Received: 17 August 2023 | Accepted: 18 August 2023

DOI: 10.1002/nau.25273

CLINICAL ARTICLE

Neurourology Urodynamics | SUFU | WILEY

How should prospective research be designed to legitimately assess the value of urodynamic studies in female urinary incontinence?

Tufan Tarcan[1] | **Enrico Finazzi-Agrò**[2] | **Thomas M. Kessler**[3] | **Maurizio Serati**[4] | **Eskinder Solomon**[5] | **Peter F. W. M. Rosier**[6]

[1]Department of Urology, Marmara University School of Medicine and Koç University School of Medicine, Istanbul, Turkey

[2]Department of Surgical Sciences, University of Rome Tor Vergata and UOSD Urologia, Policlinico Tor Vergata, Rome, Italy

[3]Department of Neuro-Urology, Balgrist University Hospital, University of Zürich, Zürich, Switzerland

[4]Department of Obstetrics and Gynecology, University of Insubria, Varese, Italy

[5]Urology Centre, Guy's and St Thomas' NHS Trust, London, UK

[6]Department of Urology, University Medical Center Utrecht, Utrecht, The Netherlands

Correspondence
Tufan Tarcan, Department of Urology, Marmara University School of Medicine and Koç University School of Medicine, Istanbul, Turkey.
Email: bilgi@tufantarcan.com

Abstract

Aims: Since formal evidence demonstrating the value of urodynamic studies (UDS) in functional urology remains elusive, we aimed to consider how best to design robust research for this purpose in female urinary incontinence.

Methods: An expert group was convened to debate the following considerations: (a) precedents for formally proving the value of a gold standard diagnostic test, (b) key research principles, (c) defining a study population, (d) selecting endpoints, (e) defining interventional and controls arms, (f) blinding, (g) powering the study, and (h) duration of follow-up. In each case, we considered the strengths and weaknesses of different approaches in terms of scientific validity, ethical acceptability, practicality, and likelihood of bias.

Results: We agreed that unlike evaluating therapies, attempting to judge the value of a diagnostic test based on eventual treatment success is conceptually flawed. Nonetheless, we explored the design of a hypothetical randomized controlled trial for this purpose, agreeing that: (1) the study population must sufficiently reflect its real-world counterpart; (2) clinical endpoints should include not only continence status but also other lower urinary tract symptoms and risks of management; (3) participants in the interventional arm should receive individualized management based on their UDS findings; (4) the most scientifically valid approach to the control arm—empiric treatment—is ethically problematic; (5) sufficient statistical power is imperative; and (6) ≥ 2 years' follow-up is needed to assess the long-term impact of management.

Conclusions: Although a perfect protocol does not exist, we recommend careful consideration of our observations when reflecting on past studies or planning new prospective research.

KEYWORDS
female urinary incontinence, prospective research, urodynamics

1 | INTRODUCTION

More than 50 years ago, Bates and associates described the bladder as an "unreliable witness," emphasizing that, in some instances, the bladder was found to be empty when the patient claimed it to be full and vice versa.[1] Since lower urinary tract symptoms (LUTS) often do not reflect underlying pathologies, urodynamic studies (UDS) are recommended to achieve the most accurate diagnosis before potentially irreversible management is implemented.[1–3] (To note is that UDS includes both noninvasive and invasive investigations; in this article, our use of the term "UDS" refers to invasive UDS). Today, UDS are still the gold standard approach for objective evaluation of lower urinary tract dysfunction (LUTD),[4] both in research settings and in routine clinical practice to establish the diagnosis, guide selection of appropriate management and enable patient counseling about the likely clinical outcome.[4] Nonetheless, formal research evidence demonstrating the value of UDS in female urinary incontinence (UI), as well as in male LUTS, has long been desired.

