

**What is the required certainty of evidence for the implementation of novel techniques for the treatment of benign prostatic obstruction?**

Mark J Speakman<sup>a</sup>, Jean-Nicolas Cornu<sup>b</sup>, Mauro Gacci<sup>c</sup>, Christian Gratzke<sup>d</sup>, Charalampos Mamoulakis<sup>e</sup>, Thomas RW Herrmann<sup>f</sup>, Muhammad Imran Omar<sup>g</sup>, Malte Rieken<sup>h</sup>, Kari AO Tikkinen<sup>i</sup>, Stavros Gravas<sup>j</sup>.

<sup>a</sup>Department of Urology, Taunton & Somerset Hospital, Taunton, UK.

<sup>b</sup>Department of Urology, Charles-Nicolle University Hospital, Rouen Cedex, France

<sup>c</sup>Minimally Invasive and Robotic Surgery, and Kidney Transplantation, University of Florence AOUC- Careggi Hospital, Florence, Italy

<sup>d</sup>Department of Urology, Ludwig-Maximilians-University, Munich, Germany

<sup>e</sup>Department of Urology, University General Hospital of Heraklion, University of Crete Medical School, Heraklion, Crete, Greece

<sup>f</sup>Urology Clinic, Spital Thurgau AG, Frauenfeld, Switzerland and Urology and Urological Oncology, Hanover Medical School, Hanover, Germany

<sup>g</sup>Academic Urology Unit, University of Aberdeen, Aberdeen, Scotland, UK

<sup>h</sup>alta uro AG, Basel, Switzerland; University Basel, Basel, Switzerland

<sup>i</sup>Departments of Urology and Public Health, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

<sup>j</sup>Department of Urology, Faculty of Medicine, School of Health Sciences, University of Thessaly, Larissa, Greece

**Corresponding author:**

Mark J Speakman, Consultant Urological Surgeon, Department of Urology, Taunton & Somerset Hospital, Taunton, TA1 5DA. UK.

Word Count 2779

## **Abstract**

**Context:** A large number of minimally invasive techniques (MITs) have been developed for the surgical management of male lower urinary tract symptoms presumed secondary to benign prostatic obstruction (LUTS/BPO) over the last three decades. Many have not stood the test of time often because they were over promoted before there were sufficient data.

**Objectives:** The scope of this paper is to consider whether new devices, for the treatment of male LUTS/BPO, have been implemented prematurely in the past. We also examine the relative certainty of evidence (CoE) that is currently available for the newer developing technologies and make recommendations about the CoE that should be demanded in the future before widespread implementation.

**Key Messages:** This evidence must provide adequate lengths of follow up to allow proper information to be provided for patients before treatment choices are made and to be able to create recommendations in high quality guidelines such as those of the European Association of Urology (EAU). It's not just within the domain of LUTS treatments that this is important; other urological devices, such as mesh devices, have been equally 'guilty' and likewise devices in most other (surgical) specialties.

We believe that there is a need for a set of requirements built around primary randomised controlled trials (RCTs) looking at both efficacy and safety and secondary studies to confirm the reproducibility and the generalisability of the first pivotal studies. Otherwise there is a danger that a single pivotal study can be over exploited by the device manufacturers.

Studies that are needed include i) proof of concept, ii) RCTs on efficacy and safety as well as iii) cohort studies with a broad range of inclusion and exclusion criteria to confirm both reproducibility and generalisability of the benefits and harms.

It is not the purpose of this paper to make judgements about individual treatments but simply to look at different treatments to provide verification for this debate.

### **Patient Summary**

Many new device treatments have been developed over the last 20-30 years, often with inadequate medium to long-term results. Many have not stood the test of time but were heavily promoted by the manufacturers, the press, and some doctors, when they were first released, meaning that many patients had unsatisfactory results. This paper proposes minimum standards for the investigation of new treatments before their widespread promotion to patients.

### **Introduction**

It is a common perception that surgical rather than medical intervention becomes appropriate in those patients with bothersome LUTS due to benign prostatic obstruction (BPO), who are unwilling to try medical therapies, or in cases where medical therapies are ineffective or not well tolerated, and in cases of complicated LUTS. However, with the advent of more minimally invasive treatments the threshold for surgical intervention has become much lower<sup>1</sup>.

This paper is the opinion of the current EAU male LUTS guideline development group and has been developed after considerable review of the current literature on the novel treatment devices.

