Conflict of Evidence

Elsevier Editorial System(tm) for European Urology

Manuscript Draft

Manuscript Number: EURUROL-D-16-01016R2

Title: Conflict of Evidence: Resolving discrepancies when findings from Randomized Controlled Trials and Meta-analyses disagree

Article Type: Review Paper

Section/Category: Statistics in Urology (STA)

Keywords: conflict of evidence; meta-analyses; randomized controlled trials; systematic reviews; treatment guidelines

Corresponding Author: Dr. Richard J Sylvester, ScD

Corresponding Author's Institution: EAU Guidelines Office

First Author: Richard J Sylvester, ScD

Order of Authors: Richard J Sylvester, ScD; Steven E Canfield; Thomas B Lam; Lorenzo Marconi; Steven MacLennan; Yuhong Yuan; Graeme MacLennan; John Norrie; Muhammad Imran Omar; Harman M Bruins; Virginia Hernández; Karin Plass; Hendrik Van Poppel; James N'Dow

***Revision notes**

Reply to Reviewer Comments

Reviewer #1: Thank you for the opportunity to re-review this paper after revisions. Again, Sylvester et al are to be congratulated for this remarkable contribution to European Urology.

The authors have not indicated on their revised manuscript where the revised changes are, making it difficult to find them. Overall, I believe the authors have sufficiently addressed Reviewer #1's suggestions/comments.

Reply: Thank you. Both the text of the revisions and the line numbers where changes were made to the manuscript were included in the reply to the reviewers.

However for Reviewer #2, there are some inadequate responses from the authors.

Reviewer 2 point 4

- I believe that the authors should include a discussion on other types of meta-analyses, such as diagnostic test accuracy (e.g. PMID 27363387), prognostic factors (e.g. PMID 25559810), and even that of retrospective studies (e.g. PMID 24680361).

- Again, despite what the authors feel about observational data, these represent real-world comparative effectiveness data that are typically of the patients we treat and therefore such data is practical, useful and believable to us as clinicians.

- Additionally, for some rarer diseases such as UTUC, there just are not any RCTs and the best level of evidence will be a meta-analysis of all available retrospective studies.

Reply:

There are 6 important areas to consider when evaluating the validity and risk of bias in studies of prognostic factors (QUIPS): study participation, study attrition, prognostic factor measurement, study confounding, outcome measurement, and analysis and reporting. In order to minimize the risk of bias, prognostic factor studies to be included in a meta-analysis should preferably be prospective and have a protocol which addresses these topics.

Reference:

Hayden JA, van der Windt DA, Cartwright JL, Côté P, Bombardier C. Assessing bias in studies of prognostic factors. Ann Intern Med. 2013; 158:280-6.

For diagnostic test accuracy studies, QUADAS-2 provides a tool for the quality assessment of diagnostic test accuracy studies which comprises 4 domains for assessing the risk of bias: patient selection, index test, reference standard, and flow and timing. Once again, in order to minimize the risk of bias, diagnostic test accuracy studies to be included in a meta-analysis should preferably be prospective and have a protocol which addresses these issues.

Reference:

Whiting PF, Rutjes AWS, West ME. QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies. Ann Intern Med. 2011; 155:529-536.

For prognostic factor and diagnostic test accuracy systematic reviews, we agree with the reviewer that randomized controlled trials are not required, however the individual studies included in the meta-analysis should preferably be prospective in nature and have a protocol in order to minimize the risk of bias.

Since the manuscript deals primarily with discrepancies between intervention RCTs and meta-analyses, including meta-analyses involving diagnostic test accuracy and prognostic factor studies goes beyond the scope of the paper, notwithstanding the impact on the word count. Nevertheless, we have, as indicated below, added a sentence concerning them to the Discussion.

We do not agree with the reviewer, however, that non-randomized comparative studies (whether prospective or retrospective) or observational case series should be included in meta-analyses of interventions because of the high risk of bias. Included in a qualitative systematic review, yes, but not included in a quantitative meta-analysis. The reasons for this position have already been outlined in our previous responses and are further discussed below. While we accept that some referees might have a different opinion, as a guideline authority we believe that this is an extremely important principle to uphold.

For non RCT intervention effectiveness systematic reviews, one should present the results of the individual studies from a narrative point of view, in descriptive tables or even in forest plots, but the results of the individual studies should not be combined together in a formal meta-analysis to produce the diamond at the bottom of the forest plot.

Although Stroup et al (MOOSE) provide a Reporting Checklist for Authors, Editors, and Reviewers of Meta-analyses of Observational Studies, they state in the Comment:

“The application of formal meta-analytic methods to observational studies has been controversial. One reason for this has been that potential biases in the original studies, relative to the biases in RCTs, make the calculation of a single summary estimate of effect of exposure potentially misleading. Similarly, the extreme diversity of study designs and populations in epidemiology makes the interpretation of simple summaries problematic, at best. In addition, methodologic issues related specifically to meta-analysis, such as publication bias, could have particular impact when combining results of observational studies.”

For example, the paper “Overall Survival Advantage with Partial Nephrectomy: A Bias of Observational Data?” by Shuch et al (reference 48), illustrates our concerns about bias when comparing partial nephrectomy to radical nephrectomy based on non RCT studies:

“CONCLUSIONS: RN patients had similar OS compared with controls, suggesting that this treatment modality does not compromise survival. Patients undergoing PN had improved OS compared with controls, suggesting possible selection bias. The apparent survival advantage conferred by PN in SEER-Medicare case series is likely the result of selection bias involving unmeasured confounders.”

We thus feel that the risk of bias is too high in non RCT intervention effectiveness meta-analyses (where a formal risk of bias assessment of the individual studies isn’t always done) for their conclusions to directly impact on treatment recommendations and guidelines. Most readers will not be aware of their limitations. We believe that it is better to present such results in a qualitative systematic review rather than to run the risk of publishing incorrect or misleading results in a meta-analysis that may steer further research in the wrong direction or adversely impact on patient care.

- This is a paper submitted under Statistics in Urology, therefore the reply that "majority of whom do not have advanced statistical knowledge or experience" does not seem to be appropriate or accurate.

Reply: It was submitted under Statistics in Urology for the lack of a better category. The most appropriate category would have been Guidelines, however this category does not exist. The paper is aimed at clinicians and guidelines developers and not at statisticians. In any case, the majority of readers, including those who read articles under the topic of Statistics in Urology, are urologists who do not have advanced statistical knowledge or experience.

- This point should be addressed and included in the manuscript, rather than brushed aside, given the substantial proportion of systematic reviews and meta-analysis of not just RCTs, but other types of studies.

