PEER REVIEW HISTORY

ARTICLE DETAILS

TITLE (PROVISIONAL)	ROBOTIC vs. OPEN UROLOGIC ONCOLOGIC SURGERY: STUDY PROTOCOL OF A SYSTEMATIC REVIEW AND META-ANALYSIS
AUTHORS	Cacciamani, Giovanni; Gill, Karanvir; Gill, Inderbir

VERSION 1 – REVIEW

REVIEWER	Haidar Abdul-Muhsin Mayo Clinic Arizona
REVIEW RETURNED	29-Jul-2019

GENERAL COMMENTS	will be a great comprehensive review that will help urologists understands the status of robotic interventions in urology

REVIEWER	Stavros I. Tyritzis Karolinska Institutet, Stockholm, Sweden HYGEIA Hospital, Athens, Greece
REVIEW RETURNED	08-Aug-2019

GENERAL COMMENTS	The authors attempt to gather in a very promising study all data on robotic urologic oncologic surgery. The dichotomization to study periods is the most important one, I would consider changing the abbreviation of the robotic operation to RARP, RARC and RAPN, so that it is easier to follow.

REVIEWER	Bhavan Rai Freeman Hospital Newcastle, UK
REVIEW RETURNED	19-Sep-2019

GENERAL COMMENTS	The authors have chosen to perform a systemic review (SR) comparing open and robotic urologic oncological surgery. There are major issues with this from the outset: 1) This is a broad topic for an SR covering all urological cancers which are far from ideal. SR ideally should address very specific questions and clear endpoints. This review appears to attempt to publish 5 systemic reviews under one heading. Individual cancers/techniques have different outcomes measures of interest. Therefore the 5 outcome measures reported in this protocol will not be consistently applied to all techniques/cancers 2) Penetration is an outcome measure that will be reported using national databases as part of a primary research work. This, therefore, cannot be claimed to be an outcome as part of an SR. SRs are secondary research that appraises published primary research using standardised validated methods. 3) Quote “We have carried out a preliminary screening of the literature: only 3 comparative studies were identified for open vs

	robotic radical nephrectomy for kidney cancer and no comparative studies were identified comparing open vs. robotic retroperitoneal lymph node dissection for testis cancer; as, these two topics will be censored from further detailed analyses. Therefore our SR and MA will compare data on ORP vs RRP, ORC vs RRC and OPN vs RPN exclusively" The authors claim 5 cancers will be compared initially. It does appear that they have already done a literature search of sorts and have decided to exclude RPLNDs and RN. This isn't the correct way to perform an SR and casts doubt on the credibility of their approach 4) Quality assessment: Quote "All papers will be categorized according to the Oxford Level of Evidence Working Group 238 2011 levels of evidence (LOEs) for therapy studies [6] and according to the Grading of 239 Recommendations Assessment, Development and Evaluation (GRADE) system" GRADE evaluates the quality of evidence. It does not evaluate the quality of papers The authors need to state the confounders particularly for non-RCTs for individual comparators in the protocol. 5) Statistical analysis: All analysis is ideally done with a random effect model. At the very least in cases of high heterogeneity, a random effect model must definitely be used and fixed-effect model should be reserved for low heterogeneity Quote: "Due to limitations in the Review Manager v5.3 software, analysis of continuous variables is possible only when data are presented as mean and standard deviation (SD). Since some studies may report continuous variables in "median" and "interquartile range" or "min/max" range, we will use a validated mathematical method to estimate "mean" and "SD". When available, we will use data reported in a matched-pair comparison manner" The Cochrane handbook which is accessible on Review manager clearly gives instructions on how to impute data from median/range to make MA feasible The author's plan perform a meta-analysis of RCTs and non-RCTs. A meta-analysis of non-RCT's is not recommended. Data from Non-RCTs has presented graphically using the Forrest plots, but pooled analysis is not recommended. The reviewer is not certain if temporal and proficiency MA is validated and recognized methodology in SR and would suggest statistical input on this. 6) There are sentences that inadequately referenced. 7) There are a number of high-quality SR that has been published on the suggested techniques. The current review is therefore unlikely to add to the existing literature. The methodology isn't correct and would recommend the current protocol for publication.

