BMJ Open

Reliability and validity of assessment methods available in primary care for bladder outlet obstruction and benign prostatic obstruction in men with lower urinary tract symptoms: a systematic review

Tom Vredeveld ,[1,2] Esther van Benten ,[1,3] Rikie E P M Beekmans,[4] M Patrick Koops,[5] Johannes C F Ket ,[6] Jurgen Mollema ,[7] Stephan P J Ramaekers ,[2] Jan J M Pool ,[3] Michel W Coppieters ,[1,8] Annelies L Pool-Goudzwaard [1,9]

To cite: Vredeveld T, van Benten E, Beekmans REPM, *et al*. Reliability and validity of assessment methods available in primary care for bladder outlet obstruction and benign prostatic obstruction in men with lower urinary tract symptoms: a systematic review. *BMJ Open* 2022;**12**:e056234. doi:10.1136/bmjopen-2021-056234

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-056234).

Received 06 August 2021
Accepted 25 March 2022

For numbered affiliations see end of article.

Correspondence to
Tom Vredeveld;
t.vredeveld@vu.nl

ABSTRACT

Objectives To systematically review the literature regarding the reliability and validity of assessment methods available in primary care for bladder outlet obstruction or benign prostatic obstruction in men with lower urinary tract symptoms (LUTS).

Design Systematic review with best evidence synthesis.

Setting Primary care.

Participants Men with LUTS due to bladder outlet obstruction or benign prostatic obstruction.

Review methods PubMed, Ebsco/CINAHL and Embase databases were searched for studies on the validity and reliability of assessment methods for bladder outlet obstruction and benign prostatic obstruction in primary care. Methodological quality was assessed with the COSMIN checklist. Studies with poor methodology were excluded from the best evidence synthesis.

Results Of the 5644 studies identified, 61 were scored with the COSMIN checklist, 37 studies were included in the best evidence synthesis, 18 evaluated bladder outlet obstruction and 17 benign prostatic obstruction, 2 evaluated both. Overall, reliability was poorly evaluated. Transrectal and transabdominal ultrasound showed moderate to good validity to evaluate bladder outlet obstruction. Measured prostate volume with these ultrasound methods, to identify benign prostatic obstruction, showed moderate to good accuracy, supported by a moderate to high level of evidence. Uroflowmetry for bladder outlet obstruction showed poor to moderate diagnostic accuracy, depending on used cut-off values. Questionnaires were supported by high-quality evidence, although correlations and diagnostic accuracy were poor to moderate compared with criterion tests. Other methods were supported by low level evidence.

Conclusion Clinicians in primary care can incorporate transabdominal and transrectal ultrasound or uroflowmetry in the evaluation of men with LUTS but should not solely rely on these methods as the diagnostic accuracy is insufficient and reliability remains insufficiently researched. Low-to-moderate levels of evidence for most assessment methods were due to methodological shortcomings and inconsistency in the studies. This highlights the need for better study designs in this domain.

Strengths and limitations of this study

- This review consists of a broad and systematic search for literature in PubMed, Ebsco/CINAHL and Embase databases for studies on the evaluation of bladder outlet obstruction and benign prostatic obstruction in men with lower urinary tract symptoms.
- The identified literature evaluates a variety of assessment methods, thoroughly evaluated with the COSMIN checklist on all aspects of reliability and validity.
- A level of evidence was estimated based on methodological quality of the studies and precision, direction and consistency of the results.
- Studies with poor COSMIN scores were excluded from the best evidence synthesis, to strengthen conclusions and recommendations.
- Due to low methodological quality of many studies and inconsistencies in findings regarding the diagnostic accuracy, only a best-evidence synthesis was possible.

INTRODUCTION

Lower urinary tract symptoms (LUTS) include problems with storage of urine, voiding and postvoiding.[1] The prevalence of one or more of these symptoms in men over 50 years and older is 50%–75% and increases with age.[2–4] Men with LUTS often experience a reduced quality of life and reduced mental and physical health.[5 6] These symptoms are

not organ-specific and may be related to underlying pathophysiologic mechanisms.[7]

Bladder outlet obstruction (BOO) may cause LUTS in 24% of men.[8] It has different causes, but is known to frequently occur due to benign prostatic obstruction (BPO), caused by benign prostatic hyperplasia.[9] This benign growth of the prostate is harmless, until it compresses the urethra and interrupts the flow of urine. Therefore, prostate size and specific measurements of the prostate are used to evaluate BPO.

Men presenting with LUTS are often evaluated via comprehensive history taking including urological history, physical examination and questionnaires as recommended by the European Association of Urology (EAU) guideline on non-neurogenic male LUTS.[10 11] Nonetheless, in primary care, patients are frequently referred to a urologist or urology clinic for further diagnostic procedures, although this is expensive and not always warranted (eg, MRI or urodynamic studies).[11]

Accurate evaluation of men with LUTS may result in distinct treatment pathways. Less bothersome symptoms could be targeted by conservative treatment including watchful waiting, medication, pelvic floor muscle training or lifestyle changes.[12–14] These therapies can be provided by general practitioners or men's health physiotherapists. Bothersome symptoms, with a medium (>30 mL) to large (>80 mL) size prostate, may require surgery.[11] Yet, the frequent referral to urologists, even in men with a small prostate size, leads to waiting lists, increased costs of healthcare and does not always appear to be beneficial for men with LUTS.[15] Therefore, accurate assessment of the role of BOO and BPO in primary care could reduce the need for referral and allow for early treatment in primary care. However, it is unknown which methods are valid and reliable.

A recent review discouraged the use of noninvasive tests, such as uroflowmetry or penile cuff tests over pressure flow studies to diagnose BOO, although the role of BPO was not specifically researched and not all aspects of validity (eg, correlations with a criterion) and reliability were covered.[16]

Therefore, this study aimed to systematically review the literature to determine the reliability and validity of assessment methods available in primary care to evaluate BOO and BPO in men with LUTS and to provide recommendations for the best assessment methods for clinicians in primary care.

METHODS

Study design

A systematic review with best evidence synthesis was conducted, according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[17]

Search

Embase, PubMed and Ebsco/CINAHL databases were searched from inception up to 26 November 2020. A previously validated search strategy for terms on reliability, validity and reproducibility of measurements was used.[18] This search strategy was combined with relevant terms for BOO and BPO. Studies on benign prostatic hyperplasia were also included, as the term was frequently used to define its clinical symptoms, instead of its histological features. The currently accepted definitions of BOO and BPO are provided in table 1. The full search strategy was developed by medical-information specialists (JCFK and JM) and is provided for each database in online supplemental table 1A–C.

Screening and selection

Studies were included if the reliability and/or validity was investigated of assessment methods to evaluate BOO, BPO or benign prostatic hyperplasia in men with LUTS. A wide variety of tests and methods is described in the literature for the evaluation of BOO and BPO. To provide a comprehensive overview, all types of assessment methods available in primary care were considered, including all types of questionnaires and all forms of clinical examination techniques and tests (eg, ultrasound imaging, free uroflowmetry).

Exclusion criteria were invasive techniques, CT or MRI, 3D/4D-ultrasound imaging, urodynamic studies, postvoid residue measurement or studies that aimed to estimate prostate size to predict prostate cancer. Studies with assessment methods for BOO or BPO in men with LUTS

Table 1 Currently accepted definitions of BOO and BPO

Terminology	Definition	Source
BOO	This is the generic term for obstruction during voiding. It is a reduced urine flow rate with a simultaneously increased detrusor pressure.	D'Ancona *et al* 2019 in: The International Continence Society report on terminology for adult male lower urinary tract and pelvic floor symptoms and dysfunction.[1]
BPO	Is a form of BOO; and may be diagnosed when the cause of outlet obstruction is known to be benign prostatic enlargement, due to histological benign prostatic hyperplasia.	Abrams *et al* 2002, in: The Standardisation of Terminology of Lower Urinary Tract Function: Report from the Standardisation Sub-committee of the International Continence Society.[9]

BOO, bladder outlet obstruction; BPO, benign prostatic obstruction.

due to known pathology (eg, prostate cancer, neurological diseases) were also excluded.

Studies were first screened on title, then on abstract and subsequently on full text. Each full text article was independently screened by two investigators from a group of investigators (TV, EvB, REPMB, MPK, SPJR, JJMP and ALP-G). In case of disagreement, a third investigator was consulted from the same group of investigators.

Methodological quality

The methodological quality of the included studies was evaluated using the COSMIN (COnsensus-based Standards for the selection of health Measurement INstruments) checklist for measurement properties. The COSMIN checklist consists of nine sections (called: 'boxes') to score different aspects of validity, reliability and responsiveness. Each box contains 5–18 items on a 4-point scale (poor, fair, good or excellent) and is awarded an overall score per box based on the lowest scoring item within that box. Grading studies with the COSMIN checklist follows a tailor-made approach, as only the measurement properties of the assessment method researched by the study are graded.[19]

Data collection and analysis

Data extraction included: population characteristics, index and reference tests, reliability (including percentage agreement, kappa values, intraclass correlation coefficient (ICC)) and validity measures (including correlation coefficients, sensitivity and specificity, likelihood ratios).

Primary outcomes were measures for reliability (test–retest, inter-rater reliability and agreement) or hypothesis testing (construct) and criterion validity. Secondary outcomes included other measurement properties regarding reliability (internal consistency, measurement error) or validity (face validity, cross-cultural validity) as evaluated through the COSMIN checklist.

Best evidence synthesis was performed, based on criteria that include the methodological quality, imprecision, indirectness and inconsistency of results.[19] Based on these criteria the level of evidence was estimated, described in detail in table 2.[20] Per assessment method, between-study results were graded for consistency as consistent, inconsistent or indeterminate. Consistent findings were defined if the measures of validity and reliability were similar or of adjacent categories (eg, moderate-good). The level of evidence was downgraded one level if inconsistency was found.

Patient and public involvement

No patient involved.

RESULTS

Search results and grading of evidence

The search identified 7224 articles. After removal of duplicates 5644 articles remained which were screened for title. Of these, 337 articles were screened for abstract. Then, full-text screening was performed for 152 studies, of which 61 studies met all selection criteria and were scored using the COSMIN checklist. Subsequently, 37 studies received a COSMIN score of fair, good or excellent.[21–57] Twenty-four studies[58–81] received 'poor' COSMIN scores and were therefore excluded from the best evidence synthesis. Of the 37 included studies, 18 studies evaluated the assessment of BOO and 17 BPO, 2 studied aspects relevant to both BOO and BPO. A flow chart of the study selection is provided in figure 1.

With 33 scores, the COSMIN box for criterion validity was scored the most, followed by 8 scores for reliability, 6 for hypothesis testing and 4 scores for measurement

Table 2 Level of evidence rating

Level of evidence	Rating	Criteria* Consistency	Methodological quality	Total sample size
High-quality level of evidence	+++	Similar or of adjacent categories	Multiple studies with methodological quality rated as at least 'good' OR One study with methodological quality rated as excellent	≥100
Moderate-quality level of evidence	++	Similar or of adjacent categories	Multiple studies with methodological quality rated as 'fair' OR One study with methodological quality rated as 'good'	≥50
Low-quality level of evidence	+	Similar or of adjacent categories	One study with methodological quality rated as 'fair'	–
Conflicting evidence	+/-	Non-similar categories	Multiple studies	–
No evidence	?	Rated as '?'	Only studies with methodological quality rated as 'poor' OR no outcomes reported.	–

*In order to meet a level of evidence, all three criteria have to be met (consistency, methodological quality and sample size). Adapted from van Tulder *et al*,[20] in line with criteria by Mokkink *et al* and Prinsen *et al*.[99 100]

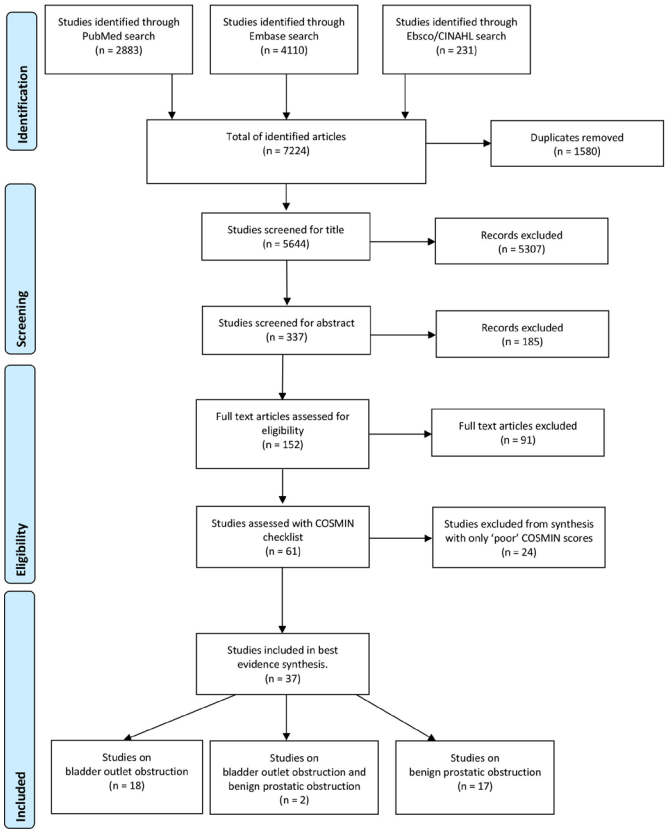

Figure 1 PRISMA flow chart of the inclusion of studies. COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

error. In none of the studies internal consistency, content validity, structural validity, cross-cultural validity or responsiveness were scored (see table 3 for BOO and table 4 for BPO).

Sensitivity, specificity, positive and negative predictive values in this section are reported as percentages. A summary of findings including scores for consistency of findings and levels of evidence is presented in table 5. Data extraction is found in online supplemental table 2A–C.

Assessment methods for BOO

Twenty studies evaluated the assessment methods for BOO. The reference tests were BOO-related measures (eg, obstruction grade number or maximum urine flow rate), rather than reference tests for prostate size.[22–25 28 29 34 35 37–39 41 44 47 50 51 55–57] An overview of the COSMIN scores per study on BOO is provided in table 3.

Transrectal ultrasound

Transrectal ultrasound to measure prostate size to indicate BOO, demonstrated poor correlations with obstruction grade numbers (r=0.22[55] and r=0.29[56]), maximum flow rate (r=0.20[39] and r=−0.11[55]) or postvoid residue (r=0.05[55] and r=0.21[39]). The measurement of peripheral zone thickness, transitional zone volume and index by transrectal ultrasound yielded comparable, poor correlations.[39] Different cut-off scores for the prostate size yielded poor to good diagnostic accuracy, including: >25 mL (sensitivity: 85%; specificity: 27%)[44] and >40 mL (sensitivity: 66%; specificity: 64%[51]), indicating obstruction and <40 mL (sensitivity: 43%; specificity: 83%[51]) and <25 mL (sensitivity: 21%; specificity: 92%[51]) indicating no obstruction. The COSMIN scores ranged from fair to good for hypothesis testing[39] and criterion validity.[44 51 55 56] Based on the COSMIN scores, consistent findings, the number of studies (n=5) and total sample size (n=1731), the level of evidence for the validity of assessment of BOO with transrectal ultrasound was graded as high (see table 5). Measures of reliability were not reported (see online supplemental tables 2A,B).

Transabdominal ultrasound

The diagnostic accuracy of transabdominal ultrasound to indicate BOO, ranged from poor to good, depending on the cut-off values for obstruction: >40 mL (sensitivity: 58%; specificity: 67%[35]) or >45 mL of prostate size (sensitivity: 86%; specificity: 26%[22]). The diagnostic accuracy was 0.658 based on the area under the curve from receiver operator curve analysis (ROC-AUC), although it appears heavily skewed data were analysed.[47] The correlations of prostate size measured by transabdominal ultrasound with maximum flow rate (r=−0.40)[57] or BOO index (r=0.24[22] to r=0.40)[35 47] from urodynamic studies were poor.

Intravesical prostatic protrusion to indicate BOO was measured by transabdominal ultrasound, with moderate to good diagnostic accuracy. Two different cut-off values to determine obstruction were used: 8 mm (sensitivity: 80%; specificity: 80%[22]) and 10 mm (sensitivity: 65%–81.6%; specificity: 40%–84.9%[24 35 47]). Correlations of intravesical prostatic protrusion correlated poorly with BOO index (r=0.59[47] to r=0.69).[22 35] The COSMIN scores ranged from fair[47 57] to excellent[22] for criterion validity for measurement of the prostate size and good[24 35] to excellent[22] for the criterion validity of the measurement of the intravesical prostatic protrusion. Based on the COSMIN scores, inconsistent findings, number of studies (n=5) and total sample size (n=536), the level of evidence for the validity of assessment of BOO with transabdominal ultrasound was graded as moderate (see table 5). No measures of reliability were reported (see online supplemental table 2B).