As a starting point, we reflected on the necessity of seeking to formally prove the value of a gold standard diagnostic test like UDS in the first place. Once a diagnostic test has been shown to reliably detect underlying abnormalities, and cut-off values have been refined through epidemiological insights, achievement of gold standard status is usually gained simply through standing the test of time in routine clinical practice. There is very little precedent for attempting to prove the value of diagnostics (especially when gold standard) via randomized controlled trials (RCTs), which are much more suited to assessing the value of therapies. In daily urology practice, digital rectal examinations and cystoscopies remain a cornerstone of diagnostic work-ups, despite their lack of formal published evidence. Similarly, in the management of urologic cancer, various types of imaging modalities—requiring trained staff and costly techniques—are generated by radiologists without randomized prospective evidence for their value. The vast majority of studies evaluating diagnostics (e.g., clinical imaging or clinical chemistry) consist of retrospective analysis of subcohorts. It is also worth noting that the value of UDS in patients with neurogenic LUTD remains undisputed by the urology community and by guideline committees, even though the evidence underpinning this[5–7] does not include RCTs. In the case of nonneurogenic LUTD, the call for formal research to demonstrate the value of UDS appears motivated by practical reasons, that is, because undertaking UDS requires competent staff, incurs time and cost and can cause some patients discomfort or embarrasment—a scenario which is hardly unique in the diagnostics field.

Over a decade ago, the desire for formal proof of the value of UDS gave rise to two RCTs, the VALUE and VUSIS-II trials, which both focused on women with "uncomplicated" stress urinary incontinence (SUI). Both trials concluded that conducting preoperative UDS does not change the 12-month clinical outcome in this population and can therefore be safely skipped.[8,9] This conclusion resulted in revisions to societal guidelines for SUI management and prompted scrutiny of the role of UDS in general. However, the VALUE and VUSIS-II studies both attracted strong criticism of their methodological approaches, leading to a fierce and prolonged debate in the literature.[10–17] Moreover, a recent systematic review and meta-analysis of RCTs intended to assess the clinical value of UDS concluded that further well-designed trials are needed.[18]

2 | METHODOLOGY

Our aim for this article was to analyse the key considerations, challenges and pitfalls of designing research to legitimately assess the value of UDS in women with UI. A multidisciplinary expert group, including three urologists, a urogynecologist, a urodynamicist and a clinical scientist, was convened for this purpose. Despite the evident conceptual drawbacks of RCTs in the diagnostics setting, we set ourselves the challenge of considering the methodology for a hypothetical RCT, collectively exploring a range of study parameters, including patient eligibility criteria, designs of interventional and control arms, study power, endpoints, blinding, subjectivity in treatment decision-making, and duration of follow-up. In all cases, we assessed how different design choices might impact the validity and acceptability of the findings. To note is that, even though most of the examples and data we cite relate to SUI, we believe the principles broadly apply to all of female UI.

3 | RESULTS

3.1 | Key research principles and reflections on the existing literature

First and foremost, we agreed that assessing the value of a diagnostic test cannot be done in the same way as investigating the value of a treatment. With a new diagnostic test, value can be determined by comparing the test findings with those of a reference (gold standard) test, which has already proved its worth in determining a patient's pathophysiological status. However, since UDS

are themselves the gold standard in highlighting abnormal physiology, there is no available comparator apart from what is known about normal physiology. This predicament has led researchers to conclude that the only way forward is to assess whether greater health gain is achieved by management based on UDS findings compared to management based on tests which omit UDS. However, this approach poses a significant conceptual problem, since it fundamentally relies on the success of treatments in a field where treatment outcomes are rather variable and there are no "standard" treatments. In the case of female SUI, the recent banning of synthetic mid-urethral sling surgery in some countries has meant there isn't even a universal gold standard surgical approach. Hence, setting out to prove the value of a well-established diagnostic test like UDS is a profoundly challenging mission.

