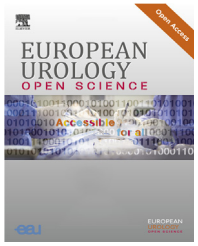
EUROPEAN UROLOGY OPEN SCIENCE 69 (2024) 89–99

available at www.sciencedirect.com
journal homepage: www.eu-openscience.europeanurology.com



European Association of Urology

Trial Protocol

Protocol for a Series of Systematic Reviews and Network Meta-analyses of Randomized Controlled Trials of Medications for Patients with Overactive Bladder Symptoms

Henk van der Worp [a,†], *Angie K. Puerto Nino* [b,c,†,*], *Marco H. Blanker* [a], *OAB Investigators,* [‡] *Romina Brignardello-Petersen* [d], *Gordon H. Guyatt* [d,e], *Kari A.O. Tikkinen* [b,d,f,g,*]

[a] *Department of Primary and Long-term Care, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands;* [b] *Faculty of Medicine, University of Helsinki, Helsinki, Finland;* [c] *Division of Urology, Department of Surgery, University of Toronto, Toronto, Canada;* [d] *Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada;* [e] *Department of Medicine, McMaster University, Hamilton, Canada;* [f] *Department of Urology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland;* [g] *Department of Surgery, South Karelian Central Hospital, Lappeenranta, Finland*

Article info

Article history:
Accepted August 6, 2024

Associate Editor:
Véronique Phé

Keywords:
Overactive bladder
Network meta-analysis
Multiple comparisons
Treatment
Clinical outcomes
Risk of bias
Urology

Abstract

Multiple randomized controlled trials (RCTs) have examined first-line pharmacological agents such as anticholinergics and β3 agonists for the management of overactive bladder symptoms (OAB). Although earlier systematic reviews and (network) meta-analyses aimed to summarize the evidence, a substantial number of trials were not included, so a comprehensive and methodologically rigorous evaluation of the comparative effectiveness of all first-line pharmacological treatments is lacking. We aim to conduct a series of systematic reviews and network meta-analyses (NMAs) for a comprehensive assessment of the effectiveness and safety of first-line pharmacological treatments for OAB. Eligible studies will include RCTs comparing anticholinergics and β3 agonists to one another or to placebo in adults with OAB or detrusor overactivity. Pairs of reviewers with methodological training will independently evaluate candidate studies to determine eligibility and extract relevant data. We will incorporate patient-important outcomes,

[‡] OAB Investigators: Arnav Agarwal, Yoshitaka Aoki, Scott R. Bauer, Neera Bhatnagar, Kostiantyn Bolsunovskyi, Gerhard H. te Brummelstroete, Jovita L. Cárdenas, Rufus Cartwright, Dean Elterman, Leyla Eryuzlu, Pramila Gaudel, Rachel E. Gutschon, Tomas L. Griebling, Sanna K. Hallamies, Katriina Heikkilä, Zain Izhar, Tuomas P. Kilpeläinen, Saqib Khurshid, Sanna M. Laakkonen, Yung Lee, Luciane C. Lopes, Anne Loohuis, Alexander Mendieta, Henri Mynttinen, Hanieh Salehi-Pourmehr, Alexey Pryalukhin, Mikolaj Przydacz, Fahad J. Siddiqui, Riikka Tähtinen, Robin W.M. Vernooij, Philippe D. Violette, Lambertus P.W. Witte, Koji Yoshimura.
[†] These authors are joint first authors.
[*] Corresponding authors. Faculty of Medicine, University of Helsinki, Biomedicum 2B, P.O. Box 13, Tukholmankatu 8 B, 00290 Helsinki, Finland.
E-mail addresses: angie.puerto-nino@helsinki.fi (A.K. Puerto Nino), kari.tikkinen@helsinki.fi (K.A.O. Tikkinen).

https://doi.org/10.1016/j.euros.2024.08.006

Muscarinic antagonists
Adrenergic β3 receptor agonists

including urinary urgency episodes, urgency incontinence episodes, any type of incontinence episodes, urinary frequency, nocturia, and adverse events. We will conduct the NMAs using a frequentist framework and a graph theory model for each outcome. Analysis will follow rigorous methodologies, including handling of missing data and assessment of the risk of bias. We will conduct sensitivity and subgroup analyses and will apply the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach to rate evidence certainty. Our approach aims to address the knowledge gap in the treatment of OAB by synthesizing evidence from RCTs worldwide. We will employ robust statistical methods, including frequentist NMA, to general clinically relevant and patient-important insights. Sensitivity and subgroup analyses will enhance the robustness and generalizability of our findings. Our reviews strive to inform evidence-based decisions in the management of OAB, to ultimate improve patient outcomes. Our study results may guide health policy decisions, such as reimbursement policies, and future studies in functional urology. The protocol for the review series is registered on PROSPERO as CRD42023266915.

1. Introduction

Overactive bladder symptoms (OAB) affect hundreds of millions of people worldwide [1,2] and account for substantial health care expenditure [3,4]. OAB is often bothersome and associated with a reduction in quality of life and work productivity [5,6]. As the risk of developing OAB increases with age, OAB prevalence is likely to increase with the ageing of the population [7,8].

Behavioral and lifestyle interventions have limited long-term benefits in reducing OAB [9,10]. These interventions are often used in conjunction with pharmacological therapy, including anticholinergics and β3 agonists. Even though first-line pharmacological treatments are recognized and recommended in guidelines worldwide, clinicians do not currently have a clear basis for recommending one medication or dose over another [11–15]. Table 1 presents the major guidelines addressing OAB treatment.

Multiple randomized controlled trials (RCTs) have studied the first-line pharmacological treatments for OAB. In addition, many meta-analyses are available. However, earlier systematic reviews and (network) meta-analyses included only a proportion of the RCTs completed. This situation has probably occurred for many reasons, including restrictions on eligibility criteria and specific patient populations. We are therefore planning to conduct a comprehensive analysis of all the available trial data to assess the efficacy and adverse effects of first-line OAB pharmacological treatments. Our aim is to encompass a wide range of RCTs and provide a thorough analysis to establish the impact of first-line OAB medications on patient-important benefits and harms in adults with a diagnosis of OAB.