One study compared transabdominal ultrasound measured bladder weight to pressure flow studies, to indicate infravesical obstruction, and demonstrated good sensitivity (85.3%) and specificity (87.1%).[38] With a fair COSMIN score for criterion validity from one study with a sample size of n=65, the level of evidence was low for the measurement of bladder weight using transabdominal ultrasound to indicate obstruction (see table 5). Measures of reliability were not reported (see online supplemental table 2B).

Transperineal ultrasound uroflowmetry

Transperineal ultrasound uroflowmetry with a radio frequency reflection measurement to evaluate BOO was

Table 3 Included studies on BOO with at least one COSMIN score: fair, good or excellent

Measurement property		Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Cross-cultural validity	Criterion validity	Responsiveness
COSMIN score box		**Box A**	**Box B**	**Box C**	**Box D**	**Box E**	**Box F**	**Box G**	**Box H**	**Box I**
First author	**Year**									
Abdel-Aal[22]	2011	–	–	–	–	–	–	–	Excellent	–
Aganovic[23]	2019	–	–	–	–	–	–	–	Good	–
Al-Mosawi[24]	2020	–	–	–	–	–	–	–	Good	–
Arif[25]	2016	–	–	Poor	–	–	–	–	Fair	–
Chan[28]	2012	–	Good	–	–	–	Poor	–	Fair	–
Chen[29]	2019	–	–	–	–	–	–	–	Fair	–
Harding[34]	2004	–	–	–	–	–	–	–	Fair	–
Hossain[35]	2012	–	–	–	–	–	–	–	Good	–
Kim[37]	2020	–	–	–	–	–	Poor	–	Good	–
Kojima[38]	1997	–	–	–	–	–	–	–	Fair	–
Kwon[39]	2015	–	Fair	–	–	–	Good	–	–	–
Matzkin[41]	1996	–	Poor	–	–	–	Fair	–	–	–
Oelke[44]	2007	–	–	–	–	–	–	–	Good	–
Reddy[47]	2019	–	–	–	–	–	–	–	Fair	–
Reynard[48]	1996	–	Poor	–	–	–	–	–	Good	–
Salinas[50]	2003	–	–	–	–	–	–	–	Fair	–
Steele[51]	2000	–	–	–	–	–	–	–	Good	–
Venrooij[55]	1996	–	–	–	–	–	–	–	Fair	–
Venrooij[56]	2004	–	–	–	–	–	–	–	Good	–
Zhou[57]	2012	–	–	–	–	–	–	–	Fair	–

BOO, bladder outlet obstruction; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments.

Table 4 Included studies on BPO with at least one COSMIN score: fair, good or excellent

Measurement property		Internal consistency	Reliability	Measurement error	Content validity	Structural validity	Hypothesis testing	Cross-cultural validity	Criterion validity	Responsiveness
COSMIN score box		**Box A**	**Box B**	**Box C**	**Box D**	**Box E**	**Box F**	**Box G**	**Box H**	**Box I**
First author	**Year**									
Aarnink[21]	1996	–	–	–	–	–	–	–	Fair	–
Baltaci[26]	2000	–	–	–	–	–	–	–	Fair	–
Carballido[27]	2011	–	–	–	–	–	–	–	Good	–
David[30]	2020	–	–	–	–	–	–	–	Good	–
Demir[31]	2016	–	–	–	–	–	–	–	Fair	–
De Nunzio[32]	2015	–	–	–	–	–	–	–	Good	–
Güzelsoy[33]	2016	–	–	–	–	–	–	–	Fair	–
Kim[36]	2014	–	–	–	–	–	Fair	–	–	–
Kwon[39]	2015	–	Fair	–	–	–	Good	–	–	–
Malemo[40]	2011	–	–	Poor	–	–	–	–	Good	–
Narayanamurthy[42]	2020	–	–	–	–	–	–	–	Good	–
Nathan[43]	1996	–	Poor	Poor	–	–	Fair	–	–	–
Prassopoulos[45]	1996	–	Fair	Fair	–	–	–	–	Fair	–
Rathaus[46]	1991	–	–	–	–	–	–	–	Fair	–
Roehrborn[49]	2001	–	Fair	–	–	–	–	–	Fair	–
Stravodimos[52]	2009	–	–	–	–	–	–	–	Fair	–
Styles[53]	1988	–	Poor	–	–	–	–	–	Good	–
Su[54]	2013	–	–	–	–	–	–	–	Good	–
Venrooij[55]	1996	–	–	–	–	–	–	–	Fair	–

BPO, benign prostatic obstruction; COSMIN, COnsensus-based Standards for the selection of health Measurement INstruments.

Table 5 Summary of findings

Method	Studies	Consistency of findings	Conclusion	Level of evidence
Assessment methods for bladder outlet obstruction (BOO)				
Transrectal ultrasound (TRUS)	Kwon *et al* 2016[39] Oelke *et al* 2007[44] Steele *et al* 2000[51] Venrooij *et al* 1996[55] Venrooij *et al* 2004[56]	+	Validity The prostate size measured with transrectal ultrasound to indicate BOO showed poor correlations with urinary flow parameters and obstruction grade number. Prediction of BOO, based on different cut-off points (25 mL or 40 mL) for prostate size or prostate dimensions, showed a sensitivity and specificity from poor to good.	+++
	–	?	Reliability Reliability was not reported	?
Transabdominal ultrasound (TAUS)	Abdel-Aal *et al* 2011[22] Al-Mosawi *et al* 2020[24] Hossain *et al* 2012[35] Reddy *et al* 2019[47] Zhou *et al* 2012[57]	–	Validity Several studies evaluated the use of transabdominal ultrasound to measure prostate size to indicate BOO. Sensitivity ranged between moderate to good and specificity between poor and moderate. Intravesical prostatic protrusion as predictor of BOO showed a ranging sensitivity and specificity, however poor correlations with maximum urine flow rate and bladder outlet obstruction index were demonstrated.	++
	–	?	Reliability Reliability was not reported.	?
TAUS	Kojima *et al* 1997[38]	+	Validity Bladder weight to indicate infravesical obstruction demonstrated good sensitivity and specificity.	+
	–	?	Reliability Reliability was not reported.	?
Transperineal ultrasound uroflowmetry	Arif *et al* 2016[25]	+	Validity The use of transperineal ultrasound, by using a specific radio reflection measurement during voiding, was used to predict BOO showed a good area under the curve and good sensitivity and specificity.	+
	–	?	Reliability Reliability was not reported.	?
Uroflowmetry at home	Chan *et al* 2012[28]	+	Validity Low-quality evidence was found for the use of a device for uroflowmetry at home, with a moderate to good sensitivity and specificity to indicate BOO related surgery, depending on the chosen cut-off point of maximum urinary flow rate.	+
	Chan *et al* 2012[28]	+	Reliability Agreement was good for repeated measurements, with an optimum found at 10 repeated measurements.	+
Uroflowmetry	Chen *et al* 2019[29] Oelke *et al* 2007[44] Reynard *et al* 1996[48] Venrooij *et al* 2004[56]	+	Validity The maximum urinary flow measured through free uroflowmetry indicate BOO yielded poor to good sensitivity and poor to good specificity, depending on the cut-off value used (in mL/s). Using the mean value of multiple voids increases accuracy. However, correlations of flow rate and mean voided volume with obstruction grade numbers were poor.	+++
	–	?	Reliability Reliability was not reported.	?

Continued

Table 5 Continued

Method	Studies	Consistency of findings	Conclusion	Level of evidence
Penile Compression Manoeuvre	Aganovic *et al* 2019[23]	+	Validity The manual compression manoeuvre during urination showed a moderate to good sensitivity and specificity to predict BOO, with good diagnostic accuracy. Correlations with urodynamic studies were poor.	++
	–	?	Reliability Reliability was not reported.	?
Penile Cuff Uroflowmetry	Harding *et al* 2004[34] Kim *et al* 2020[37] Salinas *et al* 2003[50]	–	Validity Sensitivity ranged from moderate to good and specificity from fair to moderate to detect obstruction from non-obstruction, with good diagnostic accuracy.	+
	–	?	Reliability Reliability was not reported.	?
Combination of assessment methods	Venrooij *et al* 2004[56]	+	Validity A bladder outlet obstruction number was calculated based on prostate volume, maximum urinary flow rate and mean voided volume from free uroflowmetry. Good area under the curve values were found, although poor correlations with the criterion were reported.	++
	–	?	Reliability Reliability was not reported.	?
Questionnaires: International Prostate Symptom Score (IPSS) or American Urology Association questionnaire (AUA)	Chan *et al* 2012[28] Matzkin *et al* 1996[41] Steele *et al* 2000[51] Venrooij *et al* 1996[55] Venrooij *et al* 2004[56]	–	Validity The IPSS was used to indicate BOO by a few studies, compared with different values for the maximum urinary flow rate (<10,<15,<19 mL/s), the sensitivity was moderate to poor and specificity moderate to good. However, correlations of the IPSS with urinary flow parameters were found to be poor. Two studies assessed the identical AUA questionnaire without the separate Quality of Life question included in the IPSS, with 24-hour uroflowmetry recording and detrusor pressure at maximum flow as predictor of BOO and found poor correlations.	++
	–	?	Reliability Reliability was not reported.	?
Assessment methods for benign prostatic obstruction (BPO)				
Digital rectal examination (DRE)	Carballido *et al* 2011[27] Roehrborn *et al* 2001[49] Su *et al* 2013[54]	+	Validity The use of digital rectal examination to measure prostate size showed moderate to good sensitivity and specificity based on a cut-off value of 30 mL. Correlations with transrectal ultrasound measured prostate size were poor to good.	+++
	Roehrborn *et al* 2001[49]	+	Reliability Inter-rater reliability of digital rectal examination was found to be showed ranging reliability from poor to good, based on different grading scales used.	+
TAUS	Demir *et al* 2016[31] Güzelsoy *et al* 2016[33] Malemo *et al* 2011[40] Prassopoulos *et al* 1996[45] Stravodimos *et al* 2009[52] Styles *et al* 1988[53]	–	Validity Fair to good sensitivity and good specificity were found to measure prostate size with transabdominal ultrasound compared with transrectal ultrasound, with a cut-off prostate volume of >80 mL. Correlations with transrectal ultrasound or enucleated adenoma weight were good.	++

Continued

Table 5 Continued

Method	Studies	Consistency of findings	Conclusion	Level of evidence
	Prassopoulos *et al* 1996[45]	+	Reliability One study described the reliability of TAUS assessment as good with interobserver error of lower than 5%.	+
Transperineal ultrasound	Rathaus *et al* 1991[46]	+	Validity One study evaluated transperineal ultrasound to measure prostate volume and found a good correlation with enucleated adenoma weight.	+
	–	?	Reliability Reliability was not reported.	?
TRUS	Aarnink *et al* 1996[21] Baltaci *et al* 2000[26] David *et al* 2020[30] Demir *et al* 2016[31] Güzelsoy *et al* 2016[33] Kim *et al* 2014[36] Narayanmurthy *et al* 2020[42] Nathan *et al* 1996[43] Stravodimos *et al* 2009[52]	+	Validity Several formulas can be used to calculate prostate size after measurement of the prostate with transrectal ultrasound. Good correlations were found with the regularly used step planimetry method or manual outline formula. Correlations with the enucleated adenoma weight were good. The diagnostic accuracy for the transitional zone index by transrectal ultrasound showed good sensitivity and poor to good specificity, based on the cut-off values. Transitional zone index or volume assessed by transrectal ultrasound correlated moderate to good with enucleated tissue weight.	+++
	Kwon *et al* 2016[39] Prassopoulos *et al* 1996[45]	+	Reliability The inter-rater reliability for the assessment of peripheral zone volume by transrectal ultrasound was good and showed low interobserver error.	++
Combination of assessment methods	De Nunzio *et al* 2015[32]	+	Validity One study predicted BPO using a nomogram based on free-flowmetry and transrectal ultrasound assessment of transitional zone volume. Moderate sensitivity and specificity to indicate BPO were found.	++
	–	?	Reliability Reliability was not reported.	?
Questionnaire	Carballido *et al* 2011[27] Kwon *et al* 2016[39] Nathan *et al* 1996[43] Venrooij *et al* 1996[55]	+	Validity IPSS scores were correlated with prostate size and other prostate zones, including transitional zone volume, index and peripheral zone thickness. All correlations were found to be poor. The sensitivity and specificity were fair to moderate.	+++
	–	?	Reliability Reliability was not reported.	?

Consistency (between-study results): +=consistent, ?=indeterminate, -=inconsistent.
Levels of evidence: +++=high-quality evidence, ++=moderate-quality evidence, +=low-quality evidence, +/-=conflicting evidence, ?=no evidence.

compared with pressure flow studies, and demonstrated a high ROC-AUC of 0.96, and good sensitivity (88%) and specificity (95%).[25] The COSMIN score was fair for criterion validity. Based on the COSMIN scores from one study with a sample size of n=45, the level of evidence was graded as low to evaluate BOO using transperineal ultrasound uroflowmetry (see table 5). No measures of reliability were reported (see online supplemental table 2B).

Uroflowmetry at home

Uroflowmetry at home with a compartment-meter showed a good sensitivity (79%–99%) and specificity (68%–90%[28]) compared with different maximum flow rates (<10 mL/s, <15 mL/s, <19 mL/s[28]) measured by uroflowmetry at the clinic. The agreement between scores was the highest after 10 measurements (kappa: 0.84).[28] The COSMIN scores were fair for criterion validity and good for reliability. Based on

these COSMIN scores from one study with a sample size of n=186, the level of evidence was graded as low to indicate BOO based on uroflowmetry at home (see table 5, see online supplemental table 2B).

Uroflowmetry

Uroflowmetry to indicate BOO demonstrated a moderate to good sensitivity (68%–99%) and poor to moderate specificity (39%–73%). A higher flow rate cut-off point (7, 10, 15 mL/s) resulted in a higher sensitivity and lower specificity to identify non-obstructed.[44] Comparable sensitivity and specificity trade-offs were found to identify obstruction based on maximum flow rate below cut-off values (8, 10, 12, 15 mL/s). Mean values of maximum flow rate of three or four voids would increase diagnostic accuracy.[48] Poor correlations were found for the maximum urinary flow compared with the Abrams-Griffiths number (τ=−0.41), urethral resistance factor (τ=0.26) and Schäfer's obstruction grade (τ=−0.43).[56] The COSMIN scores for criterion validity were good.[44 48 56]

Based on uroflowmetry, Chen *et al* developed a new nomogram to detect obstructed (≤10 mL/s) from non-obstructed (≥15 mL/s) male patients.[29] Compared with the Abram-Griffiths obstruction number from urodynamic studies the sensitivity (81%) and specificity (91%) were good with an ROC-AUC of 86%.[29] The COSMIN score was fair for criterion validity.

Based on the COSMIN scores, the consistency of findings, number of studies (n=4) and the total sample size (n=1001), the level of evidence was graded as high for validity for uroflowmetry to identify BOO (see table 5). No measures of reliability were reported (see online supplemental table 2B).

Penile cuff uroflowmetry and penile compression manoeuvre

The method of manual penile compression by the patient to interrupt the flow midstream was analysed by one study.[23] This manoeuvre was performed while voiding in a uroflowmetry device. An index was created based on two points from the recorded flow. With a cut-off index value of 96.4%, BOO based on Schäfer's obstruction grade could be predicted with a moderate sensitivity of 74% and good specificity of 94%.[23] The level of evidence was graded as moderate for the validity of using a penile compression manoeuvre to indicate BOO, based on a good COSMIN score for criterion validity from a single study with a sufficient sample size (n=135). No measures of reliability were reported for this method (see table 5).

Uroflowmetry using an automated inflating penile cuff was analysed by three studies.[34 37 50] One study evaluated a similar compression index as beforementioned and found a moderate sensitivity (78%) and good specificity (84%) at an index cut-off value of 160% to predict BOO.[34] The sensitivity for the measured maximum urinary flow with the penile cuff to predict BOO was good (80%[37] and 100%[50]) with a fair to good specificity of (56%[50] and 100%[37]).

Based on the fair[34 50] or good[37] COSMIN scores for criterion validity, inconsistency of findings, number of studies (n=3) and total sample size (n=253) the level of evidence was graded as low for the validity to indicate BOO using uroflowmetry with an automated inflating penile cuff (see table 5). Measures of reliability were not reported (see online supplemental table 2B).

Combination of assessment methods

One study developed and evaluated a BOO number based on prostate size measured by transrectal ultrasound, mean voided volume and maximum urinary flow rate from uroflowmetry. It demonstrated poor correlations (r=0.48 to r=0.52) with obstruction indices. The overall diagnostic accuracy, through ROC-AUC analysis was good with reference values: Abram-Griffith's number (0.83), Schäfer's obstruction grade (0.82) and urethral resistance factor (0.87).[56] Measures of reliability were not reported. Based on a good COSMIN score for criterion validity from one study with a sufficient sample size (n=160), the level of evidence was graded as moderate for the validity of using a combination of assessment methods for the indication of BOO (see table 5, online supplemental table 2B).

