Statistics in Urology

Key Steps in Conducting Systematic Reviews for Underpinning Clinical Practice Guidelines: Methodology of the European Association of Urology

Thomas Knoll\(^a,1\), Muhammad Imran Omar\(^b,1,\ast\), Steven Maclennan\(^b\), Virginia Hernández\(^c\), Steven Canfield\(^d\), Yuhong Yuan\(^e\), Max Bruins\(^f\), Lorenzo Marconi\(^g\), Hein Van Poppel\(^h\), James N’Dow\(^b\), Richard Sylvester\(^i\),

**EAU Guidelines Office Senior Associates Group Authorship**\(^2\)

\(^a\) Department of Urology, Sindelfingen-Boeblingen Medical Center, University of Tübingen, Sindelfingen, Germany; \(^b\) Academic Urology Unit, University of Aberdeen, Aberdeen, Scotland, UK; \(^c\) Department of Urology, Hospital Universitario Fundación Alcorcon, Madrid, Spain; \(^d\) Division of Urology, University of Texas McGovern Medical School, Houston, TX, USA; \(^e\) Department of Medicine, McMaster University, Hamilton, ON, Canada; \(^f\) Department of Urology, Radboud University Medical Center, Nijmegen, The Netherlands; \(^g\) Department of Urology, Coimbra University Hospital, Coimbra, Portugal; \(^h\) Department of Urology, University Hospital Gasthuisberg, Katholieke Universiteit Leuven, Leuven, Belgium; \(^i\) EAU Guidelines Office, Brussels, Belgium

**Article info**

**Article history:**
Accepted August 17, 2017

**Associate Editor:**
James Catto

**Keywords:**
Systematic review
Meta-analysis

**Abstract**

**Context:** The findings of systematic reviews (SRs) and meta-analyses (MAs) are used for clinical decision making. The European Association of Urology has committed increasing resources into the development of high quality clinical guidelines based on such SRs and MAs.

**Objective:** In this paper, we have summarised the process of conducting SRs for underpinning clinical practice guidelines under the auspices of the European Association of Urology Guidelines Office.

**Evidence acquisition:** The process involves explicit methods and the findings should be reproducible. When conducting a SR, the essential first step is to formulate a clear and answerable research question. An extensive literature search lays the foundation for evidence synthesis. Data are extracted independently by two reviewers and any disagreements are resolved by discussion or arbitration by a third reviewer.

**Evidence synthesis:** In SRs, data for particular outcomes in individual randomised controlled trials may be combined statistically in a meta-analysis to increase power when the studies are similar enough. Biases in studies included in a SR/MA can lead to either an over estimation or an under estimation of true intervention effect size, resulting in heterogeneity in outcome between studies. A number of different tools are available such as Cochrane Risk of Bias assessment tool for randomised controlled trials. In circumstances where there is too much heterogeneity, or when a review has included nonrandomised comparative studies, it is more appropriate to conduct a narrative synthesis. The GRADE tool for assessing quality of evidence strives to be a structured and transparent system, which can be applied to all evidence, regardless of

---

\(^1\) These authors are joint first authors.

\(^2\) The EAU Guidelines Office Senior Associates Group Authorship members are: Saeed Dabestani, Tobias Gross, Fabian Hofmann, Arjun Nambiar, Marc Schneider, Thomas Seisen, Lisette t’Hoen, Thomas van den Broeck, Erik Veskimae.

\(*\) Corresponding author. European Association of Urology, University of Aberdeen, Health Sciences Building (Second Floor), Foresterhill, University of Aberdeen, Aberdeen AB25 2ZD, UK.

E-mail addresses: m.i.omar@abdn.ac.uk, driman@yahoo.com (M.I. Omar).
quality. A SR not only identifies, evaluates, and summarises the best available evidence, but also the gaps to be targeted by future studies.

Conclusions: SRs and MAs are integral in developing sound clinical practice guidelines and recommendations.