Table 1 – Major guidelines for the management of overactive bladder symptoms

Guideline	Level/strength of recommendation	Certainty of evidence [a]	Magnitude of effect reported	Level/strength of recommendation	Certainty of evidence	Magnitude of effect reported	Threshold for magnitude of effect reported [b]
	By medication class			By medication type			
AUA/SUFU 2024 Non-neurogenic OAB in adults	AC/β3A: Strong	Grade A	No	NR	NR	No	No
EAU 2024	AC: Strong	1a	No	NR	NR	Yes	No
Non-neurogenic female LUTS	β3A: Strong	1a	No	NR	NR	No	No
EAU 2024	AC: Strong	2	No	NR	NR	No	No
Non-neurogenic male LUTS	β3A: Weak	2	No	NR	NR	No	No
NICE 2019 Urinary Incontinence in women	NR	NR	No	NR	NR	No	No
NICE 2010 [c] LUTS in men	NR	NR	Yes	NR	NR	No	No

AUA = American Urological Association; SUFU = Society of Urodynamics, Female Pelvic Medicine & Urogenital Reconstruction; EAU = European Association of Urology; NICE = National Institute for Health and Care Excellence; OAB = overactive bladder; LUTS = lower urinary tract symptoms; AC = anticholinergic; β3A = β3 agonist; NR = not reported.

[a] Grade A = very confident that the true effect lies close to that of the estimate of the effect; 1a = evidence obtained from meta-analysis of randomized controlled trials; 2 = evidence obtained from one well-designed controlled study without randomization or from at least one other type of well-designed quasi-experimental study.

[b] Guideline established a threshold to assess if the actual impact of the medication on the outcomes of interest is patient-important/clinically relevant.

[c] The guideline was updated in 2015 but there were no changes to the 2010 recommendations for the anticholinergics and β3 agonists.

2. Design

We adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis Protocols (PRISMA-P) guidance to report this protocol [16]. Our series of articles will establish the impact of OAB medications on patient-important benefits and harms in adults with OAB. Given the multiple pathophysiological reasons for OAB symptoms and their influence on medication response, we will perform the analysis for different patient groups, such as the older population, OAB associated with lower urinary tract symptoms suggestive of benign prostatic obstruction, and neurogenic OAB [17]. The analysis for each patient population will be presented as a separate publication within the series. The study protocol has been registered in the International Prospective Register of Systematic Reviews (PROSPERO CRD42023266915). No protocol amendments have been carried out since inception of the protocol. Possible future amendments will be recorded in the PROSPERO registry.

2.1. Eligibility criteria

We will include RCTs that compared anticholinergics or β3 agonists for OAB to placebo, anticholinergics, β3 agonists, or combinations of medications including anticholinergics and β3 agonists. Eligible anticholinergic medications are oxybutynin, tolterodine, propiverine, trospium, solifenacin, darifenacin, fesoterodine, emepronium, propantheline, and imidafenacin (Fig. 1). Eligible β3 agonists are mirabegron, solabegron, and vibegron (Fig. 1). We will incorporate any dose and type of formulation (eg, immediate release, extended release, and transdermal). We will include trials enrolling adults (>95% of patients at least 18 yr of age) with a diagnosis of OAB or detrusor overactivity with follow-up of at least 4 wk. We will include RCTs published since inception of the search databases. In the event of new medications being tested in RCTs with results published after the initial search date and before conclusion of our research project, we will also incorporate these results.

2.2. Outcomes to assess benefits and harms

Several outcomes have been used in RCTs on OAB medications. However, many of these outcomes have limited importance to patients. We will limit our analyses to outcomes that have been shown to be important to patients [18,19]. Our outcomes for benefits will be the change from baseline in the number of episodes per day in:

- Urinary urgency;
- Urinary urgency incontinence;
- Any type of incontinence;
- Average times of voiding per day;
- Nocturia.

We will also assess adverse events. We reviewed the list of adverse events included in published systematic reviews and we created a comprehensive list of adverse events (Table 2 [20,21]).

2.3. Literature search

In collaboration with an experienced information specialist (Neera Bhatnagar), we developed a comprehensive search strategy. Our search encompasses three electronic databases: CINAHL, EMBASE, and MEDLINE. The search strategy comprises keyword searches for the different medication names and includes keywords that restrict the search to randomized trials (Supplementary material). We will also undertake a review of systematic reviews to identify articles that may have been missed in the initial literature search. We will also search in abstract books from major urology and incontinence conferences, including the annual meetings of the American Urological Association (AUA), the International Incontinence Society, the European Association of Urology (EAU), and the International Urogynaecological Association. We will search the abstracts for potential unpublished studies and their data. We will not impose any language restrictions and will remove the duplicates before the screening phase.

2.4. Study selection, application of eligibility criteria, and data extraction processes

We will create and pilot test standardized data forms for screening of articles and extraction of data before their use for training of the research team. Training for reviewers will involve written and video guidance, followed by online meetings to cover important concepts and address questions. After training, reviewers will use the forms to assess the eligibility of study reports and to extract data from eligible articles. To ensure a high level of agreement, reviewers will carry out pilot screening and data extraction exercises and receive feedback. Reviewers will repeat the pilot screening exercises until they achieve a high level of agreement.

We will conduct screening and data extraction using DistillerSR [22]. Pairs of reviewers will work independently to scan all titles, abstracts, and full-text articles (Fig. 2). To obtain information from publications in languages other than English, two reviewers proficient in the respective language will review and extract the data. Adjudicators (one of the lead authors and/or a clinician-methodologist) will resolve any disagreements regarding screening and data extraction.