Reply: As indicated in the paper's title, the scope and subject of the paper is to resolve discordant findings between RCTs and meta-analyses. It was not our intent to deal with prognostic factor or diagnostic test accuracy meta-analyses, but only with intervention meta-analyses. Nevertheless, in accordance with the reviewer's comments, the following modifications been made to the manuscript:

Lines 321 – 323:

It is important to reiterate that combining observational studies in general, and even comparative non-randomized studies with RCTs in an intervention MA, may produce unreliable results and is not considered valid.

In addition, the following text has been added in lines 330 – 333:

Although non RCTs can be included in SRs, we have emphasized that only RCTs should be included in intervention MAs. RCTs are not required for prognostic factor and diagnostic test accuracy MAs, however the studies included in these MAs should preferably be prospective in nature and based on a protocol to minimize risk of bias.

Reviewer #2:

The authors have addressed most of my concerns. While I disagree with some of their responses, like those to questions #3, #4, and particularly #6, I think their responses are well thought out and certainly reasonable. I do feel, however, that these responses sure smell like those coming primarily from individuals that do not treat many patients.

Reply: Seven of the 14 co-authors are urologists who regularly treat patients.

The ITT vs PP analysis problem in my opinion clearly is best managed by presenting both results. How can one argue otherwise (i.e. for a less complete revelation of the data)?

Reply: Unfortunately the risks associated with the results of a PP analysis are not often presented in the paper. Nevertheless, lines 93 - 96 have been modified as follows:

In some RCTs, not all participants receive their randomized intervention; they may, for example, cross-over to the other randomized treatment, in which case a per-protocol analysis may also provide useful information.

Similarly, the hierarchy of evidence is actually not based in evidence! I wonder what would happen if we randomized patients to be treated by statistical robots or by experienced physicians? I bet the robots miss the boat because of the innumerable immeasurables that physicians, and not data-analysts, recognize and utilize. There is a reason some MDs get better results than others, and it is not better access to trial data.

Reply: Yes, quality of results by MD or by institution is an important topic, and variations in outcomes may be linked to pre-existing experience, education, training, one's innate ability to learn and adapt, institutional support and other elements of the learning curve. See, for example, the conclusions of the following paper:

https://www.ncbi.nlm.nih.gov/pubmed/12074794

Take Home Message

New or existing RCT data can lead to conflicts with MA data. In this paper, we present examples of, and explore reasons for, such conflicts. Guidance is provided to guideline developers on how to assess conflicting data in such circumstances to help determine which source is more reliable. For guideline organizations, both within and outside of urology, having a well-defined and robust process to deal with such conflicts is essential to improve the quality of their guidelines.

Conflict of Evidence: Resolving discrepancies when findings from Randomized Controlled Trials and Meta-analyses disagree

Richard J. Sylvester, EAU Guidelines Office, Brussels, Belgium

Steven E. Canfield, Division of Urology, University of Texas McGovern Medical School, Houston, Texas, USA

Thomas B. L. Lam, Academic Urology Unit, University of Aberdeen, Aberdeen, Scotland, UK

Lorenzo Marconi, Department of Urology, Coimbra University Hospital, Coimbra, Portugal

Steven MacLennan, Academic Urology Unit, University of Aberdeen, Aberdeen, Scotland, UK

Yuhong Yuan, Department of Medicine, McMaster University, Hamilton, ON, Canada

Graeme MacLennan, Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland, UK

John Norrie, Health Services Research Unit, University of Aberdeen, Aberdeen, Scotland, UK

Muhammad Imran Omar, Academic Urology Unit, University of Aberdeen, Aberdeen, Scotland, UK

Harman M. Bruins, Department of Urology, Radboud University Medical Center, Nijmegen, The Netherlands

Virginia Hernández, Department of Urology, Hospital Universitario Fundacion Alcorcon, Madrid, Spain

Karin Plass, EAU Central Office, Guidelines Office, Arnhem, The Netherlands

Hendrik Van Poppel, Department of Urology, University Hospital Gasthuisberg, Katholieke Universiteit Leuven, Leuven, Belgium

James N'Dow, Academic Urology Unit, University of Aberdeen, Aberdeen, Scotland, UK

Abstract: 299 words

Text: 3898 words

Keywords: conflict of evidence; meta-analyses; randomized controlled trials; systematic reviews; treatment guidelines

Abstract

Context: Clinicians and treatment guideline developers are faced with a dilemma when the results of a new, large, well conducted, randomized controlled trial (RCT) are in direct conflict with the results of a previous systematic review (SR) and meta-analysis (MA).

Objective: To explore and discuss the possible reasons for disagreement in the results from SRs/MAs and RCTs and to provide guidance to clinicians and guideline developers for making well informed treatment decisions and recommendations in the face of conflicting data.

Evidence Acquisition: The advantages and limitations of RCTs and SRs/MAs are reviewed. Two practical examples which have a direct bearing on EAU guidelines treatment recommendations are discussed in detail to illustrate the points to be considered when conflicts exist between the results of large RCTs and SRs/MAs.

Evidence Synthesis: RCTs are the gold standard for providing evidence of the effectiveness of interventions, however concerns over an RCT's internal and external validity may limit their applicability on clinical practice. SRs/MAs synthesize all evidence related to a given research question but two urological examples show that the validity of their results depends on the quality of the individual studies, the clinical and methodological heterogeneity of the studies, and publication bias.

Conclusions: Although SRs/MAs can provide a higher level of evidence than RCTs, the quality of the evidence from both the RCT and the SR/MA should be investigated when their results conflict to determine which source provides the better evidence. Guideline developers should

have a well-defined and robust process to assess the evidence from MAs and RCTs when such conflicts exist.

Patient Summary: We discuss the advantages and limitations of using data from randomized controlled trials and systematic reviews/meta-analyses in informing clinical practice when there are conflicting results and provide guidance on how such conflicts should be dealt with by guideline organizations.

Take Home Message

New or existing RCT data can lead to conflicts with MA data. In this paper, we present examples of, and explore reasons for, such conflicts. Guidance is provided to guideline developers on how to assess conflicting data in such circumstances to help determine which source is more reliable. For guideline organizations, both within and outside of urology, having a well-defined and robust process to deal with such conflicts is essential to improve the quality of their guidelines.

Tweets

Clinicians: SRs/MAs theoretically provide a higher LE than RCTs, but their quality needs scrutiny in case of conflict #eauguidelines

Patient summary: High level scientific publications should be interpreted with caution when there are conflicting results #eauguidelines

1. Introduction

The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research [1].

