PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (http://bmjopen.bmj.com/site/about/resources/checklist.pdf) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

<table>
<thead>
<tr>
<th>TITLE (PROVISIONAL)</th>
<th>Need to clamp indwelling urinary catheters before removal after different durations: A systematic review and meta-analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>AUTHORS</td>
<td>Ma, Sumin; Gu, Jiayi; Fan, Xiaoyan</td>
</tr>
</tbody>
</table>

VERSION 1 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Medina-Polo, Jose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Univ Hosp 12 Octubre, Urology</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>05-May-2022</td>
</tr>
<tr>
<td>GENERAL COMMENTS</td>
<td>The manuscript is valuable research. Clamping bladder catheters is a routine step before removal in many cases. The revision of the topic is necessary. The correction is adequately conducted. However, some points may be reviewed. In the results section of the abstract, the author should state the values for each alternative. Clamping or not clamping is not included in the text; just the data are mentioned. The introduction is a description of state of the art. The study is well designed. The authors stated that there is no language restriction. However, it seems that only English and Chinese articles were included. The results, Discussion and tables are correct.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Çulha, Yeliz</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Istanbul University-Cerrhopasa, Fundamentals of Nursing</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>06-Jun-2022</td>
</tr>
<tr>
<td>GENERAL COMMENTS</td>
<td>In the literature, there are studies on clamping before the catheter is removed, but the results of the studies differ on how long it should be clamped and what the effects are. I think the study will contribute to the literature. I congratulate you on addressing a clear gap in current knowledge about clamping indwelling urinary catheters before removal. However, I have few my suggestion: 1. PRISMA used when reviewing the studies should be added in method section. 2. PubMed search string can be added as a table with terms (combination of terms and how many publications were found) 3. I couldn't understand why gender was considered in data extraction. What was your purpose here? Or how do you think it will make a difference? There are already sample differences in the studies. Some studies have been done with prostate patients. 4. 10.page at line 36, in the conclusion section, I do not think that the</td>
</tr>
</tbody>
</table>
In addition, the use of free drainage before urinary catheter removal during clinical practice somewhat reduced nurse workload and difficulty is appropriate.

**REVIEWER**

Hajebrahimi, Sakineh  
Iranian Evidence Based Medicine Centre of Excellence, Tabriz University of Medical Sciences, Tabriz, Iran.

**REVIEW RETURNED**

03-Oct-2022

**GENERAL COMMENTS**

Introduction:  
The authors should describe the rationale for the review in the context of existing knowledge. Considering the published papers previously on this topic, why did the authors not consider an umbrella review to conclude the accurate decision on the need to clamp indwelling urinary catheters before removal after different durations.  
As the authors mentioned in the introduction “from 2016 to 2021, several systematic reviews analysed the need to clamp urinary catheters in patients with short-term indwelling urinary catheters (duration ≤ 14 days, with the majority of the study population being postoperative general surgery patients with an indwelling duration ≤ 7 days), but the conclusions were inconsistent”, why they did not perform an umbrella review in this regard?

Methods:  
Present the full search strategies for some databases.  
Why did you consider time limitations for your search?  
It is suggested to define the primary and secondary outcomes clearly.  
How did you evaluate the publication bias? Most of the included studies were in Chinese, regarding databases selection of this country.  
Why did not the authors perform a comprehensive search? Searching unpublished literatures as well as gray literature is missed.

Results:  
Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.  
In outcomes assessment, what was the effect size measurement? Please define.

Discussion:  
Discuss implications of the results for practice, policy, and future research.  
Conclusion:  
In this section, the authors should not provide a citation. Please revise it.

**REVIEWER**

Güru nlü Alma, Özlem  
Mugla Sitki Kocman Universitesi Istatistik Bolumu, Fen Fak

**REVIEW RETURNED**

10-Oct-2022

**GENERAL COMMENTS**

No further comments.

**REVIEWER**

Daignault-Newton, Stephanie  
University of Michigan, Biostatistics

**REVIEW RETURNED**

28-Oct-2022
GENERAL COMMENTS

This study conducted a meta-analysis of randomized controlled trials and quasi-experimental studies to estimate the effect of bladder training by clamping among patients with indwelling urinary catheter. Multiple outcomes were examined including time to first voiding.

Methods:

Statistical Analysis: please list variables that were analyzed with the statistical terms. Examples of abbreviations include IV, M-H, and the statistical terms.

Page 5, Lines 54-56 in the methods: an ordered outcome that required a proportional odds model is described. Please clarify which outcome this is, what method was used, and the impact on the results. It would be preferred that other software be used that has capabilities to do the analysis correctly.

Results:

Page 7, Lines 38-41: The first statement does not make sense. 13 studies report the outcome and 2 do not. This only adds up to 15 and there are 17 studies. Additionally, this outcome only included 11 studies in the analysis. The references do not add up either. Please clarify and correct.

Page 8, Lines 5-17: The reporting of this outcome is not ordered well. The first statement reports the statistically significant combined effect size p-value but the effect size and the confidence interval are not provided until the last sentence of the paragraph. Please move the last sentence associated with the first sentence. Then follow with the subgroups that are statistically different and the results of each.