REVIEWER	Francesco Prete University Medical School of Bari, Bari, Italy
REVIEW RETURNED	30-Sep-2019

GENERAL COMMENTS	Thank you for asking me to review this protocol I found it mostly coherent with its aims and in general appropriately supported by research plan. Regarding assessment of risk of bias for randomized and nonrandomized studies respectively, I understand that Authors will likely use the Cochrane RoB2 ans ROBINS-I tool, this could be

	specified in the Quality Assessment section of the manuscript.

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
Reviewer Name: Haidar Abdul-Muhsin
Institution and Country: Mayo Clinic Arizona
Please state any competing interests or state 'None declared': none

Please leave your comments for the authors below
will be a great comprehensive review that will help urologists understands the status of robotic interventions in urology

REPLY TO REVIEWER #1

We thank the reviewer for the comment. We agree with the reviewer. We believe this new information is of considerable interest and practical use to the general medical community at large, who need to be aware of contemporary surgical trends, thereby to better advise patients seeking care for urologic cancers.

_____

Reviewer: 2
Reviewer Name: Stavros I. Tyritzis
Institution and Country: Karolinska Institutet, Stockholm, Sweden
HYGEIA Hospital, Athens, Greece
Please state any competing interests or state 'None declared': "None declared"

Please leave your comments for the authors below
The authors attempt to gather in a very promising study all data on robotic urologic oncologic surgery. The dichotomization to study periods is the most important one, I would consider changing the abbreviation of the robotic operation to RARP, RARC and RAPN, so that it is easier to follow.

REPLY TO REVIEWER #2

We thank the reviewer and agree with him. The chronologic time of publication may impact the reported outcomes of urologic oncologic surgery (Temporal sensitivity meta-analysis). Similarly, surgical case volumes may impact the outcomes of urologic oncologic surgery; because of presumed differences in surgical 'proficiency', low-volume centers may deliver inferior outcomes compared to high-volume centers (proficiency sensitivity meta-analysis). These 2 sensitivity metanalysis will be helpful to understand the impact of the caseload on the outcomes of interest. If agreeable to the Editor, we would respectfully state that Robotic Radical Prostatectomy, Robotic Radical Cystectomy and Robotic Partial nephrectomy will be reported as RRP, RRC and RPN respectively.

_____

Reviewer: 3

Reviewer Name: Bhavan Rai

Institution and Country: Freeman Hospital

Newcastle, UK

Please state any competing interests or state 'None declared': None

Please leave your comments for the authors below

1) This is a broad topic for an SR covering all urological cancers which are far from ideal. SR ideally should address very specific questions and clear endpoints. This review appears to attempt to publish 5 systemic reviews under one heading. Individual cancers/techniques have different outcomes measures of interest. Therefore, the 5 outcome measures reported in this protocol will not be consistently applied to all techniques/cancers

REPLY to Comment 1: is sub-divided and provided for each individual sentence within Comment 1

Reviewer Comment 1.1: *This is a broad topic for an SR covering all urological cancers which are far from ideal. SR ideally should address very specific questions and clear endpoints.*

REPLY to comment 1.1: We fully agree with the reviewer that "SR ideally should address very specific questions and clear endpoints**"**. Our systematic review (SR) and meta-analysis (MA) does precisely that! At the outset, we clearly list our 5 key questions and then methodically answer each one. Contrary to the reviewer's subjective comment, SR of a "broad topic" is actually valuable. The main advantage of a broad scope is that it gives a comprehensive summary of the evidence and a unique opportunity to explore consistency of findings (and therefore generalizability) across different types of participants. (*Chapter 2.3.a - Broad versus narrow reviews; Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019*).

Reviewer Comment 1.2: This review appears to attempt to publish 5 systemic reviews under one heading. Individual cancers/techniques have different outcomes measures of interest.

