Aug 23, 2019

John Fletcher, MD
Associate Editor
BMJ

Dear Dr. Fletcher,

Thank you very much for considering our paper and giving us the opportunity to revise it. We are very grateful for the reviewers' and editor's constructive comments which greatly improved our paper. We have made extensive revisions to our manuscript, especially in the sections on statistical models and reports; we also modified our figure and tables to conform to journal requirements. All the changes in the revised manuscript are highlighted in red. We have provided a point-by-point response to reviewers' comments. We hope that the revised version will now meet the standard of publication of *BMJ*.

Thank you again for offering the opportunity of publishing our manuscript.

Sincerely yours,

Shi-Wei Huang

Chung-You Tsai

Chi-Shin Tseng

Ming-Chieh Shih

Yi-Chun Yeh

Yeong-Shiau Pu

Kuo-long Chien

Yu-Kang Tu

All correspondence to: Prof Yu-Kang Tu

Institute of Epidemiology and Preventive Medicine

College of Public Health

National Taiwan University

Taipei Taiwan.

5F, NO. 17, Hsu-Chow Road, Taipei, 100, Taiwan

Tel: +886 2 3366-8039

Fax: +886-2-2351-1955

E-mail: yukangtu@ntu.edu.tw

Response to the editors:

Detailed comments from the meeting:

1. This is a clinical topic relevant to many of our readers and your review appears to have a clear message.

Response: Thank you for the positive comment.

2. Our main reservation at the moment is the complex and somewhat unclear presentation of the results. Please see our statistician's report for guidance on revision.

Response: We have carefully revised the text and the presentation of results according to the statistician's suggestions.

3. Please can you provide an illustration or perhaps a text box that describes how each of the procedures is carried out. We appreciate this will be obvious to surgeons but our general readers may appreciate a little more explanation.

Response: We understand that the readers of BMJ are general physicians. We have now provided a graphical explanation for the three types of surgery in figure 1.

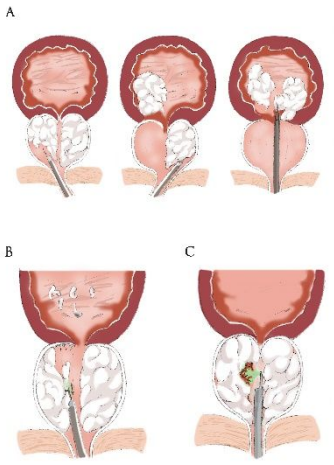


Figure 1: Different endoscopic surgical methods for benign prostate hyperplasia. A: enucleation methods: peeling the whole prostate adenoma from prostate capsule using end-firing laser fiber or designed bipolar loop, then morcellating the adenoma with a shaver; B: resection methods: resecting the enlarged prostate adenoma with monopolar or bipolar resection loop piece by piece; C: vaporization methods: vaporizing the enlarged prostate adenoma with side-firing laser fiber or mushroom-like bipolar electrode.

4. Our patient editor noted that there is no PPI declaration or dissemination plan. Please include these in your revision.

Response: Thank you for the comment. We have included the following PPI declaration in our revised manuscript:

PPI declaration

This research was done without patient involvement. Patients were not invited to comment on the study design and were not consulted to develop patient-relevant outcomes or interpret the results. Patients were not invited to contribute to the writing or editing of this document for readability or accuracy. (page 14 lines 3-7)

5. Please can you explain or discuss a little more to what extent patients have a real choice about surgery? Will it be down to what the local surgeons offer or are good at?

Thank you for the comment. We have added the following paragraph to the revised manuscript:

Urologists prefer using resection or enucleation methods in large prostates (>70-80 ml), using laser modalities or vaporization methods in patients with increased bleeding risk. Moreover, financial considerations, such as whether the cost is fully reimbursed by insurance companies, the availability of required equipment, and surgeons' skill and experience also affect the choice of surgical method. As shared decision-making has become the norm in clinical care, clinicians should discuss the benefits and risks of different surgical approaches with their patients. (page 26 lines 2-8)

6. Please can you elaborate a little how the outcome measures in your review translate into symptomatic improvement for patients? How noticable is a change in QMax of 3 ml/sec?

Thank you for your comment. In the medical treatment of BPH, α-blocker with/without 5-α reductase inhibitor can improve Qmax by around 0.9-2.4 ml/sec

compared with placebo, which was previously considered clinically significant. Hence, the change in Qmax on 3 ml/sec is clinically meaningful for patients undertaking BPH treatment. We have added the following paragraph to the revised manuscript:

In the medical treatment of BPH, α-blocker with or without 5-α reductase inhibitor can improve Qmax around 0.9-2.4 ml/sec compared with placebo, which was considered clinically significant.[26 27] In our study, enucleation methods were found to improve Qmax by 1.71-1.98 ml/sec and 4.12 to 4.82 ml/sec more at 6-12 and 24-36 months after surgery when compared with vaporization. Hence, the difference in Qmax between enucleation and vaporization methods was clinically meaningful. (page 22 lines 11-16)

7 Please revise your paper to respond to all of the comments by the reviewers. Their reports are available at the end of this letter, below.