Once reviewers have conducted screening by title and abstract and full text and determined the eligibility of an article, two trained reviewers will independently extract data items, including the study setting, type of trial (parallel, crossover, factorial), funding, inclusion and exclusion criteria, study duration (or follow-up time), number of participants enrolled in each of the arms, baseline characteristics of participants, doses, and route of administration. For our outcomes of interest, we will collect the number of participants included in the analysis, the measure of central tendency and dispersion for each outcome at baseline, the end of follow-up period, and the change from baseline to the end of follow-up period. In addition, we will collect information on the number of patients experiencing any adverse event or any serious adverse events or withdrawing from the study because of adverse events. We will extract

Approval	Medication	Formulation	Doses
1960s	Emepronium[a]	Tablet	Not available
1974	Propantheline	Tablet	7.5 mg, 15 mg
1975	Oxybutynin	Tablet	5 mg
1979	Oxybutynin Propiverine[b]	Syrup Tablet	5 mg / 5 ml Not available
1997	Term "overactive bladder" was introduced		
1998	Tolterodine Oxybutynin	Tablet ER tablet	1 mg, 2 mg 5 mg, 10 mg, 15 mg
2000	Tolterodine	ER capsule	2 mg, 4 mg
2003	Oxybutynin	ER transdermal	3.9 mg
2004	Darifenacin Solifenacin Trospium	ER tablet Tablet Tablet	7.5 mg, 15 mg 5 mg, 10 mg 20 mg
2007	Trospium	ER capsule	60 mg
2008	Fesoterodine	ER tablet	4 mg, 8 mg
2009	Oxybutynin	Transdermal gel	10% (100 mg / packet)
2011	Imidafenacin[c]	Tablet	0.1 mg
2012	Mirabegron	ER tablet	25 mg, 50 mg
2020	Solifenacin Vibegron	Suspension Tablet	1 mg / ml 75 mg
2021	Mirabegron	ER suspension	8 mg / ml

Fig. 1 – Timeline of approval of overactive bladder medications by year. Antimuscarinics are shown in blue, and β3 agonists in green. Approval year and medication information are from the US Food and Drug Administration unless otherwise indicated. An application for solabegron was accepted on September 4, 2018. However, it has not been approved. [a] Emepronium information from the UK. [b] Propiverine information from Germany. [c] Imidafenacin information from Japan.

Table 2 – List of adverse events in alphabetical order by system [a]

Cardiovascular	**Integumentary**	**Reproductive**
Chest pain	Dry skin	Breast discomfort
Electrocardiograph abnormalities	Edema	Erectile dysfunction
Hypertension	Pruritus	Vaginal spotting
Hypotension	**Muscular**	**Respiratory**
Palpitations	Fatigue	Cough
Tachycardia	**Nervous**	Influenza
Digestive	Anxiety	Nasal congestion
Abdominal distension	Back pain	Nasopharyngitis
Abdominal pain	Blurred vision	Respiratory events
Alanine aminotransferase increase	Cognition impairment	Upper respiratory tract infection
Constipation	Confusion	White blood cells urine positive
Diarrhea	Depression	**Urinary**
Dyspepsia	Dizziness	Dysuria
Dry mouth	Dry eye	Postvoid residual volume
Flatulence	Fall	Red blood cell positivity for urine
Gamma-glutamyl transferase increase	Headache	Urinary tract infection
Nausea	Insomnia	Urinary retention
Stomach discomfort	Keratoconjunctivitis	
Vomiting	Lethargy	
Endocrine	Nasal dryness	
Blood triglycerides increased	Psychotic episodes	
Hematopoietic	Somnolence	
Decrease in white cell count	Syncope	
Hyponatremia	Visual impairment	

[a] The classification by anatomic system was generated on the basis of the hierarchical anatomic classification schema for the prediction of phenotypic side effects [20]. In an approach involving iterative discussion and consensus-building and informed by prior literature [21], we estimated the minimally important difference for postvoid residual volume (PVR) to be 50 ml in this context. While a PVR of 50 ml may not be patient-important in everyday practice, we accounted for the possibility that most of the trials excluded both patients with higher PVR at baseline and patients with high risk of developing an increase in PVR or urinary retention and did not have long-term follow-up (to develop an increase in PVR). We therefore considered a lower MID for PVR than what would typically be used in routine clinical practice.

the data items for each publication included after full-text screening, regardless of whether more than one publication corresponds to a single RCT or not. This approach ensures that we collect all relevant data, even if it is mentioned in one publication and not in another.

2.5. Organization of RCTs

Identification of independent trials can be challenging for various reasons, such as (1) the absence of registration numbers in some articles and (2) the variability of registration systems in RCTs. To ensure that we do not miss any trials and to avoid counting several articles from the same trial as independent, we will assess all publications included in the review according to the following process. We will identify articles containing the primary results from each RCT, as well as those with supplementary results, such as articles with partial results and conference abstracts. The identification of primary and secondary publications will be based on several variables, including NCT number (ClinicalTrials.gov identifier) or any other trial registration ID, year of publication, authors and their affiliations, the number of participants enrolled, the treatments compared, the number and location of study sites, and descriptive variables for each arm. If required, we will also contact the authors of the original publications. For instance, for cases in which only conference abstract(s) are available, we will contact the authors to request a full summary of the study.

2.6. Risk-of-bias assessment

Pairs of reviewers will independently evaluate the risk of bias of eligible RCTs using a modified version of the Cochrane Collaboration risk-of-bias tool developed by the Clinical Advances Through Research and Information Translation (CLARITY) Group at McMaster University [23] (Table 3). Our assessment will specifically focus on determining the adequacy of blinding and whether the allocation sequence was adequately generated and concealed in the RCT (Table 3). We will also examine the frequency of loss to follow-up and missing data, selective outcome reporting, and any other potential source of bias on an outcome-specific basis (Table 3). If two or more domains are classified as being at high risk of bias, the overall risk of bias will be rated as high. We will subsequently use the overall risk of bias for assessment of the evidence certainty.

2.7. Preparation and transformation of outcome measures for analyses

The data analysis requires information for each of the outcomes, specifically the mean change as a measure of central tendency and the standard deviation (SD) as the measure of dispersion of the change from baseline. For publications that do not include such estimates, we will use the following strategies to prepare the data for analysis.