Patient summary: Clinical practice guidelines should be evidence based, and systematic reviews and meta-analyses are essential in their production. We have discussed the key steps of conducting systematic reviews and meta-analyses in this paper.

© 2017 Published by Elsevier B.V. on behalf of European Association of Urology.

1. Introduction

In 1979, Archibald Cochrane, a Scottish doctor, proposed: “It is surely a great criticism of our profession that we have not organised a critical summary, by specialty or subspecialty, adapted periodically, of all relevant randomised controlled trials” [1]. Cochrane was one of the founding fathers of evidence-based medicine. He highlighted and advocated the importance of critically summarising the findings of research studies and was passionate about this cause. He designated the systematic review (SR) as a method providing such a summary [1]. Cochrane’s vision ultimately led to the development of the Cochrane Collaboration in 1993.

The findings of SRs and meta-analyses (MAs) are used for clinical decision making and are integral in developing sound clinical practice guidelines and recommendations. The European Association of Urology (EAU) has committed increasing resources into the development of high quality clinical guidelines based on such SRs and MAs [2–4]. The Guidelines Office has completed 21 SRs and 32 reviews are ongoing. These high quality systematic reviews are used for making guideline recommendations [5]. Supplementary Table 1 summarises the Guideline recommendation(s) underpinned by the selected SRs conducted under the auspices of the EAU Guidelines Office. In this paper, we have summarised the process of conducting SRs under the auspices of the EAU Guidelines Office. The process is also graphically illustrated in Figure 1.

2. Evidence acquisition

2.1. What is a randomised controlled trial?

A randomised controlled trial (RCT) is a type of study in which participants are randomly assigned to two or more groups in order to assess interventions. A well designed RCT is adequately powered and follows the principles and recommendations of the CONSORT statement [6].

2.2. What is a SR?

A SR identifies, selects, synthesises, and appraises studies that meet prespecified inclusion criteria to answer a research question. The process involves explicit methods and the findings should be reproducible. A SR not only identifies,
evaluates, and summarises the best available evidence, but also the gaps to be targeted by future studies [7].

2.3. What is a MA?

In SRs, data for particular outcomes in individual RCTs may be combined statistically in a MA to increase power when the studies are similar enough [7]. A MA may thus be able to answer a clinical question which is unanswerable by individual studies due to inadequate power. A MA should not be done outside of the SR setting because it is unlikely that all relevant trials will have been identified and therefore a biased or misleading pooled estimate may result. The term MA was coined by an American statistician, Gene Glass [8].

2.4. What is a traditional review?

A traditional review is usually written by an expert and provides a summary or an overview of a topic. Unlike a SR, the topic of a traditional review is often broad and less precisely defined. The methods used to conduct traditional reviews are not standardised. The search strategy is usually not stated, the review is confined to well-known articles [9], and the author’s personal beliefs may influence the overall conclusion [10]. If a MA is attempted in this situation, an exaggerated and spuriously precise estimate may be obtained and treated with greater credibility than it is due. The main features of SRs are covered throughout this report and the key differences between a SR and traditional review are summarised in Table 1.

2.5. Formulating a clear, well-designed research question and writing a review protocol

When conducting a SR, the essential first step is to formulate a clear and answerable research question. This clarifies the objectives and the study inclusion and exclusion criteria. A straightforward way to do this is to break the question down into constituent parts using the PICO framework, which stands for Participants, Intervention, Comparator, and Outcomes [11].

It is necessary to be transparent and explicit about which populations, intervention, comparator, and outcomes are to be included or excluded. This is important for not only designing the search strategy but is also helpful for systematic reviewers who do not necessarily have in-depth clinical knowledge on the topic. Furthermore, it serves as a blueprint to inform study screening and data extraction. To illustrate: “Should I perform partial nephrectomy on a patient with kidney cancer?” is a meaningful question, particularly for urologists. However, it is vague and not informative for deciding which studies should be included. A more scientific way for phrasing this question is: “In patients with localised renal cell carcinoma (P), how effective is partial nephrectomy (I), when compared with radical nephrectomy (C), in improving overall survival (O)?”