Thank you for the comment. We have revised our paper and responded to all the comments from the reviewers.

Response to Reviewer #1 Comments:

-There are various grammatical/syntactical errors throughout the paper, for example:
line 16: should read "the year 2000"
line 19: should read "benign prostatic hyperplasia"; should be corrected throughout the manuscript

Response: Thank you for the comment. We have sent our revised manuscript for English editing and corrected all the errors.

-I am confused by the authors use/meaning of "blood clot tamponade" throughout the paper.

Response: Thank you for the comment. We have changed “blood clot tamponade” to “clot retention” (blood clot retention in the bladder)

-The authors appear to lump bipolar electrocautery into the enucleation group during their discussion. This should be avoided as bipolar TUR is not traditionally regarded

as an "enucleation" technique.

Response: Thank you for the comment. In the past, bipolar electrocautery was regarded as a resection technique. However, with the use of specialized designed loops, bipolar electrocautery can be used for resection, vaporization, or enucleation. Therefore, we considered bipolar TURP, bipolar enucleation, and bipolar vaporization as separate techniques and compared them in our network meta-analysis. We have revised the text to avoid confusion.

-Lastly, it would have been nice to see comparisons between the various laser techniques as I'm not convinced that monopolar TURP remains the "gold standard" in outlet procedures. While this comment is not directly related to the manuscript, the authors should consider this in future studies

Response: Thank you for the comment. We will consider this approach in future studies.

Response to Reviewer #2 Comments:

Comments:
This manuscript tried to describe and compare different new surgical techniques in the efficacy and safety of BPH by applying frequentist network meta-analysis. I believed that Authors should spend lots of time on data collection and analysis, while in my opinion, the work still has no novelty in terms of methodology and conclusion. I do not think the manu. should be suitable for publication in the BMJ. Anyway, I have some comments or suggestions as following:

1. The updated searching date needs to avoid missing searches to the greatest extent(In the manuscript, the searching date was only updated to Mar. 2018).

Response: Thank you for the comment. We have extended the search date to March 2019 and included four more new trials into our study. The surgical treatment of BPH is an important topic, as BPH is a common condition. The conclusions of our updated analyses with the inclusion of the four new trials remain unchanged, indicating that our results are robust. The revised paragraph on literature search reads as follows:

To identify published and unpublished trials, we used electronic databases including Pubmed (inception - March 2019), Embase (inception - March 2019), and Cochrane

clinical trial registers (inception-March 2019) (page 9 lines 6-9)

2. In Supplementary figure 2, the blank area is supposed to be yellow, indicating unclear risk of bias? Maybe, authors are not familiar with the statistical software.

Response: The blank area means that we did not undertake the appraisal of the particular domain for a specific trial. We evaluated several primary and secondary outcomes, both long- and short-term, and the follow-up length was different across the included trials. For instance, the domain of attrition bias was not evaluated and kept blank if the follow-up length was shorter than 3 months. This was explained in supplementary Figure 2:

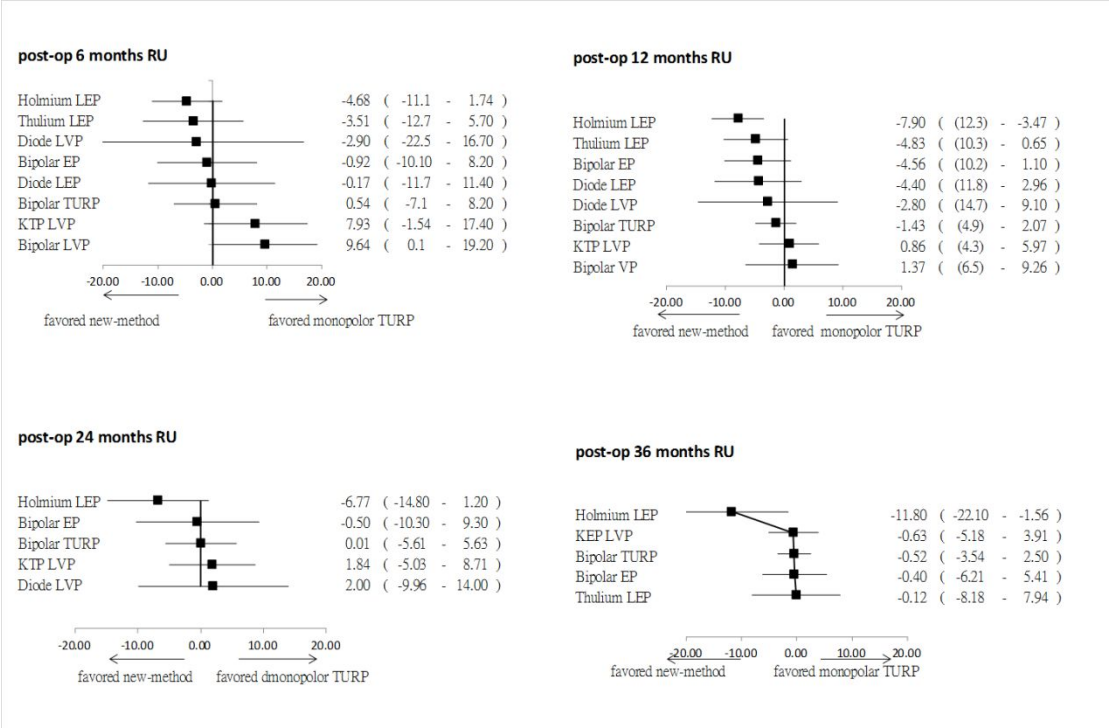
We used the Cochrane Collaboration's risk of bias tool to appraise study quality. Each item was adjudicated within each study and the results are represented in a risk of bias table. We considered the study as low risk in blinding of participants if patients were kept unware of which surgical method they received; blinding of outcome assessment was considered as present if the assignment of treatments was unknown to the assessor of functional outcomes (IPSS and Qmax); low risk of selective reporting was considered as present if complications were evaluated using a pre-specified form such as the modified Clavien-Dindo classification or a detailed report of complications. We only evaluated the domain of incomplete outcome data if the follow-up period was greater than 3 months. The domain of a trial was kept blank if it was not assessed. (supplementary Figure 2)

3. The authors mentioned the search strategy of grey literature by manual search in the method part. However, the information was missed in the parts of flow diagram and results.

Response: Thank you for the comment. We hand searched abstracts of relevant conferences and references in published meta-analysis articles and relevant articles. However, these articles could all be found in Embase. Hence, we did not add any other article for analysis from the manual search. This was explained in Supplement Figure 1.

4. It seems that the authors selectively reported surgical outcomes, for the key outcome of PVR was not included in the final analysis, while it was mentioned and planed in the protocol.

Response: We have added the results of PVR to supplementary Figure 8, which are similar to those of Qmax. Multiple outcomes were evaluated at different time points after BPH surgical treatment. We selected IPSS and Qmax at 6, 12, 24, 36 months after surgery as the main outcomes in our protocol. PVR is an important objective outcome for BPH treatment; however, it is less frequently reported in randomized clinical trials. While there were 51, 52, 18, and 14 trials reporting 6- , 12-, 24-, and 36-month Qmax after surgery, respectively, there were only 30, 31, 8, and 5 trials reporting PVR. (supplementary figure 8)



5. The authors have included IPSS and Qmax as the primary outcomes, however, it is suggested that the primary outcomes should be consisted of both outcomes of efficacy and safety. How about the primary endpoints of safety? Please describe in the context. In addition, the total number of primary outcomes is suggested no more than three (see Chochrane Handbook).

Thank you for the comment. Both efficacy and safety are important in the evaluation of surgical treatments for BPH. The most important safety outcomes for endoscopic BPH surgery are TUR syndrome (hyponatremia) and intra/post-operative bleeding. However, TUR syndrome is no longer a problem with the use of new surgical techniques (bipolar or laser modalities). Therefore, blood loss remains the major safety concern. We chose Hb-decline, clot retention, duration of catheterization, and the need for blood transfusion as surrogates of bleeding problems and considered

these four parameters as secondary outcomes. We chose the 12-month IPSS and Qmax after surgery as primary outcomes and the 6-, 24-, and 36-month IPSS and Qmax as secondary outcomes and rephrased the text.

Response: We chose the postoperative 12 months Qmax and IPSS as primary outcomes and other clinical measurements as secondary outcomes. (page 10 lines 8-10)

6. Patients and Urologists often face the challenge in selecting the optimal intervention among various treatments of BPH, so GRADE approach is suggested to classify evidence into different recommendation levels based on single outcome.

Response: Thank you for the comment. We have now used the GRADE approach to rate the evidence into different recommendation levels in Supplementary Table 8.