A large number of MITs have been developed for the management of male LUTS/BPO, in the last three decades or so. From the middle of the 1980s onwards attempts to use balloon

dilation of the prostate, prostatic stents, prostate hyperthermia and cryotherapy, Neodymium:YAG based therapies of the prostate like interstitial laser coagulation (ILC) or visual laser ablation of the prostate (VLAP), transurethral needle ablation (TUNA), transurethral microwave therapy (TUMT) and high intensity focused ultrasound (HIFU) have all had strong advocates, but over time many devices have not stood the test of time and have fallen out of favour. These may have disappeared for a variety of reasons, such as lack of long-term efficacy, adverse events, poor cost effectiveness or the development of newer versions (the ‘industrial/commercial circle’), which again dilutes the data available on the latest or current version of the device.

It is important to bear in mind, from the start, that patients’ and clinicians’ expectations of outcomes may differ. For instance, patients (compared to clinicians) may prefer treatment options that are less effective if they result in a smaller risk of complications or a faster return to normal activities. There is little doubt that there is a desire from patients for MITs.

Unfortunately, sometimes the eagerness for these new MITs, both in the popular press – the commonest source of direct patient information – but sometimes also from the early adopters of the treatment, has meant that this enthusiasm has often outstripped the clinical evidence.

Early and aggressive company marketing can also over-power clinical evidence. Unlike pharmaceutical interventions, implementation of novel devices and techniques into clinical practice is not governed by the same strict regulatory mechanisms<sup>2</sup>.

Therefore, the quality of the studies must remain high: i) there must be adequate control arms (for instance, compared to sham treatment or to other established comparable therapies), ii) outcomes that are patient-important, including patient reported outcome measurements (PROMS), and iii) include a relevant study population that is generalisable.

We also need to remember that behind the tool itself, there is also a concept: ablation, resection, enucleation... therefore a new device for resection is not a new device for vaporisation. Some of the principles behind these 'failed' treatments have been born again using newer or more powerful devices. It is therefore important that unlike their predecessors, sufficient controlled trials are carried out before their widespread application. Certain concepts however, such as (laser) enucleation have resulted in a paradigm change in the surgical management of LUTS/BPO.

---

### **Long-term unsuccessful treatments**

In 1988, transurethral balloon dilatation of the prostate was reported as a safe and simple procedure which could be performed on an outpatient basis using topical anaesthesia and sedation with minimal morbidity<sup>3</sup>. Two years later however, another series claimed that significant improvement in objective measurement of outflow obstruction was noted in only two out of fourteen patients<sup>4</sup>

Another example was TUNA where again early results promised significant improvements in both subjective and objective outcomes: "TUNA is a promising, anaesthesia-free alternative treatment for men with symptomatic BPH"<sup>5</sup>. Yet a few years later a study reported that "TUNA produced an unsatisfactory clinical result"<sup>6</sup>. The overall retreatment rate after TUNA<sup>TM</sup> was 19% based on an analysis of seventeen non-comparative studies (median follow-up unreported; only three out of seventeen studies had follow-up exceeding two years; a rate considerably higher than that seen with TURP<sup>7</sup>. A very high re-intervention rate for BPO was also reported in another French study of TUNA<sup>8</sup>

Many of these and similar studies, in their conclusions, promised prospective randomised studies to determine the role of treatment 'x' in the management of LUTS, but these rarely materialised.

### **How to assess Certainty of Evidence**

The CoE in urological clinical research literature is generally very low, and the reporting is inadequate<sup>9</sup>. The Idea, Development, Exploration, Assessment, Long-term (IDEAL) study recommendations have been established as a potential solution by serving as guidelines tailored to surgical research; a platform for systematic data generation from well-designed, conducted, and reported trials; and provides a regulatory protective framework against potential harms of novel procedures before incorporation into practice<sup>10</sup>.

*The Grading of Recommendations Assessment, Development and Evaluation (GRADE)* is a transparent approach for assessing the CoE (also known as quality of evidence). GRADE approach is adopted and/or endorsed by more than 100 organizations including Cochrane and World Health Organization. The EAU Guidelines Office use modified GRADE for rating guidelines recommendations.