Page 8, Lines 28-48: The 3 studies that were analyzed descriptively (lines 32-42) provide p-values for each study but do not provide the effect sizes associated with the clamped and free drainage groups. Please provide the clinical result, not just p-values and z-scores. Please edit your results and when reporting the outcome, provide the effect size with the corresponding 95% confidence interval and the associated p-value. Please leave the z-score out of the written results.

Page 10, Lines 34-36 states a reduced nurse workload in the conclusion. This is the first statement about this in the manuscript. If this is a conclusion, then it needs to be discussed prior to the conclusion in the manuscript and supporting evidence.

REFERENCE

Yates, Jonathan
University of Pittsburgh, Medicine

REVIEW RETURNED

31-Oct-2022

This manuscript is appropriate. The main methods and results of this manuscript are appropriate. The figures report the results well.

The following need to be addressed:

1. In the figures, all abbreviations included need to be defined.

2. Methods:
   - Data extraction: please list variables that were analyzed with the statistical terms.
   - Statistical Analysis: please list variables that were analyzed with the statistical terms.
   - Examples of abbreviations include IV, M-H, and the statistical terms.

The meta-analysis results need to include the effect size and 95% confidence interval in addition to the p-value on line 45. Conclusion: Statement on page 10 lines 34-36 states a reduced nurse workload in the conclusion. This is a conclusion, then it needs to be discussed prior to the conclusion in the manuscript and supporting evidence.

The main methods and results of this manuscript are appropriate.
urinary retention, UTI, re-catheterization, and residual urine volume. For the most part, the manuscript was written clearly and adhered to PRISMA reporting guidelines, but there are some issues in the methods and presentation of results. More specifically:

- In the abstract conclusion, the statement that “bladder training by clamping did not reduce” any of the outcomes is not entirely accurate since the results show that clamping is not significantly different from free drainage in most outcomes but it did increase incidence of UTI. The last sentence captures the effect on UTI, but stated in favor of free drainage, the comparison group which was not clearly identified as the comparator in the abstract.
- This study performed stratified analyses by catheter duration using a cutpoint of 7 days. This should be stated in the methods section instead of the introduction. In addition, the choice of the cutpoint is vaguely attributed to the literature. The rationale for using 7 days versus others is not entirely clear.
- A descriptive analysis approach was chosen if the effect sizes cannot be combined. However, what specific criteria was used to determine which effect sizes cannot be combined is unknown.
- The choice of modeling approach (fixed vs random effects) relied on the test for heterogeneity, an approach that is not advocated in Cochrane handbook (section 10.10.4.1).
- It is curious why no pre-planned sensitivity analyses were conducted. Were the authors 100% confident that the assumptions they used were correct?
- There was no pre-specified plan to handle missing data or information. Are there no missing data across the studies?
- Other than for catheter duration, there was also no pre-specified plan to explore and identify sources of heterogeneity.
- The “other means” used to obtain 12 studies were not clearly described.
- Please clarify the unit of the effect estimates for time to first voiding (hours, days, etc) in Figure 3.
- In sensitivity analyses, 3 studies were excluded due to “methodological differences”. What these differences are were not clearly described.
- There were very few studies for the residual volume outcome (Figure 7), which raises concern about whether a meta-analytic pooling of effects is appropriate for this outcome.
- For UTI among those with duration <=7 days, it is curious why the p-value (=0.13, not significant) provides inconsistent evidence with the confidence interval that excludes 1 (1.42 to 2.02, significant).

**VERSION 1 – AUTHOR RESPONSE**

**Reviewer 1:**

**Comment 1.** In the results section of the abstract, the author should state the values for each alternative. Clamping or not clamping is not included in the text; just the data are mentioned.

**Reply 1:** Thank you for your comments and suggestions! We have supplemented the related values for each alternative, such as effect size, P-value, 95%CI and I². We have clarified which is the clamping group or not clamp group in the result.

**Change in revised manuscript:** Page 2, Line 17-26 in RED
Comment 2. The introduction is a description of the state of the art.

Reply 2: Thank you for your kind reminder. In the Introduction, we focus on the characteristics and features of different technologies, and further improve the introduction. We hope to meet your expectations.

Change in revised manuscript: Page 3, Line 2-6, Page 3, Line 24-30, Page 3, Line42-44 and Page 4, Line 1-7 in RED.

Comment 3. The study is well designed. The authors stated that there is no language restriction. However, it seems that only English and Chinese articles were included.

Reply 3: Thanks for your praise. This question was our state error. We just included the English and Chinese articles and did the revision.

Change in revised manuscript: Page 2, Line 11 in RED.

We really appreciate your positive and insightful suggestions on our manuscript. We are looking forward to your further suggestions.

Reviewer 2:

Comment 1. PRISMA used when reviewing the studies should be added in method section.

Reply 1: Thank you for your comments and suggestions! We have added PRISMA in the method section.

Change in revised manuscript: Page 4, Line 10-11 in RED

Comment 2. PubMed search string can be added as a table with terms (combination of terms and how many publications were found)

Reply 2: Thank you for your suggestions! We have added the search strategy of PubMed database in Supplemental Material Table S1.

Comment 3. I couldn't understand why gender was considered in data extraction. What was your purpose here? Or how do you think it will make a difference? There are already sample differences in the studies. Some studies have been done with prostate patients.