Reflecting on the existing literature, it is clear that diagnoses based on UDS findings often differ from provisional clinical diagnoses, even in female SUI patients considered "uncomplicated,"[19–21] but evidence that this correlates with better clinical outcomes has proved elusive. A Cochrane review[19] analyzed the available data for 1036 women in seven trials, of whom 526 received UDS. In three of these trials, women in the urodynamic arms were more likely to have their management changed than women in the control arm (17% and 3%, respectively). Yet no evidence was apparent that the modified management plans impacted post-treatment health outcomes—such as incontinence, quality of life or economic outcomes—compared with women who did not undergo UDS. This paradox rightly led the authors to conclude that larger and more definitive trials are needed, but it should also be noted that, in some of the trials included in this Cochrane review, surgery had already been chosen as the management approach, irrespective of UDS findings, and the review was unable to include the outcomes of patients not treated surgically.

In the VALUE study, it is notable that the clinical diagnosis was altered by preoperative UDS in as many as 56% of the participants.[8] There were two key reasons for this: UDS detected unsuspected voiding dysfunction in 10% of patients and also ruled out filling phase abnormalities in a third of patients for whom this had been clinically suspected owing to overactive bladder (OAB) syndrome. (To note is that voiding dysfunction and OAB are both clinical diagnoses; details of the corresponding urodynamic findings of detrusor underactivity and overactivity were not provided in the VALUE study publication, which also did not report how the change of clinical diagnosis affected preoperative patient counseling). Yet, despite the high number of diagnoses that were changed by UDS, this again did not translate into a difference in treatment success between the UDS arm and the control arm. In this case, this can hardly be deemed a paradox, since almost all patients in the VALUE study ended up with the same treatment (mid-urethral sling) whichever arm they were assigned to and irrespective of their UDS findings (see Section 3.4).

In our opinion, the most useful insight from the VALUE study was the detection of clinically unsuspected voiding dysfunction by UDS in 10% of patients.[22] The presence of voiding dysfunction, usually only identifiable with UDS, is significant as it predicts a higher rate of adverse postsurgical outcome.[10,23] Indeed, the VALUE study itself confirmed that the success of surgery within this subgroup was lower than that in the remaining study population (62.1% vs. 78.3%).[8,22] Although statistically borderline ($p = 0.064$)—likely due to the underpowering of the trial (only half of the trial population was included in the subgroup analysis)—this is a clinically significant difference and highlights the diagnostic and prognostic importance of UDS for decision-making and patient counseling in this subgroup. (A similar prevalence of voiding dysfunction in uncomplicated female SUI (13.4%) was found in a large database study conducted by Serati et al.,[15] who also noted this dysfunction in 22.5% of complicated cases). Clearly, careful trial design is imperative for future research aiming to elucidate the relationship between UDS and clinical outcomes.

3.2 | Defining the trial population and eligibility criteria

For real world applicability, the value of a diagnostic test like UDS should ideally be studied in a patient population that represents the full spectrum of a condition and its natural heterogeneity. The VALUE and VUSIS-II trials were criticized for restricting their study populations to uncomplicated (or index) female SUI patients, who represent only about one-third of the real-world female SUI population (according to the definition of "uncomplicated" used in these trials).[15,24,25] In addition, while authors of the VUSIS-II study have argued that the strength of their research was its focus on a "homogenous group of women with predominant SUI,"[6] others countered that, by aiming to exclude women with conditions such as low leak point pressure, low maximum urethral closure pressure, pelvic organ prolapse, previous failed surgery and voiding dysfunction, the VUSIS-II (and VALUE) studies focused on the subpopulation least likely to benefit from UDS.[26] Indeed, a large, multicentre database study ($n = 2053$)

subsequently demonstrated that UDS are indeed of greater diagnostic value in "complicated" versus "uncomplicated" female SUI patients.[15]

In an ideal world, participants for our hypothetical trial should be "all comers" with primary female UI (urgency, stress, mixed, and OAB), excluding only those with neurogenic dysfunction, evident anatomical/hormonal abnormalities or prior urological surgery. On the other hand, a broad and heterogeneous study population creates methodological challenges. Some individuals will be much more obvious candidates for UDS than others and—whatever form the control arm takes—it would also be more difficult to ensure the interventional and control arms are well matched. There are therefore arguments for defining the trial population more tightly than an "all comers" approach—SUI or OAB only, for example—but not to the point that it bears little resemblance to its real-world counterpart.