Of course, more clinically or methodologically justifiable detail will need to be provided to define limits on age, disease stage, which approaches to partial and radical nephrectomy are included, which other (secondary) outcomes should be included, and search date cut-offs. The types of study designs to be included should be specified. Cochrane SRs are traditionally confined to evidence from RCTs. However, it is not always ethical to randomise participants for comparing interventions and observational studies play a crucial role in this scenario.

SRs should be protocol driven and prospectively registered in a database, such as PROSPERO [12] or the Cochrane Library, to help counter publication bias and selective outcome reporting by giving a permanent record of the a priori methods. PROSPERO is a database maintained by the University of York. PROSPERO provides a free registration and publication of systematic review protocols and is available at the following link: https://www.crd.york.ac.uk/PROSPERO/. Registration of protocols also means that others can see if a review is currently underway and work is not duplicated. Crucially, in addition to the PICO elements, a protocol outlines what actions will be taken in which circumstances, such as how particular types of data will be handled in the analyses, the conditions under which MAs or narrative synthesis will be undertaken, which subgroups will be analysed, and how risk of bias (RoB) will be assessed.

<table>
<thead>
<tr>
<th>Table 1 – Differences between a systematic review and traditional review</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Systematic review</strong></td>
</tr>
<tr>
<td><strong>Review question</strong></td>
</tr>
<tr>
<td><strong>Protocol</strong></td>
</tr>
<tr>
<td><strong>Methods</strong></td>
</tr>
<tr>
<td><strong>Literature search</strong></td>
</tr>
<tr>
<td><strong>Critical appraisal</strong></td>
</tr>
<tr>
<td><strong>Synthesis</strong></td>
</tr>
<tr>
<td><strong>Findings/conclusion</strong></td>
</tr>
</tbody>
</table>
3. Evidence synthesis

3.1. Literature search

Extensive literature search lays the foundation for evidence synthesis. With increasing numbers of published research reports and numerous bibliographic databases, it is challenging to efficiently identify relevant studies, especially when research topics are complex and study designs are not limited to RCTs.

Conducting a literature search to identify relevant reports from databases and other resources requires rigorous search techniques. When designing searches, high sensitivity and high specificity should be aimed for. However, no search strategy is 100% perfect; high sensitivity may result in a huge number of irrelevant records and there is always a trade-off. It is advisable to involve an information scientist for designing search strategies.

Literature searches for SRs should involve at least two bibliographic databases and be as extensive as possible in order to identify all relevant studies and increase the generalisability of the results. According to Methodological Expectations for Cochrane Intervention Reviews, Cochrane Central Register of Controlled Trials, MEDLINE, (and EMBASE and Cochrane Review Group’s Specialized Register when available) are the minimum databases to be searched for a Cochrane SR [13]. Most published SRs now search at least MEDLINE, EMBASE, and Cochrane library databases. Databases relevant to the topic of the review (eg, CINAHL for nursing-related topics), national and regional databases, and subject-specific databases should also be considered. Trials registers such as ClinicalTrials.gov, the World Health Organisation International Clinical Trials Registry Platform portal, and other sources (eg, pharmaceutical industry and regulatory agencies websites) should be searched for unpublished and ongoing studies. Other sources include dissertations, theses, conference abstracts and proceedings, and grey literature (information produced in electronic and print formats not controlled by commercial publishing) databases. Searching the bibliographies of identified relevant studies and previous SRs on the same topic is a good approach for reducing the chance of missing relevant studies. Study authors and organisations should be contacted for missing information on unpublished or ongoing studies [7,13].