Reply 3: Thank you for your comments and suggestions. Because we know that bladder volume and function have a difference in gender from many literatures. We considered gender as an influencing factor. So, we extracted gender information in order to further analysis in the part of the discussion and found that the male- and -female rate of gender was 3:2 in this systematic review, which seemed to be a little different. As for the outcome of residual urinary volume, all of three articles included in the meta-analysis were male patients with prostate disease, and there was no statistical significance between the clamping group and free drainage. So, it is worth investigating further.
whether or not to clamp in men with pre-existing prostate dysfunction. And, we hope further studies recognize and explain what is the effect of clamping indwelling urethral catheters in different gender.

**Change in revised manuscript:** Page 11, Line 22-29 and Page 11, Line 47-50 in RED.

**Comment 4:** 10-page at line 36, in the conclusion section, I do not think that the sentence "In addition, the use of free drainage before urinary catheter removal during clinical practice somewhat reduced nurse workload and difficulty" is appropriate.

**Reply 4:** Thanks for your advice. We agree with your opinions. It’s uncertain about it when writing the manuscript. We have deleted it.

We really appreciate your positive and insightful suggestions on our manuscript. We are looking forward to your further suggestions.

**Reviewer 3:**

**Comment 1.** Introduction: The authors should describe the rationale for the review in the context of existing knowledge.

**Reply 1:** Thanks to the reviewer to point out this problem. We reorganized the logic to describe the rationale of the review. We described the existing dispute on whether or not to clamp urinary catheters in different durations in detail and the current state of indwelling urinary catheter care. These details are elaborated in the Introduction section.

**Change in revised manuscript:** Page 3, Line 2-5, Page 3, Line 23-29, Page 3, Line 40-44 and Page 4, Line 1-4 in RED.

**Comment 2.** Considering the published papers previously on this topic, why did the authors not consider an umbrella review to conclude the accurate decision on the need to clamp indwelling urinary catheters before removal after different durations. As the authors mentioned in the introduction “from 2016 to 2021, several systematic reviews analysed the need to clamp urinary catheters in patients with short-term indwelling urinary catheters (duration ≤ 14 days, with the majority of the study population being postoperative general surgery patients with an indwelling duration ≤ 7 days), but the conclusions were inconsistent”, why they did not perform an umbrella review in this regard?

**Reply 2:** Thank you for your comments and suggestions. There are many reasons for not conducting an umbrella review. First, the quality of many systematic reviews was low which may influence the quality of the umbrella review. Then, there are different definitions of short and long-term indwelling urethral catheters in different systematic reviews and very hard to conduct a subgroup analysis. Some meta-analyses mixed all duration of indwelling urethral catheters to discuss the effect of clamping. Additionally, some meta-analyses included literature from different time points. Due to the above reasons, we feel more firmly that umbrella evaluation cannot explain these issues clearly.
Therefore, we make this meta-analysis from a smaller entry point. Our aim is to discuss the different effects of indwelling urinary catheter duration <= 7 days and > 7 days.

Comment 3. Methods: Present the full search strategies for some databases.
Reply 3: Thanks for your reminder. We have supplemented the search strategies for PubMed, Cochrane Library, CNKI, and CINAHL in detail.
Change in revised manuscript: See Supplemental Material Table S1.

Comment 4. Why did you consider time limitations for your search?
Reply 4: Thank you for your comments. Because many previous studies such as Oberst's research in 1981, Williamson's research in 1982, and Bergman's research in 1987 have been included and analyzed in other studies, and indeed so far away from where we are now. According to the argument of the article “An indwelling urinary catheter for the 21st century”, catheters in the 21st century may well be more expensive but they can improve the care of millions of patients and reduce the enormous costs of managing the complications associated with the indwelling Foley catheter. So, we just think included the most recent studies with the year 2000 as the boundary.

Comment 5. It is suggested to define the primary and secondary outcomes clearly.
Reply 5: Thank you for your suggestions. We have supplemented the definition of the primary and secondary outcomes clearly. The primary outcome was the incidence of UTI, and the secondary outcomes included time to first voiding, the incidence of urinary retention, the incidence of re-catheterization, and residual urine volume. The UTI was defined as bacteriuria accompanying fever, frequent or painful urination, and burning sensation during urination without other foci of infection and measured by evaluation of subjective symptoms or laboratory, including pain, discomfort, and burning upon micturition.
Change in revised manuscript: Page 4, Line 35-39 in RED

Comment 6. How did you evaluate the publication bias? Most of the included studies were in Chinese, regarding databases selection of this country.
Reply 6: Thank you for your comments and suggestions. Although we only searched Chinese and English databases, and many of the included literatures were in China, our search strategy was detailed and relevant grey literatures were included, further reducing publication bias. In addition, we verified the results by funnel plot and publication bias test, and found no evidence of publication bias. As shown below, our funnel plot is basically symmetric.
Figure 1 The funnel plot of urinary retention  Figure 2 The funnel plot of the incidence of UTI

Figure 3 The funnel plot of the time to first void  Figure 4 The funnel plot of the residual urine volume

Figure 5 The funnel plot of the incidence of re-catheterization

Change in revised manuscript:  Page 10, Line 38-45 in RED

Comment 7. Why did not the authors perform a comprehensive search? Searching unpublished literature as well as grey literature is missing.

Reply 7: According to your suggestion, we re-completed the search strategy and included the grey literature in this search such as conference abstracts/proceedings, government information, reports and dissertations. Detailed search strategies are in Supplemental Material Table S1.