3.3 | Selecting endpoints

Since most general urologists want to know whether adding UDS into the patient pathway makes a difference to clinical outcomes, previous trials investigating the value of UDS have used continence as their endpoint. Notably however, this has been studied only over the short term (see Section 3.8) and little attention has been paid to other potentially undesirable outcomes of management such as voiding difficulties, storage dysfunction or urinary tract infections. More significantly, there is—as already stated—a conceptual problem in judging the value of a diagnostic test based on the outcome of a treatment. Since the aim of performing UDS is to guide management decision-making, it seems more logical to judge UDS on its ability to influence treatment choice, rather than on eventual outcome. For example, research designed to prove the hypothesis that the pathophysiological insights provided by UDS leads to modification or even cancellation of surgical interventions, would be highly worthwhile and conceptually valid. If choice of management is our endpoint, two study arms are not needed, as each participant would simply be assessed to see if this choice changes after UDS compared to before.

Several single-arm prospective cohort studies of this nature have already been done. For example, a prospective study conducted in a US academic referral center ($n = 285$) reported that UDS findings changed treatment plans in 42.5% of cases, most commonly related to surgery (35.0%).[20] Another single-arm prospective study performed in a tertiary referral center ($n = 102$) reported that UDS resulted in a change in treatment plan in as many as 78% of patients.[21] It is worth noting that, in the diagnostics field, cohort studies are accepted as Level 1b evidence by the Oxford Classification of Evidence,[27] reflecting the view that, in many diagnostic scenarios, RCTs may be impractical, unethical and methodologically unfit to demonstrate potential benefits. Nonetheless, RCT supremacy is a deeply held perception, so, even when RCTs have significant limitations and methodological controversies, they tend to be held in higher esteem than well conducted cohort studies. In addition, clinical outcomes are instinctively valued above other endpoints. If clinical outcomes are to be the chosen endpoints, it's important that they include not only continence but also other LUTS (given that treatments can impact those), adverse events, disease-specific quality of life, and, ideally, costs.

3.4 | The interventional arm

It seems self-evident that the treatment approach for patients in the interventional arm of our hypothetical RCT should be selected, and tailored, according to their individual urodynamic profiles (with knowledge of clinical presentation and medical history). Thus, the most illogical aspect of the VALUE and VUSIS-II trials is that virtually all participants appeared pre-destined to receive a mid-urethral sling, irrespective of UDS findings. In the VALUE trial, over 90% of patients in both arms ended up with a mid-urethral sling, even though UDS-based diagnosis differed from the initial clinical diagnosis in over half of the interventional arm patients.[8]

The study protocols for VALUE and VUSIS-II stated that treatment should be decided "according to guidelines." This is also a problematic approach since functional urology guidelines guide decision-making based on clinical syndromes only, rather than on the specific pathophysiological dysfunctions (or combinations of these) that UDS can reveal in each individual.[28]

We have the same concern about the FUTURE study,[29] which is assessing the value of UDS in female OAB by randomizing participants to undergo UDS + clinical assessment or clinical assessment alone. The protocol states that "in both groups, women will be offered standard treatments for refractory OAB as per defined treatment pathways and in accordance with the national guidelines." Our worry is that—as with previous trials—patients in the interventional arm of this study will not receive individualized management that addresses the specific pathophysiologies identified by UDS because the practice standards provide no guidance for this and thus a true comparison between the interventional and control arms won't be achieved.

Currently, there are no best practice recommendations for how to treat each pathophysiology (or combination of pathophysiologies) revealed by UDS. Thus, a prerequisite for our hypothetical trial might be to give precise direction on this, matching pathophysiologies to management approaches. (Achievement of more precise staging and grading of LUTD and identification of defined diagnostic cohorts, with best practice management guidance for each, is indeed a desirable research mission in its own right). The alternative for the interventional arm of our trial is to leave participating clinicians free to treat the patients in this arm as they see fit, as long as their decision is clearly based on UDS findings. This would, of course, introduce greater subjectivity into the study, but the protocol could mandate recording of a rationale for the choice of treatment in each participant.