REPLY to comment 1.2: We evaluate outcomes of open versus robotic surgery for 4 cancers. We seek to examine the state-of-the-field of open versus robotic urologic oncologic surgery over the past 20 years (2000-2020) by assessing the highest-volume oncologic surgeries: radical prostatectomy (RP) for prostate cancer, radical cystectomy (RC) for bladder cancer, partial nephrectomy (PN) and radical nephrectomy (RN) for kidney mass and retroperitoneal lymphnode dissection (RLND) for testicular cancer. In our comprehensive SR and MA, studies comparing different surgical procedures are not combined in the same analysis. (*Chapter 2.3.2; 'Lumping' versus 'splitting' Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019).*

As clearly stated in our Methods section our SR will do **SEPARATE** analyses for each surgical procedure and for each outcome of interest. For each procedure type, we will methodically compare open versus robotic surgery as

regards five key questions: penetrance in the field, peri-operative data, procedural morbidity, oncologic outcomes, functional outcomes and financial cost (page 8, lines 199-202)

We agree that individual cancers/techniques may indeed have "different outcomes measures of interest", however, please note that the overwhelming majority of our identified outcome measures are identical between the 4 cancers studied, thus lending themselves beautifully to a SR like the one we propose. These ***majority of similar outcomes measures (n=28)*** include:

a) Baseline characteristics: age, BMI, ASA score, sex, pT stage;

b) Intra-operative characteristics: operative time, blood loss, transfusion rates, hospital stay, readmissions, and complications (overall, early [<30 days], late [31-90 days], minor [Clavien <3], major [Clavien ≥3]; and six specific complication categories [wound issues, CVS, GI, DVT/PE, infectious, anastomotic issues]);

c) Pathologic characteristics: positive margin status, lymph node yield for prostatectomy/cystectomy and retroperitoneal lymphnode dissection;

d) Post-operative characteristics: recurrence rate, cancer-specific mortality, overall mortality;

d) Financial issues: intra-operative, post-operative

However, those characteristics that differ between the 4 cancers will be analyzed separately. These ***minority of dissimilar outcomes measures (n=9)*** include:

a) Baseline characteristics: prostatectomy (PSA, Gleason score for biopsy and prostatectomy specimen); cystectomy (neoadjuvant chemotherapy status); partial nephrectomy (tumor size, RENAL score, laterality);

b) Intra-operative characteristics: None

c) Pathologic characteristics: None

d) Post-operative characteristics: prostatectomy (continence, erectile dysfunction); partial nephrectomy (eGFR outcomes)

d) Financial issues: None

Reviewer Comment 1.3: Therefore, the 5 outcome measures reported in this protocol will not be consistently applied to all techniques/cancers.

REPLY to comment 1.3: We respectfully disagree. As clearly shown above, the 5 key questions (not "outcomes measures") reported in our protocol will indeed be applied consistently across all 4 techniques/cancers.

We would also like to make a few general comments. Our submitted protocol rigorously follows the Cochrane methodology to perform a Systematic review**.** A feasible review is one that asks a question that the authors are capable of addressing using the evidence available. SRs can focus on broad questions, or be more narrowly defined. There are advantages and disadvantages of each. A SR should focus on questions that are important to people

making decisions about health or health-care (*Chapter 2 - Determining the scope of the review and the questions it will address; Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019*). The main advantage of a broad scope is that it gives a comprehensive summary of the evidence and a unique opportunity to explore consistency of findings (and therefore generalizability) across different types of participants. (*Chapter 2.3.a - Broad versus narrow reviews; Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019*).

Reviews that examine multiple interventions and aim to identify which might be the most effective can be broader and more challenging than those looking at single interventions. The former can also be the most useful for end users, where decision making involves selecting from a number of intervention options (*Chapter 2.2 - Aims of reviews of interventions; Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019*).

Moreover, in order to properly organize our large study and to formulate our research questions (page 6, lines 140-156), we will strictly adhere to the Population, Intervention, Comparators, Outcomes, Timing and Setting (PICOTS) format (*Chapter 2.3 - Defining the scope of a review question; Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019*; *Samson et al. J Gen Intern Med. 2012 Jun;27 Suppl 1(Suppl 1):S11-19*). (page 21, Table 1)

2) Penetration is an outcome measure that will be reported using national databases as part of a primary research work. This, therefore, cannot be claimed to be an outcome as part of an SR. SRs are secondary research that appraises published primary research using standardised validated methods.