We further applied the Grading of Recommendation Assessment, Development and Evaluation (GRADE) approach to assessing evidence quality regarding the primary outcomes, which was considered to be critical in clinical decision-making. (page 11 lines 1-4)

Response to the Reviewer #3 Comments:

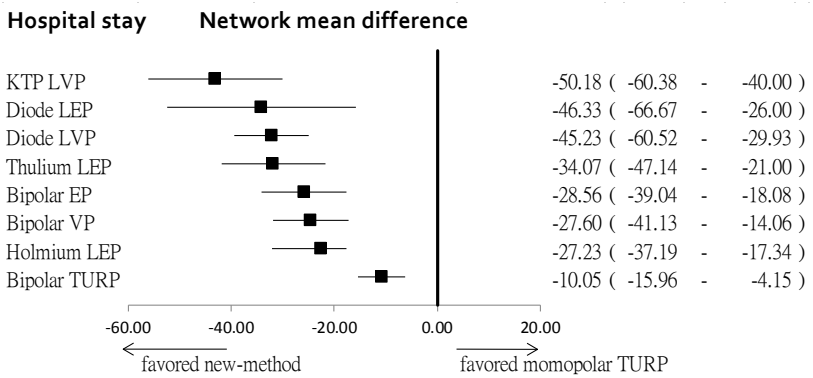
Comments:
BPH is a common condition that affects aging men. As authors point out, there are several surgical treatments available for it. The authors compare and rank 9 surgical treatments. The authors are congratulated for undertaking this substantial project. The study is well done.

A few points to consider.

MAJOR:
1. Did any of the trials included in this network meta-analysis report on length of hospital stay? This should be considered an endpoint too, if available.

Response: Thank you for the comment. We agree that length of hospital stay is also an important endpoint. While 91 trials reported catheterization duration, only 53 trials reported the length of hospital stay. We extracted the corresponding data and added

the results of our analysis in Supplementary figure 8. The treatment ranking for the results of length of hospital stay was similar to that of catheterization duration.

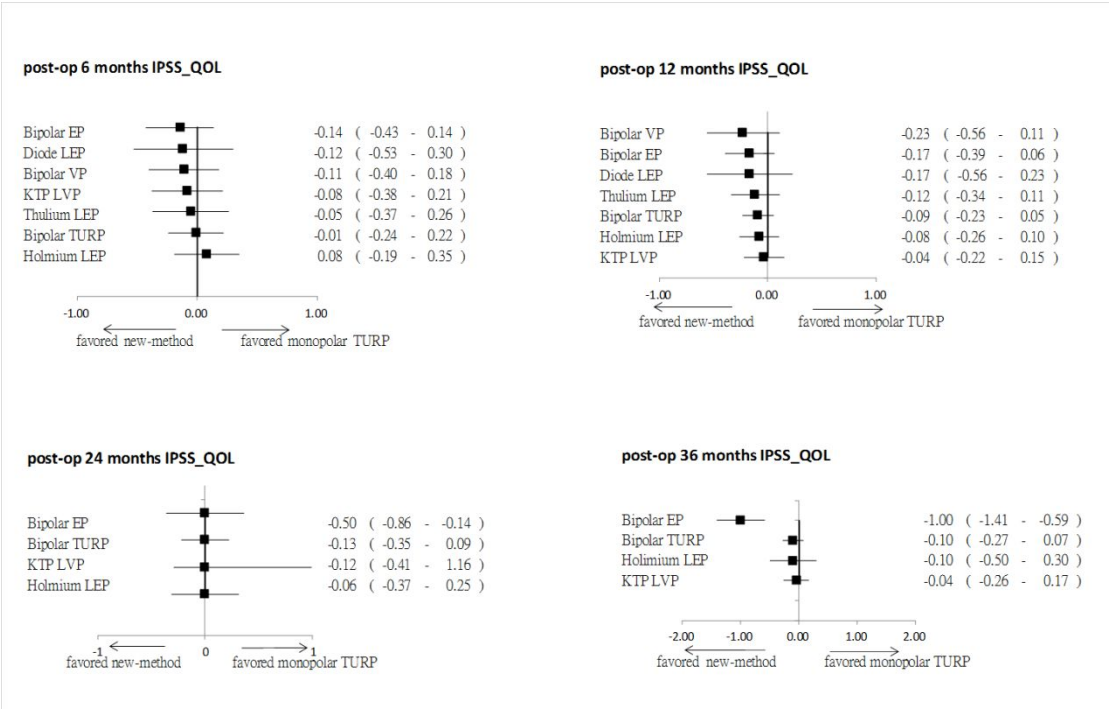


2. Did any of the trials included in this network meta-analysis report on the individual IPSS scores? This maybe helpful as nocturia is a multifactorial entity and does not always relate to BPH. Further, certain IPSS questions pertaining to 'urgency' and 'frequency' after treatment of BPH may reflect permanent detrusor overactivity (from BPH or not) and thus may falsely inflate IPSS scores, and mask the effectiveness of the treatment. Therefore looking at individual scores is sometimes helpful.