3.5 | The control arm

Assuming a control arm is felt to be needed, defining the control arm for our trial raises several thorny problems of its own. As we see it, there are three options:

- **Option A**: *Patients in the control arm would all be treated empirically (based on clinical presentation/office evaluation alone). Either UDS would not be done at all or physicians would be blinded to UDS results*. Even though empirical treatment for female UI remains all too common in everyday practice, we believe this option would not be acceptable to research ethics committees. The only way to make it more ethically acceptable is to restrict the study population to those where the use of UDS is less established. However, doing this risks making the trial findings less generalizable and could also undermine the research objective, as the study population would likely consist of less complicated patients, where UDS is already known to be less valuable. Ultimately the results of any study focusing on a specific subpopulation cannot be extrapolated beyond that subpopulation.
- **Option B**: *The control arm would consist of "typical practice" at each participating center, that is, probably doing UDS in some individuals and not in others, and treating according to local "habits."* This option would create a "real life" control arm which would likely be acceptable to ethics committees, but it would be subjective and potentially unrepresentative. The choice of participating centers and participating physicians could also skew the results as some are much more inclined to use UDS to guide treatment in routine practice than others. In addition, the choice of management based on UDS insights would be too variable to generate statistically reliable results, except in the case of very large-scale studies.
- **Option C**: *The comparator would be historical data on success rates in treating female urinary incontinence (i.e., the study could be a single-arm study, with no control arm or randomization)*. Although expert consensus about prevailing management success rates is stated in current clinical practice guidelines and elsewhere in the literature, the problem with this approach is that, wherever the historical data come from, they could be criticized as unrepresentative. Furthermore, the surgical learning curve and patient care advances over time would need to be taken into account. As previously discussed, a single-arm cohort study would also carry less status than an RCT, whether or not this view is justifiable.

Deciding the best approach to the control arm was probably the most contentious issue that our expert group faced. Despite extensive discussion, we were unable to reach a consensus on which approach would be best as all have notable drawbacks. Purely for the purposes of further academic debate (e.g., on topics like blinding), the rest of this article assumes that a control arm would be included in our hypothetical trial.

3.6 | Blinding

In an effort to create a truly legitimate comparison between UDS-based management and non-UDS-based management, appropriate blinding is clearly important. Although it is impractical to blind physicians to a patient's allocated arm or to the results of UDS in the interventional arm (given that treatment choice in the interventional arm needs to be based on UDS findings), patients themselves could conceivably be blinded to both. With regard to the control arm, we believe it's imperative that physicians be blinded to UDS results to avoid the risk of being influenced by them. In the VUSIS-II trial, for example, where the control arm patients were all allocated to undergo surgery, the authors conceded that lack of physician blinding to UDS results could have affected the type of surgery selected and even the surgical technique (e.g., tensioning of the sling). Similarly, in the VALUE study—where physicians were also unblinded—the initially chosen approach to midurethral sling surgery (retropubic or transobdurator) changed from one to the other in 18 patients based on UDS findings,[8] even though the protocol design intended UDS results to be disregarded. In our hypothetical trial—if randomized—blinding should be used wherever practical to ensure that

the interventional arm is treated according to the pathophysiologies identified by UDS whereas the control arm is not.

3.7 | Powering the trial

Ensuring sufficient statistical power is another key concern. The INVESTIGATE 1 study—a feasibility pilot for a future RCT intended to assess the value of UDS for patients with SUI or stress-predominant MUI[30]—showed that, based on UDS insights, 15% fewer patients would be assigned to undergo surgery than if based on clinical evaluation alone. However, this pilot study was not powered to show significance and the authors concluded that 450 patients would be needed in each arm to demonstrate outcome superiority, which is more than the populations of the VALUE and VUSIS-II studies combined. In addition, as noted earlier, the VALUE trial reported a clear but nonstatistically significant ($p = 0.064$) trend towards a poorer outcome of surgery in the subgroup with urodynamically detected voiding dysfunction than the rest of the study population. Had the trial been adequately powered, this finding could have been statistically confirmed.[31,32]