Questionnaire

The International Prostate Symptom Score (IPSS) was the only identified questionnaire used to detect BOO. It demonstrated poor correlations with maximum flow rate, postvoid residue and obstruction grade numbers (r=−0.07 to r=0.06).[55] Different IPSS item-score cut-off values to indicate obstruction (maximum flow rate <10, <15, <19 mL/s[28]) demonstrated poor to moderate sensitivity (25%–74%) and moderate to good specificity (55%–86%) compared with uroflowmetry.[28] A lower flow rate cut-off point resulted in a lower sensitivity and higher specificity. The IPSS studied as the American Urological Association questionnaire, yielded poor correlations with uroflowmetry recordings,[41] detrusor pressure at maximum flow (r=0.18[51]) and obstruction grade numbers (r=0.15 to r=0.16[56]). The COSMIN scores were fair for hypothesis testing[41] and ranged from fair[28 55] to good[51 56] for criterion validity. Combined with the inconsistency of findings, number of studies (n=5), total sample size (n=788) and COSMIN scores, the level of evidence was graded as moderate for the validity of detection of BOO using questionnaires (see table 5). However, no measures of reliability were not reported (see online supplemental table 2B).

Assessment methods for BPO

Nineteen studies evaluated assessment methods related to prostate size or intravesical prostatic protrusion, peripheral zone volume and transitional zone volume or transitional zone index to evaluate BPO due to benign prostatic enlargement, unless specified otherwise.[21 26 27 30–33 36 39 40 42 43 45 46 49 52–55] An overview of the COSMIN scores per study on BPO is provided in table 4.

Digital rectal examination

The diagnostic accuracy of digital rectal examination to determine prostate size varied between studies. With a cut-off value of ≥30 mL for the prostate size, the sensitivity was good (94%[54]) and specificity moderate (78%[54]). Prostate size measured by digital rectal examination performed by a general practitioner and a urologist as reference test, demonstrated a poor correlation (k=0.28[27]). Correlations with prostate size measured with transrectal ultrasound were poor to good (r=0.56 to r=0.72).[49] The COSMIN scores were fair[49] and good[27 54] for criterion validity. Including the COSMIN score, consistent findings, the number of studies (n=3) and total sample size (n=1067) the level of evidence for validity of digital rectal examination to determine prostate size was high (see table 5).

One study established the reliability of digital rectal examination to be poor to good, based on the grading scale used (ICC: 0.58–0.86).[49] The COSMIN score for reliability was fair, based on one study with a sufficient sample size (n=121). Therefore, the level of evidence was low for the reliability of digital rectal examination (see online supplemental table 2A,C).

Transabdominal ultrasound

The diagnostic accuracy of prostate size measured by transabdominal ultrasound was good, compared with transrectal ultrasound, with a cut-off prostate volume of ≤80 mL (sensitivity: 95%, specificity: 96%).[40] Another study investigated a cut-off prostate volume of <80 cc compared with a specimen weight of <80 with fair sensitivity (56%) and specificity (100%).[52] Transabdominal ultrasound measured prostate size correlated good with transrectal ultrasound measured prostate size (r=0.82[53] to r=0.98[40 45]) and good with enucleated tissue weight (r=0.73[33] to r=0.82[31 52]). Correlations with transitional zone volume measured by transrectal ultrasound were good (r=0.95).[45] The COSMIN scores for criterion validity were fair[31 33 45 52] to good.[40 53] Due to the inconsistency of findings, number of studies (n=6) and total sample size (n=395), the level of evidence was graded as moderate for the validity of measuring prostate size using transabdominal ultrasound (see table 5, online supplemental table 2C).

A low interobserver error was demonstrated (5%), with a fair score on the COSMIN box for reliability and measurement error.[45] Based on the COSMIN scores from a single study with a study sample of n=95, the level of evidence for reliability was graded as low for prostate size measurements with transabdominal ultrasound (see table 5, online supplemental table 2A).

Transperineal ultrasound

Good correlations were demonstrated for prostate size measured by transperineal ultrasound, compared with enucleated tissue weight (r=0.89).[46] Based on a fair COSMIN score for criterion validity from a single study with a sample size of n=80, the level of evidence was graded as low for the measurement of prostate size using transperineal ultrasound (see table 5). Measures of diagnostic accuracy and reliability were not reported (see online supplemental table 2C).

Transrectal ultrasound

Prostate size measured by transrectal ultrasound correlated poor to good with enucleated tissue weight (r=0.67[42] to r=0.95[26 30 31 33 52]) One study found a mean difference of −12.5 g underestimation of prostate volume measured by transrectal ultrasound, with levels of agreement between −38 and 13 g.[42] Other studies assessed the criterion validity of various transrectal ultrasound formulas or outline methods to calculate and measure prostate size. Correlations with the frequently used ellipsoid formula or step planimetry method were good (r=0.88[43] to r=0.96[21]) and non-significant differences[36] between measurements of size were found.

The transitional zone volume measured by transrectal ultrasound correlated good (r=0.87[30] to r=0.97[33 52]) with enucleated tissue weight, and subgrouping for men with a larger prostate size slightly increased the correlation.[30] Correlations with transrectal ultrasound measured total prostate volume were good (r=0.82[21] and r=0.96[33]). Good sensitivity (93%) and fair specificity (61%) to identify volumes under 80 cc with a specimen weight under 80 grams were found.[52]

Calculation of the transitional zone index, based on prostate zones measured with transrectal ultrasound, yielded good sensitivity (91%–100%) and poor to good specificity (19%–91%) for different cut-off values, although reference values were unclear.[33] For the transitional zone index, poor correlations (r=0.55) with resected tissue weight were found.[33]

The COSMIN scores were fair[21 26 31 33 36 52] to good[30 42] for criterion validity and one study scored fair for hypothesis testing.[43] Including the COSMIN scores, consistency of findings, number of studies (n=9) and total sample size (n=1854), the level of evidence was graded as high for the validity of the measurement of prostate size using transrectal ultrasound (see table 5).

The interrater scores for peripheral zone thickness were good (ICC: 0.87[39]), and low interobserver error (4%)[45] was found with fair COSMIN scores for reliability. Based on the COSMIN scores from two studies with an overall sample (n=1104) and consistent findings, the level of evidence was graded as moderate for the reliability of transrectal ultrasound (see online supplemental table 2C).

Combination of assessment methods

One study combined transitional zone volume measured by transrectal ultrasound and free uroflowmetry, to calculate a nomogram-based index score. Compared with a Schäfer's obstruction grade of ≥3 to indicate BPO (as described by the study), a moderate diagnostic accuracy (sensitivity: 74%; specificity: 79%) was demonstrated, with a good ROC-AUC of 0.76.[32] With a good COSMIN score for criterion validity from a single study with a sufficient sample size (n=449), level of evidence was graded as moderate for the validity of a combination of assessment methods to indicate BPO

(see table 5). Measures of reliability were not reported (see online supplemental table 2C).

Questionnaires

The construct validity (hypothesis testing) of the subdomain and total score of the IPSS was assessed by correlation with prostate size, peripheral or transitional zone volume and transitional zone index. The COSMIN score for hypothesis testing was good.[39] All correlations were poor (r=−0.17 to r=0.15).[39 43 55] Sensitivity and specificity of the IPSS to match the urologist's final diagnosis was moderate (sensitivity: 58%; specificity: 59%) and slightly increasing with age in the model (sensitivity: 57%; specificity: 64%).[27] The COSMIN scores for criterion validity were fair[43 55] to good.[27] Including the COSMIN scores and the consistency of findings, number of studies (n=4) and total sample size (n=1916) the level of evidence for the validity of IPSS to determine BPO was graded as high (see table 5). Measures of reliability were not reported (see online supplemental table 2C).

DISCUSSION

Statement of principal findings

Through this review a variety of assessment methods were identified to assess BOO and to identify the role of BPO in men with LUTS due to BOO. To evaluate BOO, transrectal or transabdominal ultrasound and uroflowmetry were identified to be the most adequate assessment methods for clinicians in primary care. These methods showed moderate to good diagnostic accuracy and are supported by moderate to high quality of evidence. However, compared with urodynamic studies, correlations were poor.

The IPSS questionnaire is frequently used to assess severity of symptoms in men with BOO.[11] Nonetheless, the IPSS should not be recommended to identify BOO due to poor to moderate sensitivity and specificity and poor correlations with parameters from urodynamic studies. Noninvasive uroflowmetry related methods such as home-uroflowmetry, penile cuff or penile compression manoeuvres can be supported by low to moderate quality of evidence for their validity. Correlations of these methods with urodynamic studies were moderate to good.

In men with BOO, the role played by BPO could be adequately assessed by transrectal and transabdominal ultrasound to measure prostate size. Studies suggest the importance of the intravesical prostatic protrusion, specific enlargement of the prostate towards the bladder floor, which showed higher correlations and diagnostic accuracy[24 35 47 57] with BOO compared with the total prostate volume.[82] Similarly, the transitional zone volume could also be assessed through these ultrasound modalities, as some studies reported higher correlations with reference tests compared with total prostate volume.[26 30 33] Although these prostate parameters could be evaluated with transabdominal or transrectal ultrasound, they were less frequently researched, and may be best evaluated after initial inspection of total prostate volume to improve the evaluation of BPO.

Digital rectal examination may be used to determine prostate volume, although correlations with reference tests were lower compared with transabdominal and transrectal ultrasound. Also, reliability ranged from poor to good, depending on the classification scale to estimate size. Therefore, outcomes of digital rectal examination to indicate BPO should be interpreted with care.

Clinical implications and therapeutic strategies

Following the EAU guidelines on non-neurogenic male LUTS, one of the main goals in men presenting with LUTS is to obtain a diagnosis that may help outline the multifactorial causes of LUTS.[11] A male patient presenting in primary care with LUTS may undergo thorough urological history taking and additional transrectal, transabdominal ultrasound or uroflowmetry to unravel the specific role of BOO and BPO. Subsequently, low-cost and effective therapeutic strategies could start early, initiated in primary care. These could include watchful waiting, pelvic floor muscle training, support of self-management, lifestyle advice or pharmacological treatment, including α1 adrenoceptor antagonists and 5α-reductase inhibitors.[11 83] These options may suffice for some men to address their level of bother and could prevent unwanted surgery or invasive and burdensome diagnostic procedures in specialised urology clinics.[84]

Strengths and weaknesses of the study

Thirty-seven of the 61 studies scored at least one COSMIN-box 'fair', 'good' or 'excellent'. Nonetheless, this systematic review was hindered by the lack of definitions of measured constructs, minimal to no description of patient characteristics or how the diagnosis of BOO, BPO or benign prostatic hyperplasia was established.

Thirty-three of the included studies evaluated the criterion validity of assessment methods. The COSMIN describes criterion validity as the degree to which the scores of an instrument are an adequate reflection of a 'gold standard'.[85] In the evaluation of BOO, urodynamic studies were often used to calculate a BOO Index (or: Abram-Griffiths number), Schäfer's obstruction grade or International Continence Society nomogram to define obstruction.[22–25 29 34 35 37 38 47 48 50 51 55–57] However, an overestimation or underestimation of the diagnostic accuracy could be present. Men with a score of >40 on the BOO Index are defined as obstructed and below 20 are non-obstructed.[86] Men with scores in between are described as equivocal, which some studies excluded from analysis, while other studies included these men in the non-obstructed group.

Also, in the evaluation of BPO, transrectal ultrasound was frequently used as reference test to measure prostate size. This is questionable, as it shows moderate values for the inter-rater reliability for smaller prostate sizes (<30 mL).[87] If a poor reference test is used, the index test

may incorrectly classify results as false positives or false negatives.[88]

These limitations lead to undefinable populations, unclear classification of the disease, and arbitrary categorisation of participants. The variability in prevalence in the study samples may lead to a underestimation or overestimation of the diagnostic accuracy and introduces heterogeneity among the selected studies.[89 90] Accordingly, this may be the main reason for the inconsistency of results found for some assessment methods which led to downgrading the level of evidence. Consequently, only a best-evidence synthesis was performed, not performing an in-depth meta-analysis to pool the diagnostic accuracy from selected studies.[91]

Another reason for the inconsistency of results could be the poor use of statistical methods in some primary studies. For example, incorrect use of Cohen's kappa statistic to correlate instruments or Pearson's correlations in heavily skewed data.[29 47] These studies could not be excluded from synthesis, as the lowest COSMIN score for incorrectly applied statistical methods is 'fair'.

These limitations threaten the reported diagnostic accuracy of considered methods and therefore compromise their overall added value to the evaluation of men with LUTS suspected of BOO.

Strengths and weaknesses in relation to other studies

Current guidelines on the evaluation of men with LUTS describe the use of questionnaires and suggest digital rectal examination and ultrasound measurements.[11 92 93] The present review underlines the recommendation for the use of transabdominal or transrectal ultrasound, while digital rectal examination may be less informative. Ultrasound assessment methods require time and certain experience.[94] The cost of ultrasound systems used to be limiting, although affordable good quality portable systems and promising hand-held devices connected to tablets address this limitation. Questionnaires, like the IPSS, provide accurate insight into the severity of complaints and should solely be used for that purpose.[95 96]

A review by Malde *et al* did not recommend uroflowmetry to indicate BOO.[16] Our review found poor to good diagnostic accuracy for uroflowmetry, home-uroflowmetry or combinations of such measures, only with a low to moderate level of evidence. In our opinion these methods look promising and may help to identify BOO in men presenting in primary care. However, conclusive diagnosis should not be derived solely from the outcomes of these methods.

Unanswered questions and future research

Overall, for a conclusive diagnosis the assessment methods in this review do not have sufficient diagnostic accuracy to be used as stand-alone tests. Therefore, adequate history taking and clinical reasoning of the clinician are required in the evaluation of male patients with LUTS suspected of BOO. A cluster of measurements including questionnaires and other assessment methods may be more promising than comparing single outcomes to a criterion. Additionally, Bossuyt *et al* presented several diagnostic pathways, that may prove helpful in the evaluation of men with LUTS and BOO in primary care.[97] Many of the included studies compared methods with a criterion test or measured construct, without in-depth analysis of pretesting probability or post-testing probability. New studies should further research the best possible diagnostic pathway, by implementing quick and easy triage tests with high sensitivity, followed by add-on tests with high specificity. This could help to improve accurate referrals and select appropriate treatment pathways with less misclassification of men with BOO.

Future studies on the diagnostic accuracy of noninvasive assessment methods should use the COSMIN-taxonomy, which provides helpful definitions and tools to perform accurate research.[85] To make meta-analysis possible in the future, researchers should not only follow adequate methodology, but also describe the results according to reporting guidelines of diagnostic studies, such as the Standards for the Reporting of Diagnostic Accuracy Studies (STARD).[98]

CONCLUSION

Transrectal or transabdominal ultrasound and uroflowmetry are the most adequate methods to evaluate BOO based on BPO in primary care and may be beneficial for an adequate referral to specialised urology care. Digital rectal examination could be used to evaluate men with BPO, although caution is needed when interpreting the results. Devices and methods related to uroflowmetry, that is, home-uroflowmetry and penile cuff or compression methods may be inconclusive in the evaluation of BOO. Included questionnaires that measure symptoms of severity of LUTS are not recommended to detect BOO or BPO. Overall, this review was hindered by suboptimal methods of the included studies and unclear definitions and imprecise presentation of results in many of the included studies, often resulting in lower levels of evidence.

Author affiliations
[1]Human Movement Sciences, Faculty of Behavioural and Movement Sciences, Amsterdam Movement Sciences, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[2]Centre of Expertise Urban Vitality, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands
[3]HU University of Applied Sciences Utrecht, Institute of Movement Sciences, Utrecht, The Netherlands
[4]Physiotherapy Practice Emmastraat, Enschede, The Netherlands
[5]Physiotherapy Practice De Werfheegde, Haaksbergen, The Netherlands
[6]Medical Library, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands
[7]Medical Library, HU University of Applied Sciences Utrecht, Utrecht, The Netherlands
[8]Menzies Health Institute Queensland, Griffith University, Brisbane and Gold Coast, Queensland, Australia
[9]SOMT University of Physiotherapy, Amersfoort, The Netherlands

Contributors JCFK and JM created the search for the databases. TV, EvB, REPMB, MPK, SPJR, JJMP and ALP-G contributed to the screening of titles, abstracts and full texts and the COSMIN grading of the selected studies. TV extracted the data from the studies. TV, EvB, SPJR, MWC and ALP-G drafted the manuscript, which all

authors revised and approved for publication. ALP-G accepts full responsibility for the finished work and/or the conduct of the study as guarantor, had access to the data, and controlled the decision to publish.

Funding This work was supported by the Scientific Committee on Physiotherapy of the Dutch Royal Physiotherapy Association (KNGF). Author TV was supported by the Dutch Organisation for Scientific Research (Nederlandse Organisatie voor Wetenschappelijk Onderzoek-NWO), grant number: 023013042.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information.