Change in revised manuscript:  Page 2, Line 39 in RED

Comment 8. Results: Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.

Reply 8: Thank you for your comments and suggestions. We have supplemented the certainty in the body of evidence for each outcome. The GRADE profile software was used to assess the quality of evidence for each outcome, which criteria comprised the risk of bias, inconsistency, indirectness, inaccuracy, and publication bias. The quality of evidence was classified as high, moderate, low, or
very low. Finally, the quality of evidence was low for Urinary retention, the incidence of urinary tract infection, time to first void, and residual urine volume. The quality of evidence was moderate for the incidence of re-catheterization. The detailed reasons were available in the footnotes.

**Comment 9.** In outcomes assessment, what was the effect size measurement? Please define.

**Reply 9:** Thank you for your comments. We have defined the effect size of outcomes measurement. All continuous variables were pooled using mean difference (MD) and standardized mean standardized (SMD) with 95%CI. For dichotomous outcomes, we pooled the numbers of outcomes to calculate a risk ratio (RR) with 95%CI. A descriptive analysis was used for ordered outcomes data (residual urine volume). The I² statistic was used to assess the statistical heterogeneity. The heterogeneity might not be important when 0% < I² < 40%; the heterogeneity may be moderate when 30% < I² < 60%; the studies were considered substantial heterogeneity when 50% < I² < 90%; the studies were viewed considerable heterogeneity when 75% < I² < 100%.

**Change in revised manuscript:** Page 5, Line 18-30 in RED.

**Comment 10.** Discussion: Discuss implications of the results for practice, policy, and future research.

**Reply 10:** Thank you for your comments and suggestions. We rewrote the implication part from the aspect of practice, policy and future research.

**Change in revised manuscript:** Page 12, Line 39-46 in RED.

**Comment 11.** Conclusion: In this section, the authors should not provide a citation. Please revise it.

**Reply 11:** Thank you for your reminder! We have deleted the citation.

**Change in revised manuscript:** Page 12, Line 30-36

Thank you for your thoughtful and constructive comments. We have further modified and improved the manuscript according to the comments, which we believe has greatly improved the quality of our manuscript. Thank you again. If we missed anything, please be sure to let us know. We look forward to your further advice and guidance again.

**Reviewer 5:**

**Comment 1.** In the figures, all abbreviations included need to be defined. Examples of abbreviations include: IV, M-H, and the statistical terms.

**Reply 1:** Thank you for your comments and suggestions! We have added the definition of abbreviations to forest plots including IV, M-H and CI.

**Comment 2.** Methods: Data extraction: please list the outcomes extracted.

**Reply 2:** Thank you for your suggestions! Data extraction was carried out by two authors independently using a pre-designed data extraction sheet in Microsoft Excel. The extracted data consisted of study characteristics (name of the first author, publication year and country), patient
characteristics (gender, sample size and type of disease), intervention characteristics (duration of indwelling urinary catheter and removal time) and outcomes indicators.

**Change in revised manuscript:** Page 4, Line 42-44 and Page 5, Line 1-5.

**Comment 3.** Statistical Analysis: please list variables that were analyzed with the statistic used.

**Reply 3:** Thank you for your reminder! We have listed the statistic used. All continuous variables were pooled using mean difference (MD) and standardized mean standardized (SMD) with 95%CI. For dichotomous outcomes, we pooled the number of cases to calculate a risk ratio (RR) with 95%CI. A descriptive analysis was used for ordered outcomes data (residual urine volume) according to the recommendations of the Cochrane Handbook for Systematic Reviews of Interventions.

**Change in revised manuscript:** Page 5, 18-22.

**Comment 4.** Page 5, Lines 54 - 56 in the methods: an ordered outcome that required a proportional odds model is described. Please clarify which outcome this is, what method was used, and the impact on the results. It would be preferred that other software be used that has capabilities to do the analysis correctly.

**Reply 4:** Thank you for your comments and suggestions! Two studies (Gong and Liu) reported residual urine volume by the ordered outcome which divided symptoms into four objective levels respectively. Gong’s study classified residual urine volume into four objective levels (0-50ml, 50-100ml, 100-200ml, >200ml). Liu’s study classified the level according to the patient’s subjective feelings. According to the recommendations of Cochrane, the short-ordered data should be merged into adjacent categories and converted into dichotomous data. However, we thought that we were unable to define the criteria of residual urine volume. So, we adopted the descriptive analysis method to deal with it. In addition, one study (Chen SZ) measured residual urine volume after 24 hours of removing catheters, which had clinical heterogeneity compared with the other three studies. Then, the other three studies were pooled for analysis which used continuous variables to report residual urine volume and measured the residual urine volume after removing indwelling urinary catheters immediately.

**Change in revised manuscript:** Page 10, Line 10-23.

**Comment 5.** All of the statistics reported using abbreviations need to have the statistic defined and what the statistic is assessing or testing. An example is I^2.

**Reply 5:** Thank you for your suggestions. We have defined all reported statistical information including I^2, SMD, MD and 95%CIs.

**Change in revised manuscript:** Page 5, Line 22-30 in RED.

**Comment 6.** Results: Time to first void paragraph (Page 7 Lines 38-41): The first statement does not make sense. 13 studies report the outcome and 2 do not. This only adds up to 15 and there are 17
studies. Additionally, this outcome only included 11 studies in the analysis. The references do not add up either. Please clarify and correct.