Response: We agree that some symptoms may falsely inflate the IPSS. The surgical treatment of BPH mainly resolves symptoms related to voiding and storage induced by poor voiding. As the included trials did not report individual IPSS values, we were unable to analyze the effects of surgical treatment on individual scores.

3. In the sensitivity analysis, the authors should strongly consider including IPSS QoL. This is probably the most important component of the IPSS score that governs patients' decision to undergo treatment and postoperative satisfaction.

Response: Thank you for the comment. We have added IPSS QoL to our article (supplementary Figure 8). IPSS QOL is an important subjective outcome for BPH treatment. However, it is a global satisfactory parameter with a single score ranging from 0 to 7. When comparing multiple active treatments, it is not a very sensitive outcome. Besides, fewer studies reported it as the primary or secondary outcome. Compared with 50, 51, 17, and 13 trials reporting the 6-month, 12-month, 24-month, and 36-m IPSS, respectively, only 28, 26, 9, and 6 trials reported IPSS QOL at the same time points. Hence, we chose IPSS as our main outcome.



MINOR:

1. Please clarify this statement in the results "However, only seven methods reported outcomes for 24-36-month postoperative follow-up, and these were predominantly pairwise comparisons of bipolar TURP with monopolar TURP."

Response: Thank you for the comment. We have modified the statement as follows:

Outcomes for the postsurgical 24-36 months' follow-up were only available for seven of the nine surgical methods compared in our network meta-analysis. The majority of trials performed a pairwise comparison between bipolar and monopolar TURP or compared holmium LEP and KTP LVP with bipolar or monopolar TURP. (page 16 lines 4-7)

2. I could not find in the text what was considered a significant p-value? Was it <0.05? If so, then should the authors consider using correction for multiple corrections such as Bonferroni etc?

Response: Thanks for your comment. Studies with multiple comparisons inevitably inflate type 1 error. However, there is no generally accepted method to correct for the inflated type 1 error in network meta-analyses. Although we used a p-value <0.05 for statistical significance, we avoided taking a naïve approach to interpreting the p-values by focusing on the estimated effect sizes and confidence intervals. Additionally, the differences between active treatments are usually small and

statistically non-significant. Hence, we also included treatment ranking and SUCRA in the results to help readers with interpretation of differences between treatments.

Response to the Reviewer #4 Comments:

Recommendation:

Comments:
Very good paper reporting the results of a meta-analysis about a hot topic.
The only one limitation of this study is that the authors did not analyze early postoperative urinary symptoms such as urgency, or post-micturition pain: this factor represents the main concern of some of the reported procedures. The authors stressed this limitation in the discussion.

Response: Thank you for the comment. For the coagulation treatment methods, these symptoms could be severe and persist for a while. In contrast, these symptoms are usually mild, can be relieved with medications, and improve more quickly in patients treated with the new surgical techniques.

Reviewer: 5

Recommendation:

Reviewer: 5

Recommendation:
Comments:
Thank you for the opportunity to review this interesting paper, on clearly an important topic. I have been through this from a statistical perspective, and have a number of comments and suggestions for improvement, as follows.

Response: We would like to thank Prof. Riley for his careful reading of our paper and constructive comments.

1) The authors should use PRISMA-NMA, not PRISMA (http://www.prisma-

statement.org/Extensions/NetworkMetaAnalysis.aspx)

Response: We have changed the form from PRISMA to PRISMA-NMA.

2) I-squared is not a test of heterogeneity, and indeed is a poor direct measure of heterogeneity. (Rucker G, Schwarzer G, Carpenter JR, et al. Undue reliance on I(2) in assessing heterogeneity may mislead. BMC Med Res Methodol 2008;8:79)

Response: We also agree that I-squared is not a good measure for heterogeneity, especially in rare-event scenarios, and only included it due to its popularity. We have replaced our statistical model in pairwise comparisons with one-stage meta-analysis using the exact binomial likelihood and offered tau-squared to avoid misleading the readers. The revised text reads as follows:

"for binary variables, we conducted a one-stage meta-analysis using a generalized linear mixed model with the exact binomial likelihood." (page 11 lines 17-18)

"We conducted the arm-based network meta-analysis using generalized linear mixed models[14] with a REML approach. No imputation for zero-cell counts was performed. All analyses for binary data were undertaken using the GLIMMIX procedure of the SAS software version 9.4 (SAS Institute, Cary, NC) with the Laplace integration method." (page 12 lines 12-14)

3) The authors use Stata and the mvmeta module; do they actually mean they used the network module (which uses mvmeta in the background)?