When the median and range or interquartile range are available and when it is reasonable to assume a normal distribution, we will follow the method of Wan et al [24] to calculate the mean and SD. If the median (m), minimum (a), maximum (b), and sample size (n) are given, we will calculate the mean and SD as follows:

$$\text{Mean} \approx \frac{a+2m+b}{4}$$

$$\text{SD} \approx \frac{b-a}{2\Phi^{-1}\left(\frac{n-0.375}{n+0.25}\right)}$$

We will estimate the mean and SD with the following formulas when the first (q_1) and third (q_3) quartiles are also provided:

$$\text{Mean} \approx \frac{a+2q_1+2m+2q_3+b}{8}$$

$$\text{SD} \approx \frac{b-a}{4\Phi^{-1}\left(\frac{n-0.375}{n+0.25}\right)} - \frac{q_3-q_1}{4\Phi^{-1}\left(\frac{0.75n-0.125}{n+0.25}\right)}$$

When q_1 and q_3 are available but the minimum and maximum are not available, we will use the following formulas:

$$\text{Mean} \approx \frac{q_1+m+q_3}{3}$$

$$\text{SD} \approx \frac{q_3-q_1}{2\Phi^{-1}\left(\frac{0.75n-0.125}{n+0.25}\right)}$$

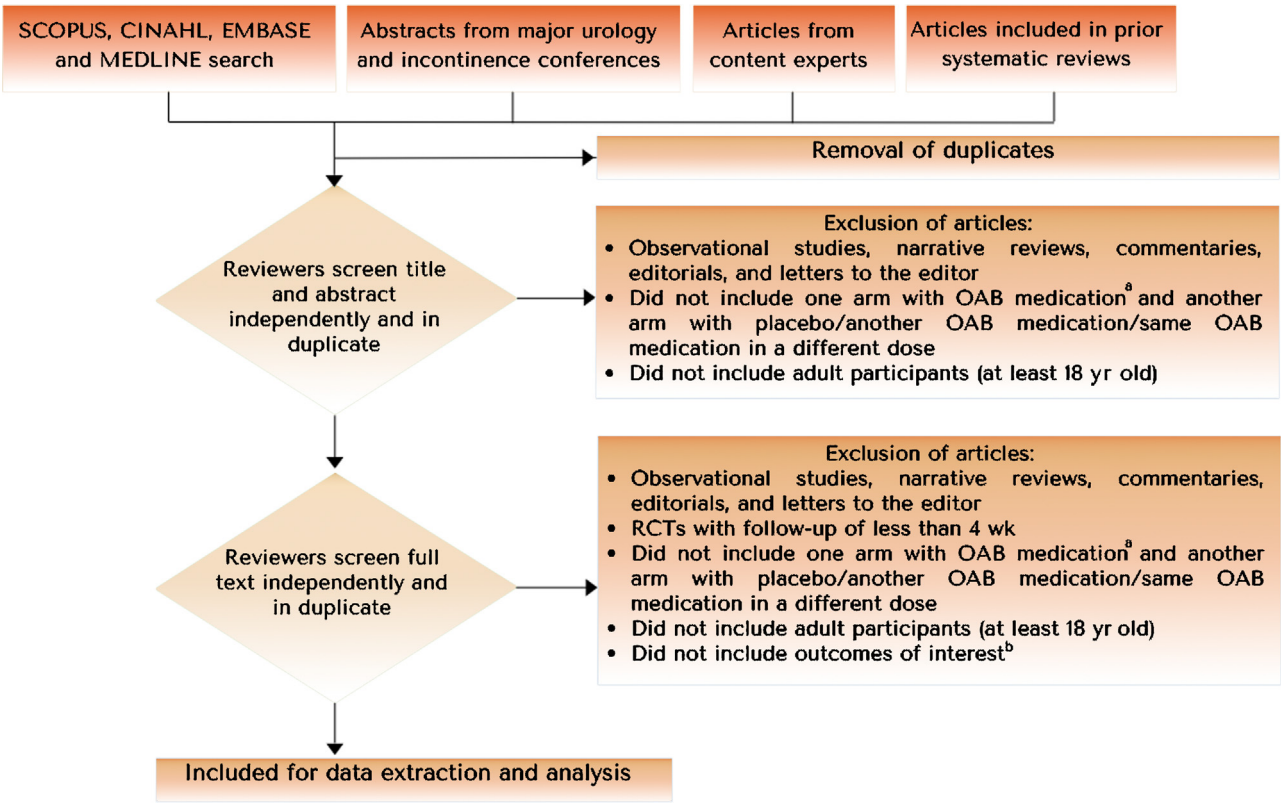


Fig. 2 – Study flow chart. OAB = overactive bladder symptoms; RCT = randomized controlled trial. [a] OAB medication includes oxybutynin, tolterodine, propiverine, trospium, solifenacin, darifenacin, fesoterodine, emepronium, propantheline, imidafenacin, mirabegron, solabegron, and vibegron. [b] Outcomes of interest include episodes of urinary urgency, urinary urgency incontinence, any type of incontinence, average times of voiding, nocturia, and adverse events.

Table 3 – Risk of bias assessment for randomized trials on medications for overactive bladder symptoms

Domain	Description
Allocation sequence	Process by which participants are assigned to different treatment groups **Examples of risk of bias:** Low: Random number table or computer generator High: Rules based on dates or record numbers
Allocation concealment	Procedures used to prevent individuals from knowing in advance the participants' treatment group assignation **Examples of risk of bias:** Low: Centralized computerized randomization, or consecutive numbered, opaque, sealed envelopes High: Unsealed, transparent, or nonconsecutive labeled envelopes
Blinding	Technique used to prevent individuals from knowing which treatment group a participant has been assigned to **Examples of risk of bias:** Low: Complete blinding of participants and trial personnel High: No blinding, or the blinding protocol could have been broken
Loss to follow-up [a]	Inability to collect outcome data from a subset of participants who were initially enrolled in the trial **Examples of risk of bias:** Low: No missing data or proper use of imputation techniques High: Improper use of imputation techniques or missing outcomes that impact the effect estimate
Selective outcome reporting	Practice of reporting only a subset of the outcomes measured in a trial **Examples of risk of bias:** Low: The protocol is accessible and all relevant predetermined outcomes have been reported as originally intended High: Results do not include all the predetermined primary outcomes or include primary outcomes that were not originally designated for the trial, or primary outcomes were analyzed using tools, methodologies, or data subsets that were not initially specified
Other sources of risk of bias [a]	Presence of any other flaws in the design, development, or analysis that may systematically affect the validity of the trial results **Examples of risk of bias:** Low: Voiding diary completed for ≥3 d at each time point or adverse events assessed at each time point whether participants mentioned or not High: Terminated prematurely for a data-dependent process or demonstrated major disparities in the baseline characteristics