Reply 6: Thank you for your comments and suggestions. We reclarified it. Thirteen studies reported this outcome, two of them reported it using median and quartile, and the value cannot be converted to mean and standard deviation. Maybe it doesn’t conform to the normal distribution. At last, we excluded it and included 11 studies in the analysis.

Change in revised manuscript: Page 9, Line 27-30 in RED

Comment 7. The statistics reported from each meta-analysis model needs to provide the clinical result, not just p-values and z-scores. Please edit your results and when reporting the outcome, provide the effect size with the corresponding 95% confidence interval and the associated p-value. Please leave the z-score out of the written results.

Reply 7: Thanks for your advice. We edited the result of the meta-analysis and added all outcomes’ effect size with 95%CIs and the associated p-value.

Change in revised manuscript: Page 9, Line 12-54 and Page 10, Line 1-23 in RED.

Comment 8. Incidence of urinary tract infections (Page 8, lines 5-17): The reporting of this outcome is not ordered well. The first statement reports the statistically significant combined effect size p-value but the effect size and the confidence interval are not provided until the last sentence of the paragraph. Please move the last sentence associate with the first sentence. Then follow with the subgroups that are statistically different and the results of each.

Reply 8: Thank you for your comments and suggestions. We have adjusted the sequence in this paragraph. We described the pooled heterogeneity first. Then analyzed the pooled results with associated effect size and confidence interval. Then, we discussed the result of the subgroup analysis.

Change in revised manuscript: Page 9, Line 12-23 in RED.

Comment 9. The 3 studies that were analyzed descriptively (lines 32 - 42) provide p-values for each study but do not provide the effect sizes associated which would provide meaning to the outcome by treatment group. Please add the effect sizes and confidence intervals between the clamped and free drainage groups to this paragraph.

Reply 9: Thank you for your comments and suggestions. We added the effect size of Liu’s study. However, we didn’t find the correlated effect sizes in the Chen and Gong studies which just reported the P-value and the number of cases.

Change in revised manuscript: Page 10, Line 10-18 in RED.

Comment 10. The meta-analysis results need to include the effect size and 95% confidence interval in addition to the p-value on line 45.
Reply 10: Thank you for your reminder. We have added the effect size and 95% confidence interval in addition to the p-value. The pooled result showed that, clamping the urinary catheter had no significant effect on improving residual urine volume compared with free drainage (MD = -0.36, 95% CI -4.17 to 3.44, P=0.85).

Change in revised manuscript: Page 10, Line 20-22 in RED.

Comment 11. Conclusion: Statement on page 10 lines 34 - 36 states a reduced nurse workload in the conclusion. This is the first statement about this in the manuscript. If this is a conclusion, then it needs to be discussed prior to the conclusion in the manuscript and supporting evidence.

Reply 11: Thank you for your insightful comments. We have deleted it. This statement was an opinion derived from Zou's study. We also consider this statement as an inappropriate conclusion. Thank you again for your reminding.

Change in revised manuscript: Page 12, Line 31-37.

Thank you for your thoughtful and constructive comments. We have further modified and improved the manuscript according to the comments, which we believe has greatly improved the quality of our manuscript. Thank you again. If we missed anything, please be sure to let us know. We look forward to your further advice and guidance again.

Reviewer 6:

Comment 1. In the abstract conclusion, the statement that "bladder training by clamping did not reduce" any of the outcomes is not entirely accurate since the results show that clamping is not significantly different from free drainage in most outcomes but it did increase incidence of UTI. The last sentence captures the effect on UTI, but stated in favor of free drainage, the comparison group which was not clearly identified as the comparator in the abstract.

Reply 1: Thank you for your comments and suggestions. This was our misrepresentation. Our original meaning was that bladder training by clamping the urinary catheter will not bring any beneficial outcome. We extremely agree with your opinions. And, we revised the part of the conclusion and stated the comparison clearly as follows. Bladder training by clamping indwelling urinary catheters may increase the incidence of urinary tract infection and lengthen the time to first void in patients with durations of indwelling urinary catheters ≤ 7 days. But, the effect of clamping training on durations of indwelling urinary catheters > 7 days was uncertain.

Change in revised manuscript: Page 2, Line 24-27 in RED.

Comment 2. This study performed stratified analyses by catheter duration using a cutpoint of 7 days. This should be stated in the methods section instead of the introduction. In addition, the choice of the cutpoint is vaguely attributed to the literature. The rationale for using 7 days versus others is not entirely clear.
Reply 2: Thank you for your suggestions. We have moved the stratified analyses to the method part. The reasons why we chose 7 days as the cutpoint were as follows. First, there are many controversies on whether clamping the urinary catheter between short and long-term indwelling urinary catheters. The definition regarding the duration of short-term and long-term indwelling urinary catheters is not uniform. Then, the average duration of patients with short indwelling urinary catheters was 7 days according to the literature and clinical practice, especially post-operator patients. So, we didn't adopt the short-term indwelling urinary catheters 14 days as the cutpoint and thought the cutpoint of 7 days possesses more clinical reference value.

Change in revised manuscript: Page 5, Line 33-40 in RED

Comment 3. A descriptive analysis approach was chosen if the effect sizes cannot be combined. However, what specific criteria was used to determine which effect sizes cannot be combined is unknown.