Content experts are very important in designing advanced search strategies [7]. By working closely with information specialists, they can assist with identifying search concepts, and suggest keywords as well as key references, while revising and improving preliminary searches. Too many different search concepts should be avoided. Sometimes, some concepts (eg, outcomes) may not be well described in the title or abstract of an article and are often not well indexed with controlled vocabulary terms; it might be better to search only for the population and intervention. However, a broad range and wide variety of search terms should be combined with “OR” within each concept. Both free-text words (including spelling variants, synonyms, related terms, opposites, plurals, acronyms, truncations, wildcards, and proximity operators) and appropriate subject headings should be used (eg, MeSH and Emtree). Ensure the Boolean operators “AND,” “NOT,” and “OR” are used correctly. An example of a search strategy, along with an explanation, is presented in Figure 2. Strategies must be translated for every database and each interface (eg, PUBMED and OvidSP are two different interfaces for the same database MEDLINE). Specially designed high sensitivity and tested filters can be used when appropriate. There is no need to use an RCT filter when searching Cochrane Central Register of Controlled Trials because all the records are thoroughly and correctly indexed. Applying a publication date, publication format, or language restrictions in the search strategy should be justified and included in the report [7].

The search strategy should be documented and reported transparently in the protocol and in the review. It is necessary to report the sources searched (databases and interfaces), dates and limits, and the numbers found from each source [14,15]. Key search terms can be reported in the methods section. The full search strategies for one or more electronic databases can be submitted as an appendix and published as supplementary files available online to readers.

3.2. Abstract and full-text screening

After the literature search has been conducted, the list of retrieved abstracts is screened for studies that may be eligible for inclusion. Abstract screening is performed using an a-priori developed study screening form. This form contains the inclusion and exclusion criteria for each PICO element as reported in the study protocol. Following this form will lead to the decision whether to exclude or include the abstract. In case no decision can be made based on the information in the abstract, it will be marked as “unclear.” Unclear abstracts are included at this stage, as it requires full-text retrieval and review to make the final decision. Abstract screening must be performed independently by (at least) two reviewers. The combined results will subsequently be reviewed for disagreements, and a third independent reviewer can be consulted for resolving conflicts.

Full-text screening follows the same principles as abstract screening. A difference, however, is that full-text papers cannot be marked as “unclear” but require a definitive decision whether to include or exclude the study. Also, for excluded full-text papers, reviewers must keep track of the reason for exclusion. Upon completion, the final number of included and excluded studies (with the reason for exclusion) is reported in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses flow diagram as illustrated in Figure 3.

3.3. Data extraction from included studies

One of the most important and time-consuming parts of a SR is data extraction. Data are extracted independently by two reviewers and any disagreements are resolved by discussion or arbitration by a third reviewer [16]. Table 2 summarises the key information that should be extracted.
from the studies included in a SR. Each review question is unique and therefore requires development of a separate form. A pilot tested electronic or paper version of the data extraction form includes entries for study characteristics along with relevant results and RoB assessment. The data extraction form should first be piloted on two to three studies.

The setting and timing of the study, participants, and disease characteristics (such as age, sex, comorbidities, diagnostic criteria, staging, and prognostic factors) that may influence intervention effects or external validity should be collected. Relevant information on interventions include surgical techniques, drug doses and frequency, or routes of delivery. It is important to clearly predetermine the outcome measures that will be collected in terms of their definition (eg, measurement method, scale, and threshold), timing, and unit(s) of measurement. This is crucial as many outcomes may be reported using multiple definitions (eg, biochemical recurrence after radical prostatectomy), measurement scales (eg, multiple urinary symptom questionnaires to access the severity of urinary incontinence), measurement method (eg, healthy parenchymal volume loss after partial nephrectomy measured by computed tomography volumetric analysis or estimated by surgeon), or different points in time (eg, cancer-specific survival at 1 yr, 5 yr, and 10 yr). Standardised sets of outcomes, known as “core outcome sets” represent the minimum that should be measured and reported in all clinical trials of a specific condition making it easier for the results of trials to be combined [17]. The numerical data required for MA are not always available and sometimes other statistics or graphical information can be collected and converted into the required format.

