Peer Review File

Article information: https://dx.doi.org/10.21037/tau-23-642

Reviewer A

Comment 1:

1. The primary endpoint is vague and unclear

In the methods section it states USSQ scores and adverse events. In the objective it was efficacy and safety of mirabegron. The primary endpoint should choose one drug, state superiority vs equivalency in one USSQ category. Everything else can be secondary endpoints.

Reply 1: The authors thank you for your kind comments on our work. As you said, the description of the primary endpoint in the manuscript is vague and unclear. In this revision, we have made a detailed description of the primary endpoint.

Changes in the text: The primary outcomes were the USSQ urinary symptom and body pain scores related to SRSs treated with mirabegron, as well as adverse events.

Comment 2:

2. Time and surgery are significant event modifiers that are not consistent across the different RCTs. Therefore, there is much higher intransitivity than would be expected at surface glance.

For example, Abdelaal's trial has USSQs filled out at day of stent removal which ranged from 1.5 to 8 weeks. Also, the RCT included complex and uncomplex ureteroscopy. While this worked for the individual RCT as the complexity and stent duration could be assessed for significance, it then becomes harder to compare such a study's medication effects to another. Tae's trial was uncomplex URS + USSQ and stent removal was at 2 weeks. Yavuz's was uncomplex URS + USSQ + stent at 4 weeks.

As per Lingeman et al (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3164810/) study we can see that time effect has a significant impact on USSQ scores independent of medications.

Furthermore, the RCTs absolutely must be screened to control for ureteroscopic complexity which changes the nature of the surgery, recovery and medication requirements. An endopyelotomy has significantly and intentionally more ureteral tissue damage than a standard uncomplicated ureteroscopy. The NMA should be more rigorously restricted and then the NMA re-run.

It is interesting that mirabegron has not been found to be significant for USSQ urinary symptoms compared to flomax on the forest plot. I believe that if the primary and secondary endpoints are clarified and the inclusion criteria for the NMA are expanded to control for time and type of ureteroscopy, this manuscript would be fantastic.

Reply 2: Thank you for your comment, which greatly improved our manuscript quality. As you mentioned, time and surgery are significant event-modifying factors.

Irani et al. found that using a 10 cm linear visual analog scale (VAS) to assess symptom tolerance revealed significant improvement over time for certain symptoms such as dysuria and hematuria(DOI: 10.1046/j.1464-410x.1999.00154.x). Lee et al. compared patients who had ureteral stent placement for 3 days versus 2 weeks post-ureteroscopy lithotripsy, with 84% reporting urgency, lower abdominal discomfort, and other urinary system symptoms, showing no difference in symptoms over time(Lee JH, Park HJ, Cho JM, Han KH, Kang JY, Jeong JY, et al. Is the 3-day stenting sufficient for ureteroscopic lithotripsy? Korean J Urol. 2003;44:1011–1014.). Recent research indicates more severe urinary symptoms in the 7-day stent group compared to the 3-day stent group, with statistically significant differences(DOI: 10.1089/end.2023.0189). Another study using IPSS, VAS, and QoL showed worsened urinary system symptoms and pain post-double-J ureteral stent placement for all patients. VAS and QoL scores significantly decreased at 9 months, and IPSS total score significantly decreased at 12 months (DOI: 10.5213/inj.2010.14.2.93). It remains unclear whether ureteral stent-related symptoms gradually improve over time, worsen before improving, or have no significant effect.

In their 2003 study, Joshi et al. introduced the USSQ questionnaire and validated it through a survey conducted four weeks after ureteral stent placement.(DOI: 10.1097/01.ju.0000049198.53424.1d). This is consistent with the findings reported by Damiano et al., who noted that complications of double-pigtail stents occur within the first four weeks after placement, including discomfort, bladder irritative symptoms, hematuria, bacteriuria with or without clinical urinary tract infection, fever, and flank pain(DOI: 10.1159/000065563). Present-day studies also commonly use this time point (within four weeks). Considering this, we revisited the studies included in our analysis and excluded the trial by Abdelaal. Although their average follow-up time was 2.8 weeks, the follow-up duration ranged from 1.5 to 8 weeks.

As you mentioned, compared to standard uncomplicated ureteroscopy, intracavitary incision surgery results in significantly greater damage to the ureteral tissue. Therefore, we further refined our inclusion criteria to include patients who underwent ureteral stent placement after ESWL, URSL, or PCNL. Upon reevaluation, we excluded the trial by Abdelaal and conducted a new network meta-analysis. Based on the updated meta-analysis, all figures and tables have been revised. Additionally,

modifications in the manuscript are highlighted in yellow.

Changes in the text:

In the inclusion and exclusion criteria section:

Randomized controlled trials that met all of the following criteria were included: (1) Drug treatments for ureteral stent-related symptoms should contain mirabegron, tamsulosin, and solifenacin. (2) Patients underwent ureteral stent placement after extracorporeal shock wave lithotripsy, ureteroscopic lithotripsy, or percutaneous nephrolithotomy.(3) Ureteral stent symptom questionnaire (USSQ) was used to evaluate all outcomes before removal of stents. (4) Crossover trials, dose titration studies, daily dosing studies, and studies for which full text was not available were excluded.

In the conclusion section:

In conclusion, our network meta-analysis reveals that compared with traditional drugs for relieving SRSs, mirabegron performs best in terms of alleviating body pain, sexual matters, and adverse events, with little difference in urinary symptoms and general health. Further high-quality prospective double-blinded RCTs are required to provide sufficient evidence supporting our observations.