Response: We did use the network module (which uses mvmeta in the background) in our analysis. We have revised the text as follows:

"For continuous variables such as functional outcomes and perioperative parameters, we performed a contrast-based network meta-analysis using Stata (Stata Corp., College Station, TX) through a network module based on the 'mvmeta' command for multiple treatment comparisons with the restricted maximum likelihood (REML) approach. Between-study variances were equalized, correlations were set to 0.5, and confidence intervals were estimated based on asymptotic error variance and normal distribution." (page 12 lines 1-7)

4) If relevant, how were multiple intervention effects from the same study handled in

the analysis (i.e. was their correlation accounted for)?

Response: We agree that the correlation from multiple intervention effects from the same study could be an issue. However, to take it into account requires individual patient data or the included studies have to report the correlations between multiple treatment effects. As both were not available in our network meta-analysis, we therefore did not consider the correlations between multiple intervention effects.

5) What assumptions were made about the specification of the between-study variance matrix components? E.g. were between-study variances made equal and correlations set to 0.5, as is standard?

Response: The between-study variances were made equal and correlations were set to 0.5 as standardized in the *network* module in Stata. We have added a short statement to our manuscript:

"For continuous variables such as functional outcomes and perioperative parameters, we performed a contrast-based network meta-analysis using Stata (Stata Corp., College Station, TX) through a network module based on the 'mvmeta' command for multiple treatment comparisons with the restricted maximum likelihood (REML) approach. Between-study variances were equalized, correlations were set to 0.5, and confidence intervals were estimated based on asymptotic error variance and normal distribution." (page 12 lines 1-7)

6) Was a random effects meta-analysis used in the network meta-analysis, as in the pair-wise analyses? Was the uncertainty of between-study variance estimates accounted for when deriving subsequent CIs for summary results? E.g. using Hartung-Knapp Sidik-Jonkman approach?

Response: Thank you for the comments. We used a random-effects model in both the pairwise and network meta-analysis. The Hartung-Knapp Sidik-Jonkman approach adjusts for uncertainty of heterogeneity variance estimates in pairwise meta-analyses. While this approach has been extended to the network meta-analysis, it has not been implemented in Stata *network* module yet. Therefore, we reported our results using the default method, REML, of the *mvmeta* command, which uses the asymptotic error variance and normal distribution to construct the confidence intervals. We agree that further efforts should be made to develop robust methodologies that take the uncertainty of heterogeneity variance estimates into account in network meta-

analyses.

7) What estimation method was used for the network meta-analyses? REML?

Response: Thanks for the comments. We used REML as the estimation method. We have rephrased our article to make this clearer:

"For continuous variables such as functional outcomes and perioperative parameters, we performed a contrast-based network meta-analysis using Stata (Stata Corp., College Station, TX) through a network module based on the 'mvmeta' command for multiple treatment comparisons with the restricted maximum likelihood (REML) approach. Between-study variances were equalized, correlations were set to 0.5, and confidence intervals were estimated based on asymptotic error variance and normal distribution." (page 12 lines 1-7)

8) STATA should be Stata

Response: We have changed STATA to Stata.

9) "We applied a 0.5 zero-cell correction only in the pairwise meta-analysis as a default of the Stata meta command but not in the network-meta-analysis to obtain a more unbiased estimation." – I don't think adding 0.5 in the pair-wise analysis is as appropriate as using the Sweeting correction. (Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. Stat Med 2004;23(9):1351-75)

Moreover, I do not think the 2-stage framework is correct when outcomes are rare, and a 1-stage model is more exact and appropriate. That is, the mvmeta module in Stata requires treatment effect estimates and their variances to be calculated for each study, and these are then pooled in a meta-analysis. However, when the event rate is low, there is a concern that such effect estimates are not normally distributed and variances are poorly estimated. This, a one-stage network meta-analysis that uses the exact binomial likelihood might be preferred. Did the authors consider this, or evaluate if their conclusions are robust to this issue?

See for example:

1. Riley RD, Jackson D, Salanti G, Burke DL, Price M, Kirkham J, et al. Multivariate and network meta-analysis of multiple outcomes and multiple

treatments: rationale, concepts, and examples. BMJ. 2017;358:j3932.
2. Salanti G, Higgins JP, Ades AE, Ioannidis JP. Evaluation of networks of randomized trials. Stat Methods Med Res. 2008;17(3):279-301.

Response: Thank you for this comment. We have changed our pairwise meta-analytic model to one-stage meta-analysis using the exact binomial likelihood. The text has been revised as follows:

"for binary variables, we conducted a one-stage meta-analysis using a generalized linear mixed model with the exact binomial likelihood.
." (page 8 lines 5-7)

"We conducted the arm-based network meta-analysis using generalized linear mixed models[14] with a REML approach. No imputation for zero-cell counts was performed. All analyses for binary data were undertaken using the GLIMMIX procedure of the SAS software version 9.4 (SAS Institute, Cary, NC) with the Laplace integration method." (page 12 lines 10-14)

10) Page 12: met-analysis should be meta-analysis

Response: The text has been amended.