Reply 3: Thank you for your suggestions. This outcome was reported in two types: continuous data and order data. Three studies (Zheng Yong, Moon and Büyükyilmaz) which adopted continuous data to report and measured the residual volume after the first void were included in the meta-analysis. We excluded Chen’s study which measured residual urinal volume after removing the urinary catheter 24 hours, which had clinical heterogeneity compared with the other three studies. Liu’s study divided residual urine volume into four levels according to the patient’s subjective feelings. Gong’s study classified the residual urine volume into four objective levels (0-50ml, 50-100ml, 100-200ml, >200ml). So, it’s very hard to convert to the same dichotomous variable. Thus, we reported it descriptively.

Change in revised manuscript: Page 10, Line 10-18 in RED.

Comment 4. The choice of modeling approach (fixed vs random effects) relied on the test for heterogeneity, an approach that is not advocated in Cochrane handbook (section 10.10.4.1).

Reply 4: Thank you for your suggestions. We extremely agree with your advice. In the process of statistical analysis, we recognized the heterogeneity from many aspects, such as clinical heterogeneity, methodological heterogeneity and statistical heterogeneity. In addition, we conducted a sensitivity analysis to find the heterogeneity and found the rationale explanation.

Change in revised manuscript: Page 5, Line 26-30 in RED.

Comment 5. It is curious why no pre-planned sensitivity analyses were conducted. Were the authors 100% confident that the assumptions they used were correct?

Reply 5: Thanks for your advice. We supplemented the plan of sensitivity analysis and conducted the sensitivity analysis for the outcome of time to first urine. The heterogeneity dropped from 64% to 0% after excluding some studies considering the result of the sensitivity analysis and methodological heterogeneity. The pooled result of other outcomes showed that there was no statistical heterogeneity.
Change in revised manuscript:  Page 5, Line 26-30 and Page 9, Line 32-38 in RED.

Comment 6. There was no pre-specified plan to handle missing data or information. Are there no missing data across the studies?
Reply 6: Thank you for your comments and suggestions. Our pre-specified plan was to contact the authors to request the data if there were data missing. In trials reporting mean values without standard deviations (SDs) but with P values or 95% CI, we performed data conversion by using Excel document.
Change in revised manuscript:  Page 5, Line 2-4 in RED.

Comment 7. Other than for catheter duration, there was also no pre-specified plan to explore and identify sources of heterogeneity.
Reply 7: Thank you for your reminder! In the process of statistical analysis, we identified the heterogeneity mainly through I² Statistics, sensitivity analyses and the clinical perspective. The method to identify sources of heterogeneity has been added to the part of statistical analysis.
Change in revised manuscript:  Page 5, Line 23-29 in RED.

Comment 8. The "other means" used to obtain 12 studies were not clearly described.
Reply 8: Thank you for your reminder. We obtained the 12 studies through the reference of retrieval studies.

Comment 9. Please clarify the unit of the effect estimates for time to first voiding (hours, days, etc) in Figure 3.
Reply 9: Thank you for your comments and suggestions! The unit of the time to first voiding is the hour. We have added it under figure 4a, figure 4b and figure 4c. Figure 3 is moved to Figure 4 because of the reordering.
Change in revised manuscript:  See Figure File.

Comment 10. In sensitivity analyses, 3 studies were excluded due to “methodological differences”. What these differences are were not clearly described.
Reply 10: Thank you for your reminder! Because 3 studies were excluded due to the higher bias in random sequence generation. The other studies stated the random sequence generation clearly.
Change in revised manuscript:  Page 9, Line 27-29 in RED.

Comment 11. There were very few studies for the residual volume outcome (Figure 7), which raises concern about whether a meta-analytic pooling of effects is appropriate for this outcome.
Reply 11: We are so grateful for your kind question. The testing of model assumptions can be difficult due to the limited number of studies. The advanced statistic methods (Hartung-Knapp-Sidik-Jonkman)
can only be suitable for providing statistical inferences in the circumstance of high accuracy effect sizes. So, this method may not suitable for this outcome. But the heterogeneity was low, and the risk of the conclusion may be low.


Comment 12. For UTI among those with duration <=7 days, it is curious why the p-value (=0.13, not significant) provides inconsistent evidence with the confidence interval that excludes 1 (1.42 to 2.02, significant).

Reply 12: Thank you for your comments. We thought that the p-value of 0.13 means the heterogeneity of the first group. The p< 0.00001 is the effect size.

Change in revised manuscript: See Figure File Figure 3.

We really appreciate your positive and insightful suggestions on our manuscript. We are looking forward to your further suggestions.