3.8 | Duration of follow-up

As stated earlier, previous RCTs assessing the value of preoperative UDS have opted for midurethral sling (either via retropubic or transobturator route) as the 'default' surgical approach[8,9] but, according to the Cochrane review, no study has followed up participants over the long term.[19] In both the VALUE and VUSIS-II trials, outcomes were not monitored beyond 12 months. In our view, any duration of follow-up that is less than 2 years should be considered short-term. Logistic regression has shown that women with mid- and long-term follow-up were around half as likely as their short-term counterparts to report treatment success after mid-urethral sling procedures, where "short term" was defined as any follow-up less than 3 years.[33]

In addition, when the definition of surgical success was expanded in the VALUE trial to include a negative stress test at a bladder volume of 300 mL at 12 months, the success rates were only around 70% in both arms.[8] Although these success rates are within the range expected in current clinical practice, they are not particularly impressive and raise further questions, both about the selection of patients for surgery and the durability of surgical success. It's also important to note that, following the banning of mid-urethral synthetic sling surgery in some countries due to safety concerns with the permanent mesh implants,[34] new surgical techniques for female SUI are emerging, and/or old techniques re-emerging, where there is less familiarity with their outcomes. For all these reasons, long-term follow-up (no less than 2 years) is important for our hypothetical trial.

4 | CONCLUSIONS

- Evaluation of diagnostic tests like UDS cannot be done in the same way as evaluation of treatments and is especially challenging when the test is a gold standard diagnostic modality. Judging the value of a diagnostic test according to the outcome of a treatment is conceptually problematic and requires very careful trial design to minimize the numerous pitfalls and confounding factors. The RCT route, in particular, is fraught with issues in the diagnostics field.
- The patient population, or subpopulation, for a trial designed to assess the value of UDS should be sufficiently reflective of its real-world counterpart.
- If endpoints are clinical outcomes, they should not only include continence status but also other LUTS, as well as risks and costs of management, and the trial should have sufficient statistical power.
- Patients in the interventional arm should receive individualized management based on their UDS findings and not pre-assigned to a treatment as previous trials have done, nor treated 'according to guidelines', which currently do not allow for individualization based on UDS-detected pathophysiology.
- The most scientifically valid approach to the control (non-UDS) arm would be to assign patients to empiric treatment but this is unlikely to be acceptable to ethics committees.
- The follow-up duration should be long enough (at least 2 years) to assess the long-term impact of management, including the detection of new or worsened voiding symptoms or increased voiding frequency that patients may find difficult or bothersome.

Due to all the methodological issues we have outlined, we conclude that a perfect RCT protocol for assessing the value of UDS in female UI does not exist. Nevertheless, much better studies than those that currently exist are certainly achievable and many research groups and clinicians have called for them. It is our hope that future researchers wishing to take up this formidable challenge will heed our advice when embarking on their trial designs.

AUTHOR CONTRIBUTIONS

All authors contributed to the debate and discussions; the manuscript was written by Tufan Tarcan and Peter F. W. M. Rosier; all authors reviewed and contributed comments on the draft.

ACKNOWLEDGMENTS

We thank Karen Lipworth for writing and editorial support.

CONFLICT OF INTEREST STATEMENT

Tufan Tarcan is a speaker/adviser for Recordati, Astellas, and Laborie. Enrico Finazzi-Agro is a speaker/advisor for Recordati, Pierre Fabre, and Laborie. All the other authors have no conflict of interest.

DATA AVAILABILITY STATEMENT

Not applicable.