REPLY (to comment 2): We agree with the reviewer. Now, we clarify that our penetration analysis is clearly distinguished from our actual SR and MA. Thus, **penetration is no longer "an outcome as part of an SR" in our paper.** Penetrance is assessed by reviewing a national database (Premiere database). We do believe that penetrance, although not a part of our SR and MA, is nevertheless important complementary information. This is because the principal goal of our study is to understand if and how the face of urologic oncologic surgery might have changed in the past 20 years. Therefore, penetrance in the field is an important issue. In conclusion, our principal analysis remains a SR and MA comparing open vs robotic urologic oncologic surgery for various outcomes. Additionally, we will perform a distinct and separate analysis by interrogating the Premiere database to evaluate penetrance. (page 7, lines 158-170)

3) Quote "We have carried out a preliminary screening of the literature: only 3 comparative studies were identified for open vs robotic radical nephrectomy for kidney cancer and no comparative studies were identified comparing open vs. robotic retroperitoneal lymph node dissection for testis cancer; as, these two topics will be censored from further detailed analyses. Therefore, our SR and MA will compare data on ORP vs RRP, ORC vs RRC and OPN vs RPN exclusively"
The authors claim 5 cancers will be compared initially. It does appear that they have already done a literature search of sorts and have decided to exclude RPLNDs and RN. This isn't the correct way to perform an SR and casts doubt on the credibility of their approach.

REPLY (to comment 3): We understand the reviewer's point. As such, we will include RPLND and RN in our initial analysis as well. Whether we ultimately include RPLND and RN in our final SR and MA depends on how much comparative data is available in the literature for RPLND and RN.

That said, we fail to see how doing a preliminary analysis to assess feasibility with strict adherence to the Cochrane guidelines somehow "casts doubt on the credibility" of our approach.

A “preliminary” search is not a “definitive” “search” and does not necessarily imply a selection bias**.** As we state in our paper and in our Prospero protocol (CRD42017064958), we initiated our SR on September 1, 2016 and the search will be finalized on December 31, 2019. As reported in our PROSPERO protocol, preliminary search and piloting our study selection process have begun. These 2 phases are necessary to gain an idea of the amount and quality of research evidence available on a particular topic before a final research question and the scope of the SR can be finalized. None of the four authoritative guidance entities (*Cochrane Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019); PROSPERO Guideline for Authors; Methods and Guide for Effectiveness and Comparative Effectiveness Review of the Agency for Healthcare Research and Quality (AHRQ); Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P))* indicate that performing a preliminary search is an incorrect way to perform an SR. As such, the reviewer’s statement that our preliminary search “casts doubt on the credibility” of our approach is not supported by either Cochrane, PROSPERO, AHRQ or PRISMA .

Finally, there is precedent in the published BJM Open protocols for SR and MAs that have employed a preliminary analysis to inform the definitive analysis *(Abdi F, et al. Protocol for systematic review and meta-analysis: hop (Humulus lupulus) for menopausal vasomotor symptoms BMJ Open 2016;6:e010734; and Jong MC, et al. Mapping the concept, content and outcome of wilderness therapy for childhood cancer survivors: protocol for a scoping review BMJ Open 2019;9:e030544).*

As an aside, the reviewer mistakenly states that “the authors claim 5 cancers will be compared initially”. Our SR does not seek to compare “cancers”; rather, we will compare robotic versus open surgery for cancer.

Since our scrrening phase will end on 10 January 2020, the present protocol will take into account two surgical approaches, open and robotic, as follows: for prostate cancer - open radical prostatectomy (ORP) vs robotic radical prostatectomy (RRP); for bladder cancer - open radical cystectomy (ORC) vs robotic radical cystectomy (RRC); for kidney cancer - open partial nephrectomy (OPN) vs robotic partial nephrectomy (RPN) and open radical nephrectomy (ORN) vs robotic radical nephrectomy (RRN); and, open retroperitoneal lymph node dissection (ORPLND) vs. robotic retroperitoneal lymph node dissection (RRPLND) for testis cancer. (page8, lines 184-189)

4) Quality assessment:
Quote “All papers will be categorized according to the Oxford Level of Evidence Working Group 2011 levels of evidence (LOEs) for therapy studies and according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system”
GRADE evaluates the quality of evidence. It does not evaluate the quality of papers
The authors need to state the confounders particularly for non-RCTs for individual comparators in the protocol.

REPLY (to comment 4): We agree that “GRADE evaluates the quality of evidence. It does not evaluate the quality of papers”. To our knowledge, our paper does not state anything to the contrary.