Treatment recommendations in European Association of Urology (EAU) Guidelines are underpinned, whenever possible, by the results of systematic reviews (SR)/meta-analyses (MA) and large randomized controlled trials (RCT). According to the 2009 Oxford Centre for Evidence Based Medicine, SRs of RCTs (with or without a meta-analysis) that are free of worrisome variations (heterogeneity) in results between individual studies provide the highest level of evidence (LE), 1a, whereas individual RCTs with a narrow confidence interval provide the next highest LE, 1b [2]. As SRs can provide a higher LE than RCTs, the results of SRs are generally considered to take precedence when developing treatment recommendations.

The quality of the results of a SR/MA depends on the quality of the included studies. Kjaergard et al [3] found a correlation between methodologic quality and discrepancies in the results of large and small RCTs included in MAs. Intervention effects were exaggerated in small trials with inadequate allocation sequence generation, inadequate allocation concealment and no double blinding.

Discrepancies have also been noted between large RCTs and previously published MAs on the same subject [4-6]. In 12 large RCTs carried out subsequent to 19 MAs addressing the same question, LeLorier et al [7] found that the results of subsequent RCTs results disagreed with those of earlier MAs 35% of the time.

To illustrate these points and provide guidance to guideline developers in dealing with conflicting data from different sources, two examples which have a direct bearing on EAU Guidelines treatment recommendations are presented. In the first example, the EAU Guidelines Office has recently been confronted with the results of a large RCT which found no beneficial effect of medical expulsive therapy (MET) on stone passage, contrary to results of previous meta-analyses which formed the basis for treatment recommendations [8]. In the second example, which compares the efficacy of partial versus radical nephrectomy for localized renal tumors, discordance between the results of the meta-analysis and the only available RCT are investigated [9,10].

2. Advantages and Limitations of Randomized Controlled Trials

As summarized in Table 1, RCTs have a number of advantages and limitations.

Advantages of RCTs

RCTs are the gold standard for providing evidence on the effectiveness of interventions [11-12]. Randomization balances, on the average, the distribution of both known and unknown prognostic factors at baseline in the intervention groups, thereby minimizing selection bias when assigning patients to treatments. Although adjusting for baseline covariates used in the randomization process can improve statistical power, complex adjustment procedures such as propensity score weighting are not usually required when comparing outcomes.

Patients are selected, treated, followed and assessed according to a common protocol testing a specific hypothesis. Blinding of participants and physicians to the allocated intervention may be possible to minimize performance bias, and is especially important when assessing outcomes [13]. Quality control measures and external review of key parameters maximize study quality.

Limitations of RCTs

RCTs can be challenging to design (randomization and blinding), conduct (poor recruitment, loss to follow up), analyze (missing data) and report (patient exclusions).

RCTs require an adequate sample size and follow-up to have sufficient power to detect clinically relevant differences between interventions [14]. In practice, many clinical trials do not meet their pre-specified power requirements so a conclusion of 'no significant difference' in outcome should not be interpreted as meaning that two or more treatments are equivalent in effect. Sample size estimation requires data about expected differences and variability of the primary outcome. Often these data are unknown or only available from observational studies prone to bias.

Although analyses using the intention-to-treat principle can provide an unbiased estimate of the treatment effect, this assumes that there are no differences in follow-up or missing outcome data that may bias the treatment comparison [15]. In some RCTs, not all participants receive their randomized intervention; they may, for example, cross-over to the other randomized treatment, in which case a per-protocol analysis may also provide useful information. Various analysis strategies exist, depending on whether the objective is to estimate treatment efficacy (the intervention effect under perfect conditions, in which case intent to treat can dilute the size of the treatment effect) or effectiveness (the real-world intervention effect with 'imperfect' compliance).

An RCT with double blinding, little missing data and good compliance will have a high internal validity, but if an RCT recruits only a very select population, the external validity (generalizability) may be low. This can happen due to overly restrictive inclusion/exclusion

criteria or including only expert clinicians in select sites [16]. Single-center RCTs typically have lower external validity compared with multicenter RCTs which allow the comparison of results between centers.

Finally, robust, adequately powered RCTs with long term follow up are difficult to organize, expensive and resource-intensive. Thus many RCTs focus on short-term or surrogate outcomes, the clinical significance of which is often uncertain. Any short-term benefits might not be maintained over longer time horizons which are more relevant to patients, clinicians and policy makers [17].

3. Advantages and Limitations of Systematic Reviews and Meta-analyses

Table 2 outlines the advantages and limitations of SR/MAs.

Advantages of SR/MAs

A SR is a literature review focused on a research question that tries to identify, appraise, select and synthesize all research evidence relevant to that question.

SRs are *a priori* defined in a PICO (Participant, Intervention, Comparator, Outcome) based protocol outlining the study inclusion criteria. They are the only transparent and replicable form of literature review that provide a rigorous and critical qualitative appraisal of the evidence related to an intervention. SRs explore the findings of individual studies, draw attention to their differences and identify sources of bias [18].

A MA is a statistical technique for quantitatively combining the data from two or more separate RCTs asking the same or a similar question [19]. They should only be done as part of a SR, otherwise it is a combined analysis, susceptible to study selection bias. Two different types of

meta-analyses exist: literature-based or aggregate data (AD) MAs and individual patient data (IPD) MAs [20, 21].

MAs provide an overall estimate of the size of the treatment effect, giving due weight to the size of the individual RCTs. They are useful when individual studies are underpowered, yield inconclusive or conflicting results, or when an overall, more precise estimate of the size of the treatment effect is required. MAs increase the power to detect moderate but clinically meaningful differences in treatment outcome and assess if the treatment effect is similar across different studies or types of patients [22]. They are useful in exploring the effects of an intervention in subgroups of patients, especially in IPD MAs [20, 21].

SRs and MAs are vital for guideline developers, healthcare providers, patients, researchers and policy makers in order to guide clinical practice, research and healthcare policies [23].

Limitations of SR/MAs

The validity of a MA depends on the quality of the systematic review upon which it is based. SRs and MAs have a number of potential limitations including poor quality of included studies, heterogeneity, and publication bias.

The literature summary provided in a SR and the results of a MA are only as reliable as the quality of the included studies. Although IPD meta-analyses and multicenter RCTs can be analyzed using the same statistical techniques for clustered data, where the clusters are studies and centers, respectively, there may be important clinical and methodological heterogeneity between the studies in a MA since they are not carried out based on a common protocol. The studies may be heterogeneous regarding patients included, the intervention or the assessment of treatment outcome. Although heterogeneity in treatment effect can be better investigated

in IPD MAs, the primary studies should be similar enough to be combined, otherwise genuine differences in effects may be obscured [24,25]. Since institutions participating in a multicenter study are supposed to treat, follow up and assess patients according to a common protocol, there is potentially a greater degree of standardization and higher quality data in multicenter clinical trials as compared to studies included in meta-analyses.