ORCID iDs
Tom Vredeveld http://orcid.org/0000-0003-0563-817X
Esther van Benten http://orcid.org/0000-0002-5094-0047
Johannes C F Ket http://orcid.org/0000-0002-1909-3150
Jurgen Mollema http://orcid.org/0000-0001-8941-5867
Stephan P J Ramaekers http://orcid.org/0000-0002-8702-3461
Jan J M Pool http://orcid.org/0000-0002-9240-4488
Michel W Coppieters http://orcid.org/0000-0002-3958-4408
Annelies L Pool-Goudzwaard http://orcid.org/0000-0001-8524-2516

REFERENCES

1 D'Ancona C, Haylen B, Oelke M, *et al*. The International Continence Society (ICS) report on the terminology for adult male lower urinary tract and pelvic floor symptoms and dysfunction. *Neurourol Urodyn* 2019;38:433–77.
2 Irwin DE, Milsom I, Hunskaar S, *et al*. Population-based survey of urinary incontinence, overactive bladder, and other lower urinary tract symptoms in five countries: results of the EPIC study. *Eur Urol* 2006;50:1306–15.
3 Coyne KS, Sexton CC, Thompson CL, *et al*. The prevalence of lower urinary tract symptoms (LUTS) in the USA, the UK and Sweden: results from the epidemiology of LUTS (EpiLUTS) study. *BJU Int* 2009;104:352–60.
4 Egan KB. The epidemiology of benign prostatic hyperplasia associated with lower urinary tract symptoms: prevalence and incident rates. *Urol Clin North Am* 2016;43:289–97.
5 Robertson C, Link CL, Onel E, *et al*. The impact of lower urinary tract symptoms and comorbidities on quality of life: the BACH and UREPIK studies. *BJU Int* 2007;99:347–54.
6 Boyle P, Robertson C, Mazzetta C, *et al*. The relationship between lower urinary tract symptoms and health status: the UREPIK study. *BJU Int* 2003;92:575–80.
7 Chapple CR, Wein AJ, Abrams P, *et al*. Lower urinary tract symptoms revisited: a broader clinical perspective. *Eur Urol* 2008;54:563–9.
8 Irwin DE, Kopp ZS, Agatep B, *et al*. Worldwide prevalence estimates of lower urinary tract symptoms, overactive bladder, urinary incontinence and bladder outlet obstruction. *BJU Int* 2011;108:1132–8.
9 Abrams P, Cardozo L, Fall M, *et al*. The standardisation of terminology of lower urinary tract function: report from the standardisation sub-committee of the international continence society. *Am J Obstet Gynecol* 2002;187:116–26.
10 Braeckman J, Denis L. Management of BPH then 2000 and now 2016 - From BPH to BPO. *Asian J Urol* 2017;4:138–47.
11 Gravas S, Cornu JN, Drake MJ. EAU guidelines on management of non-neurogenic male LUTS guidelines: management of non-neurogenic male lower urinary tract symptoms (LUTS), INCL. benign prostatic obstruction (BPO), Arnhem, EAU Guidelines Office 2018 http://uroweb.org/guidelines/compilations-of-all-guidelines/
12 Strope SA. Evidence-based guidelines in lower urinary tract symptoms secondary to benign prostatic hyperplasia and variation in care. *Curr Opin Urol* 2018;28:262–6.
13 Newman DK, Guzzo T, Lee D, *et al*. An evidence-based strategy for the conservative management of the male patient with incontinence. *Curr Opin Urol* 2014;24:553–9.
14 Mobley D, Feibus A, Baum N. Benign prostatic hyperplasia and urinary symptoms: evaluation and treatment. *Postgrad Med* 2015;127:301–7.
15 Errando-Smet C, Müller-Arteaga C, Hernández M, *et al*. Healthcare resource utilization and cost among males with lower urinary tract symptoms with a predominant storage component in Spain: the epidemiological, cross-sectional MERCURY study. *Neurourol Urodyn* 2018;37:307–15.
16 Malde S, Nambiar AK, Umbach R, *et al*. Systematic review of the performance of noninvasive tests in diagnosing bladder outlet obstruction in men with lower urinary tract symptoms. *Eur Urol* 2017;71:391–402.
17 Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
18 Terwee CB, Jansma EP, Riphagen II, *et al*. Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Qual Life Res* 2009;18:1115–23.
19 Terwee CB, Mokkink LB, Knol DL, *et al*. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res* 2012;21:651–7.
20 van Tulder M, Furlan A, Bombardier C, *et al*. Updated method guidelines for systematic reviews in the Cochrane collaboration back review group. *Spine* 2003;28:1290–9.
21 Aarnink RG, de la Rosette JJ, Huynen AL, *et al*. Standardized assessment to enhance the diagnostic value of prostate volume; Part I: morphometry in patients with lower urinary tract symptoms. *Prostate* 1996;29:317–26.
22 Abdel-Aal A, El-Karamany T, Al-Adl AM, *et al*. Assessment of noninvasive predictors of bladder outlet obstruction and acute urinary retention secondary to benign prostatic enlargement. *Arab J Urol* 2011;9:209–14.
23 Aganovic D, Kulovac B, Bajramovic S, *et al*. Penile compression release index revisited: evaluation and comparison with other noninvasive tools in the prediction of bladder outlet obstruction in men with benign prostatic enlargement. *Med Arch* 2019;73:81.
24 Al-Mosawi HQ, Al-Aridy HM, Hameed HG. Accuracy of intravesical prostate protrusion measured by trans-abdominal ultrasound in predicting bladder outlet obstruction. *Int J Pharm Res* 2020;12:1150–6.
25 Arif M, Groen J, Boevé ER, *et al*. Noninvasive diagnosis of bladder outlet obstruction in patients with lower urinary tract symptoms using ultrasound decorrelation analysis. *J Urol* 2016;196:490–7.
26 Baltaci S, Yagci C, Aksoy H, *et al*. Determination of transition zone volume by transrectal ultrasound in patients with clinically benign prostatic hyperplasia: agreement with enucleated prostate adenoma weight. *J Urol* 2000;164:72–5.
27 Carballido J, Fourcade R, Pagliarulo A, *et al*. Can benign prostatic hyperplasia be identified in the primary care setting using only simple tests? results of the diagnosis improvement in primary care trial. *Int J Clin Pract* 2011;65:989–96.
28 Chan C-K, Yip SK-H, Wu IPH, *et al*. Evaluation of the clinical value of a simple flowmeter in the management of male lower urinary tract symptoms. *BJU Int* 2012;109:1690–6.
29 Chen H, Song B, Li L. A novel noninvasive topological mathematical nomogram for diagnosing male bladder outflow obstruction. *J Biomater Tissue Eng* 2019;9:1205–10.
30 David RA, Badmus TA, Salako AA, *et al*. Diagnostic performance of transrectal ultrasound for prostate volume estimation in men with benign prostate hyperplasia. *Int J Clin Pract* 2020;74:1–8.
31 Demir A, Karadağ MA, Çeçen K, *et al*. Abdominal or transrectal ultrasonographic prostate volume and cystoscopic prostatic

urethral length measurements to determine the surgical technique for prostatectomy in patients with benign prostate hyperplasia. *J Urol Surg* 2016;3:119–22.
32 De Nunzio C, Lombardo R, Gacci M, *et al*. The diagnosis of benign prostatic obstruction: validation of the young academic urologist clinical nomogram. *Urology* 2015;86:1032–6.
33 Güzelsoy M, Kirkali Z. Role of transition zone index in the prediction of clinical benign prostatic hyperplasia. *J Urol Surg* 2016;3:114–8.
34 Harding CK, Robson W, Drinnan MJ, *et al*. An automated penile compression release maneuver as a noninvasive test for diagnosis of bladder outlet obstruction. *J Urol* 2004;172:2312–5.
35 Hossain AKMS, Alam AKMK, Habib AKMK, *et al*. Comparison between prostate volume and intravesical prostatic protrusion in detecting bladder outlet obstruction due to benign prostatic hyperplasia. *Bangladesh Med Res Counc Bull* 2012;38:14–17.
36 Kim SB, Cho I-C, Min SK. Prostate volume measurement by transrectal ultrasonography: comparison of height obtained by use of transaxial and midsagittal scanning. *Korean J Urol* 2014;55:470–4.
37 Kim KS, Choi YS, Bae WJ, *et al*. Comparison of penile cuff test and conventional urodynamic study prior to photoselective vaporization of prostate for benign prostate hyperplasia using a 120 W GreenLight high performance system laser. *J Clin Med* 2020;9:1189.
38 Kojima M, Inui E, Ochiai A, *et al*. Noninvasive quantitative estimation of infravesical obstruction using ultrasonic measurement of bladder weight. *J Urol* 1997;157:476–9.
39 Kwon JK, Han JH, Choi HC, *et al*. Clinical significance of peripheral zone thickness in men with lower urinary tract symptoms/benign prostatic hyperplasia. *BJU Int* 2016;117:316–22.
40 Malemo K, Galukande M, Hawkes M, *et al*. Validation of supra-pubic ultrasonography for preoperative prostate volume measurement in sub-Saharan Africa. *Int Urol Nephrol* 2011;43:283–8.
41 Matzkin H, Greenstein A, Prager-Geller T, *et al*. Do reported micturition symptoms on the American Urological Association questionnaire correlate with 24-hour home uroflowmetry recordings? *J Urol* 1996;155:197–9.
42 Narayanamurthy V, Mishra K, Mahran A, *et al*. Inter-imaging accuracy of computed tomography, magnetic resonance imaging, and transrectal ultrasound in measuring prostate volume compared to the anatomic prostatic weight. *Turk J Urol* 2020;46:50–6.
43 Nathan MS, Seenivasagam K, Mei Q, *et al*. Transrectal ultrasonography: why are estimates of prostate volume and dimension so inaccurate? *Br J Urol* 1996;77:401–7.
44 Oelke M, Höfner K, Jonas U, *et al*. Diagnostic accuracy of noninvasive tests to evaluate bladder outlet obstruction in men: detrusor wall thickness, uroflowmetry, postvoid residual urine, and prostate volume. *Eur Urol* 2007;52:827–35.
45 Prassopoulos P, Charoulakis N, Anezinis P, *et al*. Suprapubic versus transrectal ultrasonography in assessing the volume of the prostate and the transition zone in patients with benign prostatic hyperplasia. *Abdom Imaging* 1996;21:75–7.
46 Rathaus V, Richter S, Nissenkorn I, *et al*. Transperineal ultrasound examination in the evaluation of prostatic size. *Clin Radiol* 1991;44:383–5.
47 Reddy SVK, Shaik AB. Non-invasive evaluation of bladder outlet obstruction in benign prostatic hyperplasia: a clinical correlation study. *Arab J Urol* 2019;17:259–64.
48 Reynard JM, Peters TJ, Lim C, *et al*. The value of multiple free-flow studies in men with lower urinary tract symptoms. *Br J Urol* 1996;77:813–8.
49 Roehrborn CG, Sech S, Montoya J, *et al*. Interexaminer reliability and validity of a three-dimensional model to assess prostate volume by digital rectal examination. *Urology* 2001;57:1087–92.
50 Salinas J, Vírseda M, Arredondo F. Validity of cuff-uroflow as a diagnostic technique for bladder outlet obstruction in males. *Scand J Urol Nephrol* 2003;37:316–21.
51 Steele GS, Sullivan MP, Sleep DJ, *et al*. Combination of symptom score, flow rate and prostate volume for predicting bladder outflow obstruction in men with lower urinary tract symptoms. *J Urol* 2000;164:344–8.
52 Stravodimos KG, Petrolekas A, Kapetanakis T, *et al*. TRUS versus transabdominal ultrasound as a predictor of enucleated adenoma weight in patients with BPH: a tool for standard preoperative work-up? *Int Urol Nephrol* 2009;41:767–71.
53 Styles RA, Neal DE, Powell PH. Reproducibility of measurement of prostatic volume by ultrasound. Comparison of transrectal and transabdominal methods. *Eur Urol* 1988;14:266–9.
54 Su MZ, Lenaghan D, Woo HH. Dichotomous estimation of prostate volume: a diagnostic study of the accuracy of the digital rectal examination. *World J Mens Health* 2013;31:220–5.
55 van Venrooij GE, Boon TA. The value of symptom score, quality of life score, maximal urinary flow rate, residual volume and prostate size for the diagnosis of obstructive benign prostatic hyperplasia: a urodynamic analysis. *J Urol* 1996;155:2014–8.
56 van Venrooij GEPM, Eckhardt MD, Boon TA. Noninvasive assessment of prostatic obstruction in elderly men with lower urinary tract symptoms associated with benign prostatic hyperplasia. *Urology* 2004;63:476–80.
57 Zhou Y, Liu D, Wang J. Application of transabdominal ultrasonography in diagnosis of bladder outlet obstruction in patients with benign prostatic hyperplasia. *J Jilin Univ Med Ed* 2012;38:788–91.
58 Ajayi I, Aremu A, Olajide A, *et al*. Correlation of transrectal and transabodominal ultrasound measurement of transition zone volume with post-operative enucleated adenoma volume in benign prostatic hypertrophy. *Pan Afr Med J* 2013;16:149.
59 al-Rimawi M, Griffiths DJ, Boake RC, *et al*. Transrectal ultrasound versus magnetic resonance imaging in the estimation of prostatic volume. *Br J Urol* 1994;74:596–600.
60 Braeckman JG, Figuera FC, Vanwaeyenbergh JG, *et al*. Reproducibility of transrectal ultrasound of prostatic disease. *Scand J Urol Nephrol Suppl* 1991;137:91–3.
61 Collins GN, Raab GM, Hehir M, *et al*. Reproducibility and observer variability of transrectal ultrasound measurements of prostatic volume. *Ultrasound Med Biol* 1995;21:1101–5.
62 Da Luz Junior PN, Koff WJ, Berger M. Correlation between urinary flow rate, quality of life and international prostate symptom score [i-pss] in patients with benign prostatic hyperplasia (bph). *Braz J Urol* 2001;27:353–7.
63 Ding YY. Use of the Uroflow study in the diagnosis of bladder outlet obstruction in elderly men. *Med J Malays* 1998;53:209–16.
64 Eri LM, Thomassen H, Brennhovd B, *et al*. Accuracy and repeatability of prostate volume measurements by transrectal ultrasound. *Prostate Cancer Prostatic Dis* 2002;5:273–8.
65 Feneley MR, Dunsmuir WD, Pearce J, *et al*. Reproducibility of uroflow measurement: experience during a double-blind, placebo-controlled study of doxazosin in benign prostatic hyperplasia. *Urology* 1996;47:658–63.
66 Kiely EA, Hartnell GG, Gibson RN, *et al*. Measurement of bladder volume by real-time ultrasound. *Br J Urol* 1987;60:33–5.
67 McArdle F, Clarkson B, Robson W, *et al*. Interobserver agreement for noninvasive bladder pressure flow recording with penile cuff. *J Urol* 2009;182:2397–403.
68 Mombelli G, Picozzi S, Messina G, *et al*. Free uroflowmetry versus “do-it-yourself” uroflowmetry in the assessment of patients with lower urinary tract symptoms. *Int Urol Nephrol* 2014;46:1915–9.
69 Nathan MS, Mei Q, Seenivasagam K, *et al*. Comparison of prostatic volume and dimensions by transrectal and transurethral ultrasonography. *Br J Urol* 1996;78:84–9.
70 Okafor PI, Wild SR, Beynon LL, *et al*. Progress in transrectal ultrasonography for prostatic disease. *Br J Urol* 1983;55:721–5.
71 Oranusi C, Nwofor A, Mbonu O. Correlation between international prostate symptom score and uroflowmetry in patients with benign prostatic hyperplasia. *Niger J Clin Pract* 2016;20:454–8.
72 Pascual EM, Polo A, Morales G. Usefulness of bladder-prostate ultrasound in the diagnosis of obstruction/hyperactivity in males with BPH. *Arch Esp Urol* 2011;94:897–903.
73 Pridgeon S, Harding C, Newton D, *et al*. Clinical evaluation of a simple uroflowmeter for categorization of maximum urinary flow rate. *Indian J Urol* 2007;23:114.
74 Porru D, Scarpa RM, Prezioso D, *et al*. Home and office uroflowmetry for evaluation of LUTS from benign prostatic enlargement. *Prostate Cancer Prostatic Dis* 2005;8:45–9.
75 Rademakers KLJ, van Koeveringe GA, Oelke M, *et al*. Ultrasound detrusor wall thickness measurement in combination with bladder capacity can safely detect detrusor underactivity in adult men. *World J Urol* 2017;35:153–9.
76 Singla S, Garg R, Singla A, *et al*. Experience with uroflowmetry in evaluation of lower urinary tract symptoms in patients with benign prostatic hyperplasia. *J Clin Diagn Res* 2014;8:NC01-3.
77 Solomon DJ, Van Niekerk JP. Ultrasonography should replace intravenous urography in the pre-operative evaluation of prostatism. *S Afr Med J* 1988;74:407–8.
78 Stavropoulos N, Christodoulou K, Chamilos E, *et al*. Evaluation of patients with benign prostatic hypertrophy: IVU versus ultrasound. *J R Coll Surg Edinb* 1988;33:140–2.
79 Szabó V, Sobel M, Légrádi J, *et al*. Estimation of the size of prostatic adenoma by ultrasonography. A preliminary report. *Eur Urol* 1976;2:226–8.
80 Tewari A, Indudhara R, Shinohara K, *et al*. Comparison of transrectal ultrasound prostatic volume estimation with magnetic resonance