[a] Assessment by type of outcome (benefits and harms),

When only the within-group standard error (SE) is available, we will calculate the SD by multiplying the square root of the total number of participants (N) by the SE [25]:

$$SD = \sqrt{N} \times SE$$

If only the within-group confidence interval (CI) is available, we will calculate the SD in two steps. First, SE will be calculated by subtracting the lower CI limit from the upper CI limit, divided by two times the z or t value. We will choose either the z or t value depending on their availability in the publication. Second, we will determine the SD by multiplying the square root of the total number of participants by the SE calculated in step 1 [25]:

$$SE = \frac{\text{Upper CI limit} - \text{lower CI limit}}{2 \times (z\text{value})} \text{ or SE} = \frac{\text{Upper CI limit} - \text{lower CI limit}}{2 \times (t\text{ value})}$$

$$SD = \sqrt{N \times SE}$$

These two formulas are limited to the availability of the z or t value. If these values are not available, we will follow the recommendations of the *Cochrane Handbook*. In this case, the SD is derived by dividing the width of the CI by 5.15 (for 99% CI), 3.92 (for 95% CI), or 3.29 (for 90% CI), and then multiplying the result by the square root of the total number of participants [26]:

$$SD = \sqrt{N} \times (\text{upper CI limit} - \text{lower CI limit}) / 3.92$$

When only pre- and post-measures are available, we will calculate the mean change by subtracting the baseline mean from the final mean. In this case, we will determine the SD of the mean change using a correlation coefficient (CC) [27]. We will assume a moderate CC of 0.5:

$$\text{Mean change} = \text{Final mean} - \text{baseline mean}$$

$$SD = \sqrt{\text{Baseline SD}^2 + \text{final SD}^2 - (2 \times CC \times \text{baseline SD} \times \text{final SD})}$$

When the mean change and pre- and post-measures are available, we will calculate the missing SD for the mean change assuming a CC of 0.5 using the following formula [27]:

$$SD = \sqrt{\text{Baseline SD}^2 + \text{final SD}^2 - (2 \times CC \times \text{baseline SD} \times \text{final SD})}$$

When there are no pre- and post-measures, measure of dispersion, or correlation coefficient reported, we will replace the missing SD with the median SD from all the RCTs that have the same outcome and arm [27]. First, we will select the SD from all the RCTs that have the same outcome and arm. Then, we will calculate the median for those selected SDs. Finally, we will treat the median SD as if it was observed in the RCTs.

3. Statistical methods

We will use the same statistical methods for all articles in our series of systematic reviews and meta-analysis articles. For each outcome, we will conduct random-effects NMA using a frequentist framework and a graph theory model [28]. Table 4 shows the main concepts for NMA [29]. We will perform a separate NMA for each outcome. Each medication will be represented by a separate node. We will include all doses and administration routes for the same medication in a single treatment node. When an intervention includes more than one active medication, it will be included as a separate node. When an intervention includes one active medication and placebo, it will be included within the same node for the active medication.

For efficacy outcomes (urinary urgency, urinary urgency incontinence, any type of incontinence, average times of voiding, and nocturia), we will report each estimate as a mean difference and corresponding 95% CI. For safety outcomes, we will report the absolute risk difference (RD) and corresponding 95% CI. We will use automated generation of node-splitting models to assess local incoherence and to obtain indirect estimates. We will report the direct, indirect, and network estimates for each comparison for each outcome. We will present the risk of bias results, forest plots for all direct comparisons, and a network plot for each outcome [30]. We will refrain from generating and reporting treatment rankings, such as p scores. Instead, we will use a grouping approach for the medications in categories from the most to the least effective and from the least to the most harmful, as described in Section 5 [31]. We will use the *Netmeta* package in R v4.0.0 (RStudio, Boston, MA, USA) to conduct our meta-analyses [32].

3.1. Sensitivity analyses

We will impute data for studies with missing outcome data [27,33]. We will impute values by replacing them with data from studies with similar characteristics as described in Section 2.7. We will conduct a sensitivity NMA by excluding imputed data to test the impact of imputation on the overall findings.

We will assess the width of the CI for all the estimates in each comparison. If we detect that the CI for the network

Table 4 – Network meta-analysis concepts for the systematic reviews [29]

Concept	Components
NMA assumptions	Transitivity • If two OAB medications have been compared in separate studies with a common comparator, indirect comparisons between these medications will be made by linking the two studies through the common comparator. • The indirect comparisons will be free of bias if the effect modifiers of the direct comparisons have a similar distribution. Consistency • Degree of agreement in the treatment estimates across different OAB trials.
Network plot	The geometry of our network will comprise the following elements: • Node color: type of OAB medication • Node size: number of participants randomized to each medication • Edge thickness: number of studies comparing two medications

OAB = overactive bladder

estimate is notably broader in comparison to the direct-estimate sparse network for one or more comparisons, we will infer that the assumption of common between-study heterogeneity across the network is not viable. Therefore, we will conduct a sensitivity analysis using frequentist fixed-effect models. In this case, the sensitivity analysis will assume the role of the principal analysis [34].

3.2. Subgroup analyses

We will assess the inconsistency of the direct estimates by looking at the similarity of the point estimates and the overlap of the CI. This assessment will be complemented by estimating the variance between studies using the I^2 statistic and the χ^2 test for heterogeneity. While lower I^2 values will raise only minimal concerns, higher I^2 values will increase concerns. In case of substantial inconsistency, we will explore the possible reasons for inconsistency via subgroup analyses.

We have identified the following potential effect modifiers [35,36]: (1) type of medication: antimuscarinics or β3 agonist (effect expected to be greater with antimuscarinics); (2) funding: industry-funded or not industry-funded (effect expected to be greater in industry-funded studies); and (3) end of follow-up time: short-term or long-term. We will conduct subgroup analyses for these effect modifiers. We will use the ICEMAN tool to determine the credibility of these analyses [37]. If these subgroups analyses do not reveal an explanation for substantial inconsistency, we will rate down for inconsistency when rating the certainty of the evidence. Moreover, we considered age as a potential effect modifier and conducted a survey among urologists who frequently treat patients with OAB symptoms (Supplementary material). However, owing to the variability in the results and the lack of consensus, we decided not to incorporate age as a potential effect modifier.

3.3. Secondary analysis

We will conduct a secondary analysis to determine whether different administration routes and doses yield varying effects. In this secondary analysis, each dose and administration route will be represented by a separate node in the network. We will include doses used in clinical practice around the globe. To gain insights into the doses used in clinical practice, we will examine clinical practice guidelines from major urological associations (including the AUA/Society of Urodynamics, Female Pelvic Medicine & Urogenital Reconstruction, EAU, and National Institute for Health and Care Excellence) and official documentation from drug regulators. We will also seek input from urologists across all continents identified via a convenience sample by surveying them regarding their clinical practices (covering all the doses identified in the RCTs).