If bias is present in the individual studies included in a MA, MAs will compound these errors and produce a biased result. The risk of bias (RoB) on the outcomes in each study should be systematically assessed and sensitivity analyses performed to examine the effect of RoB on the conclusions. Observational and non-randomized comparative studies in SRs of interventions should not be included in MAs because the MA may provide very precise but spurious results due to confounding and patient selection bias.

Only a non-random proportion of research projects ultimately reach publication in an indexed journal and become readily identifiable for systematic reviews. Statistically significant, 'positive' results favoring an intervention are more likely to be published, published quicker and published in higher impact journals, leading to publication bias [26]. When these trials are pooled together in a MA, this may lead to an exaggeration of the treatment effect. Begg and Egger have both proposed tests along with funnel graphs and plots to detect publication bias, however they have limited power in small meta-analyses, for example those including less than 10 studies [27]. In order to minimize publication bias, authors should perform a comprehensive systematic literature search, looking not only for published trials in various electronic databases, but also search trial registries for unpublished studies and conference abstracts or proceedings [18].

4. The Results of a Randomized Controlled Trial are in conflict with the Results of a Systematic Review/Meta-analysis

It is not uncommon for the results of a large RCT to appear to be inconsistent with evidence from SRs/MAs. The most extreme is when an intervention thought to be beneficial is demonstrated to be harmful in a large RCT [9,10]. More commonly, an RCT may show a treatment to be ineffective, or less effective than that found in a previous MA, or perhaps only effective in a subpopulation of patients. Assuming the conflicting RCT was of high quality, a number of issues should be explored to try to explain the discrepancies.

Quality of the systematic review

The starting point is the methodological quality of the SR. AMSTAR and DART checklists [28-30] allow readers to judge a review's quality by focusing on the essential components of a well-conducted SR. Items include the comprehensiveness of the search strategy, a description of the characteristics of included studies and an assessment of their scientific quality. A poor quality SR/MA may produce biased results that conflict with a large RCT.

Small study effects and publication bias

Small study effects and publication bias can individually and jointly produce results in a SR/MA that conflict with a large RCT. Studies have shown that small RCTs can exaggerate intervention effects due to shortcomings in methodological rigor which may then introduce bias [3]. Small studies that find statistically significant (but unrealistically large) treatment effects are more likely to be published than negative studies and then included in an SR and MA, leading to publication bias. Both of these phenomena can be investigated using funnel plots [31].

Heterogeneity

Heterogeneity within a SR/MA can arise from many sources, including the population recruited (age, sex, disease severity, etc.), the intervention(s) and control treatments, and the definition and timing of outcome measurements. If studies included in a SR/MA differ substantially from a subsequent large RCT, then judgement is required on whether similar findings should be expected.

Another source of heterogeneity is differences in the methodological quality of the included studies. Deficiencies in the generation and concealment of the allocation sequence, adherence to treatment, handling of missing data, and outcome assessment can all introduce bias in the outcomes reported in the included studies [18]. Bias may then be propagated in meta-analyses through the pooling of biased study effects, thus contributing to different estimates of effectiveness between a SR/MA and subsequent large RCTs. Nevertheless, since a MA is generally seen to have a higher LE than a single RCT, the results of a poor quality MA may have more impact than a well-conducted RCT.

Heterogeneity should be assessed using both clinical knowledge and statistical methods. If substantial heterogeneity from any source is suspected, random effects models are recommended, however the pooling of data and estimation of an overall treatment effect may be inappropriate with any statistical model in the presence of heterogeneity. Meta-regression is a useful tool to explore the relationship between RCT effect sizes and characteristics on a study level [32], however IPD are required for assessment on a patient level [21, 33]. Appropriate statistical modelling may show that after correcting for sources of bias and heterogeneity,

discrepancies between SR/MA and definitive RCTs are reduced. Whatever the approach, interpretation of results is less straightforward when heterogeneity is present.

In order to provide guidance to clinicians and guideline developers when there is a conflict of results between a large RCT and a SR/MA, a practical checklist of points to consider is provided in Table 3.

5. Examples of discrepancies between findings from meta-analyses and large randomized controlled trials

Medical expulsive therapy

Five SRs and MAs on the management of uncomplicated symptomatic ureteric stones using medical expulsive therapy (MET) were published in the past 10 years [34-38]. All five suggested that alpha blockers and nifedipine were more effective in increasing the spontaneous passage of ureteric stones compared to control (risk ratios ranging from 1.45-1.59). The reviews identified numerous sources of potential bias which limited the strength of evidence and the authors concluded an urgent need to conduct a large, robust, multicenter RCT to address these shortcomings. Pickard et al [8] published the results of such an RCT in 1167 patients and found no evidence that either tamsulosin or nifedipine increased the rate of spontaneous stone passage compared with placebo. Results were consistent across subgroup and sensitivity analyses.

We compare the Pickard et al RCT [8] to the meta-analysis with the most studies, Seitz et al [36], to explore and discuss discordant findings. Most RCTs included in Seitz's meta-analysis were small and recruited from a single-center; only 6 of 35 (17%) recruited more than 100

patients. The majority had low internal validity and only one RCT reported allocation concealment. As small RCTs may report larger effect sizes compared to larger RCTs, a meta-analysis of small RCTs can lead to biased estimates of treatment effects [39]. Seitz also found evidence of publication bias which can lead to an overestimation of treatment effects and compromise the validity of the meta-analysis findings [40].

There was evidence of clinical heterogeneity in Seitz's review concerning the patient inclusion criteria, stone characteristics, intervention, treatment in the control group, and outcome measurement. In the MA, the primary outcome of being stone-free was inconsistently defined, assessed using different imaging modalities, and measured at a variety of time points. In Pickard, the primary outcome was need for further intervention within 4 weeks of randomization, which is compared here to being stone-free. In the control group, 80% of patients were stone-free in the Pickard RCT whereas in Seitz, the stone-free rates ranged from 4% to 78%, which highlights the potential impact of the heterogeneity in the included studies. With contrasting primary outcomes and different baseline event rates in the control groups, it is not surprising that the RCT and the MA reported discordant findings. The choice of primary outcome is clearly of paramount importance in any trial. Heterogeneity in the conduct, design and reporting of trials in this MA makes pooled treatment effects difficult, if not impossible, to interpret.