Oxford LOEs alone are not sufficient to rate the level of evidence and the strength of recommendation of every single study. Although this approach helps to justify study selection, the Oxford LOEs may mean different things to different readers, and novel or hybrid approaches are not easily accommodated. This can lead to anomalous rankings.

For this reason, in order to provide as much clarity as possible, we will categorize each paper (reposting findings for each study, which represent the body of the evidence) as follows:

a) LOE for each study according to the Oxford Level of Evidence Working Group 2011 for therapy studies
b) GRADE of the recommendation of each study according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (*Guyatt, Gordon, et al. "GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables." Journal of Clinical Epidemiology 64.4 (2011): 383-394*)

We wish to clarify our methodology further. After completing the meta-analysis process, we will grade the evidence according to the GRADEpro *(Chapter 12.2.1 The GRADE approach Cochrane Handbook for Systematic Reviews of*

Interventions version 6.0 (updated July 2019). As reported in our Protocol manuscript (Pages 15-16, Lines 371-391) the quality of scientific evidence and outcomes will be assessed using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach. GRADE will offer a specific definition of the quality of evidence that is different in the context of making recommendations and in the context of summarizing the findings of a systematic review.

According to the GRADE methodology, we will rate evidence for a given outcome by upgrading or downgrading the evidence. Indications for upgrading the quality of evidence include having a large effect size and dose-response gradient. Indications for downgrading the quality of evidence include serious risk of bias, serious inconsistency between studies, serious indirectness, serious imprecision and likely publication bias. (*GRADEPro, https://gdt.gradepro.org/app/handbook/handbook.html#h.wlsfq2lmj0gb. GRADEpro. and (Guyatt, Gordon, et al. "GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables." Journal of clinical epidemiology 64.4 (2011): 383-394.)*

For this reason, in order to provide as much clarity as possible, we will categorize each paper as follows

a) LOE for each study according to the Oxford Level of Evidence Working Group 2011 for therapy studies (page 11, lines 256-264)

b) GRADE of the recommendation of each study according to the Grading of Recommendations Assessment, Development and Evaluation (GRADE) system (*Guyatt, Gordon, et al. "GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables." (J. Clin. Epidemiol. 2011; 64.4: 383) (page 16, lines 882-402)*

5) Statistical analysis

1. All analysis is ideally done with a random effect model. At the very least in cases of high heterogeneity, a random effect model must definitely be used and fixed-effect model should be reserved for low heterogeneity.

REPLY (**to comment 5.1):** We agree that, in the presence of heterogeneity, a random-effects MA weights the studies relatively more equally than a fixed-effect analysis. However, when there is concern about the influence of small-study effects on the results of a MA, a comparison between random and fixed effect should be assessed.

According to the *Cochrane Handbook for Reviews and Meta-analysis* (*chapter 10.4.4.1 Comparing fixed and random-effects estimates Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)* "If the random-effects estimate is more beneficial, review authors should consider whether it is reasonable to conclude that the intervention was more effective in the smaller studies. If the larger studies tend to be those conducted with more methodological rigor, or conducted in circumstances more typical of the use of the intervention in practice, then review authors should consider reporting the results of meta-analyses restricted to the larger, more rigorous studies. Formal evaluation of such strategies in simulation studies would be desirable."

"Note that formal statistical comparisons of the fixed- and random-effects estimates of intervention effect are not possible, and that it is still possible for small-study effects to bias the results of a meta-analysis in which there is no evidence of heterogeneity, even though the fixed- and random-effects estimates of intervention effect will be identical in this situation".

At this phase of our study, since the present paper describes our protocol of the SR, we do not know which effect would be used. For this reason, as we reported in our manuscript and according to the Cochrane Handbook, random

and fixed effects will be used in case of the presence or absence of heterogeneity, respectively. (pages 11-12, lines 267-286)

2. Quote: “Due to limitations in the Review Manager v5.3 software, analysis of continuous variables is possible only when data are presented as mean and standard deviation (SD). Since some studies may report continuous variables in “median” and “interquartile range” or “min/max” range, we will use a validated mathematical method to estimate “mean” and “SD”. When available, we will use data reported in a matched-pair comparison manner”
The Cochrane handbook which is accessible on Review manager clearly gives instructions on how to impute data from median/range to make MA feasible

REPLY (to comment 5.2): For continuous variables, sample mean and standard deviation are used to perform MA of clinical trials. However, sometimes, the reported results may only include the sample size, median, range and/or IQR. To combine these results in MA, we need to estimate the sample mean and standard deviation (SD) from them.