4. Certainty-of-evidence assessment

The Grading of Recommendations Assessment, Development, and Evaluation (GRADE) Working Group provides guidelines and tools for assessing the certainty of the evidence (also known as quality of evidence) [30,34,38]. We will use the GRADE approach for rating the overall certainty of the evidence at the outcome level for each of the comparisons and each of the estimates (direct, indirect, and network) [30]. As an initial step, we will assess the risk of bias, publication bias, inconsistency, and indirectness for all direct estimates.

We will establish the certainty of the indirect evidence for each comparison on the basis of the most dominant first-order loop and intransitivity. The most dominant first-order loop is defined as the loop with "only one common comparator between the two medications being compared" and contributes the most to the indirect estimate [30]. Ratings for the certainty of indirect evidence will be the lowest of the certainty ratings for the direct comparisons with the most dominant first-order loop [30].

We will use the direct and indirect estimates to calculate the network estimates according to the GRADE guidance and using automated sheets [30]. If the direct estimate has a high certainty rating and contributes the most to the network estimate, we will not rate the indirect estimates [39]. First, we will choose the rating of evidence that contributes the most to the network estimate [30]. Second, we will assess for incoherence as explained in Section 3.2. Third, we will analyze imprecision on the basis of upper and lower CI limits for the network estimates [40]. Using a minimally contextualized approach, we have selected a minimally important difference (MID) value for each benefit outcome as our threshold (Table 5) [41]. If the CI crosses any of the selected thresholds, we will rate down for imprecision [40]. If the CI does not cross the selected thresholds, we will evaluate the effect size and the optimal information size (OIS) according to the GRADE guidance [40]. If the effect size is modest or the OIS is met, we will not rate down for imprecision. Then we will establish the final certainty of the evidence for each comparison using the automated tool from the GRADE Working Group [30].

Table 5 – MID values for benefit outcomes [a]

Outcome	MID
Daily urinary urgency episodes	−1.7
Daily urinary urgency incontinence episodes	−1.0
Any type of daily urinary incontinence episodes	−1.0
Average times of voiding per day	−1.5
Nocturia episodes	−1.0

MID = minimally important difference.

[a] We took MID data for urinary urgency, urinary urgency incontinence, and urinary frequency (average times of voiding per day) from an international multicenter study [41]. Using data from the same international multicenter study and a population-based study [18], we used a comprehensive discussion and consensus building to estimate MID values for any type of daily urinary incontinence and nocturia.

5. Interpretation of results

Once we complete assessment of the certainty of the evidence for each comparison, we will draw our conclusions according to the minimally contextualized approach [31]. As the initial step, we will select the most connected node within the network as the reference intervention [31]. In case two or more nodes are highly connected, we will select the medication with higher certainty as the reference inter-

vention. We hypothesize that the most connected intervention in our network will be placebo. We will use the same thresholds established for judging imprecision as our decision thresholds.

We will develop the first classification based on a comparison with the reference intervention [31]. If the 95% CI crosses the decision threshold, we will place the intervention in the same group of the reference intervention. Conversely, if the 95% CI does not cross the decision threshold, we will classify the intervention as more or less effective/harmful than the reference intervention. We expect to organize all the medications included in our NMA in the following categories: medications less effective than the reference intervention; medications similar to the reference intervention; and medications more effective than the reference intervention. For safety outcomes, we will classify the medications as follows: medications more harmful than the reference intervention; medications less harmful than the reference intervention; and medications similar to the reference intervention.

We will develop a second classification based on comparisons of the pairs of medications within each of the categories established during the first classification [31]. We will again use the same decision threshold described above. If a medication is more effective than another medication within the same category, the more effective medication will be moved to a higher category. We will apply the same procedure to distinguish between medications within each category until we are not able to establish further groupings. We will subsequently use the level of certainty of the evidence for each medication in comparison to the reference intervention to divide the intervention into two groups: (1) high or moderate certainty; and (2) low or very low certainty [31]. As the final step, we will review the pairwise comparisons not previously evaluated, checking for consistency with the rest of the comparisons [31]. We will present the final interpretation of the results as a summary table.

6. Summary

Despite numerous RCTs and systematic reviews investigating OAB medications, there is a lack of comprehensive analyses that could inform evidence-based recommendations for OAB treatment. As prior systematic reviews on this topic have not included all the RCTs available for OAB treatments, our aim is to fill this gap by conducting comprehensive systematic reviews and NMAs encompassing all published RCTs worldwide.

Our study methods adhere to rigorous standards, including PRISMA-P and GRADE guidance. The comprehensive literature search spanning multiple electronic databases, conference abstracts, information from content experts, and published systematic reviews is a strength of our study that is aimed at minimizing publication bias. The meticulous study selection and data extraction processes, including the use of standardized forms and pilot testing, should enhance the reliability of our results. We carefully defined our eligibility criteria to encompass the various medications, doses, and formulations used most frequently in OAB management. The inclusion of both anticholinergic and β3 agonist medications, along with consideration of different patient populations, enhances the applicability of our findings. In addition, the outcomes selected focus on those directly relevant and important to patients.

Our statistical methods provide a robust framework for synthesizing evidence. On the basis of previous studies and simulations, frequentist and Bayesian frameworks and their models for NMAs show little to no significant difference in their performance, with overlapping results in most circumstances [28]. Therefore, we have selected a frequentist approach. Similarly, we will not report treatment rankings because these rankings have several limitations [42]. Specifically, they fail to include the magnitude of differences between the interventions in their calculation, present challenges when the NMA includes multiple relevant outcomes, and, most importantly, do not consider the evidence certainty [42]. Instead, we will interpret our results using the minimally contextualized approach. This approach offers a comprehensive evaluation of the results and their certainty, while making the results easier to understand and implement.

We have considered sensitivity analyses to assess the impact of missing data and secondary analyses to assess the influence of follow-up time and administration routes and doses on our results. To the best of our knowledge, this is one of the first, if not the first, systematic review of OAB to conduct these types of analyses. Our methods also include strategies to deal with missing data. Many previous systematic reviews have only analyzed the data that were available, without addressing missing data. We will need to use substitute values from the same outcome and arm when no measure of dispersion is available. While widely used, this strategy neglects to account for the uncertainty for both imputed values and results [27]. By conducting data imputation and sensitivity analyses for missing data, we aim to evaluate how the patient-important outcomes might change. The inclusion of subgroup analyses by medication type and funding source adds further depth to our findings.