VERSION 2 – REVIEW

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Hajebrahimi , Sakineh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Iranian Evidence Based Medicine Centre of Excellence, Tabriz University of Medical Sciences, Tabriz, Iran.</td>
<td></td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>17-Dec-2022</td>
</tr>
<tr>
<td>GENERAL COMMENTS</td>
<td>All of my comments have responded satisfactorily.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Daignault-Newton, Stephanie</th>
</tr>
</thead>
<tbody>
<tr>
<td>University of Michigan, Biostatistics</td>
<td></td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>30-Dec-2022</td>
</tr>
<tr>
<td>GENERAL COMMENTS</td>
<td>The authors have improved the manuscript with the response to the reviews and the edits that have been made. The following concerns now exist based upon the revised version of the manuscript. 1. The methods section mis-names the standardized mean difference (SMD) in the method's statistical analysis section (page 5, line 19) and calls it &quot;standardized mean standardized&quot;. Please correct this. 2. Method's statistical analysis section: The definitions of heterogeneity for cutpoints of I² overlap. Please describe what is used for your study specifically and use the reference to justify it. 3. Methods, statistical analysis section: The authors state what qualified for a fixed effects model. &quot;Studies with I² ≤ 60% or P &gt; 0.1 were categorized as mildly heterogeneous and analyzed the use of a fixed-effects model.&quot; What analysis was used if the heterogeneity did not meet this criterion? Please add to the methods if all endpoints were analyzed with the fixed-effects model because they were mildly heterogeneous or if there were some that were not, what method of analysis was used. 4. A sensitivity analysis is mentioned but the statistical methods that...</td>
</tr>
</tbody>
</table>
were used and the decisions made based upon the results need to be described. Additionally, this sensitivity analysis was performed on only 1 endpoint, and it was a secondary endpoint. Why was the sensitivity analysis not performed on the primary endpoint? Why was this secondary endpoint chosen?

5. The sensitivity analysis figure (Figure 5) is poorly done. The x-axis should include a value of '0' which indicates the value where no difference is found for the endpoint so the reader can determine which and how many analyses result in significant or non-significant findings from the sensitivity analysis. The coloring with the black background is a poor choice.

6. The funnel plot was made for only 1 endpoint. It may be useful to provide the funnel plot for the secondary endpoint "time to first void".

7. Figure 6: There is not a sub-group for this analysis. Thus, the subtotal is not necessary in this figure as it displays the same information as the Total. Please remove the duplicate information from this figure.

8. Please edit the following sentence from the results section, residual volume subsection (Page 10, lines 14-16) to read that there is no difference between the clamped and free drainage groups. Since there is not statistical significance, the clamped group is NOT higher than the drainage group. Sentence to edit:

"Residual urine volume in patients with percutaneous nephrolithotomy was slightly higher in the clamped group than in the free drainage group, with no statistically significant difference ($t=1.370, P=0.087$)."

9. Certainty of evidence section needs to be written with complete sentences. The outline does not seem appropriate for the quality for this journal.

10. Conclusion: Please remove the word "may" from the first statement in the conclusion and provide the control arm description ("free drainage") at the end of the statement. Your results are statistically significant which mean that in your study you can conclude that clamping does increase the risk of UTI and lengthens the hours to first voiding compared to free drainage.

11. The use of the abbreviation "CI" is used in the abstract and the methods prior to defining. Please define it in the abstract and then use the abbreviation.

12. Consider renaming the endpoint "time to first void" as "hours to first void" so the time unit is clear throughout the manuscript.

REVIEWER
Yabes, Jonathan
University of Pittsburgh, Medicine

REVIEW RETURNED
31-Dec-2022

GENERAL COMMENTS
The manuscript may benefit from further proof reading to correct a few grammatical/typo errors. Nonetheless, the authors were very responsive to reviewer comments. As a result, this manuscript has substantially improved. My concerns from the previous submission were adequately addressed.

VERSION 2 – AUTHOR RESPONSE

Reviewer 5:
Comment 1. The methods section mis-names the standardized mean difference (SMD) in the method’s statistical analysis section (page 5, line 19) and calls it "standardized mean standardized". Please correct this.

Reply 1: Thank you for your careful suggestions. We have corrected the misname of the standardized mean difference.

Change in revised manuscript: Page 5, Line 16 in RED.

Comment 2. Method's statistical analysis section: The definitions of heterogeneity for cutpoints of I^2 overlap. Please describe what is used for your study specifically and use the reference to justify it.

Reply 2: Thank you for your comments and suggestions. We did note the overlap of I^2. And, we restated the specific method in dealing with the heterogeneity and added related reference. The definition of I^2 is a rough guide according to the Cochrane Handbook for Systematic Reviews of Interventions. If I^2 was ≤60% or P-value was >0.1, the study was categorized as moderate statistically heterogeneous, and a fixed effect model was used to analyse. Otherwise, the effect size was pooled using a random-effects model if heterogeneity could not be explained and I^2 was >60% or P-value was ≤0.1. Besides, a sensitivity analysis was conducted to explore the potential sources of heterogeneity and stability of the results using STATA software (version 17.0; Stata Corp, College Station, TX, USA).

Change in revised manuscript: Page 5, Line 23-28 in RED.

Comment 3. Methods, statistical analysis section: The authors state what qualified for a fixed effects model. "Studies with I^2 ≤ 60% or P > 0.1 were categorized as mildly heterogeneous and analyzed the use of a fixed-effects model." What analysis was used if the heterogeneity did not meet this criterion? Please add to the methods if all endpoints were analyzed with the fixed-effects model because they were mildly heterogeneous or if there were some that were not, what method of analysis was used.

Reply 3: Thanks for your advice. If some endpoints didn’t meet the criterion, we adopted the sensitivity analysis to explore the source of the heterogeneity. And, if the heterogeneity can’t be explained, we adopted the randomized effect model to analyze the data. We have added the related content to the manuscript.