3.4. RoB assessment

Biases in studies included in a SR can lead to either an overestimation or an underestimation of true intervention effect size, resulting in heterogeneity in outcome between studies.

3.4.1. RCTs

In SRs of RCTs, the RoB in each study is assessed for each outcome using the domains (judged as either low, high, or unclear) in the Cochrane Collaboration’s RoB Tool [7,18]:

- Selection bias, which includes random sequence generation bias and allocation concealment bias.
- Performance bias due to knowledge of the allocated intervention by study personnel and participants.
- Detection bias in assessing the outcome due to knowledge of the allocated intervention.
- Attrition bias due to incomplete data and exclusion from analyses.
- Reporting bias due to selective outcome reporting.
- Other sources of bias.

RevMan is then used to create a RoB summary graph to visually illustrate the RoB (low, high, or unclear) for each domain in each study [19]. A draft version of a revised tool to assess the RoB in RCTs (RoB 2.0) has recently been proposed [20].

Table 2 – Information to consider in data extraction form

<table>
<thead>
<tr>
<th>Data extraction form field</th>
<th>Information to consider in data extraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reviewer identification</td>
<td>Review author ID; date</td>
</tr>
<tr>
<td>Study identification</td>
<td>Study ID; report ID; citation; author contact details; publication yr; country; source of data</td>
</tr>
<tr>
<td>Methods</td>
<td>Study design; setting; enrolment period; no. of centres</td>
</tr>
<tr>
<td>Participant characteristics</td>
<td>Total no.; age; sex; co-morbidity; ethnicity; no. lost to follow-up</td>
</tr>
<tr>
<td>Disease characteristics</td>
<td>Staging (eg, TNM); severity (eg, IPSS score); biological behaviour (eg, Gleason score); surgical complexity (eg, R.E.N.A.L. score); method of diagnosis</td>
</tr>
<tr>
<td>Interventions&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Surgical technique; surgeon experience/volume; drug dose, route of delivery and length</td>
</tr>
<tr>
<td>Diagnostic test characteristics&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Description of the reference standard, index test, comparator, manufacturer; interpreter of diagnostic test</td>
</tr>
<tr>
<td>Prognostic factor characteristics&lt;sup&gt;c&lt;/sup&gt;</td>
<td>Dose, level, duration of exposure; method of measurement</td>
</tr>
<tr>
<td>Outcome</td>
<td>Outcome definition (incl. unit of measurement, scale, assessor, time point of measurement)</td>
</tr>
<tr>
<td>Results to include in a meta-analysis</td>
<td>Dichotomous outcomes: no. of events/no. of participants</td>
</tr>
<tr>
<td></td>
<td>Continuous outcomes: mean value and SD in each intervention group</td>
</tr>
<tr>
<td></td>
<td>Time-to-event outcomes: HR (with 95% CI)</td>
</tr>
<tr>
<td></td>
<td>Diagnostic test performance outcomes: TP, FP, TN, FN</td>
</tr>
<tr>
<td>Risk of bias</td>
<td>Cochrane RoB tool for RCTs or other tools such as QUIPS and QUADAS-2</td>
</tr>
<tr>
<td>Other</td>
<td>Review author comments</td>
</tr>
</tbody>
</table>

CI = confidence interval; FN = false negative; FP = false positive; HR = hazard ratio; ID = identification; IPSS = International Prostate Symptom Score; MA = randomised controlled trial; RoB = risk of bias; SD = standard deviation; TN = true negative; TP = true positive.

<sup>a</sup> For comparative effectiveness of interventions systematic reviews.

<sup>b</sup> For diagnostic test accuracy systematic reviews.

<sup>c</sup> For prognostic factor systematic reviews.