11) "We evaluated the potential inconsistencies… " – more details are needed on what criteria they used to confirm consistency or inconsistency. These results should also be provided in the main text, as this is a fundamental part of a network meta-analysis.

Response: We have revised the text regarding this inconsistency and included the results in the main text as suggested:

"We evaluated potential inconsistencies between direct and indirect evidence within the network meta-analysis using the design-by-treatment interaction model[15] and side-splitting method.[16] The design-by-treatment interaction model provides a global assessment of consistency across the entire network. Side splitting method separated evidence on a particular comparison into direct and indirect evidence and then assessed their differences." (page 12 line15 –page 13 line1)

“We found no evidence of global inconsistency in any primary or secondary outcomes using the design-by-treatment interaction models except catheterization duration. After removing the single trial comparing bipolar TURP with thulium LEP, the inconsistency was no longer observed. No substantial inconsistency between direct and indirect comparisons was observed by side-splitting models (see supplementary Table 9).” (page20 line8-13)

12) It is not clear if the meta-regression described in the methods relates to the network meta-analysis or the pair-wise meta-analysis.
Regardless, meta-regression is very prone to study-level confounding, so I would class these as an exploratory analysis. In particular, the association of mean prostate volume and overall treatment effect is at the ecological level – what we really need is the association between individual prostate volume and individual treatment response.
This could only be ascertained from IPD and within-trial information, and so I strongly suggest the meta-regression of prostate volume is downplayed.

A nice paper in the BMJ on this recently is Fisher (Fisher DJ, Carpenter JR, Morris TP, et al. Meta-analytical methods to identify who benefits most from treatments: daft, deluded, or deft approach? BMJ 2017;356:j573). Also see: Hua H, Burke DL, Crowther MJ, et al. One-stage individual participant data meta-analysis models: estimation of treatment-covariate interactions must avoid ecological bias by separating out within-trial and across-trial information. Stat Med 2017;36(5):772-89. doi: 10.1002/sim.7171

Response: We agree that meta-regression of prostate volume should be viewed as an exploratory analysis. The meta-regression described in the Methods section was used in our network meta-analysis. In practice, the treatment efficacy of benign prostatic hyperplasia largely depends on the patient's prostate size. Therefore, whether prostate size moderates the relative effect difference between treatments is of great clinical importance. Moreover, during data acquisition, we noted the differences in the range of prostate size among the included studies, which may cast doubt on the validity of pooling these study results. This issue is best addressed with meta-regression, and we agree that using aggregated data in meta-regression may introduce ecological bias. However, since we did not have access to individual patient data, we reckoned that the best we could do to tackle this issue was to perform a meta-regression on the mean prostate volume of each study. Still, we thank the reviewer for pointing out the potential bias that may arise from using aggregated data in meta-regression. We have

now downplayed the results by putting the meta-regression results in the supplementary table 10 and modified our statements to avoid misleading the readers. We have also discussed this limitation in our article:

"Sixth, we used the mean prostate size of each article in our meta-regression on the relation between prostate sizes and treatment efficacy, because we did not have individual patient data. This may have led to ecological bias and increase proneness to study-level confounding." (page 27 lines 16-19)

13) Multiple time-points are considered. Was the correlation across time-points accounted for? Or was a separate network meta-analysis done at each time-point? If the latter, then were most time-points available in most studies, such that missing time-points is not a big issue?

Response: We performed separate network meta-analyses for each time point. The correlation between the time points was not considered, since the correlation between results of different time points were not reported in the included studies and individual patient data were unavailable. Although different studies chose different follow-up periods, we used the postoperative 12-month IPSS and Qmax as the primary outcomes, since the 12-month follow-up was viewed as the shortest follow-up length to prove the efficacy of BPH surgical treatment according to the 2013 EAU (European association of urology) guideline: Management of Non-Neurogenic Male LUTS (https://uroweb.org/guideline/treatment-of-non-neurogenic-male-luts/). Post-operative 6-month data was viewed as the short-term surgical outcome. Post-operative 24- and 36-month follow-up were also important to evaluate the mid-term outcomes; however, only a few articles had these longer follow-up periods. Therefore, these time points (6, 24, and 36 months) served as the secondary outcomes.

14) I find Table 2 hard to follow. Why are the authors using dichotomised values of prostate volume here?