Reviewer B

Comment 1:

I would like to congratulate the authors for their work in conducting an excellent meta-analysis of stent related symptom relief from pharmacotherapy with a focus only on studies that utilized a validated USSQ Questionnaire. I do have a few comments however that mainly relate to the specific nuances of the USSQ that need to be considered when performing such a meta-analysis:

1) The authors use the phrase "Compared with the blank control" often to determine the relief the medications provide for ureteral stent symptoms. What do the authors mean by this? Do they mean compared to an RCT patient in the placebo arm of the study? Or do they mean compared to the patient's baseline status? It is important as the USSQ attempts to noramlize an individual's USSQ score by having the patient fill in the questionnaire 1 week post stent removal to essentially determine an individual's "baseline". This language is somewhat confusing and needs clarification.

Reply 1: The authors thank you for your kind comments on our work. We are sorry for

the trouble caused by this language confusion. In the manuscript, what we want to express is that compared with the patients in the placebo arm of the study. In this revision, we have revised this point to avoid confusion. In addition, we have revised all the figures.

Changes in the text:

In the results section:

When placebo was used as the comparative reference, three treatments, mirabegron[MD,−1.36; 95%CI (−2.36,−0.37)], tamsulosin [MD,−1.12; 95%CI (−1.81,−0.43)] and solifenacin [MD,−1.35;95%CI (−2.01, −0.69)] were associated with a statistically significant improvement.

When placebo was used as the comparative reference, tamsulosin [MD, −0.41; 95%CI (−0.82, −0.01)] and solifenacin [MD, −0.45; 95%CI (−0.85, −0.04)] had shown remarkable efficacy in relieving body pain.

Compared to the placebo, mirabegron [MD, −0.50; 95%CI (−0.95, −0.04)] and solifenacin [MD, −0.48; 95%CI (−0.82, −0.15)] showed a significant advantage in terms of general health.

Comment 2:

2) How did the authors integrate their quality assessment in regards to the size of patient populations? I would note that the USSQ often requires at least 64 patients per arm of the study to adequately compare two different interventions (i.e., 128 in total). This is in order to detect a difference of 15%, 30% and 25% in the mean index scores for the urinary symptom, pain and general health domains of the USSQ respectively with 80% power. I would note some of the studies listed had less than 64 patients per arm of the study. Indeed, it's interesting that some studies compared up to 3 different interventions and only had around 40 patients (e.g., El-Nahas 2015 in Supp Table).

Can the authors provide more detail regarding how they performed their meta-analysis in regards to the possibility that some studies in this meta-analysis were underpowered and thus potentially not able to be considered a "valid" use of the USSQ?

Reply 2: The author thanks you for your rigorous scientific attitude. As you said, USSQ often requires at least 64 patients per arm of the study to adequately compare two different interventions, and some included original research may not be regarded as "effective" use of USSQ. However, researchers in this field are not all as rigorous as you. For example, in a recent study, the researcher proposed that the sample size of 100 patients (50 per group) was sufficient for 80% power to detect a 20% difference in each USSQ domain score(DOI: 10.1089/end.2023.0189). In the study, the number of participants in each group is calculated to be 21(DOI：10.1007/s00345-015-1544-1).

In addition, in a recent meta-analysis published in *European Urology Focus*(doi:10.1016/j.euf.2021.10.002),This phenomenon still exists(40 people in one arm).We didn't find bias through various methods (such as Cochrane risk-of-bias 2 tool, inconsistency detection and the comparison-adjusted funnel plot), But in the current meta-analysis methodology, there is no method to evaluate the impact of patient population size. Due to the limited research on mirabegron in the treatment of stent-related symptoms, with more focus on overactive bladder syndrome, there are not many studies that meet the "effective" use of USSQ for our research. Therefore, in this revision,we emphasize in the limitations section the potential bias caused by insufficient sample size in the original study affecting the "effective" use of USSQ, and strongly encourage further high-quality randomized controlled trials to improve the quality of meta-analysis on the treatment of ureteral stent-related symptoms.

Changes in the text: Fourth, although all the included studies used USSQ, the sample size of per arm in some studies were not enough (at least 64 patients) to detect the differences in the mean index scores for each domain of the USSQ with 80% power, which may lead to potential bias. We strongly encourage further high-quality randomized controlled trials to improve the quality of meta-analysis on the treatment of SRSs.

Comment 3:

3) How did the authors normalize their meta-analysis in regards to the types of stents used across different studies, hospitals and time periods? Did the authors only limit their study to Double-J stents with specific designs? I would appreciate more detail here.

Reply 3: The authors thank you for your kind comments. Our research does not impose specific restrictions on ureteral stents. Polyurethane double J stents are the most widely used type of ureteral stent in the clinic, and they are also used in our included studies. However, for special types of stents (such as metallic ureteral stents), whether the use of mirabegron could improve stent-related symptoms has not been reported.

Previous studies have shown that the size and length of ureteral stent could also affect symptoms(For example, DOI: 10.1111/luts.12259 and DOI：10.5173/ceju.2017.1533).

In the real scenario in everyday clinical practice, clinicians frequently provide personalized treatment based on each patient's condition. Therefore, it is difficult to achieve "standardization" in clinical practice. This was also largely not achieved in previous meta-analyses(DOI:10.1016/j.euf.2021.10.002). Although our research has low heterogeneity and high credibility, this does indeed lead to potential biases. In this revision, we have made revisions to the limitations section, emphasizing potential biases that may exist.

Changes in the text: Second, different stent sizes and lengths, different outcome

assessment times, and different use of analgesics may lead to potential biases. Nevertheless, this heterogeneity could better represent the real scenario in every day clinical practice.

Reviewer C

Comment 1:
The researchers have compiled the data well.
We look forward to good research in the future.
Reply 1: The authors thank you for your kind comments on our work.