Partial versus radical nephrectomy

In an EORTC RCT involving 541 patients with a solitary T1-T2 N0 M0 renal tumor $\leq$ 5 cm, 21 patients progressed, 9 after radical nephrectomy (RN) and 12 after partial nephrectomy (PN). An intent to treat analysis found an overall survival (OS) advantage in favor of RN (HR = 1.5, p =

0.03), however only 12 of the 117 deaths were due to kidney cancer, 4 on RN and 8 on PN [10]. Subsequently, Kim et al published a SR and MA including some 41,000 patients which found statistically significant improvements in both OS (HR = 0.81, $p < 0.001$) and disease specific survival (DSS) (HR = 0.71, $p < 0.001$), but this time in favor of PN [9]. How can this discordance be explained?

The Kim meta-analysis has a number of limitations. Firstly, the 38 included trials were mostly retrospective, single center studies. The only RCT was the EORTC study. No information was provided about the distribution of follow up or patient characteristics by treatment group (T category when > T1, tumor size, grade, cell type, or renal function). Consequently, the observed differences in survival may not be directly due to differences in treatment efficacy. In addition, it is not clear to which patients the results can be generalized. Lastly, there was significant heterogeneity in the size of the treatment effect across the studies so the overall estimate of the HR is not meaningful. Nevertheless, the EORTC RCT also had limitations and should be interpreted cautiously: 55 patients crossed over to the other randomized treatment, 140 patients were clinically or pathologically ineligible and there were few cancer related events.

The MA found that PN was associated with a decreased risk of severe chronic kidney disease (CKD), however the EORTC study only found a reduced incidence of at least moderate renal dysfunction, not of advanced kidney disease or renal failure, and this was not associated with a corresponding difference in survival [41]. The studies in the MA did not always specify the status of the contralateral kidney whereas in the EORTC study the contralateral kidney had to be normal.

Critical information regarding the biases of the studies included in the SR were not made explicit since a GRADE approach to assess the quality of evidence was not done [42]. The quality of the studies in the SR and heterogeneity of results call into question the validity of the conclusions of the MA which should thus be viewed with skepticism. The same year, another SR suggested that localised RCCs are best managed by PN where technically feasible. However, the evidence base had significant limitations due to studies of low methodological quality and high risks of bias [43].

Further non-randomized studies have found improved survival with PN [44,45] and a reduction in the risk of cardiovascular events relative to RN [46], however patients chosen for PN had a higher baseline likelihood of long-term survival [47,48]. In another study, only stage-II CKD patients had a decreased risk of developing significant renal impairment on PN [49]. More recently, a SR and MA of 21 non randomized comparative studies in patients with clinical T1b and T2 renal tumors found better tumor control and survival with PN as compared to RN [50], but it is subject to the same biases as the Kim MA.

Taking into account all available efficacy data and a perceived advantage in renal function, the 2016 EAU Guidelines recommend, with several exceptions, that localized renal cancers are better managed by PN than with RN.

6. Discussion

It is generally accepted that a high quality SR of RCTs and associated MA can provide a higher level of evidence than a single RCT addressing the same question [2]. It can be problematic,

however, when the results of the MA are in direct conflict with the RCT, making it difficult for guideline organizations to interpret the evidence and issue recommendations.

Guideline groups should follow well-defined methodological rules to assess the studies in these situations. RCTs should be appraised on their internal and external validity using established tools [51]. The conflicting SR/MA should be appraised in the same fashion, to determine the methodological quality of the review, the quality of the included studies, inconsistency within the studies, unexplained heterogeneity, and likelihood of publication bias using tools such as AMSTAR [28,29] and DART [30]. In some cases, the discrepancy may be due to errors in the MA in applying study eligibility criteria or even data extraction [52], hence the need for a SR/MA protocol and strict quality control.

When MAs include many small underpowered studies, especially combined with likely presence of publication bias, there is immediate concern for over-inflation of, or completely erroneous, effect size measurement. Additionally, when a great degree of heterogeneity exists in the MA which cannot be easily accounted for, the results may be highly unreliable. In this regard, IPD MAs provide a better platform for assessing and explaining heterogeneity than aggregate data MAs.

Two examples were discussed in this manuscript to illustrate the assessment process. In the case of MET for ureteric stones, a large, high quality RCT [8] contradicted many well established MAs which pointed to a benefit with this therapy. Analysis of a representative MA [36] revealed the inclusion of many small RCTs, poor internal validity, significant study heterogeneity and likely publication bias. When such MA concerns are present, a single high quality RCT may be considered as having the higher LE. For guideline organizations, this

process can be used to justify a change in recommendations based on methodologically sound principles.

Radical versus partial nephrectomy provides a more complex example. The MA [9] included only a single RCT, which was the study in conflict with its own results. The other included studies were all retrospective, which in general provide a lower LE. Risk of bias was poorly assessed, and significant study heterogeneity was present. It is important to reiterate that combining observational studies in general, and even comparative non-randomized studies with RCTs in an intervention MA, may produce unreliable results and is not considered valid. In light of all this, the single RCT [10] in this circumstance might provide more guidance than the MA if it was of significantly high quality. However, this RCT also had some methodology concerns, so the comparison is not so simple.

Instead of automatically assigning a higher LE to SR/MAs which conflict with RCTs, these examples have shown that the quality of the evidence and the RoB of studies included in SRs/MAs should be assessed to determine which source provides the better evidence.

Although non RCTs can be included in SRs, we have emphasized that only RCTs should be included in intervention MAs. RCTs are not required for prognostic factor and diagnostic test accuracy MAs, however the studies included in these MAs should preferably be prospective in nature and based on a protocol to minimize risk of bias.

Despite the availability of MAs and RCTs, and also in cases where high level evidence does not exist, we may still not know what the best treatment is. The GRADE system, which takes into account the quality of evidence (high, moderate, low, very low) for critical outcomes, provides strengths of recommendations (strong, weak) for or against a treatment to aid clinicians in their

practice when consensus is not possible [42,53]. A decision curve approach, which takes into account a patient's values and preferences, may also be used to help choose between the different treatment options.

7. Conclusions

New or existing RCT data can lead to conflicts with MA data. In this paper, we present examples of, and explore reasons for, such conflicts. Guidance is provided to guideline developers on how to interpret conflicting data in such circumstances to help assess which source is more reliable.

For guideline organizations, both within and outside of urology, having a well-defined and robust process to deal with such conflicts is essential to improve guideline quality.

Financial disclosures/Conflicts of interest: none

Funding/Support: There was no financial or material support for this academic research study.

8. References

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996; 312:71-2.

2. Oxford Centre for Evidence-based Medicine – Levels of Evidence (March 2009). http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/.

3. Kjaergard LL, Villumsen J, Gluud C. Reported Methodologic Quality and Discrepancies between Large and Small Randomized Trials in Meta-Analyses. Ann Intern Med. 2001; 135:982-9

4. Borzak S, Ridker PM. Discordance between meta-analyses and large-scale randomized, controlled trials. Ann Intern Med. 1995; 123:873-7.

