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

TITLE (PROVISIONAL)	Diagnostic Performance of Image Technique-Based Transurethral Resection for Non-muscle Invasive Bladder Cancer: Systematic Review and Diagnostic Meta-analysis
AUTHORS	Chen, Changhao; Huang, Hao; Zhao, Yue; Liu, Hao; Sylvester, Richard; Lin, Tianxin; Huang, Jian

VERSION 1 – REVIEW

REVIEWER	Angelo Naselli San Giuseppe Hospital, Multimedica Group, Milan, Italy
REVIEW RETURNED	02-Dec-2018

GENERAL COMMENTS	Authors should be congratulated for their work. It complies with the most stringent rules of modern meta-analysis technique. Diagnostic superiority of NBI and HAL versus WLC is clearly showed. However, I have one minor concern. Authors declare that most of the cancers were non-muscle invasive. Is it possible to have the exact percentage? Is it possible to make subgroup analysis according to the grade of the tumours (if not please explain why)? For example, the clinical impact of a patient found with Cis or Ta/T1 high grade or G2/G3 tumour undetected from WLC is different from one additional Ta low grade or G1 tumour found in the context of a multifocal disease. Discussion should be amended accordingly.

REVIEWER	Simone Ferrero IRCCS Ospedale Policlinico San Martino, Genova, Italy
REVIEW RETURNED	11-Dec-2018

GENERAL COMMENTS	The Aithors should be congratulated for the interesting manuscript.

REVIEWER	J GAL Antoine Lacassagne Center, France
REVIEW RETURNED	10-Jun-2019

GENERAL COMMENTS	Interesting article to read. Stringent methodology.The authors discuss the strengths and weaknesses of their study. (The lack of data on important clinical variables)

REVIEWER	Tomohiro Shinozaki Tokyo University of Science, Japan
REVIEW RETURNED	11-Jun-2019

GENERAL COMMENTS	1. I was confused with the objective of the study. In Introduction, the authors pointed out the limitation of WLC's prediction of

	recurrence after the screening (1st paragraph). The following paragraphs thus call for alternative imaging techniques including 5-ALA, HAL, and MBI that may help diagnosis made by WLC. Namely, reference 13 demonstrated the increase in detection rates of future recurrence via MBI (at the cost of increase in false-positives). Please clarify the following 2 points: a. Why could WLC serve as a gold standard in this study despite the abovementioned low detection rate? What the purpose of this study? b. Is it necessary to assess the predictive performance for future recurrence (like reference 13) rather than the accuracy of "cross-sectional" diagnosis? 2. One of the inclusion criteria (p. 6) is having AUROC. That means each study has dichotomous diagnosis according to gold standard (i.e., WLC) and continuous measurements of 5-ALA, HAL, or MBI. I have 2 questions regarding this. a. How were AUROC data summarized in the analysis? b. There should be different levels of sensitivity and specificity within the same study. How did the authors select the sensitivity and specificity values in this study? 3. I did not find the definition of diagnostic odds ratio (DOR). Further, related above comment, how did the authors select DOR from each study possibly with multiple cut-points for calculating sensitivity and specificity. 4. Please more specifically explain the "diagnostic meta-analysis" adpted in Stata (p. 8). 5. Also, the methodology of SROC analysis supported by RevMan should be unclear to most readers. I suppose SROC assumes some kinds of parametric distribution in the estimation procedure (otherwise, it seems difficult to define and estimate AUC). This type of assumption should be critically examined. The estimation of AUC should be specified. 5. The authors used the term "sensitivity analysis" for the analysis of the subset of included studies. As the methodology is not different from the main analysis, it should be stated as "subgroup analysis" or "analysis of the subset of studies" etc.

VERSION 1 – AUTHOR RESPONSE

Reviewer #1:

Authors should be congratulated for their work. It complies with the most stringent rules of modern meta-analysis technique. Diagnostic superiority of NBI and HAL versus WLC is clearly showed. However, I have one minor concern. Authors declare that most of the cancers were non-muscle invasive. Is it possible to have the exact percentage? Is it possible to make subgroup analysis according to the grade of the tumours (if not please explain why)? For example, the clinical impact of a patient found with Cis or Ta/T1 high grade or G2/G3 tumour undetected from WLC is different from one additional Ta low grade or G1 tumour found in the context of a multifocal disease. Discussion should be amended accordingly.

Response: Thank you for the comments. We presented the exact percentage in table 1 in revised version. It is true that tumor grade subgroup analysis could better demonstrate the diagnostic performance of new imaging technique. However, the tumor grade information could not be retrieved in included studies. We amended the discussion accordingly in revised version. Please refer to the text (page 12, lines 9-11).

Reviewer#2
The Authors should be congratulated for the interesting manuscript.
Response: Thank you for the positive comments.