Response: In clinical practice and urology research, prostate size is usually divided into small, medium, and large. Small prostates are those smaller than 30 or 40 ml, and large prostates are those greater than 60, 70, or 80 ml. The risk/benefit of BPH surgery is not directly related to prostate size. Patients with medium-sized prostates benefit the most from surgery due to the improvement in symptoms and a relatively lower risk of surgical complications. Clinicians are usually uncertain whether the urinary symptoms of patients with small prostates are caused by BPH, so the benefit

of surgery (such as symptom relief or maximal flow rate improvement) has been debated. When the prostate is greater than 60-80 ml, the risk of surgery increases abruptly due to a greater risk of intra-operative bleeding and incomplete removal of prostatic adenoma by resection or vaporization techniques. As there was a reduced number of trials with a mean prostate size greater than 80 ml, we therefore used 70 ml as the cut-off value to dichotomize the prostate size in the meta-regression. We found that our conclusions remained unchanged if we used 60 ml as the cut-off value to define large prostate. We have carefully revised table 2 and supplementary table 10 to improve readability.

References:

Rapisarda S, Russo GI, Osman NI, et al. The use of laser as a therapeutic modality as compared to TURP for the small prostate </= 40ml: a collaborative review. *Minerva Urol Nefrol*. 2019. doi: 10.23736/S0393-2249.19.03350-2.
Demirdag C, Citgez S, Tunc B, et al. The Clinical Effect of Bipolar and Monopolar Transurethral Resection of the Prostate More Than 60 Milliliters. *Urology*. 2017;98: 132-137.
Zhu L, Chen S, Yang S, et al. Electrosurgical enucleation versus bipolar transurethral resection for prostates larger than 70 ml: a prospective, randomized trial with 5-year follow up. *J Urol*. 2013;189:1427-1431.
Li K, Wang D, Hu C, et al. A Novel Modification of Transurethral Enucleation and Resection of the Prostate in Patients with Prostate Glands Larger than 80 mL: Surgical Procedures and Clinical Outcomes. *Urology*. 2018;113:153-159

15) Sometimes in the text the comparator group is difficult to identify

Response: We have revised the text to make it clearer what the comparator group is.

16 We need ranking plots added, and information about mean rank and SUCRAs, to help summarise the network meta-analysis results in more detail.

Response: We have added ranking plots, mean rank, and SUCRAs to supplementary figure 5 and table 6.

17) For the continuous outcomes, we need more details on whether the effect estimates were appropriately derived from analysis of covariance (i.e. after adjusting for baseline) in each trial, as this is the best method.[1] If not, then were effect estimates based on change scores or final value only? And if so, how might this influence the findings?

Response: For continuous outcomes in this network meta-analysis, the effect estimates of all included studies were based on differences in post-treatment values between treatments, which this is standard practice in BPH research. Failure to adjust for baseline values in a properly randomized clinical trial is unlikely lead to bias, since the baseline values should be similar in all treatment arms (what we meant by bias is the average estimated difference in treatment effects for a large amount of studies or a large sample size). However, it may result in slightly lower power and favor the null conclusion as shown in the note by Vickers and Altman. Since we only included randomized controlled trials, using the final value only is still valid. We agree that adjusting for baseline values would be desirable, but none of the studies reported results adjusting for baseline value, nor do we have access to individual patient data.

Vickers AJ, Altman DG. Statistics notes: Analysing controlled trials with baseline and follow up measurements. Bmj. 2001;323(7321):1123-4.

18) Abstract conclusion says: ““The efficacy of vaporization in large prostates seems Questionable” – no results in the abstract relates to this point as far as I can tell? Also, see my comment about the concern of meta-regression of prostate volume above.

Response: Thanks for the comments. We have deleted this sentence.

19) Moreover, the definition of large is arbitrary. “In the large prostate group (mean PV >70 gm), … “ – we need to be looking at prostate volume as a continuous variable within trials before making strong conclusions

Response: As we have explained in our response to comment (14), prostate size is commonly categorized into small, medium, and large in clinical practice and urology research. Small prostate is defined as below 30 or 40 ml, and large size is defined as a prostate volume greater than 60, 70, or 80 ml. Based on this convention, we chose 70 ml as the cut-off value to define large prostate size and undertook a sensitivity analysis to see whether using 60 ml as the cut-off value would alter our conclusion. We found that our conclusion was robust. As we explained in our response to comment 12, we have downplayed the results of meta-regression and modified our statements to highlight the potential limitations of our approach.

1. Jhanwar A, Sinha R, Bansal A, et al. Outcomes of transurethral resection and holmium laser enucleation in more than 60 g of prostate: A prospective randomized study. Urology Annals 2017;9(1):45-50
19) Moreover, the definition of large is arbitrary. “In the large prostate group (mean PV >70 gm), … “ – we need to be looking at prostate volume as a continuous variable within trials before making strong conclusions