5. Flather MD, Farkouh ME, Pogue JM, Yusuf S. Strengths and Limitations of Meta-Analysis: Larger Studies May Be More Reliable. Controlled Clinical Trials 1997; 18:568-79.

6. DerSimonian R, Levine RJ. Resolving discrepancies between a meta-analysis and a subsequent large controlled trial. JAMA 1999; 282:644-70.

7. LeLorier J, Gregoire G, Benhaddad A, Lapierre J, Derderian F. Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 1997; 337:536-42.

8. Pickard R, Starr K, MacLennan G et al. Medical expulsive therapy in adults with ureteric colic: a multicentre, randomised, placebo-controlled trial. Lancet 2015; 386:341–9.

9. Kim SP, Thompson RH, Boorjian SA et al. Comparative Effectiveness for Survival and Renal Function of Partial and Radical Nephrectomy for Localized Renal Tumors: A Systematic Review and Meta-Analysis. J Urol 2012; 188:51-7.

10. Van Poppel H, Da Pozzo L, Albrecht W et al. Prospective, Randomised EORTC Intergroup Phase 3 Study Comparing the Oncologic Outcome of Elective Nephron-Sparing Surgery and Radical Nephrectomy for Low-Stage Renal Cell Carcinoma. Eur Urol 2011; 59:543-52.

11. Byar DP, Simon RM, Friedewald WT et al. Randomized Clinical Trials — Perspectives on Some Recent Ideas. N Engl J Med 1976; 295:74-80.

12. Sibbald B, Roland M. Understanding controlled trials: Why are randomized controlled trials important? BMJ 1998; 316:201.

13. Hansson L, Hedner T, Dahlöf B. Prospective Randomized Open Blinded End-point (PROBE) Study. A novel design for intervention trials. Blood Press 1992; 1:113-9.

14. Julious SA. Sample Sizes for Clinical Trials. Chapman and Hall/CRC, 2009, 328 pages.

15. Armijo-Olivo S, Warren S, David Magee D. Intention to treat analysis, compliance, drop-outs and how to deal with missing data in clinical research: a review. Physical Therapy Reviews 2009; 14: 36-49.

16. Rothwell PM. External validity of randomised controlled trials: “To whom do the results of this trial apply?” Lancet 2005; 365: 82–93.

17. Fleming TR, DeMets DL. Surrogate End Points in Clinical Trials: Are We Being Misled? Ann Intern Med 1996; 125:605-13.

18. Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions, Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available from http://handbook.cochrane.org.

19. Egger M, Smith GD, Phillips AN. Meta-analysis: principles and procedures. BMJ 1997; 315: 1533-7.

20. Riley RD, Lambert PC, Abo-Zaid, G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ 2010; 340: c221.

21. Tudur Smith C, Marcucci M, Nolan SJ et al. Individual participant data meta-analyses compared with meta-analyses based on aggregate data (Review). Cochrane Database of Systematic Reviews 2016, Issue 9. Art. No.: MR000007. DOI: 10.1002/14651858.MR000007.pub3.

22. Mulrow CD. Rationale for systematic reviews. BMJ 1994; 309: 597-9.

23. Institute of Medicine (US). Committee on Standards for Developing Trustworthy Clinical Practice Guidelines; Graham R, Mancher M, Miller Wolman D, et al., editors. Clinical Practice Guidelines We Can Trust. Washington (DC): National Academies Press (US); 2011.

24. Higgins J, Thompson SG. Quantifying heterogeneity in a meta-analysis. Statistics in Medicine 2002; 21: 1539-58.

25. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327: 557-60.

26. Thornton A, Lee P. Publication bias in meta-analysis: its causes and consequences. J Clin Epidemiol 2000; 53:207-16.

27. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple graphical test. BMJ 1997; 315: 629-34.

412 28. Shea BJ, Grimshaw JM, Wells GA. Development of AMSTAR: a measurement tool to assess
413 the methodological quality of systematic reviews. BMC Med Res Methodology 2007; 7:10.
414 29. Shea BJ, Hamel C, Wells GA et al. AMSTAR is a reliable and valid measurement tool to assess
415 the methodological quality of systematic reviews. J Clin Epidemiol 2009; 62:1013-20.
416 30. Diekemper RL, Ireland BK, Merz LR. Development of the Documentation and Appraisal
417 Review Tool for systematic reviews. World J Meta-Anal 2015; 3: 142-50.
418 31. Sterne JA, Egger M and Smith GD. Investigating and dealing with publication and other
419 biases in meta-analysis. BMJ 2001; 323: 101-5.
420 32. Thompson SG, Higgins JP. How should meta-regression analyses be undertaken and
421 interpreted? Statistics in Medicine 2002; 21: 1559-73.
422 33. Thompson SG, Higgins JPT. Can meta-analysis help target interventions at individuals most
423 likely to benefit? Lancet 2005; 365: 341–46.
424 34. Hollingsworth JM, Rogers MA, Kaufman SR et al. Medical therapy to facilitate urinary stone
425 passage: a meta-analysis. Lancet 2006; 368: 1171–9.
426 35. Campschroer T, Zhu Y, Duijvesz D, Grobbee DE, Lock MTWT. Alpha blockers as medical
427 expulsive therapy for ureteral stones. Cochrane Database Syst Rev 2014; 4: CD008509.
428 36. Seitz C, Liatsikos E, Porpiglia F, Tiselius H-G, Zwergel U. Medical therapy to facilitate the
429 passage of stones: what is the evidence. Eur Urol 2009; 56: 455–71.
430 37. Singh A, Alter HJ, Littlepage A. A systematic review of medical therapy to facilitate passage
431 of ureteral calculi. Ann Emerg Med. 2007; 50:552-63.
432 38. EAU/AUA Nephrolithiasis Guideline Panel. 2007 Guideline for the Management of Ureteral
433 Calculi. https://www.auanet.org/education/guidelines/ureteral-calculi.cfm

434 39. Schulz K, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of
435 methodological quality associated with estimates of treatment effects in controlled clinical
436 trials. JAMA 1995; 273:408–12.

437 40. Driessen E, Hollon SD, Bockting CL, Cuijpers P, Turner EH. Does Publication Bias Inflate the
438 Apparent Efficacy of Psychological Treatment for Major Depressive Disorder? A Systematic
439 Review and Meta-Analysis of US National Institutes of Health-Funded Trials. PLoS ONE 10(9):
440 e0137864. doi:10.1371/journal.pone.0137864

441 41. Scosyrev E, Messing EM, Sylvester R, Campbell S, Van Poppel H. Renal Function After
442 Nephron-sparing Surgery Versus Radical Nephrectomy: Results from EORTC Randomized Trial
443 30904. Eur Urol 2014; 65: 372 –7.