Reviewer#3
Interesting article to read. Stringent methodology. The authors discuss the strengths and weaknesses of their study. (The lack of data on important clinical variables).
Response: Thank you for careful read and thoughtful comments. The lack of data on important clinical variables, such as grade and stage of disease, primary vs recurrent disease and intravesical instillation settings, may introduce clinical heterogeneity and prevent further sensitivity analyses. We have attempted to minimize biases by applying rigorous selection criteria during the design phase of our study, standardizing data extraction and performing several subgroup analyses to evaluate the robustness of our findings. Please refer to the text (page 5, lines 11-23; page 7, lines 3-18; page 10, lines 7-12).

Reviewer#4
Comments 1: I was confused with the objective of the study. In Introduction, the authors pointed out the limitation of WLC's prediction of recurrence after the screening (1st paragraph). The following paragraphs thus call for alternative imaging techniques including 5-ALA, HAL, and MBI that may help diagnosis made by WLC. Namely, reference 13 demonstrated the increase in detection rates of future recurrence via MBI (at the cost of increase in false-positives). Please clarify the following 2 points:
a. Why could WLC serve as a gold standard in this study despite the abovementioned low detection rate? What the purpose of this study?
b. Is it necessary to assess the predictive performance for future recurrence (like reference 13) rather than the accuracy of "cross-sectional" diagnosis?
Response: Thank you for the comments.
a. WLC has been the standard method for detecting urothelial cell carcinoma of bladder currently. However, the sensitivity of WLC is unsatisfactory for missing small 'satellite' tumors or carcinoma in situ. Thus, new imaging technique (photodynamic diagnosis, narrow band imaging) have been introduced to enhance bladder cancer visualization to improve diagnostic accuracy and thoroughness of resection. Several studies demonstrated new imaging technique showed superior diagnostic performance than WLC1,2. While increased false positives caused by intravesical instillation or inflammation, technical complexity and cost restricted new imaging technique application3-6. It is still uncertain which technique could better improve diagnosis accuracy of bladder cancer detection. Therefore, the purpose of this study is to explore diagnostic performance of PDD using 5-ALA or HAL, NBI against the reference standard of WLC for NMIBC. Pooled diagnostic meta-analysis showed NBI, HAL or 5-ALA showed excellent efficacy compared with WLC. NBI could potentially be the most promising diagnostic intervention for NMIBC patients with advantages in terms of simplicity, cost and reliability, which may be widely applied to detect bladder cancer as clinical routine.
b. It is necessary to assess the predictive prognosis of future recurrence and progression. However, prognostic outcomes of new imaging technique compared with WLC could not be retrieved from included studies, therapeutic effectiveness of new technique assisted TUR such as recurrence and progression could not be demonstrated in the present study. Thus, future therapeutic efficacy analysis was needed to identify promising intervention. Please refer to the text (page12, lines 2-4).

Reference
1. Kausch I, Sommerauer M, Montorsi F, et al. Photodynamic diagnosis in non-muscle-invasive bladder cancer: a systematic review and cumulative analysis of prospective studies. Eur Urol 2010;57(4):595-606. doi: 10.1016/j.eururo.2009.11.041 [published Online First: 2009/12/17]

2. Ye Z, Hu J, Song X, et al. A comparison of NBI and WLI cystoscopy in detecting non-muscle-invasive bladder cancer: A prospective, randomized and multi-center study. Sci Rep 2015;5:10905. doi: 10.1038/srep10905 [published Online First: 2015/06/06]
3. Draga RO, Grimbergen MC, Kok ET, et al. Photodynamic diagnosis (5-aminolevulinic acid) of transitional cell carcinoma after bacillus Calmette-Guerin immunotherapy and mitomycin C intravesical therapy. Eur Urol 2010;57(4):655-60. doi: 10.1016/j.eururo.2009.09.037 [published Online First: 2009/10/13]
4. Geavlete B, Multescu R, Georgescu D, et al. Narrow band imaging cystoscopy and bipolar plasma vaporization for large nonmuscle-invasive bladder tumors--results of a prospective, randomized comparison to the standard approach. Urology 2012;79(4):846-51. doi: 10.1016/j.urology.2011.08.081 [published Online First: 2012/02/22]
5. Ray ER, Chatterton K, Khan MS, et al. Hexylaminolaevulinate fluorescence cystoscopy in patients previously treated with intravesical bacille Calmette-Guerin. BJU Int 2010;105(6):789-94. doi: 10.1111/j.1464-410X.2009.08839.x [published Online First: 2009/10/17]
6. Daneshmand S, Schuckman AK, Bochner BH, et al. Hexaminolevulinate blue-light cystoscopy in non-muscle-invasive bladder cancer: review of the clinical evidence and consensus statement on appropriate use in the USA. Nat Rev Urol 2014;11(10):589-96. doi: 10.1038/nrurol.2014.245 [published Online First: 2014/09/24]