According to GRADE, the outcomes included in a systematic review (SR) are scored on a scale of 1 to 9 [low importance to critically important for decision making]. Evidence from RCTs starts as "high" certainty, whereas evidence from non-randomised studies start as "low" certainty. There can be five reasons to possibly rate down the CoE: i) methodological limitation of study design (risk of bias), ii) inconsistency, iii) indirectness, iv) imprecision and v) publication bias. There are 3 possible reasons to rate up the certainty of evidence (from non-randomised studies): i) large magnitude of effect, ii) opposing plausible residual bias or confounding, iii) dose-response gradient. The guidelines recommendations are based upon

certainty of evidence of the critical outcomes, balance of benefits and harms, patients' values, preferences and acceptability, and resource implication. By considering the above factors a “strong” or “weak” recommendation, “for” or “against” an intervention can be made<sup>11</sup>.

**Safe and appropriate implementation of new treatments:**

**Table 1 – Requirements for widespread implementation**

- Proof of concept study
- Placebo/sham comparison study:
- Randomised controlled trial against accepted alternative treatment
- Cohort studies: to understand the generalisability and potential harms
- Systematic reviews and meta-analysis of high-quality primary studies
- How many patients should be included? – a sample size determination is needed
- What should be the length of follow up?
  - Short- <12 months, medium- 12–36 months and long-term > 36 months.
- Inclusion and exclusion criteria need to allow good generalisability
- Relevant outcomes – varied and need to be clearly outlined from the start

=====

To safely and effectively introduce novel techniques in the future there needs to be a systematic way of studying these new devices.

The studies should start with the publication of a proof of concept or ‘first in man’ trial followed by a randomised controlled trial against a sham (placebo) treatment. It is likely, although not essential, that these will come from experienced centres of excellence and need to follow established criteria including the usage of risk of bias assessment tools<sup>11,12,13</sup>.

These studies should be followed by RCTs against TURP or a well-recognised and accepted MIT. The issue is what should be the comparator. For instance, an MIT like Rezum is perhaps a competitor between medical treatment and an ablative technique for BPO relief. Thus, has it to be as safe as a drug, or just safer than a more invasive ablative therapy, with a trade-off about reduced efficacy?

Ideally two RCTs would be presented; one a non-inferiority trial on efficacy and one on safety, although these could come from a single study, if it is of a sufficiently high-quality design. It would be preferable to conduct multicentre studies with several surgeons, rather than single-centre or single-surgeon studies, to confirm the reproducibility of the principle.

These should be followed by cohort studies; which will help with understanding the generalisability of the concept, allow secondary centres to gain experience and investigate for harms not necessarily shown in trials with selected study populations or short duration follow-up. These should be mandatory to have the right to penetrate the market.

Finally, come RCT-based SRs and/or meta-analyses; but this is on the proviso that there are sufficient high-quality RCTs in the first place. One reasonable question is, are systematic reviews currently done too soon before there are sufficient high quality RCTs?

The next question is; how many patients should be included? Clearly a sample size determination must be performed; this does not guarantee however, that the study will be powered for all the primary outcomes of interest. This calculation would be influenced by the expected difference between treatments and also by what we could call the ‘margin of inferiority’ that might be accepted for a less invasive treatment. Systematic reviewers are encouraged to calculate “optimal information size” (OIS). OIS is the minimum threshold of pooled estimate obtained from meta-analysis that is required for drawing reliable conclusions.

What should be the length of follow up? This has been one of the greatest weaknesses in past trials. This is amplified in a systematic review of transurethral ablative procedures (techniques that have often been evaluated better than other MITs)<sup>14</sup>. They identified 69 studies on LUT/BPO ablative surgical treatments over a 20-year period. The follow-up of these studies ranged from 1 to 100 months with a median follow up of just 12 months. Unsurprisingly, the longer the duration of follow-up, the fewer the number of trials available (Fig 1)

