








## Article

# Diagnoses and Treatment Recommendations—Interrater Reliability of Uroflowmetry in People with Multiple Sclerosis

Anke K. Jaekel <sup>1,2,\*</sup> , Julia Rieger <sup>1</sup>, Anna-Lena Butscher <sup>2</sup> , Sandra Möhr <sup>3</sup> , Oliver Schindler <sup>4</sup> , Fabian Queissert <sup>5</sup> , Aybike Hofmann <sup>6</sup> , Paul Schmidt <sup>7</sup> , Ruth Kirschner-Hermanns <sup>1,2</sup> and Stephanie C. Knüpfer <sup>1,2</sup>

- <sup>1</sup> Clinic for Urology, University Hospital Bonn, 53127 Bonn, Germany; rkirschnerhermanns@googlemail.com (R.K.-H.); stephanie.knuepfer@ukbonn.de (S.C.K.)  
<sup>2</sup> Department of Neuro-Urology, Johanniter Rehabilitation Centre Godeshoehe, 53177 Bonn, Germany  
<sup>3</sup> Clinic for Neurorehabilitation and Paraplegiology, REHAB Basel, 4055 Basel, Switzerland; s.moehr@rehab.ch  
<sup>4</sup> Clinic for Urology, University Hospital Ulm, 89070 Ulm, Germany  
<sup>5</sup> Clinic for Urology, University Hospital Münster, 48149 Münster, Germany  
<sup>6</sup> Clinic St. Hedwig, Department of Paediatric Urology, University Medical Center Regensburg, 93053 Regensburg, Germany  
<sup>7</sup> Statistical Consulting for Science and Research, Berlin Statistical Consulting for Science and Research, 13086 Berlin, Germany  
\* Correspondence: a.k.jaekel@gmail.com

**Abstract:** Background: Uroflowmetry (UF) is an established procedure in urology and is recommended before further investigations of neurogenic lower urinary tract dysfunction (NLUTD). Some authors even consider using UF instead of urodynamics (UD). Studies on the interrater reliability of UF regarding treatment recommendations are rare, and there are no relevant data on people with multiple sclerosis (PwMS). The aim of this study was to investigate the interrater reliability (IRR) of UF concerning diagnosis and therapy in PwMS prospectively. Methods: UF of 92 PwMS were assessed by 4 raters. The diagnostic criteria were normal findings (NFs), detrusor overactivity (DO), detrusor underactivity (DU), detrusor-sphincter dyssynergia (DSD) and bladder outlet obstruction (BOO). The possible treatment criteria were as follows: no treatment (NO), catheter placement (CAT), alpha-blockers, detrusor-attenuating medication, botulinum toxin (BTX), neuromodulation (NM), and physiotherapy/biofeedback (P/BF). IRR was assessed by kappa ( $\kappa$ ). Results:  $\kappa$  of diagnoses were NFs = 0.22; DO = 0.17; DU = 0.07; DSD = 0.14; and BOO = 0.18. For therapies, the highest  $\kappa$  was BTX = 0.71, NO = 0.38 and CAT = 0.44. Conclusions: There is a high influence of the individual rater. UD should be subject to the same analysis and a comparison should be made between UD and UF. This may have implications for the value of UF in the neuro-urological management of PwMS, although at present UD remains the gold standard for the diagnostics of NLUTD in PwMS.

**Keywords:** multiple sclerosis; bladder disorder; neurogenic; bladder



**Citation:** Jaekel, A.K.; Rieger, J.; Butscher, A.-L.; Möhr, S.; Schindler, O.; Queissert, F.; Hofmann, A.; Schmidt, P.; Kirschner-Hermanns, R.; Knüpfer, S.C. Diagnoses and Treatment Recommendations—Interrater Reliability of Uroflowmetry in People with Multiple Sclerosis. *Biomedicines* **2024**, *12*, 1598. <https://doi.org/10.3390/biomedicines12071598>

Academic Editor: Víctor M. Rivera

Received: 10 June 2024

Revised: 10 July 2024

Accepted: 12 July 2024

Published: 18 July 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Uroflowmetry (UF) has been a well-established diagnostic procedure in urology for many years [1]. It is used to diagnose functional disorders of the lower urinary tract in children and adults [2,3]. In this non-invasive urodynamic procedure, a urinary flow rate per time is determined using various technical principles and a flow curve is recorded while voiding urine [4]. Finally, the post-void residual (PVR) is measured by sonography of the bladder [5]. Conclusions can be drawn about the existing type of anatomical or functional disorders of the lower urinary tract from the various forms of UF curves, the micturition volume, the strength of the urine flow, and PVR [4]. The advantages of UF are its simplicity, non-invasiveness, and short duration [1].

The procedure is indicated for both neurogenic and non-neurogenic functional disorders of the lower urinary tract. It is recommended in various urological guidelines [2,3,5].

For neurogenic lower urinary tract dysfunction (NLUTD) in people with multiple sclerosis (PwMS), so far there are no explicit guideline recommendations for UF. Nevertheless, individual authors recommend uroflowmetry, even as a substitute for urodynamics [6], which is the gold standard in the diagnostics of NLUTD [7]. Others only recommend urodynamics for PwMS in the case of treatment failure or prior to surgical treatment [8]. These recommendations are based on the fact that urodynamics (UD), associated with high effort and invasiveness, have ultimately no impact on the treatment outcome. Therefore, neuro-urological expert committees discuss whether UF can play a role in reducing the need for invasive UD [9].