3.4.2. Nonrandomised comparative studies

Potential biases are greater in nonrandomised comparative studies (NRCS) than for RCTs due to the high risk of confounding. NRCS are thus always considered to be at high risk of selection bias.

In addition to assessing the above domains for RCTs, a maximum of the five most important confounders should be prespecified. For each outcome, an assessment is made of: (1) whether the confounder was considered, (2) whether its distribution was balanced, and if not then (3) whether it was controlled for in the analysis.

This information is used to reach an overall decision about the RoB for each confounder and each outcome. The RoB summary graph can also include the additional confounder information.

Recently, ROBINS-I, a tool for assessing risk of bias in nonrandomised studies of interventions has been published. This tool is not yet tested within the EAU [22].

3.4.3. Single arm studies

In SRs of single arm case series, five aspects are considered:

1. Was there an a priori protocol?
2. Was the total population included or were study participants selected consecutively?
3. Was outcome data complete for all participants and any missing data adequately explained/unlikely to be related to the outcome?
4. Were all prespecified outcomes of interest and expected outcomes reported?
5. Were primary benefit and harm outcomes appropriately measured?

If the answer to all five questions is “yes,” the study is at “low” RoB. If the answer to any question is “no,” the study is at “high” RoB.

QUADAS-2 is used for assessing RoB in diagnostic test accuracy SRS [23], while QUIPS is used for prognostic factor SRS [24].

3.4.4. RoB assessment process

Two reviewers independently assess the RoB in each study when extracting data. A third reviewer acts as an arbiter. If the RoB is unclear in any domain, an attempt is made to obtain the study protocol or to contact the study authors. RoB summary graphs are created with separate graphs for each of the three different study types. Biases are described in the SR report, emphasising areas of concern and the possible effect of bias in interpreting the results [18].

3.5. MA or narrative synthesis (including subgroup analyses, assessment of heterogeneity, and publication bias)

Combining information from different studies included in a SR is known as evidence synthesis. Depending on the homogeneity of the populations, interventions, outcome definitions, and measurements in the included studies, the evidence synthesis may be narrative or quantitative.

MA is feasible when more than one RCT reporting the same outcome is identified [25]. MA provides an overall estimate of the effect of the treatment or the intervention analysed. The effect of the treatment refers to the differences in outcomes found between an intervention and comparator. This effect, along with the precision of the estimate, needs to be extracted from all the individual studies, so that the consistency in the direction and magnitude of effect across studies can be assessed [26].

However, before meta-analysing data, a detailed assessment of the included studies is crucial to decide whether a MA of the results is not only feasible but also sensible. First, the internal validity and precision of the included studies will directly impact the quality of a MA. If the included
studies are biased, underpowered, or the individuals estimates inconsistent, then performing a MA can potentially give a spuriously precise and often meaningless summary statistic.

One should take into account that the studies included in a SR will inevitably differ. Despite clearly defined eligibility criteria, differences in the included studies will exist with regards to natural variation in the populations and fidelity to the interventions. Those variations are known as heterogeneity [27]. There are three main types of heterogeneity: (1) clinical heterogeneity, which refers to variations in the populations, interventions or the way in which the outcomes were assessed, (2) methodological heterogeneity, which refers to differences in study design and reporting biases, (3) and statistical heterogeneity, which refers to variability in treatment effects due to the way the data are analysed and reported in each study.

There are several ways to investigate heterogeneity. First, heterogeneity can be assessed by visual inspection of the forest plots. Poor overlap among confidence intervals may indicate statistical heterogeneity. The chi-square Q statistic assesses the degree to which the individual study estimates of effect size differ from the overall estimate, while the I² statistic is an estimate of the percent of variation attributable to heterogeneity, that is, the ratio of true heterogeneity to total observed variation [7,27,28]. An I² greater than 50% is considered substantial. MA and the estimation of an overall summary effect in the presence of significant heterogeneity may be misleading, but several options can be considered in this scenario. Some heterogeneity may be explained by the intervention having different effects in, for example, older populations or at higher doses of a drug. In such instances, the stratification of studies according to their characteristics or subgroup analyses in specific patient populations may be done to explore heterogeneity or to try to answer specific questions; however, this may not be possible without individual patient data. Such analyses should be regarded as exploratory and the results must be interpreted with caution [29,30]. They should be planned at the protocol stage rather than in the light of the results, otherwise the conclusions may not be reliable [31].