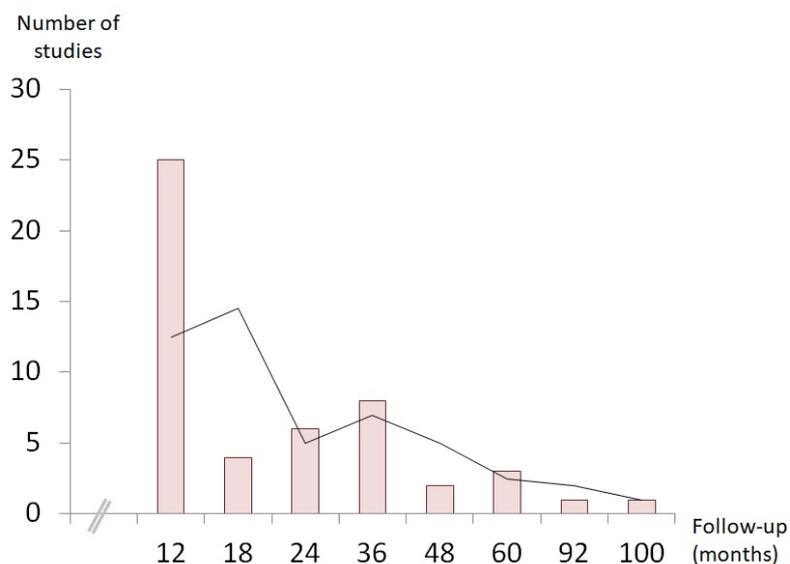


Fig 1 Follow-up of LUTS/BPO ablative studies. Adapted from Cornu JN et al. Reference 14.

As noted by the authors “...further studies are needed to provide long-term comparative data and head-to head comparisons of emerging techniques.” A reasonable suggestion therefore from the EAU non-neurogenic mLUTS Guideline Panel would be; i) short-term <12 months, ii) medium-term 12–36 months and iii) long-term > 36 months follow-up<sup>15</sup>. The outcome expectations of these studies may vary. If a therapy is very safe but with limited efficacy over time, but can be done again, it is more a trade-off with the patient. But these facts must be clear. This could be reflected in a guideline recommendation.

Inclusion and exclusion criteria need to be broad enough to allow good generalisability of the technique. Sometimes the inclusion and exclusion criteria have been too stringent and make the studies not applicable to all the population. These must therefore, include factors such as variable prostate size and configuration and patients with or without retention after failed trial without catheter. Larger post-void residual volumes could be considered but may reflect more on poor detrusor function than BPO. Patients with a range of co-morbidities need to be included, certainly in the cohort studies, to note their potentials on treatment options with less morbidity for patients with more co-morbidities. This would include patients with a variety of anticoagulant needs.

Before starting a clinical trial or a study, it is strongly recommended to check if there is an agreed minimum set of outcomes or outcome measures known as “core-outcomes set” (COS). The COS should be measured and reported by all new clinical trials and studies. “Core Outcome Measures in Effectiveness Trials” (COMET)<sup>16</sup> and the “International Consortium for Health Outcomes Measurement” (ICHOM) have developed COS for various

medical conditions. In a recent SR on the management of urinary retention in patients with LUTS/BPO, a broad search of the COMET database for a COS using the term “urology” in the disease category, yielded no directly applicable COS for the disease or treatments dealt with in that SR<sup>17</sup>. Outcomes important to all stakeholders; patients, the public, doctors and health-care decision makers need to be developed<sup>18</sup>. COS should be checked first, and if they don’t exist, then the following outcomes should be considered for measurement and reporting (Table 2). Inconsistency in the measurement of outcomes, outcome-reporting bias, adds further to the problems faced by users of research who wish to make well-informed decisions about health care.

Table 2. Potential Recordable Outcomes: -

- Symptom assessment (e.g. IPSS, bladder diary)
- Objective measurements (e.g. Flow rate, PVR and pressure flow studies)
- Quality of life questionnaire(e.g.SF-36)
- Patient satisfaction
- Erectile and Ejaculatory function (e.g. IIEF and MSHQ-EjD)
- Complications associated with the intervention
- Length of stay and other hospital factors
- Speed of return to normal activities
- Need for additional medical treatment
- Re-operation rates
- Cost effectiveness (e.g. QALYs)

The appropriateness of some of these are obvious and will depend on the device being tested. With many of the new MITs the ability to preserve both erectile and ejaculatory function has been one of their potential positive benefits. Others such as the need for supplemental medical treatment will reflect both on the success of the treatment but may, in the case of overactive bladder symptoms, not be a sign of device failure. Overall however we have to recognise that concentrating on patient important-factors such as return to all normal activities can be more important to patients than a dramatic improvement in flow rate. The minimally important difference (MID) is the smallest change in a clinical outcome that patients perceive as important, either beneficial or harmful, which could result in the patient or doctor considering a change in treatment<sup>19</sup>. It has always represented a challenging concept in the management of LUTS. It is generally accepted that patients will perceive a difference if their IPSS improves or deteriorates by 3 points or more, however, this varied considerably in the original paper and depended greatly on the baseline symptom score<sup>20</sup>. In addition, the MID is not an immutable characteristic, but may vary by population and context<sup>21</sup>. The authors emphatically support the need for a clinical significance aspiration and not merely statistical significance.