444 42. Guyatt GH, Oxman AD, Vist GE et al. GRADE: an emerging consensus on rating quality of
445 evidence and strength of recommendations. BMJ 2008; 336: 924-6.

446 43. MacLennan S, Imamura M, Lapitan MC et al. Systematic review of oncological outcomes
447 following surgical management of localised renal cancer. Eur Urol 2012; 61:972-93.

448 44. Tan HJ, Norton EC, Ye Z, Hafez KS, Gore JL, Miller DC. Long-term survival following partial
449 versus radical nephrectomy among older patients with early-stage kidney cancer. JAMA 2012;
450 307:1629-1635.

451 45. Roos FC, Steffens S, Junker K et al. Survival advantage of partial over radical nephrectomy in
452 patients presenting with localized renal cell carcinoma. BMC Cancer 2014; 14:372-8.

453 46. Capitanio U, Terrone C, Antonelli A et al. Nephron-sparing Techniques Independently
454 Decrease the Risk of Cardiovascular Events Relative to Radical Nephrectomy in Patients with a
455 T1a–T1b Renal Mass and Normal Preoperative Renal Function. Eur Urol 2015; 67:683–9.

456 47. Tomaszewski JJ, Kutikov A. Retrospective Comparison of Cardiovascular Risk in Preselected
457 Patients Undergoing Kidney Cancer Surgery: Reflection of Reality or Simply What We Want to
458 Hear? Eur Urol 2015; 67:690–1.

459 48. Shuch B, Hanley J, Lai J et al. Overall Survival Advantage with Partial Nephrectomy: A Bias of
460 Observational Data? Cancer 2013;119:2981-9.

461 49. Woldu SL, Weinberg AC, Korets R et al. Who Really Benefits From Nephron-sparing Surgery?
462 Urol 2014; 84: 860-8.

463 50. Mir MC, Derweesh I, Porpiglia F, Zargar H, Mottrie A, Autorino R. Partial Nephrectomy
464 Versus Radical Nephrectomy for Clinical T1b and T2 Renal Tumors: A Systematic Review and
465 Meta-analysis of Comparative Studies. Eur Urol (2016), http://dx.doi.org/10.1016/
466 j.eururo.2016.08.060

467 51. Guyatt GH, Sackett DL, Cook DJ, Evidence-Based Medicine Working Group: Users' guides to
468 the medical literature II: How to use an article about therapy or prevention (A): Are the results
469 of the study valid? JAMA 1993; 270:2598-601.

470 52. Ford AC, Guyatt GH, Talley NJ, Moayyedi P. Errors in the Conduct of Systematic Reviews of
471 Pharmacological Interventions for Irritable Bowel Syndrome. Am J Gastroenterol 2010;
472 105:280–8.

473 53. Jaeschke R, Guyatt GH, Dellinger P et al. Use of GRADE grid to reach decisions on clinical
474 practice guidelines when consensus is elusive. BMJ 2008;337:a744.

Table 1: Advantages and Limitations of Randomized Controlled Trials

Advantages	**Limitations**
Randomization minimizes the influence of both known and unknown prognostic variables on treatment outcome	It may be difficult to recruit and follow up patients
RCTs can demonstrate causality	Ethical considerations may make randomization difficult
Patients are treated according to a common protocol	Required study power might not be met
Quality control of treatment and outcome assessment	Generalizability may be low
RCTs provide the strongest empirical evidence of treatment efficacy	RCTs are expensive and resource intensive

Table 2

Table 2: Advantages and Limitations of Systematic Reviews and Meta-analyses

Advantages	**Limitations**

Focused well defined clinical question with a clear objective and explicit, predefined study eligibility criteria
 	
Comprehensive literature search strategy to guarantee the identification of all potentially eligible studies
 	
Critical appraisal of all the included studies that is used to guide the analysis and conclusions
 	
Increases the power to detect differences between interventions
 	
Increases the precision of the estimate of the treatment effect
 	
Allows the comparison of treatment effects across different studies or subgroups of patients, interventions and outcomes
	Depends on the quality of the included studies
	
	Susceptible to the effects of heterogeneity of included studies
	• Clinical heterogeneity:
	  ○ Participants *(e.g. age, gender, disease severity, disease subtype, study eligibility criteria)*
	  ○ Interventions *(e.g. drug doses, duration/intensity of treatment, delivery, co-interventions, surgeon experience)*
	  ○ Outcomes *(e.g definition of outcome, outcomes reported, timing and method of measurement, follow-up duration, cut-off points)*
	• Methodological heterogeneity *(e.g different study designs, reporting bias across studies)*
	• Statistical heterogeneity
	
	Publication bias
	
	Time and resource consuming

Table 3

Table 3: Checklist of points to consider when the findings from a systematic review and meta-analysis differ with those from a large randomized controlled trial

Criteria to consider	Questions to ask	Rationale
Selection bias	Were the sequence generation and allocation concealment adequate in both the studies included in the SR/MA and the subsequent trial?	If the sequence generation was not truly random or the allocation was not effectively concealed, this can lead to exaggerated estimates in individual studies and these may be amplified in MAs.
Confounding bias	Were the groups balanced for known prognostic factors at baseline and were any imbalances controlled for in the analysis?	Imbalances in known and unknown prognostic factors are possible even in well-designed RCTs. Baseline imbalances may explain differences in estimates of effect if not controlled for in the analysis.
Performance and detection bias	Where possible, in all the studies included in the SR/MA and for the new trial, was blinding of study participants, clinicians administering the treatment, ancillary care-givers and outcomes assessors done? When blinding is not possible, could knowledge of the treatment received affect interpretation of any of the outcomes?	Some objective outcomes are unlikely to be affected by knowledge of the intervention arm, but failure to blind (particularly for subjective outcomes) may lead to an exaggeration of effect sizes in individual studies and these may be amplified in MAs.
Attrition bias	Were all dropouts documented and unlikely to be related to the treatment outcome in the studies included in the SR/MA and in the new trial?	If drop-out rates differ between the treatment arms, then the reasons may be related to the outcome of interest and may hide important outcome effects.
Reporting bias	Were all outcomes that were stated in the methods and/or protocol for all the studies included in the SR/MA and in the new trial reported in the trial report? Were all the outcomes measured appropriately (as defined in the protocol) or were deviations	Selective reporting of outcomes, or selective methods of reporting, may lead to exaggerated estimates of effect