Publication bias should also be considered. Trials showing statistically significant results are more likely to be published than those with negative findings [32]. They are also more likely to be published quickly, in more than one place, in English, in high impact, indexed journals, and cited by others [7]. Funnel plots, which plot the effect estimate of individual studies against the size of the study, may be used to identify the publication bias but these should only be used when there are at least 10 studies in the MA.

In circumstances where there is too much heterogeneity, or when a review has included NRCS, it is more appropriate to conduct a “narrative synthesis” [7,33,34]. This is a theory driven and iterative approach for summarising and imposing structure on the results of the included studies. Four main stages are involved:

1. Develop a theory of how the intervention works, why, and for whom.
2. Develop a preliminary synthesis of the findings of the included studies.
3. Explore relationships within and between studies.
4. Assess the robustness of the synthesis [34].

Essential activities include tabulating and describing the key baseline characteristics and results of individual studies. Then studies with similar populations, intervention, and outcome measurement characteristics can be grouped together and then further grouped with regards to the direction and magnitude of the results. The results may then be interpreted critically with regards to the theory, considering also the RoB and potential confounders within and across the included studies to make statements about the quality of the evidence about the intervention. Figure 4 illustrates a decision flow diagram when to meta-analyse data or to use narrative synthesis.

3.6. Quality of evidence (Grading of Recommendations, Assessment, Development, and Evaluation)

The Grading of Recommendations, Assessment, Development, and Evaluation (GRADE) Working Group, in response to the need for an improved system of evidence quality assessment, developed a tool in 2000 using input from health professionals, researchers, and guideline developers around the world [35]. The GRADE tool for assessing quality of evidence strives to be a structured and transparent system which can be applied to all evidence, regardless of quality. There are a few basic tenets of evidence quality using GRADE [36]. The final quality assessment, which applies to the body of evidence (all studies included in the SR) is reported as one of four possible levels: high, moderate, low, or very low. Foremost is that RCTs start out as high quality and that nonrandomised (observational) studies start out as low quality. According to GRADE, there are five factors which can affect overall evidence quality which starts out high in a negative way, and three factors which can affect quality which starts low in a positive way [37].

Starting at potentially high quality, RCTs are susceptible to certain biases which, when present, can raise doubt in the study findings. Once summarised, these inherent study elements inform the overall RoB of all the studies included in the analysis. This overall study “RoB” is the first of the five “negative” factors used in GRADE to assess the overall quality of the evidence. Next are items which assess how the overall evidence is represented and balanced. “Inconsistency” examines the direction of different study results. For example, if one study shows an intervention is helpful while another does not, they are inconsistent. “Indirectness” examines the details of exactly what is compared in the various studies being reviewed. If studies use different intervention methods, they may be indirectly comparing results. “Imprecision” examines the margin of potential error within all combined studies. All of these factors may raise doubts about the validity of the combined data, especially when trying to use it for a recommendation. Finally, if small studies showing negative results are missing
from the literature, potential “publication bias” becomes a concern as well [35,37,38].

If observational studies are incorporated in a SR, inherent study bias is already assumed and the overall evidence quality starts out as “low.” However, if the combined effect size of the outcome is undeniable (“large” > 50%, “very large” > 75%), if there is an obvious dose response seen, or if there is evidence that the plausible confounders would...
diminish the demonstrated effect, the quality of this evidence can actually be raised. Evidence from observational studies that has been downgraded for any reason (as explained above) should not be upgraded. The process of assessing the quality of evidence is summarised in Figure 5. Ultimately this process strives to be structured, systematic, and transparent.