imaging volume estimation and surgical specimen weight in patients with benign prostatic hyperplasia. *J Clin Ultrasound* 1996;24:169–74.
81 Uzun H, Kadioglu ME, Metin NO, *et al*. The association of postvoiding residual volume, uroflowmetry parameters and bladder sensation. *Urol J* 2019;16:403–6.
82 Zheng J, Pan J, Qin Y, *et al*. Role for intravesical prostatic protrusion in lower urinary tract symptom: a fluid structural interaction analysis study. *BMC Urol* 2015;15:86.
83 Albarqouni L, Sanders S, Clark J, *et al*. Self-Management for men with lower urinary tract symptoms: a systematic review and meta-analysis. *Ann Fam Med* 2021;19:157–67.
84 Malde S, Umbach R, Wheeler JR, *et al*. A systematic review of patients' values, preferences, and expectations for the diagnosis and treatment of male lower urinary tract symptoms. *Eur Urol* 2021;79:796–809.
85 Mokkink LB, Terwee CB, Patrick DL, *et al*. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol* 2010;63:737–45.
86 Nitti VW. Pressure flow urodynamic studies: the gold standard for diagnosing bladder outlet obstruction. *Rev Urol* 2005;7:S14–21.
87 Sech S, Montoya J, Girman CJ, *et al*. Interexaminer reliability of transrectal ultrasound for estimating prostate volume. *J Urol* 2001;166:125–9.
88 Walsh T. Fuzzy gold standards: approaches to handling an imperfect reference standard. *J Dent* 2018;74:S47–9.
89 Gagnier JJ, Moher D, Boon H, *et al*. Investigating clinical heterogeneity in systematic reviews: a methodologic review of guidance in the literature. *BMC Med Res Methodol* 2012;12:1.
90 Leeflang MMG, Rutjes AWS, Reitsma JB, *et al*. Variation of a test's sensitivity and specificity with disease prevalence. *Can Med Assoc J* 2013;185:E537–44.
91 Leeflang MMG. Systematic reviews and meta-analyses of diagnostic test accuracy. *Clin Microbiol Infect* 2014;20:105–13.
92 Spangberg A, Ahlstrand C, Bergqvist D. *Benign prostatic obstruction. A systematic review*. Stockholm: Swedish Council on Health Technology Assessment, 2011.
93 Chappe C, Bilington A, Joachim P. *The management of lower urinary tract symptoms in men*. 1st edn. London: National Clinical Guideline Centre at The Royal College of Physicians, 2010.
94 Andersen CA, Holden S, Vela J, *et al*. Point-of-care ultrasound in general practice: a systematic review. *Ann Fam Med* 2019;17:61–9.
95 D'Silva KA, Dahm P, Wong CL. Does this man with lower urinary tract symptoms have bladder outlet obstruction?: the rational clinical examination: a systematic review. *JAMA J Am Med Assoc* 2014;312:535–42.
96 D'Agate S, Wilson T, Adalig B, *et al*. Impact of disease progression on individual IPSS trajectories and consequences of immediate versus delayed start of treatment in patients with moderate or severe LUTS associated with BPH. *World J Urol* 2020;38:463–72.
97 Bossuyt PM, Irwig L, Craig J, *et al*. Comparative accuracy: assessing new tests against existing diagnostic pathways. *BMJ* 2006;332:1089–92.
98 Bossuyt PM, Reitsma JB, Bruns DE. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ* 2015;351:1–9.
99 Mokkink LB, Prinsen CA, Patrick DL. COSMIN methodology for systematic reviews of Patient-Reported Outcome Measures (PROMs) - user manual, 2018. Available: https://cosmin.nl/wp-content/uploads/COSMIN-syst-review-for-PROMs-manual_version-1_feb-2018.pdf [Accessed 15 Dec 2021].
100 Prinsen CAC, Mokkink LB, Bouter LM, *et al*. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res* 2018;27:1147–57.

Supplementary material

Supplementary Table 1A. Search strategy for Embase.com.

/exp = EMtree keywords exploded
/de = EMtree keywords not exploded
:ab,ti = words in title or abstract

Definition	Search	Query	Items found
Combination	#4	#1 AND #2 AND #3	**4.110**
Assessment methods	#3	'echography'/exp OR 'diagnostic imaging'/de OR 'bladder function'/exp OR 'urine flow rate'/exp OR 'urethra function'/exp OR 'postvoid residual urine volume'/exp OR 'urine volume'/exp OR 'digital rectal examination'/exp OR ultraso*:ab,ti,kw OR echograph*:ab,ti,kw OR urodynam*:ab,ti,kw OR 'urinary flow*':ab,ti,kw OR uroflow*:ab,ti,kw OR ((urethra* NEAR/3 pressure):ab,ti,kw) OR (('pressure flow*' NEAR/3 urine):ab,ti,kw) OR ((urin* NEAR/3 volum*):ab,ti,kw) OR ((pressure NEAR/3 profilometr*):ab,ti,kw) OR ((obstruct* NEAR/3 (urine OR outflow* OR filling)):ab,ti,kw) OR 'rectal examination*':ab,ti,kw OR ((size:ab,ti,kw OR diameter*:ab,ti,kw OR boundar*:ab,ti,kw OR weight*:ab,ti,kw OR volume*:ab,ti,kw) AND ('diagnosis'/de OR assess*:ab,ti,kw OR measur*:ab,ti,kw OR diagnos*:ab,ti,kw OR evaluat*:ab,ti,kw)) OR uroobstruct*:ab,ti,kw	**3.227.711**
Measurement properties of assessment methods	#2	'measurement precision'/exp OR 'measurement accuracy'/exp OR 'measurement repeatability'/exp OR 'diagnostic accuracy'/exp OR 'diagnostic error'/exp OR 'diagnostic test accuracy study'/exp OR 'observer variation'/exp OR 'quality control procedures'/exp OR 'receiver operating characteristic'/exp OR reliab*:ab,ti,kw OR validity:ab,ti,kw OR reproduc*:ab,ti,kw OR (feasibility:ab,ti,kw NOT 'feasibility study'/exp) OR 'internal consistenc*':ab,ti,kw OR 'observer variation*':ab,ti,kw OR 'intraobserver variation*':ab,ti,kw OR 'interobserver variation*':ab,ti,kw OR 'observer variabilit*':ab,ti,kw OR 'interobserver variabilit*':ab,ti,kw OR 'intraobserver variabilit*':ab,ti,kw OR 'measurement error*':ab,ti,kw OR interpretability:ab,ti,kw OR accura*:ab,ti,kw OR 'limit of detection':ab,ti,kw OR 'detection limit':ab,ti,kw OR 'detection limits':ab,ti,kw OR 'roc curve':ab,ti,kw OR 'roc curves':ab,ti,kw OR 'roc analysis':ab,ti,kw OR 'roc analyses':ab,ti,kw OR 'receiver operating characteristic*':ab,ti,kw OR sensitivit*:ab,ti,kw OR specificit*:ab,ti,kw OR properties:ab,ti,kw OR responsive*:ab,ti,kw OR 'false positive':ab,ti,kw OR 'false negative':ab,ti,kw OR roc:ab,ti,kw OR likelyhood*:ab,ti,kw OR likelihood*:ab,ti,kw	**6.744.869**
Population	#1	'prostate hypertrophy'/exp OR bph:ab,ti,kw OR (('prostate'/exp OR prostat*:ab,ti,kw OR voiding:ab,ti,kw) AND ('hyperplasia'/de OR 'hypertrophy'/de OR hyperplas*:ab,ti,kw OR obstruct*:ab,ti,kw OR hypertroph*:ab,ti,kw OR enlarge*:ab,ti,kw OR pressure) AND flow*:ab,ti,kw) OR (('bladder outlet' NEAR/3 obstruction*):ab,ti,kw)	**50.061**

Supplementary Table 1B. Search strategy for PubMed.

[Mesh] = Medical subject headings (MeSH)
[Mesh:NoExp] = Medical subject headings (MeSH), without explosion
[tiab] = words in title or abstract

Definition	Search	Query	Items found
Combination	#4	#1 AND #2 AND #3	2.883
Assessment methods	#3	"Ultrasonography"[Mesh] OR "diagnostic imaging"[Subheading] OR "Urodynamics"[Mesh] OR ultraso*[tiab] OR echograph*[tiab] OR urodynam*[tiab] OR urinary flow*[tiab] OR uroflow*[tiab] OR urethral pressure[tiab] OR (pressure flow*[tiab] AND urine[tiab]) OR urine volum*[tiab] OR pressure profilometr*[tiab] OR (obstruct*[tiab] AND (urine[tiab] OR outflow*[tiab] OR filling[tiab])) OR rectal examination*[tiab] OR uroobstruct*[tiab] OR ((size[tiab] OR diameter*[tiab] OR boundar*[tiab] OR weight*[tiab] OR volume*[tiab]) AND ("diagnosis"[Subheading] OR assess*[tiab] OR measur*[tiab] OR diagnos*[tiab] OR evaluat*[tiab]))	2.826.341
Measurement properties of assessment methods	#2	"Reproducibility of Results"[Mesh] OR reliab*[tiab] OR validity[tiab] OR reproduc*[tiab] OR responsive*[tiab] OR (feasibility[tiab] NOT "Feasibility Studies"[Mesh]) OR internal consistenc*[tiab] OR "Observer Variation"[Mesh] OR observer variation*[tiab] OR intraobserver variation*[tiab] OR interobserver variation*[tiab] OR observer variabilit*[tiab] OR interobserver variabilit*[tiab] OR intraobserver variabilit*[tiab] OR measurement error*[tiab] OR interpretability[tiab] OR "Sensitivity and Specificity"[Mesh] OR accura*[tiab] OR "limit of detection"[tiab] OR "detection limit" [tiab] OR "detection limits"[tiab] OR "roc curve"[tiab] OR "roc curves"[tiab] OR "roc analysis"[tiab] OR "roc analyses"[tiab] OR "receiver operating characteristic"[tiab] OR "receiver operating characteristics"[tiab] OR sensitivity[tiab] OR specificity[tiab] OR prognos*[tiab] OR properties[tiab] OR responsive*[tiab]	4.891.228
Population	#1	"Prostatic Hyperplasia"[Mesh] OR bph[tiab] OR (("Prostate"[Mesh] OR prostat*[tiab] OR voiding[tiab]) AND ("Hyperplasia"[Mesh] OR "Hypertrophy"[Mesh] OR hyperplas*[tiab] OR obstruct*[tiab] OR hypertroph*[tiab] OR enlarge*[tiab] OR pressure flow*[tiab])) OR bladder outlet obstruction*[tiab]	40.128

Supplementary Table 1C. Search strategy for Ebsco/CINAHL.

MH = Mapped Heading keyword
TI = words in title
AB = words in abstract

Definition	Search	Query	Items found
Combined	S8	S3 AND S4 AND S7	231
	S7	S1 OR S6	5.139
	S6	S2 AND S5	3.341
Population	S5	((MH "Hypertrophy") OR (MH "Hyperplasia")) OR TI (hyperplas* OR obstruct* OR hypertrophy* OR enlarge* OR “pressure flow*”) OR AB (hyperplas* OR obstruct* OR hypertrophy* OR enlarge* OR “pressure flow*”) OR KW (hyperplas* OR obstruct* OR hypertrophy* OR enlarge* OR “pressure flow*”)	85.259
Assessment methods	S4	((MH "Endosonography") OR (MH "Ultrasonography") OR (MH "Diagnostic Imaging") OR (MH "Urination") OR (MH "Urodynamics") OR (MH "Digital Rectal Examination")) OR TI (ultraso* OR echograph* OR urodynam* OR “urinary flow*” OR uroflow* OR “urethral pressure*” OR (“pressure flow*” N3 (urine OR urinary)) OR “urine volum*” OR “pressure profilometr*” OR (obstruct* N3 (urine OR outflow* OR filling)) OR “rectal examination*” OR uroobstruct* OR ((size OR diameter* OR boundar* OR weight* OR volume*) N3 (assess* OR measur* OR diagnos* OR evaluat*))) OR AB (ultraso* OR echograph* OR urodynam* OR “urinary flow*” OR uroflow* OR “urethral pressure*” OR (“pressure flow*” N3 (urine OR urinary)) OR “urine volum*” OR “pressure profilometr*” OR (obstruct* N3 (urine OR outflow* OR filling)) OR “rectal examination*” OR uroobstruct* OR ((size OR diameter* OR boundar* OR weight* OR volume*) N3 (assess* OR measur* OR diagnos* OR evaluat*)) OR KW (ultraso* OR echograph* OR urodynam* OR “urinary flow*” OR uroflow* OR “urethral pressure*” OR (“pressure flow*” N3 (urine OR urinary)) OR “urine volum*” OR “pressure profilometr*” OR (obstruct* N3 (urine OR outflow* OR filling)) OR “rectal examination*” OR uroobstruct* OR ((size OR diameter* OR boundar* OR weight* OR volume*) N3 (assess* OR measur* OR diagnos* OR evaluat*)))	498.906
Measurement and psychometric properties of measurement instruments	S3	((MH "Reproducibility of Results") OR (MH "Sensitivity and Specificity") OR (MH "Observer Bias")) OR TI (feasibility OR reliab* OR validity OR reproduc* OR responsive* OR “internal consistenc*” OR “observer variation*” OR “intraobserver variation*” OR “interobserver variation*” OR “observer variabilit*” OR “interobserver variabilit*” OR “intraobserver variabilit*” OR “measurement error*” OR interpretability OR accura* OR "limit of detection" OR "detection limit" OR "detection limits" OR "roc curve" OR "roc curves" OR "roc analysis" OR "roc analyses" OR "receiver operating characteristic" OR "receiver operating characteristics" OR sensitivity OR specificity OR prognos* OR properties OR responsive*) OR AB (feasibility OR reliab* OR validity OR reproduc* OR responsive* OR “internal consistenc*” OR “observer variation*” OR “intraobserver variation*” OR “interobserver variation*” OR “observer variabilit*” OR “interobserver variabilit*” OR “intraobserver variabilit*” OR “measurement error*” OR interpretability OR accura* OR "limit of detection" OR "detection limit" OR "detection limits" OR "roc curve" OR "roc curves" OR "roc analysis" OR "roc analyses" OR "receiver operating characteristic" OR "receiver operating characteristics" OR sensitivity OR specificity OR prognos* OR properties OR responsive*) OR KW (feasibility OR reliab* OR validity OR reproduc* OR responsive* OR “internal consistenc*” OR “observer variation*” OR “intraobserver variation*” OR “interobserver variation*” OR “observer variabilit*” OR “interobserver variabilit*” OR “intraobserver variabilit*” OR “measurement error*” OR interpretability OR accura* OR "limit of detection" OR "detection limit" OR "detection limits" OR "roc curve" OR "roc curves" OR "roc analysis" OR "roc analyses" OR "receiver operating characteristic" OR "receiver operating characteristics" OR sensitivity OR specificity OR prognos* OR properties OR	676.302

		responsive*)	
Population	S2	(MH "Prostate") OR TI (prostat* OR voiding) OR AB (prostat* OR voiding) OR KW (prostat* OR voiding)	43.361
Population	S1	(MH "Prostatic Hypertrophy") OR TI (bph OR "bladder outlet obstruction*") OR AB (bph OR "bladder outlet obstruction*") OR KW (bph OR "bladder outlet obstruction*")	4.015

Supplementary Table 2A. Reliability of assessment methods to evaluate bladder outlet obstruction and benign prostatic obstruction.