Several methods have been proposed to estimate the mean and SD from median, ranges and/or IQR. The Cochrane Handbook reports one method on how to impute data from median/range to make MA feasible.

However, there is no consensus about which method should be used (*Christopher J et al. “Dealing with missing standard deviation and mean values in meta-analysis of continuous outcomes: a systematic review.” BMC Medical Research Methodology vol. 18,1 25. 7 Mar. 2018,*) since the proposed approaches are established under the assumption that the data are normally distributed. In MAs, however, the medians and quartiles are often described when data do not follow a normal distribution.

For this reason, in our SR and MA we will use a validated mathematical method to estimate “mean” and “SD” as previously described *(Wan, X et al Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 14, 135 (2014)). (page 12, lines 279-282)*

3. The author's plan perform a meta-analysis of RCTs and non-RCTs. A meta-analysis of non-RCT’s is not recommended. Data from Non-RCTs has presented graphically using the Forrest plots, but pooled analysis is not recommended.
The reviewer is not certain if temporal and proficiency MA is validated and recognized methodology in SR and would suggest statistical input on this.

REPLY (to comment 5.3): According to the Cochrane Handbook *“Reviews of interventions have traditionally focused mainly on systematic reviews of randomized trials because they are more likely to provide unbiased information about the differential effects of alternative health interventions than Non RCTs. Reviews of Non -RCTs are generally undertaken when the question of interest cannot be answered by a review of randomized trials”* (Chapter 24.1.1 Why consider non-randomized studies of interventions?; Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)).

According to the Cochrane Handbook, an accepted justification for including Non-RCTs in a SR is “*To provide evidence of the effects (benefit or harm) of interventions that can feasibly be studied in randomized trials, but for which only a small number of randomized trials is available (or likely to be available)”.* (Chapter 24.1.1 Why consider non-randomized studies of interventions?; Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)).

Both RCTs & non-RCTs have strengths & weaknesses. Both merit inclusion in meta-analyses. Meta-analyses based on non-RCT studies produce estimates of effects similar to those of MAs based on RCTs.[1] Observational studies should not be excluded a priori, especially when there is a lack of RCTs in a specific research area.[1] Particularly for surgery, RCTs are sparse.[2] As a minor side point, we also respectfully note that the esteemed reviewer (Prof. Bhavan Rai) has himself published five SR and MAs as the main author or co-author; four of these five MAs included non-RCTs.[3-6]

1. Shrier I, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. Am. J. Epidemiol., 2007; 166.10:1203

2. Ng TT, et al. Meta-analysis in surgery: Methods and limitations. Arch Surg. 2006; 141: 1125.

3. Rai BP et al. Is cryotherapy a genuine rival to robotic-assisted partial nephrectomy in the management of suspected renal malignancy? A systematic review and meta-analysis. Urology 2018; 118: 6-11.

4. Kallidonis, et al. Critical appraisal of literature comparing minimally invasive extraperitoneal and transperitoneal radical prostatectomy: A systematic review and meta-analysis." Arab J. Urology 2017; 15.4: 267.

5. McLean A, Mukherjee A, … Rai BP. Trans-peritoneal vs. retroperitoneal robotic assisted partial nephrectomy in posterior renal tumours: need for a risk-stratified patient individualized approach. A systematic review and meta-analysis. Journal of Robotic Surgery,2019;1-9.

6. Phukan C, Mclean A, ... & Rai, BP.. Retzius sparing robotic assisted radical prostatectomy vs. conventional robotic assisted radical prostatectomy: a systematic review and meta-analysis. World J. Urol, 2019; 1-12.