3.7. Incorporation of systematic review results in guidelines

The aim of the EAU Guidelines is to assist practicing clinicians in making informed decisions in different clinical situations, taking the highest quality scientific data, their patient’s personal circumstances, values, and preferences into account. The development of the guidelines is a core activity of the EAU as part of their educational efforts and the available guidelines cover most of the urological field. A transparent production process and continuous updating are assured to ensure that these guideline documents are, and remain, of high value to their users. High quality SRs are a superb source of data when composing clinical guidelines. The EAU Guideline Office made a major step forward when deciding not only to use such SRs for the guidelines, but also to perform their own SRs based on PICO questions identified and defined by the expert panels. The SRs are performed in a standardised manner, following Cochrane methodology, with inclusion of expert urologists, associates, methodologists, statisticians, and medical librarians. The results are published in peer-reviewed journals and are included in the annual updates of the specific guideline.

3.8. Recent technological advancement

The process of conducting a SR is laborious. However, a number of technological advancements have made the process of conducting a SR easier. For example, software which can read graphs such as survival curves and help with their interpretation are available. Plot Digitizer is one such tool and is freely available online [39]. In addition, RevMan now contains an online calculator for imputing missing data. The Cochrane Collaboration has continually adopted technology to simplify the process of conducting SRs. As part of The Cochrane Collaboration’s Strategy to 2020 initiative, the Informatics and Knowledge Management section is working on a number of projects such as the PICO annotation project and the Linked Data project. A number of other organisations have also developed tools for making the process of conducting SRs more efficient. Some of the available tools are: Abstrackr, Covidence, DistillerSR, Eppi-Reviewer 4, EROS, ExaCT, Rayyan, RevMan HAL, SUMARI, and TrialState SRS 4.0 [40]. Two of these tools are recommended by The Cochrane Collaboration (Covidence and Eppi-Reviewer). These packages can speed up the review process and some assist by semi-automating key SR tasks.

4. Conclusions

Projects like AllTrials will provide access to research data [41]. This will eventually reduce research wastage and will provide a source of “Big Data.” The raw data from individual trials can be used for Individual Patient Data MA, the gold standard. Text mining technologies are also explored which may reduce the reviewers’ time through automating the key tasks such as citation screening. Data extraction with data mining technology, data-tagging, and translations have also been explored [42]. Automation of review process tasks could help us to achieve “living systematic reviews,” that is, SRs which are continuously updated incorporating relevant data as it becomes available. The use of technology in the SR process will reduce time from the completion of a study to incorporating the findings within the review, and evidence will be more reliable and readily available [42].

Author contributions: Muhammad Imran Omar had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Omar, Knoll, N’Dow, Sylvester.

Acquisition of data: Omar, MacLennan, Marconi.

Analysis and interpretation of data: Omar, Knoll, MacLennan, Hernández, Canfield, Yuan, Bruins, Marconi, Van Poppel, N’Dow, Sylvester, EAU Guidelines Office Senior Associates Group Authorship.

Drafting of the manuscript: Omar, Knoll, MacLennan, Hernández, Canfield, Yuan, Bruins, Marconi, Van Poppel, N’Dow, Sylvester.

Critical revision of the manuscript for important intellectual content: Omar, Knoll, MacLennan, Hernández, Canfield, Yuan, Bruins, Marconi, Van Poppel, N’Dow, Sylvester, EAU Guidelines Office Senior Associates Group Authorship.

Statistical analysis: None.

Obtaining funding: None.

Administrative, technical, or material support: None.

Supervision: Omar.

Other: None.

Financial disclosures: Muhammad Imran Omar certifies that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (e.g., employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: None.

Acknowledgments: Figure 5 was created by the GRADE Working Group. We are grateful to Professor Holger Schünemann, from McMaster University, for allowing us to use this figure.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.eururo.2017.08.016.

References