First author	Year	Aim of method[1)]	Patient category	Sample, n (%)	Age, mean (sd; min-max)	Measures of reliability
Assessment methods for bladder outlet obstruction: uroflowmetry at home.						
Chan[28]	2012	To assess maximum urinary flow to predict BOO.	Men with LUTS attributable to BPH.	186	65.5 (7; -)	Agreement (Kappa values) of home flowmetry scores: One, three, five, seven, nine, ten measurements in agreement with scores by the criterion (electronic uroflowmetry) 0.76, 0.79, 0.78, 0.80, 0.83, 0.84. If adjusted criterion scores are used (Qmax as ordinal categories: >19 mL/s, 15–19 mL/s, 10–15 mL/s, and <10 mL/s)': One, three, five, seven, nine, ten measurements in agreement with scores: 0.65, 0.70, 0.67, 0.70, 0.72, 0.74.
Assessment methods for benign prostatic obstruction: transabdominal ultrasound.						
Prassopoulos[45]	1996	Estimation of prostate size and transitional zone volume.	Men with BPH.	95	69.7 (11.3; 47-85)	Interobserver 'error' was 5% calculating prostate volume. Transition zone of the prostate measurement error: "less than 4%".
Assessment methods for benign prostatic obstruction: digital rectal examination.						
Roehrborn[49]	2001	Estimation of prostate volume.	Volunteers from a general urology practice.	121	60.7 (10.3; -)	Overall ICC and ICC for <40cm^3 prostate volume and >40cm^3 prostate volume. Plus grading system: Overall ICC: 0.58 (95% CI: 0.45-0.69) ICC: <40cm^3: 0.44 (95% CI: 0.28-0.58), ICC: >40cm^3: 0.52 (95% CI: 0.05-0.80) Textual Scale: Overall ICC: 0.65 (95% CI: 0.53-0.74) ICC: <40cm^3: 0.53 (95% CI: 0.38-0.65), ICC: >40cm^3: 0.63 (95% CI: 0.21-0.85) Best estimate in gr: Overall ICC: 0.78 (95% CI: 0.70-0.84) ICC: <40cm^3: 0.64 (95% CI: 0.52-0.74), ICC: >40cm^3: 0.66 (95% CI: 0.26-0.86) Sizing balls: Overall ICC: 0.64 (95% CI: 0.51-0.75) ICC: <40cm^3: 0.50 (95% CI: 0.33-0.64), ICC: >40cm^3: 0.66 (95% CI: 0.22-0.87) Concentric rings: Overall ICC: 0.63 (95% CI: 0.49-0.74) ICC: <40cm^3: 0.50 (95% CI: 0.33-0.64), ICC: >40cm^3: 0.59 (95% CI: 0.11-0.84) Lever device: Overall ICC: 0.68 (95% CI: 0.57-0.76) ICC: <40cm^3: 0.52 (95% CI: 0.35-0.66), ICC: >40cm^3: 0.51 (95% CI: 0.23-0.71) Full 3D model: Overall ICC: 0.78 (95% CI: 0.68-0.85) ICC: <40cm^3: 0.67 (95% CI: 0.52-0.77), ICC: >40cm^3: 0.79 (95% CI: 0.28-0.95) Final 3D model: Overall ICC: 0.86 (95% CI: 0.75-0.93) ICC: <40cm^3: 0.83 (95% CI: 0.69-0.92), ICC: >40cm^3: 0.69 (95% CI: 0.0-0.94)
Assessment methods for benign prostatic obstruction: transrectal ultrasound.						
Kwon[39]	2016	Estimation of peripheral zone thickness and related prostate size	Men with LUTS/BPH.	1009	62.0 (10.0; -)	ICC for peripheral zone thickness for inter-rater agreement (two raters): 0.896 (95% CI: 0.883-0.908)

		parameters.				
Prassopoulos[45]	1996	Estimation of prostate size and transitional zone volume.	Men with BPH.	95	69.7 (11.3; 47-85)	Interobserver 'error' was 4% calculating prostate volume. Transition zone of the prostate measurement error: "less than 4%".

BOO = bladder outlet obstruction, BPH = benign prostatic hyperplasia, BPO = benign prostatic obstruction, ICC = intraclass correlation coefficient, LUTS = lower urinary tract symptoms, sd = standard deviation, 95% CI = 95% confidence interval, [1)] = Aim of method extracted from study and summarized by review authors.

Supplementary Table 2B. Validity of assessment methods – Assessment methods for bladder outlet obstruction.

First author	Year	Aim of method[1)]	Reference test	Patient category	Sample, n (%)	Age, mean (sd; min-max)	**Measures of criterion validity**: sensitivity, specificity, area under the curve (95% confidential interval) **and construct validity:** correlations (p or 95% confidential interval).
Transrectal ultrasound (TRUS)							
Kwon [39]	2016	Estimation of peripheral zone thickness and related prostate size parameters.	Urinary flow parameters from uroflowmetry.	Men with LUTS/BPH.	1009	62.0 (10.0; -)	Correlation TRUS prostate parameters – maximum urinary flow rate (Qmax). Total prostate volume – Qmax: r=-0.200 (p<0.01) Transitional zone volume – Qmax: r=-0.219 (p<0.01) Transitional zone index – Qmax: r=-0.196 (p<0.01) Peripheral zone thickness – Qmax: r=0.140 (p<0.01) Correlation TRUS prostate parameters – postvoid residue (PVR). Total prostate volume – PVR: r=0.214 (p<0.01) Transitional zone volume – PVR: r=0.236 (p<0.01) Transitional zone index – PVR: r=0.192 (p<0.01) Peripheral zone thickness – PVR: r=-0.154 (p<0.01)
Oelke [44]	2007	To detect BOO.	Computer urodynamic investigation, obstruction based on CHESS classification.	Men aged 40 year and older with clinical BPH, LUTS and/or prostate volume greater than 25ml.	162	median: 62 (min-max: 40-89)	Prostate volume, cut-off obstructed: ≤25 mL / >25 mL SN: 85% (95% CI: 77-93%), SP: 27% (95% CI: 18-36%), PPV: 51% (95% CI: 42-60%), NPV: 67% (95% CI: 51-83%). Diagnostic accuracy: 54% LR+: 1.16 (95% CI: 0.99-1.37), LR-: 0.56 (95% CI: 0.29-0.98) Prostate volume with obstruction: Obstructed, prostate volume in ml: median: 40 (quartiles: 29-58), Non-obstructed, prostate volume in ml: median: 32.9 (quartiles: 22-44) p-value: 0.014 ROC-AUC: 0.62 (95% CI: 0.52–0.71)
Steele [51]	2000	To predict BOO.	Multichannel urodynamic studies to obtain obstruction grade through the ICS nomogram. >2cm water per ml/s and detrusor pressure >40 cm water was defined as obstructed.	Men with LUTS.	204	66.7 (7.5; -)	Predicting bladder outlet obstruction based on prostate volume: Cut-off: ≥40gram for obstruction: SN: 0.66, SP: 0.64. Cut-off: <40 gram for non-obstruction: SN: 0.43, SP: 0.83, PPV: 0.42, NPV: 0.81. Cut-off: <25 gram for non-obstruction: SN: 0.21, SP: 0.92, PPV: 0.50, NPV 0.77.
Venrooij [55]	1996	To detect BOO and correlate	Urodynamic studies,	Men with prostatism, with	196	65.8 (7.1; 51-86)	Prostate volume Pearson's correlation:

		the prostate volume with BOO related parameters.	based on Schäfer's grade, with a classification of 0 and 1 defined as non-obstructed and ≥2 as obstructed.	and without urodynamic obstruction / possible BPH.			Prostate volume - Maximal flow: -0.19 (p=0.008) Prostate volume - Residual volume: 0.12 (not sign.) Prostate volume - Schäfer's obstruction grade: 0.29 (p<0.001) Kendall & Gibbon's correlation: Prostate volume – maximal flow: -0.11 (p=0.02) Prostate volume – residual volume: 0.05 (not significant) Prostate volume – Schäfer's obstruction grade: 0.22 (p=0.001) Note by review authors: *The authors of the study mention some variables showed a non-normal distribution and analysed the Kendall & Gibbon's correlation. In the review, we assumed the Kendall & Gibbon's correlation to be most accurate.*
Venrooij [56]	2004	To discriminate between obstructed and non-obstructed men.	Cystometry and pressure-flow studies. Analyzed according to the International Continence Society Nomogram, Schäfer's obstruction grade and URA.	Men with LUTS, suggestive of BPH.	160	65.1 (8.3; 50-85)	Kendall's and Gibbons correlation with: Abrams-Griffiths number / urethral resistance factor / Schäfer's obstruction grade. Prostate volume: 0.27 (p≤0.01) / 0.26 (p≤0.01) / 0.29 (p≤0.01)
Transabdominal ultrasound (TAUS)							
Abdel-Aal [22]	2011	To detect BOO	Pressure flow studies in patients presenting LUTS suggestive of BPO. Based on BOOI, >40 = obstructed, 20-40 = equivocal, <20 = no obstruction	Men presenting with LUTS, suggestive of benign prostatic enlargement	135	No BOO: 58.9, (4.4; 52-71) BOO: 58.4 (6.5; 50-72)	Cut-off value prostate volume >45 mL to predict obstruction: SN: 85.7%, SP: 26%, PPV: 48.6%, NPV: 72.2% Diagnostic accuracy: 50.6 ROC-AUC: 0.678 (95% CI: 0.562-0.794) LR+: 1.16, LR-: 0.549 Cut-off value intravesical prostatic protrusion >8mm to predict obstruction: SN: 80%, SP: 80%, PPV: 73.7%, NPV: 85.1% Diagnostic accuracy: 80 ROC-AUC: 0.885 (95% CI: 0.806-0.963) LR+: 4, LR-: 0.25 Spearman correlation with Bladder Outlet Obstruction Index IPP - BOOI: r = 0.595 (p<0.001) PV - BOOI: r = 0.241 (p=0.02)

Al-Mosawi [24]	2020	To detect BOO based on Intravesical prostate protrusion.	Urodynamic studies, based on BOOI index: <20 = non-obstructed, 20-40 = inconclusive, >40 = obstructed.	Men exhibiting LUTS, with confirmed BPH	63	53 (- ; 41-80)	Cut-off value IPP: >10mm or ≤10mm Note by review authors: *unclear if ≥ or > and < or ≤, deducted from text it appears it should be: ≤10mL and >10mL* Compared to BOOI obstructed and unobstructed (inconclusive (or: unequivocal) as non-obstructed) SN: 81.6% (95% CI: 65.7-92.3%), SP: 40% (95% CI: 21.1-61.3%) PPV: 67.4% (95% CI: 59.2-74.7%), NPV: 58.8% (95% CI: 38.5-76.5%) Accuracy: 65.1% (95% CI: 52-76.7%) Cut-off value PV: >40mL or <40mL Note by review authors: *unclear if ≥ or > and < or ≤, deducted from text it appears it should be: ≤40mL and >40mL.* Compared to BOOI obstructed and unobstructed (inconclusive (or: unequivocal) as non-obstructed) SN: 55.8% (95% CI: 38.3-71.4%), SP: 40.0% (95% CI: 21.1-61.3%) PPV: 58.3% (95% CI: 47.7-68.3%), NPV: 37.04% (95% CI: 24.5-62.1%) Accuracy: 37.04% (95% CI: 24.5-62.1%)
Hossain [35]	2012	Estimation of prostate volume and intravesical prostatic protrusion and diagnose BOO	Pressure flowmetry with BOO-index (BOOI): <40 BOOI = non-obstructed, >40 BOOI = obstructed.	Men with LUTS, suggestive of BPH.	50	64.3 (- ; 51-78)	Mean prostate volume non-obstructed group: 33.7 mL (sd:10.5) – obstructed group: 44.03 mL (sd: 14.32) p: <0.05. Prostate volume: ≥40 mL to predict obstruction: SN: 57.69%, SP: 66.67%, PPV: 65.21%, NPV: 59.26%. ROC-AUC: 0.70 Intravesical prostatic protrusion >10 mm to predict obstruction: SN: 69.23%. SP: 79.17%, PPV: 78.26%, NPV: 70.37%. ROC-AUC: 0.821 Spearman correlation with Bladder Outlet Obstruction Index (BOOI): PV - BOOI: 0.399 IPP - BOOI: 0.691
Kojima [38]	1997	Estimation of bladder weight as predictor of infravesical obstruction.	Pressure flowmetry to obtain a Abrams-Griffiths number, 40 was the cut-off value for obstructed and unobstructed. Grade of	Men with BPH, and a moderate to severe urinary symptoms score from the American Urological Association symptom index.	65	75 (- ; 45-89)	Bladder weight: >35 gram – ≤35 gram, compared to obstruction and no obstruction: SN: 85.3%, SP: 87.1%, PPV: 87.9%, NPV: 84.4% False-positive rate: 12.1%, False-negative rate: 15.6% Diagnostic accuracy: 86.2%

			obstruction through Schäfer's nomogram.				
Reddy [47]	2019	To detect BOO through prostate volume and IPP grade	Pressure–flow evaluation was done in all patients to calculate BOO index BOOI: >40 BOO BOOI: 20–40 equivocal, BOOI: <20 no BOO	Men with LUTS due to clinically diagnosed BPH	164	66 (9.88; -)	IPP vs BOOI *(article reports values as means)* IPP Grade I: <5 mm, grade II: 5–10 mm, grade III: >10 mm IPP Grade I - BOOI: 26.6 (sd: 11.29) IPP Grade II - BOOI: 33.93 (sd: 7.99) IPP Grade III - BOOI: 52.19 (sd: 14.51), p<0.001 IPP grade vs Maximum Flow Rate (Qmax) *(article reports values as means)* IPP Grade I: 10.31 (sd: 3.49) mL/s IPP Grade II: 8.46 (sd: 3.62) mL/s IPP Grade III: 7.29 (sd: 3.16) mL/s, p<0.001 IPP grade vs Pdet at Qmax (PdetQmax, cmH2O) *(article reports values as means)* IPP Grade I: 47.22 (sd: 18.27) cmH2O IPP Grade II: 50.85 (sd: 15.23) cmH2O IPP Grade III: 66.77 (sd: 30.83) cmH2O, p<0.001 Note by review authors: *It appears heavily skewed data were used, based on scatterplots, to calculate the following statistics.* Pearson correlation: IPP correlation with BOOI: r = 0.586 (p<0.001) PV correlation with BOOI: r = 0.374 (p=0.001) IPP on identification of BOO IPP Grade I and IPP Grade II combined and Grade III SN: Grade I/II: 92.21% (95% CI: 83.41%-96.13%), Grade III: 65.06% (95% CI: 53.81%-75.20%) SP: Grade I/II: 45.21% (95% CI: 33.52%-57.30%), Grade III: 84.93% (95% CI: 74.64%-92.23%) PPV: Grade I/II: 67.48% (95% CI: 58.45%-75.65%), Grade III: 83.08% (95% CI: 71.73%-91.24%) NPV: Grade I/II: 80.49% (95% CI: 65.13%-91.18%), Grade III: 68.13% (95% CI: 57.53%-77.51%) ROC-AUC: IPP by TAUS: 0.791, p<0.001 PV by TAUS: 0.658, p=0.002
Zhou [57]	2012	To assess BOO and intravesical prostatic protrusion	Urodynamic studies, based on BOO Index >40 as obstructed.	Men with BPH.	124	73 (- ; 54-89)	Intravesical prostatic protrusion assessed in mm Obstruction: mean: 11.05 (sd: 9.65) Non-obstruction: mean: 7.67 (sd: 7.99) Difference: (p<0.05) Spearman correlation intravesical prostatic protrusion – maximum urinary flow rate: r=-0.403, p<0.01 Spearman correlation IPP-detrusor pressure at Qmax: r=0.192, p<0.01
Transperineal ultrasound free uroflowmetry							

Arif [25]	2016	Estimation of BOO through ultrasound flowmetry.	Transperineal ultrasound catheter flowmetry	Men with LUTS, suggestive of BOO.	45	-	Bladder outlet obstruction index cut-off score: ≤40: labelled as unobstructed, >40 labelled as obstructed. SN: 88%, SP: 95% Detection of obstruction ROC-AUC: 0.961
Uroflowmetry at home							
Chan [28]	2012	To assess maximum urinary flow to predict BOO.	Electronic flowmetry in clinic and international prostate symptom score.	Men with LUTS attributable to BPH.	186	65.5 (7; -)	Home uroflowmeter 'bottom' compartment, reference: mean Qmax from uroflowmetry at clinic visit: <10 mL/s SN: 0.79 (95% CI: 0.68-0.87), SP: 0.90 (95% CI: 0.83-0.94) LR+: 7.56 (95% CI: 4.34-13.09), LR-: 0.24 (95% CI: 0.15-0.37) DOR: 32.04 (95% CI: 14.03-73.19) Home uroflowmeter 'middle' compartment, reference: mean Qmax from uroflowmetry at clinic visit: <15 mL/s SN: 0.95 (95% CI: 0.91-0.98), SP: 0.81 (95% CI: 0.69-0.89) LR+: 5.06 (95% CI: 2.89-8.86), LR-: 0.06 (95% CI: 0.03-0.12) DOR: 91.02 (95% CI: 31.23-265.23) Home uroflowmeter 'top' compartment, reference: mean Qmax from uroflowmetry at clinic visit: <19 mL/s SN: 0.99 (95% CI: 0.97-1.00), SP: 0.68 (95% CI: 0.47-0.84) LR+: 3.12 (95% CI: 1.69-5.76), LR-: 0.01 (95% CI: 0.00-0.06) DOR: 349.23 (95% CI: 40.24-3037.7)
Uroflowmetry							
Chen [29]	2019	Estimation of bladder outlet obstruction through uroflowmetry related C/Q nomogram	Urodynamic studies, based on Abram-Griffiths number. Used cut-off values not provided for obstructed, equivocal and not obstructed scores	Men with LUTS who underwent cystometry.	522	-	C/Q nomogram with P/Q urodynamic studies (UDS) Kappa value of C/Q Nomogram with urodynamic studies: 0.693 (p=0.000) SN: 0.81, SP: 0.91, PPV: 0.79, NPV: 0.84 ROC-AUC: 0.86 C/Q nomogram compared with uroflowmetry flow rate diagnosis: ≤10 ml/s: obstructed, ≥ 15 ml/s: unobstructed, 'remaining scores': equivocal. Kappa value of flow rate with urodynamic studies. 0.528 (p=0.000) SN: 0.71, SP: 0.85, PPV: 0.69, NPV: 0.80 ROC-AUC: 0.78
Oelke [44]	2007	To detect BOO	Computer urodynamic investigation, obstruction based on CHESS classification.	Men aged 40 year and older with clinical BPH, LUTS and/or prostate volume greater than 25ml.	162	Median: 62 (min-max: 40-89)	Qmax, cut-off nonobstructive/obstructive: ≥15 / <15 ml/s SN: 99% (95% CI: 96-101%), SP: 39% (95% CI: 28-49%) PPV: 59% (95% CI: 50-67%), NPV: 97% (95% CI: 91-103%) Diagnostic accuracy: 67% LR+: 1.61 (95% CI: 1.36-1.91), LR-: 0.03 (95% CI: 0-4.42) Qmax, cut-off nonobstructive/obstructive: ≥10 / <10 ml/s SN: 68% (95% CI: 57-78%), SP: 73% (95% CI: 63-82%)