Currently in the literature, bipolar-TURP is the most investigated of the newer modalities with 56 RCTs and >6500 patients! But still CoE is not perfect for many of the investigated outcomes<sup>22,23</sup> and it is much lower indeed for many of the other MIT procedures<sup>15</sup>. A simple PubMed search reveals 54 references for Urolift, 39 for Rezum, 248 for prostate artery embolisation (PAE) (because of many radiological technique papers), 42 for Aquablation and 6 for iTIND devices. The very large majority of these papers, of course are not RCTs.

As a result of the quality of the information, devices and techniques will have varying recommendations in international guidelines over the years, such as those of the EAU (Table 3)

Table 3. Emerging technologies reviewed in the EAU LUTS guideline over three iterations of the guideline from 2009 to 2019

<b>Treatment</b>	<b>EAU2009</b>	<b>EAU2015</b>	<b>EAU2019</b>
<b>Bipolar TURP</b>	ND	R	R
<b>GreenLight</b>	ND	R	R
<b>PUL</b>	ND	E	R*
<b>Intraprostatic BTX</b>	ND	E	NR

**NR:** Not Recommended **R:** Recommended **R\*:** Under conditions **ND:** Not Discussed **E:** Emerging

## **Conclusions**

New modalities therefore need to achieve similar or better improvements in patient-important outcomes – as established treatments, such as bipolar TURP, or have lower risk of complications and side effects, or similar costs and shorter hospital stays. Evidence should include studies with long-term (> 3 years) follow-up. Preferably they should also improve urodynamic parameters such as maximum urinary flow rate and postvoid residual volumes<sup>24</sup>. There is clearly a difference however between the ideal treatment which might be cost-effective, day-case, with few complications, good relief of BPO and good durability and the

ideal standard of evaluation which is all about safety and clinical effectiveness. These tradeoffs are clearly value- and preference sensitive and shared decision-making between patients and clinicians is therefore necessary.

National organisations for example, the British, French, and German Urological Associations, and the international ones like the EAU should refuse to sanction new treatments (as they do when they devise patient information leaflets and procedure specific consent forms) until an agreed minimum threshold of CoE for the implementation of novel techniques for the surgical treatment of LUTS/BPO is available.

However, we do not want an extended delay until effective treatments are fully implemented and therefore a national or better still a European initiative would allow new techniques to be used ONLY in research settings, such as structured cohort studies or registries, to collect crucial data about these early patients, and avoid or reduce the misuse under commercial pressure, particularly in private healthcare centres.

This paper presents in a transparent way how the EAU Guideline Panel on non-Neurogenic mLUTS evaluates new invasive treatments and proposes the requirements and CoE that a new therapy should meet in order to be included in the Guidelines. It is our strong belief that the paper should trigger further discussions about the optimal way to assess new technology across all surgical disciplines, but that these devices should not be broadly implemented until at least quality RCTs on both safety and efficacy with adequate follow-up are completed. Contribution from patients' organizations, device manufacturers and other stakeholders would be extremely important to formulate the optimal guidance.