There will be limitations to our reviews. These limitations are likely to reflect the weaknesses of the eligible studies. First, many RCTs have relatively short follow-up periods, which may not capture long-term efficacy or safety concerns associated with OAB medications. Chronic conditions such as OAB may require longer observation periods to assess the sustainability of treatment effects and to monitor for potential adverse events that could emerge over time. Second, although we have created a very comprehensive list of adverse events, RCTs may not fully capture the spectrum of potential harms of the treatment in real-world settings. Third, certain populations, such as patients with cognitive impairment, those with complex medical histories, and individuals of ethnic minority status, may be under-represented in the RCTs. This can limit the applicability of trial results to these groups, who may have unique treatment considerations or responses.

In conclusion, the aim of our series of systematic reviews and NMAs is to contribute valuable insights into the effectiveness and safety of first-line OAB medications. Our comprehensive assessment will support evidence-based choices

in addressing OAB that could lead to enhanced outcomes and better quality of life for patients. Our research findings have the potential to influence health care policies, advances in clinical epidemiology methods, and forthcoming studies in functional urology.

Author contributions: Henk van der Worp, Angie K. Puerto Nino and Kari A.O. Tikkinen had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: van der Worp, Puerto Nino, Blanker, Brignardello-Petersen, Guyatt, Tikkinen.
Acquisition of data: van der Worp, Puerto Nino, OAB Investigators.
Analysis and interpretation of data: van der Worp, Puerto Nino, Blanker, Brignardello-Petersen, Guyatt, Tikkinen.
Drafting of the manuscript: van der Worp, Puerto Nino, Tikkinen.
Critical revision of the manuscript for important intellectual content: van der Worp, Puerto Nino, Brignardello-Petersen, Blanker, Guyatt, Tikkinen.
Statistical analysis: van der Worp, Puerto Nino, Blanker, Brignardello-Petersen, Guyatt, Tikkinen.
Obtaining funding: Tikkinen.
Administrative, technical, or material support: None.
Supervision: Blanker, Tikkinen.
Other: None.

***Financial disclosures*:** Angie K. Puerto Nino and Kari A.O. Tikkinen certify that all conflicts of interest, including specific financial interests and relationships and affiliations relevant to the subject matter or materials discussed in the manuscript (eg, employment/affiliation, grants or funding, consultancies, honoraria, stock ownership or options, expert testimony, royalties, or patents filed, received, or pending), are the following: None.

Funding/Support and role of the sponsor: This work is supported by the Sigrid Jusélius Foundation, the Competitive Research Funding of Helsinki University Hospital (TYH2020248;TYH2022330, TYH2023236), and the Vyborg Tuberculosis Foundation. The sponsors had no direct role in the study.

Acknowledgments: We thank Drs. Behnam Sadeghirad and Karl-Erik Andersson for their advice on the meta-analysis framework and OAB medications.

Ethics approval and consent to participate: Not applicable.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.euros.2024.08.006.

References

[1] Eapen RS, Radomski SB. Review of the epidemiology of overactive bladder. Res Rep Urol 2016;8:71–6. https://doi.org/10.2147/RRU.S102441.
[2] Vaughan CP, Johnson TM, Ala-Lipasti MA, et al. The prevalence of clinically meaningful overactive bladder: bother and quality of life results from the population-based FINNO study. Eur Urol 2011;59:629–36. https://doi.org/10.1016/j.eururo.2011.01.031.
[3] Durden E, Walker D, Gray S, Fowler R, Juneau P, Gooch K. The economic burden of overactive bladder (OAB) and its effects on the costs associated with other chronic, age-related comorbidities in the United States. Neurourol Urodyn 2018;37:1641–9. https://doi.org/10.1002/nau.23513.
[4] Irwin DE, Mungapen L, Milsom I, Kopp Z, Reeves P, Kelleher C. The economic impact of overactive bladder syndrome in six Western countries. BJU Int 2009;103:202–9. https://doi.org/10.1111/j.1464-410X.2008.08036.x.
[5] Gomes CM, Averbeck MA, Koyama M, Soler R. Impact of OAB symptoms on work, quality of life and treatment-seeking behavior in Brazil. Curr Med Res Opin 2020;36:1403–15. https://doi.org/10.1080/03007995.2020.1760806.
[6] Przydacz M, Chlosta M, Dudek P, Chlosta P. Effects of overactive bladder on treatment-related behaviour and quality of life in an Eastern European country: findings from the LUTS POLAND Study. Arch Med Sci 2021:1–11. https://doi.org/10.5114/aoms/133119.
[7] Tikkinen KAO, Tammela TLJ, Rissanen AM, Valpas A, Huhtala H, Auvinen A. Is the prevalence of overactive bladder overestimated? A population-based study in Finland. PLoS One 2007;2:e195. https://doi.org/10.1371/journal.pone.0000195.
[8] Suskind AM. The aging overactive bladder: a review of aging-related changes from the brain to the bladder. Curr Bladder Dysfunct Rep 2017;12:42–7. https://doi.org/10.1007/s11884-017-0406-7.
[9] Bourcier AP, Juras JA. Behavioral modification and conservative management of overactive bladder and underactive bladder disorders. In: Martins FE, Holm HV, Sandhu J, McCammon KA, editors. Female genitourinary and pelvic floor reconstruction. Cham, Switzerland: Springer; 2022. https://doi.org/10.1007/978-3-030-71112-2_13-1.
[10] Imamura M, Williams K, Wells M, Mcgrother C. Lifestyle interventions for the treatment of urinary incontinence in adults. Cochrane Database Syst Rev 2015;2015:CD003505. https://doi.org/10.1002/14651858.CD003505.pub5.
[11] Harding CK, Lapitan MC, Arlandis S, et al. EAU guidelines on management of non-neurogenic female lower urinary tract symptoms. Arnhem, The Netherlands: European Association of Urology; 2024.
[12] Cornu JN, Gacci M, Hashim H, et al. EAU Guidelines on non-neurogenic male lower urinary tract symptoms (LUTS). Arnhem, The Netherlands: European Association of Urology; 2024.
[13] National Institute for Health and Care Excellence. Urinary incontinence and pelvic organ prolapse in women: management. NICE guideline NG123. London, UK: NICE; 2019. https://www.nice.org.uk/guidance/ng123.
[14] Cameron AP, Chung DE, Dielubanza EJ, et al. The AUA/SUFU guideline on the diagnosis and treatment of idiopathic overactive bladder. J Urol 2024;212:11–20. https://doi.org/10.1097/JU.0000000000003985.
[15] National Institute for Health and Care Excellence. Lower urinary tract symptoms in men: management. NICE guideline CG97. London, UK: NICE; 2015 https://www.nice.org.uk/guidance/cg97.
[16] Moher D, Shamseer L, Clarke M, et al. Preferred Reporting Items for Systematic Review and Meta-Analysis Protocols (PRISMA-P) 2015 statement. Syst Rev 2015;4:1. https://doi.org/10.1186/2046-4053-4-1.
[17] Peyronnet B, Mironska E, Chapple C, et al. A comprehensive review of overactive bladder pathophysiology: on the way to tailored treatment. Eur Urol 2019;75:988–1000. https://doi.org/10.1016/j.eururo.2019.02.038.
[18] Agarwal A, Eryuzlu LN, Cartwright R, et al. What is the most bothersome lower urinary tract symptom? Individual- and population-level perspectives for both men and women. Eur Urol 2014;65:1211–7. https://doi.org/10.1016/j.eururo.2014.01.019.
[19] Zhang L, Zhu L, Xu T, et al. A population-based survey of the prevalence, potential risk factors, and symptom-specific bother of lower urinary tract symptoms in adult Chinese women. Eur Urol 2015;68:97–112. https://doi.org/10.1016/j.eururo.2014.12.012.
[20] Wadhwa S, Gupta A, Dokania S, Kanji R, Bagler G. A hierarchical anatomical classification schema for prediction of phenotypic side effects. PLoS One 2018;13:e0193959.
[21] Asimakopoulos AD, De Nunzio C, Kocjancic E, Tubaro A, Rosier PF, Finazzi-Agrò E. Measurement of post-void residual urine. Neurourol Urodyn 2016;35:55–7. https://doi.org/10.1002/nau.22671.
[22] DistillerSR Inc. DistillerSR version 2.35. https://www.distillersr.com/2023.