							PPV: 69% (95% CI: 58-79%), NPV: 72% (95% CI: 63-82%) Diagnostic accuracy: 70% LR+: 2.5 (95% CI: 1.7-3.68), LR-: 0.44 (95% CI: 0.31-3.2) Qaverage, cut-off nonobstructive/obstructive ≥7 / <7ml/s SN: 89% (95% CI: 82-96%), SP: 46% (95% CI: 35-56%) PPV: 59% (95% CI: 50-68%), NPV: 83% (95% CI: 72-94%) Diagnostic accuracy: 66% LR+: 1.65 (95% CI: 1.34-2.04), LR-: 0.23 (95% CI: 0.12-1.98) ROC-AUC to detect bladder outlet obstruction: Qmax: 0.84 (95% CI: 0.78-0.91) Qaverage: 0.82 (95% CI: 0.75-0.89)
Reynard [48]	1996	To detect BOO	Pressure-flow studies using the Abrams-Griffiths nomogram to identify obstructed from non-obstructed/equivocal men	Men with LUTS, suggestive of BPO	157	Median: 68 (min-max: 50-84) *based on total sample n=165, 8 did not undergo index or reference test.*	To detect BOO, from Qmax < 8 mL/s based on: 1 void / means from 2 voids / means from 3 voids or means from 4 voids: SN: 14% / 35% / 18% / 14% SP: 85% / 97% / 98% / 98% PPV: 82% / 94% / 94% / 92% NPV: 50% / 49% / 44% / 42% To detect BOO, from Qmax < 10 mL/s based on: 1 void / means from 2 voids / means from 3 voids or means from 4 voids: SN: 71% / 49% / 39% / 29% SP: 71% / 87% / 94% / 96% PPV: 79% / 85% / 90% / 93% NPV: 61% / 53% / 50% / 47% To detect BOO, from Qmax < 12 mL/s based on: 1 void / means from 2 voids / means from 3 voids or means from 4 voids: SN: 84% / 65% / 56% / 50% SP: 50% / 74% / 87% / 91% PPV: 72% / 79% / 87% / 90% NPV: 67% / 58% / 56% / 53% To detect BOO, from Qmax < 15 mL/s based on: 1 void / means from 2 voids / means from 3 voids or means from 4 voids: SN: 95% / 85% / 80% / 76% SP: 35% / 53% / 61% / 67% PPV: 69% / 74% / 76% / 78% NPV: 81% / 70% / 67% / 63%
Venrooij [56]	2004	To discriminate between obstructed and non-obstructed men.	Cystometry and pressure-flow studies. Analyzed according to	Men with LUTS, suggestive of BPH.	160	65.1 (8.3; 50-85)	Kendall's and Gibbons correlation with: Abrams-Griffiths number / urethral resistance factor / Schäfer's obstruction grade. Maximal free flow rate: -0.41 (p≤0.01) / -0.48 (p≤0.01) /-0.43 (p≤0.01) Mean voided volume : -0.23 (p≤0.01) / -0.25 (p≤0.01) / -0.23 (p≤0.01)

			the International Continence Society Nomogram, Schäfer's obstruction grade and URA.				
Penile Compression Release Manoeuvre / Penile Compression Release Index							
Aganovic [23]	2019	To assess BOO through Penile Compression Manoeuvre.	Urodynamic studies, based on Schäfer's grade, reported as DAMPF, CLIPS, BOON2 and URA.	Men with LUTS due to BPH.	135	66.1 (7.2; 51-81)	Pearson correlation: Penile compression release index - DAMPF (continuous Schäfer variable): r=0.44 (p<0.0001) Penile compression release index to predict BOO: ROC-AUC: 85%, posttest-probability: 91.3% To predict BOO, at a penile compression release index cut-off value of 96.4%: SN: 74.3%, SP: 93.8%, PPV: 93%, NPV: 77% LR+: 9.6 (95% CI: 0.777-0.904) Number Needed to Diagnose: approx. 1.5 BOON2 to compare PCRI with: ROC-AUC: 82%, posttest-probability: 74.5% To predict BOO, at a BOON2 of > -35.3: SN: 81%, SP: 71%, PPV: 75%, NPV: 79% LR+: 2.7 (95% CI: N/A) Number Needed to Diagnose: approx. 1.9 De Long method of pair-wise comparisons of ROC-AUC: not significant for Penile compression compared with noninvasive: CLIPS or BOON2 method scoring, criterion: URA.
Penile Cuff Uroflowmetry							
Harding [34]	2004	Estimation of Bladder Outlet Obstruction through Penile Compression Release Index, based on automated penile cuff.	Pressure flow studies with and without Penile Cuff test. Abram-Griffith number greater than 40 was defined as obstructed. AG-number was combined with Penile	Men with LUTS, referred for conventional Pressure Flow studies.	101	-	Prediction of BOO based on Penile compression release index (PCRI), for PCRI > 160% SN: 78%, SP: 84%, PPV: 69%, NPV: not reported

			cuff parameters.				
Kim [37]	2020	Measurement of maximum flow rate (Qmax) and isovolumetric bladder pressure to categorize obstruction, not obstructed and equivocal groups through penile cuff test.	Urodynamic studies, scoring based on a (modified) ICS nomogram.	Men with LUTS related to BPH	59	Median and IQR 69.6 (54-89)	Penile cuff test - urodynamic studies Category: obstructed vs non-obstructed/equivocal SN: 80%, SP: 100%, PPV: 100%, NPV: 60.9% LR+: 2.6 (95% CI: 2.13-4.02), LR-: 0.23 (95% CI: 0.1-0.41)
Salinas [50]	2003	Estimation of Bladder Outlet Obstruction through penile cuff test.	Urodynamic studies, based on Abram-Griffiths number ≥40 = obstructed 20-40 = equivocal ≤20 = unobstructed	Men referred for urodynamic study on presentation of LUTS.	93	54.1 (16.1; -)	Sensitivity and specificity for predicting obstruction (exclusion of n=41 equivocal cases) SN: 100%, SP: 55.6% Diagnostic accuracy: 84.6% Pcuff.op (cmH2O): Obstructed/unequivocal vs. non-obstructed: Mean (SE) 172.92 (5.82) - 142.33 (8.77), p=0.007 Qcuff.op (ml/s): Obstructed/unequivocal vs. non-obstructed: Mean (SE) 9.43 (0.66) - 13.67 (1.42), p=0.003 *Based on calculations by authors of Systematic Review* BOO compared to Cuff outcomes No obstruction vs unequivocal based on cuff outcomes: Pcuff.OP: OR: 0.9910 (95% CI: 0.99-0.99) Qcuff.OP: OR: 1.1038 (95% CI:1.10-1.11) No obstruction vs obstruction based on cuff outcomes: Pcuff.OP: OR: 0.9835 (95% CI: 0.98-0.98) Qcuff.OP: OR: 1.3348 (95% CI:1.32-1.35)
Combinations of assessment methods							
Venrooij [56]	2004	To discriminate between obstructed and non-obstructed men.	Cystometry and pressure-flow studies. Analyzed according to the International	Men with LUTS, suggestive of BPH.	160	65.1 (8.3)	Combined measurement instruments: Bladder outlet obstruction number (BOON): prostate volume (transrectal ultrasound) – (3*maximal urinary flow rate) – (0.2*mean voided volume) Obstructed – not obstructed: BOON – Abram-Griffith: ROC-AUC: 0.83

			Continence Society Nomogram, Schäfer's obstruction grade and URA.				BOON – urethral resistance factor: ROC-AUC: 0.87 BOON – Schäfer's obstruction grade: ROC-AUC: 0.82 Kendall's and Gibbons correlation with: Abrams-Griffiths number / urethral resistance factor / Schäfer's obstruction grade. BOON 0.48 (p≤0.01) / 0.52 (p≤0.01) / 0.49 (p≤0.01)
Questionnaires to indicate BOO							
Chan [28]	2012	To assess maximum urinary flow to predict BOO.	Electronic flowmetry in clinic and international prostate symptom score.	Men with LUTS attributable to BPH.	186	65.5 (7; -)	IPSS – fifth question scores Mean IPSS score of 5th question >3, reference: mean Qmax from uroflowmetry at clinic visit: <10 mL/s SN: 0.51 (95% CI: 0.39-0.62), SP: 0.78 (95% CI: 0.70-0.85) LR+: 2.33 (95% CI: 1.54-3.54), LR-: 0.63 (95% CI: 0.49-0.81) DOR: 3.70 (95% CI: 1.95-7.04) Mean IPSS score of 5th question >2, reference: mean Qmax from uroflowmetry at clinic visit: <15 mL/s SN: 0.63 (95% CI: 0.55-0.71), SP: 0.72 (95% CI: 0.58-0.82) LR+: 2.23 (95% CI: 1.42-3.49), LR-: 0.51 (95% CI: 0.39-0.68) DOR: 4.34 (95% CI: 2.17-8.69) Mean IPSS score of 5th question >1, reference: mean Qmax from uroflowmetry at clinic visit: <19 mL/s SN: 0.74 (95% CI: 0.67-0.80), SP: 0.55 (95% CI: 0.35-0.73) LR+: 1.62 (95% CI: 1.02-2.59), LR-: 0.48 (95% CI: 0.30-0.76) DOR: 3.38 (95% CI: 1.36-8.38) IPSS – mean score for voiding (questions 1, 3, 5 and 6) Mean IPSS score for voiding, score >12, reference: mean Qmax from uroflowmetry at clinic visit: <10 mL/s SN: 0.25 (95% CI: 0.17-0.37), SP: 0.86 (95% CI: 0.79-0.91) LR+: 1.82 (95% CI: 1.00-3.34), LR-: 0.87 (95% CI: 0.74-1.01) DOR: 2.10 (95% CI: 0.99-4.46) Mean IPSS score for voiding, score >8, reference: mean Qmax from uroflowmetry at clinic visit: <15 mL/s SN: 0.49 (95% CI: 0.41-0.57), SP: 0.74 (95% CI: 0.60-0.84) LR+: 1.85 (95% CI: 1.14-3.00), LR-: 0.70 (95% CI: 0.55-0.88) DOR: 2.66 (95% CI: 1.32-5.36) Mean IPSS score for voiding, score >4, reference: mean Qmax from uroflowmetry at clinic visit: <19 mL/s SN: 0.74 (95% CI: 0.67-0.80), SP: 0.73 (95% CI: 0.52-0.87) LR+: 2.73 (95% CI: 1.37-5.43), LR-: 0.35 (95% CI: 0.24-0.51) DOR: 7.75 (95% CI: 92.85-21.1)

Matzkin [41]	1996	Correlation of uroflowmetry recordings with the AUA.	24-hour uroflowmetry	Men with enlargement of the prostate, related to BPH.	42	69 (- ; 45-83)	Correlation for AUA item with uroflowmetry recordings Questionnaire first visit / questionnaire second visit AUA Item 1 - %Frequency: t-score: 1.047 / 0.575 AUA Item 3 - %Intermittency: t-score: -0.768 / -0.516 AUA Item 5 - % Weak uroflowmetry: t-score: 0.178 / 0.467 AUA Item 7 - Nocturia: t-score: 3.167 / 2.310 The authors considered a t-score of >2 highly significant.
Steele [51]	2000	To predict BOO.	Multichannel urodynamic studies to obtain obstruction grade through the ICS nomogram. >2cm water per ml/s and detrusor pressure >40 cm water was defined as obstructed.	Men with LUTS.	204	66.7 (7.5; -)	Correlation AUA score – detrusor pressure at maximum flow as predictor of bladder outlet obstruction: r=0.18 (p>0.05)
Venrooij [55]	1996	To detect BOO and correlate the IPSS with BOO related parameters.	Urodynamic studies, based on Schäfer's grade, with a classification of 0 and 1 defined as non-obstructed and ≥2 as obstructed.	Men with prostatism, with and without urodynamic obstruction / possible BPH.	196	65.8 (7.1; 51-86)	Pearson's correlation: IPSS – maximal flow: -0.12 (not significant) IPSS – residual volume: 0.10 (not significant) IPSS – prostate volume: 0.03 (not significant) IPSS – Schäfer's obstruction grade: 0.02 (not significant) Kendall & Gibbon's correlation: IPSS – maximal flow: -0.07 (not significant) IPSS – residual volume: 0.06 (not significant) IPSS – Schäfer's obstruction grade: 0.02 (not significant) Note by review authors: *The authors of the study mention some variables showed a non-normal distribution and analysed the Kendall & Gibbon's correlation. In the review, we assumed the Kendall & Gibbon's correlation to be most accurate.*
Venrooij [56]	2004	To discriminate between obstructed and non-obstructed men.	Cystometry and pressure-flow studies. Analyzed according to	Men with LUTS, suggestive of BPH.	160	65.1 (8.3; 50-85)	Kendall's and Gibbons correlation with: AUA score – Abrams-Griffiths Number: 0.15 (p≤0.01) AUA score – Urethral resistance factor: 0.16 (p≤0.01) AUA score – Schäfer's Obstruction Grade: 0.16 (p≤0.01)

	the International Continence Society Nomogram, Schäfer's obstruction grade and URA.	

AUA = American urology association questionnaire, BOO = bladder outlet obstruction, BOOI = bladder outlet obstruction index, BOON2 = Bladder Outlet Obstruction Number 2, BPE = benign prostate enlargement, BPH = benign prostatic hyperplasia, BPO = benign prostatic obstruction, CLIPS = Clinical Prostate Score, DRE = digital rectal examination, DOR = diagnostic odds ratio, IPSS = international prostate symptom score, IQR = Interquartile range, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, mL = millilitre, mL/s = millilitre per second, NPV = negative predictive value, PPV = positive predictive value, PVR = postvoid residue, Qmax = maximal urinary flow rate, ROC-AUC = radio operator curve – area under the curve, sd = standard deviation, SN = sensitivity, SP = specificity, TAUS = transabdominal ultrasound, TRUS = transrectal ultrasound, URA = Urethral Resistance Factor, [1)] = Aim of method extracted from study and summarized by review authors.

Supplementary Table 2C. Validity of instruments – Assessment methods for benign prostatic obstruction.