Change in revised manuscript: Page 5, Line 24-31 in RED.

Comment 4. A sensitivity analysis is mentioned but the statistical methods that were used and the decisions made based upon the results need to be described. Additionally, this sensitivity analysis was performed on only 1 endpoint, and it was a secondary endpoint. Why was the sensitivity analysis not performed on the primary endpoint? Why was this secondary endpoint chosen?

Reply 4: Thank you for your insightful comments. We supplemented that the method of sensitivity analysis was leave-one-out using STATA software (version 17.0; Stata Corp, College Station, TX, USA). If the results of the meta-analysis are susceptible to significant alteration after removing
studies, it indicates that the results aren’t robust. According to the Cochrane Handbook for Systematic Reviews of Intervention, we know that sensitivity analysis is necessary according to the value of $I^2$ no matter whether it is a primary indicator or a secondary endpoint. Sensitivity analysis was conducted to explore the potential sources of heterogeneity and the stability of the results if $I^2$ was $>60\%$ or $P$ was $\leq 0.1$. Therefore, we performed the sensitivity analysis on the hours to first void which was detected substantial heterogeneity ($I^2 = 90\%$).

**Change in revised manuscript**: Page 5, Line 26-31 in RED.

**Comment 5.** The sensitivity analysis figure (Figure 5) is poorly done. The x-axis should include a value of '0' which indicates the value where no difference is found for the endpoint so the reader can determine which and how many analyses result in significant or non-significant findings from the sensitivity analysis. The coloring with the black background is a poor choice.

**Reply 5:** Thank you for your thoughtful and constructive comments. We have adjusted the background color of Figure 5 to white and added the value of '0' as the reference line to explicitly present significant of the results.

**Change in revised manuscript**: See Figure 5.

**Comment 6.** The funnel plot was made for only 1 endpoint. It may be useful to provide the funnel plot for the secondary endpoint "time to first void".

**Reply 6:** Thank you for pointing this out. We agree with your opinions. The funnel plot for the secondary endpoint "hours to first void" was provided in Figure 10.

**Comment 7.** Figure 6: There is not a sub-group for this analysis. Thus, the subtotal is not necessary in this figure as it displays the same information as the Total. Please remove the duplicate information from this figure.

**Reply 7:** Thank you for your suggestions and comments. We have deleted the duplicate information in Figure 6.

**Comment 8.** Please edit the following sentence from the results section, residual volume subsection (Page 10, lines 14-16) to read that there is no difference between the clamped and free drainage groups. Since there is not statistical significance, the clamped group is NOT higher than the drainage group.

"Residual urine volume in patients with percutaneous nephrolithotomy was slightly higher in the clamped group than in the free drainage group, with no statistically significant difference ($t=1.370, P=0.087$)."

**Reply 8:** Thank you for your suggestions. We have edited the sentence to express more clearly.

**Change in revised manuscript**: Page 10, Line 14-16 in RED.
Comment 9. Certainty of evidence section needs to be written with complete sentences. The outline does not seem appropriate for the quality for this journal.

Reply 9: Thank you for your kind comments. We have rewritten the certainty of evidence section with complete sentences and deleted the outline.

Change in revised manuscript: Page 10, Line 26-40.

Comment 10. Conclusion: Please remove the word "may" from the first statement in the conclusion and provide the control arm description ("free drainage") at the end of the statement. Your results are statistically significant which mean that in your study you can conclude that clamping does increase the risk of UTI and lengthens the hours to first voiding compared to free drainage.

Reply 10: Thank you for your reminder. We agreed with your opinions. In the Conclusion, we deleted the word "may" and added the description.

Change in revised manuscript: Page 2, Line 25-27 and Page 12, Line 25-27 in RED.

Comment 11. The use of the abbreviation "CI" is used in the abstract and the methods prior to defining. Please define it in the abstract and then use the abbreviation.

Reply 11: Thank you for your suggestions. We added the definition of "CI" in the abstract.

Change in revised manuscript: Page 2, Line 12 in RED.

Comment 12. Consider renaming the endpoint "time to first void" as " hours to first void" so the time unit is clear throughout the manuscript.

Reply 12: We wholeheartedly concurred with your views that "hours to first void" is more understandable. We have changed the name of "time to first void" to "hours to first void" in the full manuscript.

Thank you for your thoughtful and constructive comments. We tried our best to improve the manuscript according to the comments. We appreciate for your warm work earnestly, and hope that the correction will meet with approval. We are looking forward to your further suggestions.

Reviewer: 6

Comment 1: The manuscript may benefit from further proof reading to correct a few grammatical/typo errors.

Reply 1: Thank you for your sincere and constructive suggestions. Following your comments, we have carefully revised this manuscript. After that, we invited Editage language editing agency to polish the language of our revised manuscript (Job code: JQUNQ_1), and with their help, the quality of the language of our manuscript has been greatly improved.
Thank you all for all your help and looking forward to hearing from you soon.

**VERSION 3 – REVIEW**

<table>
<thead>
<tr>
<th>REVIEWER</th>
<th>Daignault-Newton, Stephanie</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>University of Michigan, Biostatistics</td>
</tr>
<tr>
<td>REVIEW RETURNED</td>
<td>30-Jan-2023</td>
</tr>
</tbody>
</table>

| GENERAL COMMENTS             | The authors have addressed all of the statistical concerns. |