[23] CLARITY Group at McMaster University. Tool to assess risk of bias in randomized controlled trials. https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-randomized-controlled-trials-distillersr.
[24] Wan X, Wang W, Liu J, Tong T. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. BMC Med Res Methodol 2014;14:135. https://doi.org/10.1186/1471-2288-14-135.
[25] Chi KY, Li MY, Chen C, Kang E. Ten circumstances and solutions for finding the sample mean and standard deviation for meta-analysis. Syst Rev 2023;12:62. https://doi.org/10.1186/s13643-023-02217-1.
[26] Higgins JPT, Deeks JJ, editors. Data extraction for continuous outcomes. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0. London, UK: Cochrane Collaboration; 2011. Chapter 7.7.3.
[27] Higgins JPT, Deeks JJ, Altman DA, editors. Missing data. In: Higgins JPT, Green S, editors. Cochrane handbook for systematic reviews of interventions version 5.1.0. London, UK: Cochrane Collaboration; 2011. Chapter 16.1.
[28] Sadeghirad B, Foroutan F, Zoratti MJ, G, et al. Theory and practice of Bayesian and frequentist frameworks for network meta-analysis. BMJ Evid Based Med 2023;28:204–9. https://doi.org/10.1136/bmjebm-2022-111928.
[29] Puerto Nino AK, Brignardello-Petersen R. How to read a network meta-analysis. Eur Urol Focus 2023;9:701–4. https://doi.org/10.1016/j.euf.2023.10.018.
[30] Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. BMJ 2023;381:e074495.
[31] Brignardello-Petersen R, Florez ID, Izcovich A, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. BMJ 2020;371:m3900. https://doi.org/10.1136/bmj.m3900.
[32] R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
[33] Guyatt GH, Ebrahim S, Alonso-Coello P, et al. GRADE guidelines 17: assessing the risk of bias associated with missing participant outcome data in a body of evidence. J Clin Epidemiol 2017;87:14–22. https://doi.org/10.1016/j.jclinepi.2017.05.005.
[34] Brignardello-Petersen R, Murad MH, Walter SD, et al. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. J Clin Epidemiol 2019;105:60–7. https://doi.org/10.1016/j.jclinepi.2018.08.022.
[35] Zhu J, Hu X, Dong X, Li L. Associations between risk factors and overactive bladder: a meta-analysis. Female Pelvic Med Reconstr Surg 2019;25:238–46. https://doi.org/10.1097/SPV.0000000000000531.
[36] Chae J, Yoo EH, Jeong Y, Pyeon S, Kim D. Risk factors and factors affecting the severity of overactive bladder symptoms in Korean women who use public health centers. Obstet Gynecol Sci 2018;61:404–12. https://doi.org/10.5468/ogs.2018.61.3.404.
[37] Schandelmaier S, Briel M, Varadhan R, et al. Development of the Instrument to assess the Credibility of Effect Modification Analyses (ICEMAN) in randomized controlled trials and meta-analyses. Can Med Assoc J 2020;192:E901–6. https://doi.org/10.1503/cmaj.200077.
[38] Puhan MA, Schünemann HJ, Murad MH, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630. https://doi.org/10.1136/bmj.g5630.
[39] Brignardello-Petersen R, Bonner A, Alexander PE, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol 2018;93:36–44. https://doi.org/10.1016/j.jclinepi.2017.10.005.
[40] Brignardello-Petersen R, Guyatt GH, Mustafa RA, et al. GRADE guidelines 33: Addressing imprecision in a network meta-analysis. J Clin Epidemiol 2021;139:49–56. https://doi.org/10.1016/j.jclinepi.2021.07.011.
[41] Frankel J, Staskin D, Varano S, et al. Interpretation of the meaningfulness of symptom reduction with vibegron in patients with overactive bladder: analyses from EMPOWUR. Adv Ther 2022;39:959–70. https://doi.org/10.1007/s12325-021-01972-8.
[42] Mbuagbaw L, Rochwerg B, Jaeschke R, et al. Approaches to interpreting and choosing the best treatments in network meta-analyses. Syst Rev 2017;6:79. https://doi.org/10.1186/s13643-017-0473-z.