First author	Year	Aim of method[1)]	Reference test	Patient category	Sample, n (%)	Age, mean (sd; min-max)	**Measures of criterion validity**: sensitivity, specificity, area under the curve (95% confidential interval) **and construct validity:** correlations (p or 95% confidential interval).
Digital rectal examination (DRE)							
Carballido [27]	2011	Diagnose the presence of BPH.	'Gold-standard' diagnosis of BPH by an urologist based on: medical history, initial assessment of symptoms, IPSS and Bother Score, PSA analysis, urinalysis, digital rectal examination, abdominal ultrasound (for prostate size and postvoid residue) and uroflowmetry.	Men with LUTS.	666	60.9 (7.9; 50-98)	Prostate size by DRE – final BPH diagnosis: p=0.123 (no correlation statistic provided) General practitioner - urologist: k=0.284 (95% CI: 0.22-0.35) General practitioner - transabdominal ultrasound prostate volume: k=0.171 (95% CI: 0.11-0.24) Urologist - transabdominal ultrasound prostate volume: k=0.624 (95% CI: 0.57-0.68) Note by review authors: *Although the word correlation is used, the values appear to be reported in kappa values. We followed the study objective, in which the urologist's assessment was referred to as the gold-standard. We assume the wrong symbol was used in the text to describe the correlation.*
Roehrborn [49]	2001	Estimation of prostate volume.	Transrectal ultrasound.	Volunteers from a general urology practice.	121	60.7 (10.3; -)	Spearman correlation: DRE – transrectal ultrasound: Assessor for DRE: attending physician / postgraduate 4-year / postgraduate 2-year Plus grade: r=0.57 / 0.64 / 0.56 Textual scale: r=0.58 / 0.59 / 0.57 Best estimate (grams): r=0.72 / 0.70 / 0.61 Sizing balls: r=0.67 / 0.62 / 0.57 Concentric rings: r=0.60 / 0.64 / 0.63 Lever device: r=0.60 / 0.59 / 0.59 Full 3D model: r=0.66 / 0.66 / 0.60 Final 3D model: r=0.75 / 0.65 / 0.67
Su [54]	2013	Estimation of prostate volume based on defined thresholds.	Transrectal ultrasound.	Men with LUTS prior to BPH-related surgery or in conjunction with prostate biopsy.	280	65 (- ; 59-71)	Cut-off value: prostate volume ≥30 mL. SN: 94.3% (95% CI: 90.1%-96.8%), SP: 78.2% (95% CI: 64.6%-87.8%) LR+: 3.97 (95% CI: 2.51-6.28) LR-: 0.08 (95% CI: 0.05-0.13) *Other cut-off values are reported: ≥50mL and ≥100mL. A higher cut-off value reportedly increased the SN and decreased the SP, although no specific values are reported.*
Transabdominal ultrasound (TAUS)							
Demir [31]	2016	Estimation of prostate size.	Resected tissue weight through open prostatectomy.	Men with LUTS, undergoing	60	68.9 (9.4; 49-85)	Mean differences: Prostatic size (in cc): TAUS: 67.81 (sd: 33.4) – TRUS: 52.61 (sd: 25.06), p=0.001 Pearson correlation: TAUS – resected tissue weight: r=0.77 (p=0.001)

				transurethral resection of prostate.			
Güzelsoy [33]	2016	Estimation of prostate size and transitional zone volume and transitional zone index.	Resected tissue weight and transrectal ultrasound.	Men with obstructive symptoms, diagnosed with clinical BPH and men with BPE without obstructive symptoms.	43	66.0 (7.9; 50-81)	Pearson correlation: TAUS prostate volume – resected tissue weight: r=0.73 (p=0.001) TAUS prostate volume – (TRUS) transitional zone volume: r=0.78 (p=0.0001) TAUS prostate volume – (TRUS) transitional zone index: r=0.54 (p=0.0001)
Malemo [40]	2011	Estimation of prostate volume.	Transrectal ultrasound.	Male patients with symptomatic BPH and IPSS score of >20, with posteroperative histologic confirmation of BPH.	50	69.7 (11.3; 51-91)	Prostate volume: ≤80 or >80 mL. SN: 95% (95% CI: 78%-99%), SP: 0.96 (95% CI: %82-99%) PPV: 80% (95% CI: 78%-99%), NPV: 95% (95% CI: 78%-99%) Spearman correlation TAUS – transrectal ultrasound: r=0.98 (p<0.001)
Prassopoulos [45]	1996	Estimation of prostate size and transitional zone volume (TZV).	Transrectal ultrasound.	Men with BPH.	95	69.7 (11.3; 47-85)	'parametrical' correlation TAUS prostate volume – TRUS prostate volume: r=0.948 (p<0.001) TAUS TZV – TRUS TZV: r=0.953 (p<0.001) (in n=76)
Stravodimos [52]	2009	Estimation of prostate volume	Transrectral ultrasound and specimen weight.	Male patients with LUTS, diagnosed with BPH.	71	72 (- ; 55-82)	Correlation: TAUS – specimen weight: r=0.82 (p<0.001) *(Calculated by review authors from the data in the article)* TAUS estimated Prostate volume (cc) – Specimen Weight (g) Accurate detection of <80cc for <80g: SN: 0.57 (95% CI: 0.43-0.70) SP: 1.00 (95% CI:0.75-1.00) PPV: 1.00 (95% CI: N/A, due to missing cases in 1 cell of the contingency table) NPV: 0.34 (95% CI: 0.28-0.41)
Styles [53]	1988	Estimation of prostate volume.	Transrectal ultrasound.	Men undergoing elective prostatectomy for	76	69 (7; -)	Spearman correlation: TAUS – transrectal ultrasound: r=0.8205 (p<0.001)

				symptoms of BOO and <15 m/ls free flow rate.			
Transperineal ultrasound							
Rathaus [46]	1991	Estimation of prostate size.	Transrectal ultrasound through: ellipsoid formula. 0.55*D1*D2*D3 D1: anteroposterior diameter D2: transverse diameter D3: cephalocaudal diameter	Men with BPH.	80	-	Correlation: transperineal ultrasound – transabdominal ultrasound, n=10: r=0.92 (p<0.001) transperineal ultrasound – specimen weight n=80: r=0.89 (p<0.001)
Transrectal ultrasound (TRUS)							
Aarnink [21]	1996	Estimation of prostate volume and transitional zone volume.	Transrectal ultrasound: manual outline method.	Men with LUTS; 'clinically benign patients'.	247	61 (- ; 28-87)	Pearson correlation for measurement methods of prostate volume: TRUS (automated volume) – TRUS (reference volume): r=0.938 (p: not reported) TRUS (off-line ellipsoid formula volume) – TRUS (reference volume): r=0.921 (p: not reported) TRUS (transverse off-line ellipsoid formula volume) – TRUS (reference volume): r=0.955 (p=not reported) Correlation transitional zone volume: Transitional zone volume – TRUS prostate volume (manual outline by urologist): r=0.82 (p=not reported)
Baltaci [26]	2000	Estimation of transitional zone volume. (TZV)	Enucleated adenoma.	Men with LUTS, scheduled to undergo prostate adenoma removal due to BPH.	48	65.7 (- ; 50-81)	Correlation: TRUS transitional zone volume (TZV) – enucleated adenoma: r=0.95 (p<0.001).
David [30]	2020	Estimation of prostate volume and transitional zone volume (TZV)	Enucleated adenoma volume.	Sub-Saharan men with BPH, undergoing surgery.	77	69.6 (7.26; 51-91)	Pearson correlation: Total sample (n = 77) TZV - Prostatic specimen volume: r = 0.865 (p=0.0000), R2 = 74.8% Prostate volume - Prostatic specimen volume: r = 0.932 (p=0.0000), R2 = 86.9% Under <100 mL prostate size: (n = 50) Prostate volume - Prostatic specimen volume: r = 0.8168 (p=N/A) Transitional zone volume - prostatic specimen volume: r = 0.6846 (p=N/A)

							Under >100 mL prostate size: (n = 27) Prostate volume - Prostatic specimen volume: r = 0.8712 (p=N/A) Transitional zone volume - Prostatic specimen volume: r = 0.7295 (p=N/A) Volumes differences by TRUS – Enucleated prostate volume Prostate volume – Enucleated prostate volume: 93.1 mL (sd: 48.9 mL) – 79.1 mL (sd: 62.9 mL) Difference: 14.0 mL (95% CI: -19.59 to -8.36) p<0.0005) Transitional zone volume – Enucleated prostate volume: 53.3 mL (sd: 28.5 mL) – 79.1 mL (sd: 62.9 mL) Difference: 25.8 mL (95% CI: 16.52-35.06), p<0.0005) Under <100 mL prostate size: (n = 50) Prostate volume – Enucleated prostate volume: 63.7 mL (sd: 19.9 mL) – 45.1 mL (sd: 23.2 mL), p = 0.0000 Transitional Zone volume – Enucleated prostate volume: 37.1 mL (sd: 15.3 mL) – 45.1 mL (sd: 23.2 mL), p = 0.0014 Under >100 mL prostate size: (n = 27) Prostate volume – Enucleated prostate volume: 147.4 mL (sd: 38.8 mL) – 142.0 mL (sd: 64.9 mL), p = 0.4467 Transitional Zone volume – Enucleated prostate volume: 83.4 mL (sd: 22.0 mL) – 142.0 mL (sd: 64.9 mL), p = 0.0000
Demir [31]	2016	Estimation of prostate size.	Resected tissue weight through open prostatectomy.	Men with LUTS, undergoing transurethral resection of prostate.	60	68.9 (9.4; 49-95)	Mean difference: Prostatic size (in cc): TAUS: 67.81 (sd: 33.4) – TRUS: 52.61 (25.06) p=0.001 Pearson correlation: TRUS – resected tissue weight: r=0.79, p=0.001
Güzelsoy [33]	2016	Estimation of prostate size and transitional zone volume (TZV) and transitional zone index (TZI).	Resected tissue weight and transrectal ultrasound.	Men with obstructive symptoms, diagnosed with clinical BPH and men with BPE without obstructive symptoms.	43	66.0 (7.9; 50-81)	Pearson correlation: TRUS prostate size – TRUS TZV: r=0.96 (p=0.0001) TRUS prostate size – TRUS TZI: r=0.56 (p=0.0001) TRUS prostate size – resected tissue weight: r=0.95 (p=0.0001) TRUS TZV – resected tissue weight: r=0.97 (p=0.0001) TRUS TZI – resected tissue weight: r=0.55 (p=0.002) Diagnostic accuracy of TZI to predict clinical BPH, unclear reference values. SN: TZI: 0.40: 97%, TZI: 0.45: 91%, TZI: 0.55-0.60: 100%, 0.25-0.35: 0% SP: TZI: 0.25: 31%, TZI: 0.30: 25%, TZI: 0.35: 19%, TZI 0.40: 91%, TZI: 0.45: 87%, TZI: 0.50: 68%, TZI: 0.55: 56%, TZI: 0.60: 54%.
Kim [36]	2014	Estimation of prostate	Transrectal ultrasound:	Men with prostate	968	58.4 (- ; 21-88)	TRUS Transaxial (index): 28.5 in mL (sd: 10.1) – TRUS Midsagittal (reference): 28.7 in mL (sd: 9.9), difference: p=0.004.

		volume.	midsagittal scanning, with prolate ellipsoid formula.	related diseases, and subgrouping of clinical benign prostatic enlargement.			
Narayanamurthy [42]	2020	Estimation of prostate volume.	Anatomical prostate weight	Men with BPH, who underwent robotic-assisted laparoscopic prostatectomy.	295	64.3 (6.3; -)	Correlation: TRUS – Anatomical prostate weight Pearson r=0.67 (95% CI: 0.60-0.73), p<0.001 Mean difference of TRUS - Anatomical prostate weight -12.5 gram (95% CI: -14.4 to -11.03) 95% levels of agreement: upper limit: 13 grams, Lower limit: -38 grams
Nathan [43]	1996	Estimation of prostate volume.	Transrectal ultrasound: step planimetry of prostate volume.	Men with symptoms of prostatic enlargement.	45	40 (- ; 43-89)	Correlation: TRUS (dimensional method) – TRUS (step planimetry) r=0.89. TRUS (largest planimetric dimensions) – TRUS (step planimetry) r=0.93. TRUS (computer enhanced dimensions) – TRUS (step planimetry) r=0.88.
Stravodimos [52]	2009	Estimation of transitional zone volume.	Transrectral ultrasound and specimen weight.	Male patients with LUTS, diagnosed with BPH.	71	72 (- ; 55-82)	Correlation: TRUS Transition zone volume – prostate specimen weight: r=0.904 (p<0.005). *(Calculated by review authors from the data in the article)* TRUS Transition Zone - Specimen weight Accurate detection of <80cc for <80g: SN: 0.93 (95% CI: 0.83-0.98) SP: 0.62 (95% CI: 0.32-0.86) PPV: 0.92 (95% CI: 0.84-0.95) NPV: 0.67 (95% CI: 0.41-0.85)
Combination of assessment methods							
De Nunzio [32]	2015	To predict BPO.	Pressure-flow studies to obtain a Schäfer's class from the Schäfer's nomogram, BPO was defined at ≥3 of a Schäfer's class.	Men with LUTS or BPE, 45 years and older.	449	61.2 (11; IQR: 61-73)	Combined measurement instruments: *Nomogram consists of: maximal flow rate from free uroflowmetry and transitional zone index* Nomogram at 80% probability for obstruction: SN: 74%, SP: 79%, PPV: 89%, NPV: 56% ROC-AUC: 0.76 (95% CI: 0.71-0.82), p=0.000
Questionnaires to indicate BPO							
Carballido [27]	2011	Diagnose the presence of BPH.	'Gold-standard' diagnosis of BPH by an urologist based on: medical history, initial assessment of	Men with LUTS.	666	60.9 (7.9; 50-98)	IPSS score - Urologist's final BPH diagnosis (not reported if cut-off scores were used, or full range of IPSS total scores) SN: 58%, SP: 59.3%, PPV: 73.5%, NPV: 42.0% Model: IPSS score and age - Urologist's final BPH diagnosis (not reported if cut-off

			symptoms, IPSS and Bother Score, PSA analysis, urinalysis, digital rectal examination, abdominal ultrasound (for prostate size and postvoid residue) and uroflowmetry.				scores were used for age or IPSS, or full range of IPSS total scores) SN: 56.8%, SP: 64.2%, PPV: 75.5%, NPV: 43.3%
Kwon [39]	2016	Estimation of peripheral zone thickness and related prostate size parameters.	Urinary flow parameters from uroflowmetry.	Men with LUTS/BPH	1009	62.0 (10.0; -)	Correlation IPSS (total score) – prostate size parameters IPSS (total score) – total prostate volume, r=0.081 (p<0.05) IPSS (total score) – transitional zone volume, r=0.098 (p<0.01) IPSS (total score) – transitional zone index, r=0.111 (p<0.01) IPSS (total score) – peripheral zone thickness, r=-0.162 (p<0.01) Correlation IPSS (voiding symptoms) – prostate size parameters IPSS (voiding symptoms) – total prostate volume, r=0.050 IPSS (voiding symptoms) – transitional zone volume, r=0.059 IPSS (voiding symptoms) – transitional zone index, r=0.074 (p<0.05) IPSS (voiding symptoms) – peripheral zone thickness, r=-0.117 (p<0.05) Correlation IPSS (storage symptoms) – prostate size parameters IPSS (storage symptoms) – total prostate volume, r=0.120 (p<0.05) IPSS (storage symptoms) – transitional zone volume, r=0.144 (p<0.01) IPSS (storage symptoms) – transitional zone index, r=0.145 (p<0.01) IPSS (storage symptoms) – peripheral zone thickness, r=-0.169 (p<0.01) Correlation IPSS (post-micturition symptoms) – prostate size parameters IPSS (post-micturition symptoms) – total prostate volume, r=0.003 (not sign.) IPSS (post-micturition symptoms) – transitional zone volume, r=0.005 (not sign.) IPSS (post-micturition symptoms) – transitional zone index, r=0.017 (not sign.) IPSS (post-micturition symptoms) – peripheral zone thickness, r=-0.073 (p<0.05)
Nathan [43]	1996	Estimation of prostate volume.	Transrectal ultrasound: step planimetry of prostate volume.	Men with symptoms of prostatic enlargement.	45	40 (- ; 43-89)	Correlation IPSS-S (unclear whether IPSS-S indicates 'score' or 'storage subscore' – prostate volume calculation methods: IPSS – DRE: 0.033 IPSS – Dimensional Method Volume through TRUS: 0.0619 IPSS – Planimetric Volume through TRUS: 0.0894
Venrooij [55]	1996	To detect BOO and correlate the IPSS with BPO related parameters.	'Prostate volume measured by transrectal ultrasonography. (TRUS)	Men with prostatism, with and without urodynamic obstruction / possible BPH.	196	65.8 (7.1; 51-86)	Pearson' correlation: IPSS - prostate volume: 0.03 (not significant) Kendall & Gibbon's correlation: IPSS – prostate volume from TRUS: 0.01 (not significant) Note by review authors: *The authors of the study mention some variables showed a non-normal distribution*

and analysed the Kendall & Gibbon's correlation. In the review, we assumed the Kendall & Gibbon's correlation to be most accurate.

AUA = American urology association questionnaire, BOO = bladder outlet obstruction, BPE = benign prostate enlargement, BPH = benign prostatic hyperplasia, BPO = benign prostatic hyperplasia, DRE = digital rectal examination, IPSS = international prostate symptom score, LR+ = positive likelihood ratio, LR- = negative likelihood ratio, mL = millilitre, NPV = negative predictive value, PPV = positive predictive value, PSA = prostate specific antigen,
sd = standard deviation, SN = sensitivity, SP = specificity, ROC-AUC = radio operator curve – area under the curve, TAUS = transabdominal ultrasound,
TRUS = transrectal ultrasound, [1)] = Aim of method extracted from study and summarized by review authors.