetljivosti. Cilj ovoga istraživanja bio je ispitati vrstu alata za procjenu kvalitete ili RP, analize osjetljivosti i pragove za kvalitetu ili RP koji se koriste u sustavnim pregledima u časopisima iz područja anesteziologije i boli visokog odjeka.

Materijali i metode: Ovo je bilo metodološko istraživanje. Analizirali smo sustavne preglede objavljene od siječnja 2005. do lipnja 2018. u časopisima koji se nalaze u prvih 25% rangiranih u kategoriji "Anesteziologija" baze Journal Citation Reports (JCR). Sustavne preglede smo prikupili s PubMed-a. Dvoje autora je nezavisno analiziralo zapise dobivene pretraživanjem, cjelovite tekstove i izvadilo podatke o upotrijebljenim alatima za kvalitetu ili RP i analizama osjetljivosti. Izvadili smo podatke o upotrijebljenim alatima za kvalitetu ili RP, vrstama analiza osjetljivosti i pragovima za kvalitetu ili RP.

Rezultati: Od 678 analiziranih sustavnih pregleda, 513 (76%) je navelo uporabu procjene kvalitete/rizika od pristranosti. Cochraneov alat za procjenu RP (N=251; 37%) i ljestvica Jadad (N=99; 15%) bili su najčešće navedeni alati za procjenu kvalitete ili RP. Meta-analiza je izvedena u 451 (66%) sustavnom pregledu, a analiza osjetljivosti u 219/451 (49%). Najčešći razlog za provođenje analize osjetljivosti bilo je istraživanje utjecaja kvalitete studije ili RP (90/219; 41%) na rezultate. Pragovi kvalitete ili RP korišteni u analizama osjetljivosti tih istraživanja bili su jasno navedeni u 47 (52%) istraživanja koja su analizu provodila. Pragovi za kvalitetu ili RP korišteni u analizama osjetljivosti bili su vrlo heterogeni i nedosljedni, čak i kada se radilo o uporabi iste vrste alata.

Zaključci: Četvrtina sustavnih pregleda navela je uporabu procjena kvalitete ili RP, a neki koji jesu, citirali su alate koji nisu namijenjeni za procjenu kvalitete ili RP. Autori koji koriste alate za procjenu kvalitete ili RP kako bi istražili snagu svojih rezultata u meta-analizama koriste vrlo heterogene pragove kvalitete ili RP u svojim analizama osjetljivosti. Nužna je metodološka dosljednost kod analiza osjetljivosti kvalitete ili RP.

10. CURRICULUM VITAE

Name and surname: Marija Franka Marušić
Date of birth: 1st of March, 1995
Place of birth: Zagreb, Croatia
Nationality: Croatian
Address: Poljička 28A, 21000 Split, Croatia
E-mail address: marija.franka@gmail.com

Education

2009-2013 High school education; Classical Gymnasium in Zagreb, Zagreb (Croatia)
2013-2014 Previous university education; University of Zagreb School of Humanities and Social Sciences, Zagreb (Croatia); Attended Comparative Literature and French Language
2014-2020 University of Split School of Medicine, Split (Croatia)

Languages

Fluent in English, French and German. Basic knowledge of Japanese.

Traineeships

2016 Internship; The Lancet, London (United Kingdom)
2018 Professional Medical Exchange; Kawasaki Medical University, Okayama (Japan)
2020 Traineeship; Centre for Global Health Research, Usher Institute, University of Edinburgh, Edinburgh (United Kingdom)

Activities

2016-2020 Active member of CroMSIC Split (Croatian Medical Students' International Committee), branch of IFMSA (International Federation of Medical Students' Associations)

Published articles

Marusic MF, Fidahic M, Cepeha CM, Farcas LG, Tseke A, Puljak L. Methodological tools and sensitivity analysis for assessing quality or risk of bias used in systematic reviews published in the high-impact anesthesiology journals. BMC Medical Research Methodology. 2020;20:121. doi: 10.1186/s12874-020-00966-4.

Awards

2018 Best oral presentation award at the ZIMS (Zagreb International Medical Summit)

11. SUPPLEMENTARY FILES

Supplementary files are available online, with the published article:

Marusic MF, Fidahic M, Cepeha CM, Farcas LG, Tseke A, Puljak L. Methodological tools and sensitivity analysis for assessing quality or risk of bias used in systematic reviews published in the high-impact anesthesiology journals. BMC Medical Research Methodology. 2020;20:121. doi: 10.1186/s12874-020-00966-4. (2019 Journal Impact Factor: 3.031).

https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-020-00966-4

Supplementary file 1:

Study protocol. This file includes the study protocol, which was defined a priori before commencement of the study.

Available at: https://figshare.com/articles/journal_contribution/Additional_file_1_of_Methodological_tools_and_sensitivity_analysis_for_assessing_quality_or_risk_of_bias_used_in_systematic_reviews_published_in_the_high-impact_anesthesiology_journals/12324731.

Supplementary file 2:

List of included studies. The file includes a list of systematic reviews/meta-analyses analyzed within this study, with their full bibliographic records.

Available at: https://static-content.springer.com/esm/art%3A10.1186%2Fs12874-020-00966-4/MediaObjects/12874_2020_966_MOESM2_ESM.docx.

Supplementary file 3:

Quality threshold for sensitivity analysis as described in the included studies. Threshold description includes all relevant information for sensitivity analysis from Methods, Results or Discussion, regardless of the part of manuscript where they were described.

Available at: https://static-content.springer.com/esm/art%3A10.1186%2Fs12874-020-00966-4/MediaObjects/12874_2020_966_MOESM3_ESM.docx.