As such, we plan to perform a MA of RCTs and non-RCTs. Since the methodological choices that may be made during a SR and MA which includes non-RCTs are complex and may affect the review findings, we decided to submit the present protocol to BMJ Open for consideration for publication (*chapter 24.1.3 The importance of a protocol for a Cochrane Review that includes non-randomized studies of interventions; Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)).*

To assess the risk of bias in the non-RCTs included in our MA, we will independently weigh the risk-of-bias for all the studies according to the Cochrane Handbook for Systematic Reviews of Interventions for including non-randomized studies. In consideration of the design of studies, we will likewise examine the risk of preassigned confounders identified as the possible predictors at the time of surgery. We will examine the publications for specific data on baseline characteristics that may have impact on outcomes of interest within their univariate analysis models *(24.5.2 Guidance and resources available to support review authors; Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). (page 12, lines288-300)*

Reviewer comment. “Data from Non-RCTs has presented graphically using the Forrest plots, but pooled analysis is not recommended”

REPLY: The use of the pooled graphical effect of a MA including non-RCTs is due to the assessment of the similarity between the studies. Forest plots allow the presentation of estimates and standard errors for each study, and in most

software (including RevMan) it is possible to omit summary estimates from the plots, or include them only for subgroups of studies. Given that effect estimates from the included studies can be expressed using consistent effect measures, Cochrane recommends that review authors display individual study results for Non-RCTs with similar study design features using Forest plots, as a standard feature. On the other hand, if the features of studies are not sufficiently similar to combine in a meta-analysis, Cochrane recommends to display the results of included studies in a Forest plot but suppressing the summary estimate *(24.5.2 Guidance and resources available to support review authors; Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019)*

However, we wish to re-emphasize that at it's current stage, the present paper is a protocol - we do not know the grade of heterogeneity between the studies.

Reviewer comment. "The reviewer is not certain if temporal and proficiency MA is validated and recognized methodology in SR and would suggest statistical input on this".

REPLY: As reported in the Cochrane Handbook "*A sensitivity analysis is a repeat of the primary analysis or meta-analysis, substituting alternative decisions or ranges of values for decisions that were arbitrary or unclear"*. (Chapter 9.7 Sensitivity analyses; Handbook for Systematic Reviews of Interventions version 6.0 (updated July 2019). In essence, our temporal and proficiency meta- analysis merely represent sensitivity analyses for two specific issues: temporal MA assesses chronologic evolution of outcomes and proficiency MA assesses impact of clinical volumes on surgical outcomes.

6) There are sentences that inadequately reference

REPLY to comment 6: Thank you. The typo regarding the missing reference has been added.

7) There are a number of high-quality SR that has been published on the suggested techniques. The current review is therefore unlikely to add to the existing literature. The methodology isn't correct and would recommend the current protocol for publication.

REPLY to comment 7*:* We acknowledge the reviewer's sentiment. Respectfully, however, we beg to differ. Our SR and MA is indeed unique in that it will comprehensively compare multiple outcomes of open versus robotic urologic oncologic surgery over the past 20 years (2000-2020). A synthesis of all available studies will identify the quality of data. (page 4, lines 89-100)

Unique and novel aspects of our paper include:

- Performance of multiple sensitivity and reverse-sensitivity analyses to control for baseline heterogeneity amongst studies; to our knowledge this is a first in the field.
- Comprehensive evaluation of two decades of published data
- Evaluation of oncologic outcomes; we present longer follow-up compared to prior MAs
- Performance of temporal MA (to assess chronologic evolution of outcomes) and proficiency MA (to assess impact of clinical volumes on surgical outcomes)
- Assess the GRADE of evidence of individual papers, and as a composite for each outcome.
- We also provide penetrance data from an administrative data-set – the Premiere database.

_____

Reviewer: 4

Reviewer Name: Francesco Prete
Institution and Country: University Medical School of Bari, Bari, Italy
Please state any competing interests or state 'None declared': None declared

Please leave your comments for the authors below
Thank you for asking me to review this protocol
I found it mostly coherent with its aims and in general appropriately supported by research plan.
Regarding assessment of risk of bias for randomized and nonrandomized studies respectively, I understand that Authors will likely use the Cochrane RoB2 ans ROBINS-I tool, this could be specified in the Quality Assessment section of the manuscript.

REPLY TO REVIEWER #4

We thank the reviewer for the comment. Rob2 and Robins I tools will be used, as reported in the paragraph Risk Bias Assessment (references reported).