ORCID

Tufan Tarcan http://orcid.org/0000-0002-3387-3524
Enrico Finazzi-Agrò http://orcid.org/0000-0002-0308-8824
Thomas M. Kessler http://orcid.org/0000-0002-1991-5919
Maurizio Serati http://orcid.org/0000-0002-8534-646X
Eskinder Solomon http://orcid.org/0000-0003-4382-0394
Peter F. W. M. Rosier http://orcid.org/0000-0003-0445-4563

REFERENCES

1. Bates CP, Whiteside CG, Turner-Warwick R. Synchronous cine-pressure-flow-cysto-urethrography with special reference to stress and urge incontinence. *Br J Urol*. 1970;42(6):714-723.
2. Blaivas JG. The bladder is an unreliable witness. *Neurourol Urodyn*. 1996;15(5):443-445.
3. Serati M, Salvatore S, Siesto G, et al. Urinary symptoms and urodynamic findings in women with pelvic organ prolapse: is there a correlation? Results of an artificial neural network analysis. *Eur Urol*. 2011;60(2):253-260.
4. De Wachter S, Gammie A, Rosier PFWM, et al. Urodynamic testing. In: Cardozo L, Rovner C, Rovner E, Wagg A, Wein A, Abrams P, eds. *Incontinence*. 7th ed., 2023:487-550.
5. Panicker JN, Fowler CJ, Kessler TM. Lower urinary tract dysfunction in the neurological patient: clinical assessment and management. *Lancet Neurol*. 2015;14:720-732.
6. Musco S, Padilla-Fernández B, del Popolo G, et al. Value of urodynamic findings in predicting upper urinary tract damage in neuro-urological patients: a systematic review. *Neurourol Urodyn*. 2018;37(5):1522-1540.
7. https://uroweb.org/guidelines/neuro-urology
8. Nager CW, Brubaker L, Litman HJ, et al. A randomized trial of urodynamic testing before stress-incontinence surgery. *N Engl J Med*. 2012;366(21):1987-1997.
9. van Leijsen SAL, Kluivers KB, Mol BWJ, et al. Value of urodynamics before stress urinary incontinence surgery: a randomized controlled trial. *Obstetr Gynecol*. 2013;121:999-1008.
10. Finazzi-Agrò E, Serati M, Salvatore S, Del Popolo G. Comments on "A randomized trial of urodynamic testing before stress-incontinence surgery" (N Engl J Med. 2012 May 24;366(21):1987-1997) from the Italian Society of Urodynamics. *Neurourol Urodyn*. 2013;32:301-302.
11. Lose G, Klarskov N. Utility of invasive urodynamics before surgery for stress urinary incontinence. *Int Urogynecol J*. 2014;25:1-3.
12. Giarenis I, Cardozo L. What is the value of urodynamic studies before stress incontinence surgery? *BJOG: Int J Obstetr Gynaecol*. 2013;120:130-132.
13. Lose G, Klarskov N. Utility of invasive urodynamics before surgery for stress urinary incontinence: response to correspondence. *Int Urogynecol J*. 2014;25:1001.
14. Lose G, Klarskov N. Re: does preoperative urodynamics improve outcomes for women undergoing surgery for stress urinary incontinence? A systematic review and meta-analysis. *BJOG: Int J Obstetr Gynaecol*. 2015;122:1147-1148.
15. Serati M, Topazio L, Bogani G, et al. Urodynamics useless before surgery for female stress urinary incontinence: are you sure? Results from a multicenter single nation database: urodynamics useless before sui surgery in female. *Neurourol Urodyn*. 2016;35(7):809-812.
16. Finazzi-Agro E, Gammie A, Kessler TM, et al. Urodynamics useless in female stress urinary incontinence? time for some sense - a European expert consensus. *Eur Urol Focus*. 2020;6:137-145.
17. Serati M, Tarcan T, Finazzi-Agrò E, et al. The bladder is an unreliable witness: the case for urodynamic investigations in female stress urinary incontinence. *Eur J Obstet Gynecol Reprod Biol*. 2020;244(244):35-37.
18. Bodmer NS, Wirth C, Birkhäuser V, et al. Randomised controlled trials assessing the clinical value of urodynamic studies: a systematic review and meta-analysis. *Eur Urol Open Sci*. 2022;44:131-141.
19. Clement KD, Lapitan MC, Omar MI, Glazener CM. Urodynamic studies for management of urinary incontinence in children and adults. *Cochrane Database Syst Rev*. 2013;2013(10):003195.
20. Suskind AM, Cox L, Clemens JQ, et al. The value of urodynamics in an academic specialty referral practice. *Urology*. 2017;105:48-53.
21. Malik RD, Hess DS, Carmel ME, Lemack GE, Zimmern PE. Prospective evaluation of urodynamic utility in a subspecialty tertiary practice. *Urology*. 2019;126:59-64.
22. Sirls LT, Richter HE, Litman HJ, et al. The effect of urodynamic testing on clinical diagnosis, treatment plan and outcomes in women undergoing stress urinary incontinence surgery. *J Urol*. 2013;189(1):204-209.
23. Lose G, Klarskov N. Preoperative voiding dysfunction is a risk factor for operative failure according to the VALUE study! *Am J Obstet Gynecol*. 2016;215(1):128.
24. Serati M, Agrò EF. Urodynamics before surgery for stress urinary incontinence: the urodynamic examination is still one of the best friends of the surgeon and of patients with stress urinary incontinence. *Eur Urol Focus*. 2016;2:272-273.