## **References**

1. Kaplan S. Re: Can long-term LUTS-BPH pharmacological treatment alter the outcomes of surgical intervention. *J Urol* 1 Jan 2018; 199: 6. <https://doi.org/10.1016/j.juro.2017.09.130>
2. Mamoulakis C. A Plea for Higher-quality Data for GreenLight Laser Technology in the Context of Surgical Benign Prostatic Obstruction Trials: The GOLIATH Study—Fact or Fiction in the Era of Evidence-based Urology? *Eur Urol* 2014; **65**: 943–946
3. Reddy PK, Wasserman N, Castañeda F et al. Balloon dilatation of the prostate for treatment of benign hyperplasia. *Urol Clin North Am.* 1988;15(3): 529-35
4. Keane PF, Charig CR, Hudd C et al. Balloon dilatation of the prostate: technique and early results. *Br J Urol.* 1990; 65(4): 354-6.
5. Schulman CC, Zlotta AR. Transurethral needle ablation of the prostate for treatment of benign prostatic hyperplasia: early clinical experience. *Urology.* 1995; 45(1): 28-33.
6. Holmes MA, Stewart J, Boulton JB et al. Transurethral needle ablation of the prostate: outcome at 1 year. *J Endourol.* 1999;13(10): 745-50.
7. Bouza, C, López T, Magro A., et al. Systematic review and meta-analysis of Transurethral Needle Ablation in symptomatic Benign Prostatic Hyperplasia. *BMC Urol*, 2006. 6: 14.
8. Lukacs B, Cornu JN, Aout M et al. Management of lower urinary tract symptoms related to benign prostatic hyperplasia in real-life practice in France: a comprehensive population study. *Eur Urol.* 2013 Sep;64(3):493-501.
9. Assel M, Sjoberg D, Elders A et al., Guidelines for Reporting of Statistics for Clinical Research in Urology. *Eur Urol* 2019; **75**: 358 – 367
10. Editorial; Surgical research: the reality and the IDEAL. *Lancet* 2009;374:1037.
11. Knoll T, Omar MI, Maclennan S et al. Key Steps in Conducting Systematic Reviews for Underpinning Clinical Practice Guidelines: Methodology of the European Association of Urology. *Eur Urol.* 2018 Feb;73(2):290-300.
12. Sterne JA, Hernán MA, Reeves BC et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ.* 2016 Oct 12;355: i4919.

13. Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions, Version 5.1.0 (updated March 2011). The Cochrane. [www.handbook.cochrane.org](http://www.handbook.cochrane.org).
14. Cornu JN, Ahyai S, Bachmann A. et al. A systematic review and meta-analysis of functional outcomes and complications following transurethral procedures for lower urinary tract symptoms resulting from benign prostatic obstruction: an update. *Eur Urol.* 2015; 67: 1066–1096
15. Gravas S, Cornu JN, Drake MJ, et al. Guidelines on the management of non-neurogenic male lower urinary tract symptoms (LUTS), incl. benign prostatic obstruction (BPO) Presented at the EAU Annual Congress Copenhagen 2018. Arnhem, The Netherlands: EAU Guide-lines Office; 2018.978-94-92671-01-1 In: <http://uroweb.org/guideline/treatment-of-non-neurogenic-male-luts/>
16. Paula R. Williamson, Douglas G. Altman, Heather Bagley The COMET Handbook: version 1.0 The Author(s) *Trials* 2017, 18(Suppl 3):280 DOI 10.1186/s13063-017-1978-4
17. Karavitakis M Kyriazis I, Omar MI et al. Management of Urinary Retention in Patients with Benign Prostatic Obstruction: A Systematic Review and Meta-analysis. *Eur Urol.* 2019 Feb 14. pii: S0302-2838(19)30092-2. doi: 10.1016/j.eururo.2019.01.046. [Epub ahead of print]
18. MacLennan S, Williamson PR, Bekema H, A core outcome set for localised prostate cancer effectiveness trials. *BJU Int.* 2017 Nov;120(5B):E64-E79. doi: 10.1111/bju.13854. Epub 2017 May 3.
19. Bradley C Johnston, Shanil Ebrahim, Alonso Carrasco-Labra. Minimally important difference estimates and methods: a protocol. *BMJ OPEN* 2015 <https://bmjopen.bmj.com/content/5/10/e007953>
20. Barry MJ, Williford WO, Chang Y, et al. Benign prostatic hyperplasia specific health status measures in clinical research: how much change in the American Urological Association symptom index and the benign prostatic hyperplasia impact index is perceptible to patients? *J Urol.* 1995 Nov;154(5):1770-4.

21. Revicki, D., Hays, R. D., Cella, D., & Sloan, J. (2008). Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *Journal of Clinical Epidemiology*, 61(2), 102- 109
22. Mamoulakis C, Ubbink DT, de la Rosette JJ. Bipolar versus monopolar transurethral resection of the prostate: a systematic review and meta-analysis of randomized controlled trials. *Eur Urol*. 2009 Nov;56(5):798-809.
23. Omar MI, et al. Systematic review and meta-analysis of the clinical effectiveness of bipolar compared with monopolar transurethral resection of the prostate (TURP). *BJU Int*. 2014 Jan;113(1):24-35.
24. Hashim H & Abrams P. Transurethral resection of the prostate for benign prostatic obstruction: will it remain the gold standard. *Eur Urol* 2015; 67: 1097-1098