Comments 2: One of the inclusion criteria (p. 6) is having AUROC. That means each study has dichotomous diagnosis according to gold standard (i.e., WLC) and continuous measurements of 5-ALA, HAL, or MBI. I have 2 questions regarding this.
a. How were AUROC data summarized in the analysis?
b. There should be different levels of sensitivity and specificity within the same study. How did the authors select the sensitivity and specificity values in this study?
Response: We are appreciated that you raised this important point.
a. We extracted data from included studies and conducted 2×2 table to re-calculate sensitivity, specificity, negative predictive value, positive predictive value, false positive rate and false negative rate for each studies. Then diagnostic meta-analysis was managed using Stata 13.0 (StataCorp, College Station, TX, USA). Random-effect model was applied in this study. Then pooled sensitivity and specificity were calculated using the bivariate model, displaying the diagnostic probabilities of individual studies with 95% confidence intervals (CIs). Then the AUROC was conducted for graphic presentation of the study results, which was plotted by RevMan.
b. We are sorry that we could not retrieve different levels of sensitivity and specificity from each included study. We extracted raw data from individual studies to conduct 2×2 table calculating sensitivity, specificity, negative predictive value, positive predictive value, false positive rate and false negative rate. Then pooled sensitivity and specificity were calculated with bivariate random-effect model. The pooled DOR and AUROC were presented. We made a wrong description in inclusion criteria section, and we revised it. Please refer to the text (page 6, lines 5-10).
Comments 3: I did not find the definition of diagnostic odds ratio (DOR). Further, related above comment, how did the authors select DOR from each study possibly with multiple cut-points for calculating sensitivity and specificity.
Response: Thank you for the comments. We have added the definition of diagnostic odds ratio (DOR) in revised version. Please refer to the text (page 8, lines 2-4). We are sorry that we could not retrieve DOR from individual studies and calculate sensitivity and specificity with definite cut-point. We extracted raw data from included studies and re-calculating sensitivity, specificity, negative predictive value, positive predictive value, false positive rate and false negative rate for each study. Diagnostic meta-analysis was then performed using Stata 13.0 (StataCorp, College Station, TX, USA). Bivariate random-effect model was conducted to pool sensitivity and specificity in forest plot. Diagnostic odds ratio (DOR) with 95% confidence intervals (CIs) of the new imaging technique was demonstrated. The present results do strongly suggest that new imaging-based technologies, NBI, are promising

diagnostic intervention for bladder cancer detection in clinical practice. We have revised the data extraction and statistical analysis in method section. Please refer to the text (page7, lines 3-18; page 7, lines 21-23, page 8, lines 1-12).

Comments 4: Please more specifically explain the “diagnostic meta-analysis” adpted in Stata.
Response: Thank you for the comments. We have revised the description of “diagnostic meta-analysis” in statistical analysis. Please refer to the text (page 7, lines 21-23; page 8, lines 1-4).
Comments 5: Also, the methodology of SROC analysis supported by RevMan should be unclear to most readers. I suppose SROC assumes some kinds of parametric distribution in the estimation procedure (otherwise, it seems difficult to define and estimate AUC). This type of assumption should be critically examined. The estimation of AUC should be specified.
Response: Thank you for the comments. The diagnostic meta-analysis was conducted by Stata13.0 (StataCorp, College Station, TX, USA). While the pooled AUROC of NBI, HAL and 5-ALA were summarized in the same figure in RevMan in accordance with Stata demonstrating same result. Please refer to the text (page 8, lines 5-6).

Comments 6: The authors used the term “sensitivity analysis” for the analysis of the subset of included studies. As the methodology is not different from the main analysis, it should be stated as “subgroup analysis” or “analysis of the subset of studies” etc.
Response: Thank you for the comments. We have revised manuscript accordingly.

VERSION 2 – REVIEW

REVIEWER	Angelo Naselli Urology Department, San Giuseppe Hospital, Multimedica group, Milan, Italy
REVIEW RETURNED	14-Aug-2019

GENERAL COMMENTS	Authors amended the paper thoroughly according to reviewers’ comments. Particularly, limitations deriving from data availability have been addressed. Now there are no major concerns to be reported. Just for clarity, the last sentence of the abstract should be rewritten.

VERSION 2 – AUTHOR RESPONSE

Reviewer’s comments:

Comments 1: Authors amended the paper thoroughly according to reviewers’ comments. Particularly, limitations deriving from data availability have been addressed. Now there are no major concerns to be reported. Just for clarity, the last sentence of the abstract should be rewritten.

Response: Thank you for the positive comments. We have revised manuscript accordingly. Please refer to the abstract section.