25. Leandro A, Marta L, Gracia G, Ariel M, Pablo VJ, Bengió RG. Prevalence of "uncomplicated" and "complicated" stress urinary incontinence in Argentinian women: assessing the role of the urodynamic study. *Eur J Obstet Gynecol Reprod Biol*. 2021;256:466-470.
26. Del Popolo G, Li Marzi V, Serati M. Do preoperative urodynamics still have a role in female stress urinary incontinence? *Neurourol Urodyn*. 2013;32:1144-1145.
27. https://www.cebm.net/2009/06/oxford-centre-evidence-based-medicine-levels-evidence-march-2009
28. Daly JO, Cartwright R. Preoperative urodynamics—self evident or evidently unnecessary? *Nat Rev Urol*. 2015;12(10):539-540.
29. Abdel-Fattah M, Chapple C, Guerrero K, et al. Female urgency, trial of urodynamics as routine evaluation (FUTURE study): a superiority randomised clinical trial to evaluate the effectiveness and cost-effectiveness of invasive urodynamic investigations in management of women with refractory overactive bladder symptoms. *Trials*. 2021;22:745.
30. Hilton P, Armstrong N, Brennand C, et al. INVESTIGATE-I (INVasive Evaluation before Surgical Treatment of Incontinence Gives Added Therapeutic Effect?): a mixed-methods study to assess the feasibility of a future randomised controlled trial of invasive urodynamic testing prior to surgery for stress urinary incontinence in women. *Health Technol Assess*. 2015;19(15):1-273.
31. Finazzi Agro E, Iacovelli V, Illiano E, Costantini E. Urodynamics before surgery for stress urinary incontinence in female patients: an open debate. *Arch Esp Urol*. 2017;70(8):691-694.
32. Nager CW. Re: comments on "A randomized trial of urodynamic testing before stress-incontinence surgery" (N Engl J Med. 2012 May 24;366(21):1987-1997) from the Italian Society of Urodynamics. *Neurourol Urodyn*. 2013;32:303-304.
33. Malek JM, Kissane LM, Martin KD, Meyer I, Powell TC, Richter HE. Short-, mid-, and long-term incontinence outcomes in women undergoing mid-urethral sling procedures: a retrospective cohort study. *Int Urogynecol J*. 2021;32(3):609-614.
34. Chapple CR, Cruz F, Deffieux X, et al. Consensus statement of the European Urology Association and the European Urogynaecological Association on the use of implanted materials for treating pelvic organ prolapse and stress urinary incontinence. *Eur Urol*. 2017;72(3):424-431.

How to cite this article: Tarcan T, Finazzi-Agrò E, Kessler TM, Serati M, Solomon E, Rosier PFWM. How should prospective research be designed to legitimately assess the value of urodynamic studies in female urinary incontinence? *Neurourol Urodyn*. 2023;42:1639-1646. doi:10.1002/nau.25273