	reasonably explained?	
Publication bias	Were funnel plots used to investigate publication bias in the SR/MA? Is the funnel plot symmetrical or is there reason to believe there is a systematic difference between published and unpublished studies? Note: this is difficult to assess when there are less than 10 RCTs contributing to a MA.	Asymmetric funnel plots raise suspicion that there are systematic differences between published and unpublished studies and that some positive or negative trials may be unpublished. The may lead to exaggerated effect sizes in a MA
Consistency and heterogeneity of outcome	Did the studies included in the SR/MA have overlapping 95% CIs for the outcome? Was variation more than would be expected by chance alone? Was the I^2 statistic <40% ? (Cochrane/GRADE rule of thumb…) Were subgroups used to explain any observed heterogeneity? Were event rates in the control group similar in the different studies? Note: Subgroups of the population, the intervention/control types, or the outcome measurement may explain heterogeneity.	If the outcomes can be shown to be more effective in certain subgroups, or with variations of an intervention (e.g. a higher dose), then this explained heterogeneity may indicate a key difference which may justify the results in the new trial. Where unexplained heterogeneity exists, then the estimate of effect is likely to be uncertain, even if precise.
Directness	Do the studies included in the SR/MA and does the new trial both directly assess the research question about the population, interventions and outcomes?	Indirect populations, interventions, surrogate outcome measures or indirect comparisons may conceal or exaggerate important differences within and between studies and may impact upon the estimate of effect.
Precision	Were the sample sizes of the studies included in the SR/MA and the new trial powered to address the outcomes of interest? Does the 95% CI in the MA include clinically judged appreciable benefit and harm?	If any of the SR/MA included trials, or the new trial were not powered to detect a clinically meaningful difference in the effect estimate, this may reduce our confidence in the estimate of effect. If the lower and upper 95% CI thresholds indicate that at one end the intervention may be beneficial, but at the other, it

		may be harmful, this will likely reduce our confidence in the estimate of effect.
Sensitivity analyses	When some studies included in a SR/MA are judged to be at high risk of bias, and others at low risk of bias, or extreme variations in the included studies' populations or interventions are apparent: did the authors conduct a sensitivity analysis to ascertain the estimates of effect on only those studies judged to be at low risk of bias?	Sensitivity analyses are different from subgroup analyses. Some studies are actively omitted as we are only interested in the results when the biased or 'different' studies are omitted.

EUROPEAN UROLOGY Authorship Responsibility, Financial Disclosure, and Acknowledgment form.

By completing and signing this form, the corresponding author acknowledges and accepts full responsibility on behalf of all contributing authors, if any, regarding the statements on Authorship Responsibility, Financial Disclosure and Funding Support. Any box or line left empty will result in an incomplete submission and the manuscript will be returned to the author immediately.

Title	Dr.	
First Name	Richard	
Middle Name	J	
Last Name	Sylvester	
Degree	ScD	(Ph.D., M.D., Jr., etc.)
Primary Phone	+322.346.4916	(including country code)
Fax Number		(including country code)
E-mail Address	richard.sylvester@skynet.be	

Authorship Responsibility

By signing this form and clicking the appropriate boxes, the corresponding author certifies that each author has met all criteria below (A, B, C, and D) and hereunder indicates each author's general and specific contributions by listing his or her name next to the relevant section.

☒ A. This corresponding author certifies that:

• the manuscript represents original and valid work and that neither this manuscript nor one with substantially similar content under my authorship has been published or is being considered for publication elsewhere, except as described in an attachment, and copies of closely related manuscripts are provided; and
• if requested, this corresponding author will provide the data or will cooperate fully in obtaining and providing the data on which the manuscript is based for examination by the editors or their assignees;
• every author has agreed to allow the corresponding author to serve as the primary correspondent with the editorial office, to review the edited typescript and proof.

☒ B. Each author has given final approval of the submitted manuscript.

☒ C. Each author has participated sufficiently in the work to take public responsibility for all of the content.

☒ D. Each author qualifies for authorship by listing his or her name on the appropriate line of the categories of contributions listed below.

The authors listed below have made substantial contributions to the intellectual content of the paper in the various sections described below.

(list appropriate author next to each section – each author must be listed in at least 1 field. More than 1 author can be listed in each field.)

_ conception and design	Sylvester, N'Dow
_ acquisition of data	Sylvester, Lam, Marconi, S. MacLennan, Y. Yuan, Van Poppel, N'Dow
_ analysis and interpretation of data	Sylvester, Canfield, Lam, Marconi, S. MacLennan, Y. Yuan, G. MacLennan, Norrie, Omar, Bruins, Hernandez, Plass, Van Poppel, N'Dow
_ drafting of the manuscript	Sylvester, Canfield, Lam, Marconi, S. MacLennan, Y. Yuan, G. MacLennan, Norrie, Omar, Bruins, Hernandez, Plass, Van Poppel, N'Dow
_ critical revision of the manuscript for important intellectual content	Sylvester, Canfield, Lam, Marconi, S. MacLennan, Y. Yuan, G. MacLennan, Norrie, Omar, Bruins, Hernandez, Plass, Van Poppel, N'Dow
_ statistical analysis	Not Applicable
_ obtaining funding	Not Applicable
_ administrative, technical, or material support	Not Applicable
_ supervision	Sylvester, N'Dow
_ other (specify)	Not Applicable

Financial Disclosure

☒ None of the contributing authors have any conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript.

OR

☐ I certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/ affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: (please list all conflict of interest with the relevant author's name):

Funding Support and Role of the Sponsor

☒ I certify that all funding, other financial support, and material support for this research and/or work are clearly identified in the manuscript.

The name of the organization or organizations which had a role in sponsoring the data and material in the study are also listed below:

Not applicable

All funding or other financial support, and material support for this research and/or work, if any, are clearly identified hereunder:

The specific role of the funding organization or sponsor is as follows:

☐ Design and conduct of the study
☐ Collection of the data
☐ Management of the data
☐ Analysis
☐ Interpretation of the data
☐ Preparation
☐ Review
☐ Approval of the manuscript

OR

☒ No funding or other financial support was received.

Acknowledgment Statement

This corresponding author certifies that:
• all persons who have made substantial contributions to the work reported in this manuscript (eg, data collection, analysis, or writing or editing assistance) but who do not fulfill the authorship criteria are named with their specific contributions in an Acknowledgment in the manuscript.
• all persons named in the Acknowledgment have provided written permission to be named.
• if an Acknowledgment section is not included, no other persons have made substantial contributions to this manuscript.
Richard Sylvester

After completing all the required fields above, this form must be uploaded with the manuscript and other required fields at the time of electronic submission.