Recommendations for the standardised performance and reporting of uroflowmetry were published by the International Continence Society in 2002 [4]. Due to the involvement of the autonomic nervous system in micturition, there are many factors influencing uroflowmetry: the environment, the patient's compliance and current situation, the micturition position, and the voiding volume [10,11]. Clear recommendations on how to perform UF were defined: a minimum bladder-filling volume should be maintained, the person should void in their usual posture, and information on the subjective representativeness of the examination should be obtained [4,12].

Another significant factor influencing UF is the rater-related individual interpretation of the examination. The rater develops a treatment proposal from his/her subjective perspective by the inclusion of clinical information, bladder diaries (BDs), and questionnaires [13].

The extent to which a measurement procedure is dependent on the influence of the examiner can be measured using reliability [14]. Interrater reliability (IRR) measures the agreement of the assessment between different raters. There are a variety of studies on IRR of UF in children and adults with different study designs [15–17]. So far, there has been no study on the IRR of UF in PwMS that relates not only to the uroflow curve itself, but also to the underlying diagnosis and treatment recommendation.

To assess the long-term significance of UF in comparison to UD regarding NLUTD in PwMS, our intention was to assess the IRR of UF regarding suspected diagnosis and treatment recommendations. We also analysed the IRR of the suspected diagnosis from medical history and BDs.

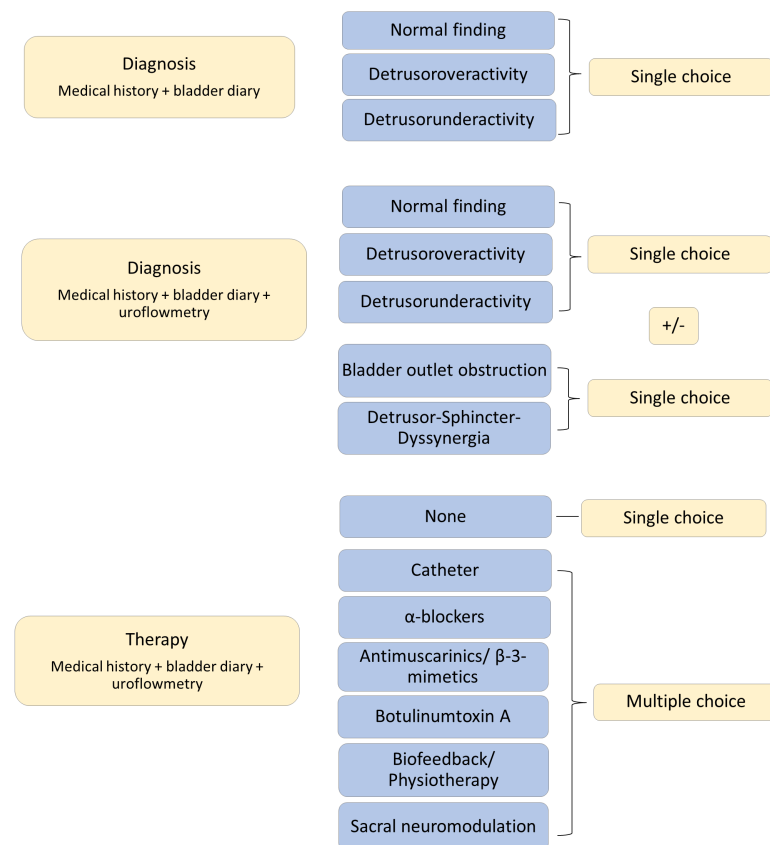
## 2. Patients and Methods

This study included all PwMS who had been presented to the neuro-urology department of an inpatient neurological rehabilitation centre for further diagnosis of NLUTD between 2017 and 2022 and who met the following requirements: written informed consent for prospective evaluation of their data, detailed urological history, a completed BD with documented drinking and micturition volumes in ml over 2 days, and uroflowmetry that was performed according to the standards of the International Continence Society. Ninety-two PwMS met these criteria. Inclusion in this study was independent of the clinical course or severity of the disease. Exclusion criteria were invalid uroflowmetry with a micturition volume of <150 mL, a BD that could not be analysed due to invalid or illegible documentation, and the absence of an informed consent form.

All UF examinations were conducted by a highly experienced neuro-urological team consisting of 2 physicians and 3 nurses. The UF was indicated after the patient had undergone an initial consultation with the physician. The examination was scheduled on a different date than the initial consultation. The UF examination procedure was explained in detail in advance. All of the participants were scheduled for UF in the same standardised time slot with a filled bladder. The examination was conducted, according to the preference of the PwMS, in a sitting or standing position in a closed quiet toilet suitable for the disabled. The measurement was performed using an MMS Nexam Pro Urodynamic System (Laborie/Medical Measurement Systems B.V., Enschede, The Netherlands) after the participants had expressed a strong desire to void urine. The measurement of PVR was

performed by transabdominal sonography with a GE LOGIQ S7 Pro (GE Ultrasound Korea Inc. Seongnam-si, Korea).

Four experienced neuro-urologists were recruited as raters. They received the following data: diagnoses of the participants, a summarised medical history, information on gender, summarised information from BDs, graphics of the UF curves with information on voided volume in ml, PVR in ml, and maximum urinary flow in ml/s. The raters were asked to assign the individual uroflowmetry findings to the most appropriate diagnostic category, considering the clinical information and the BDs, and to select a corresponding treatment recommendation. A standardised evaluation form was used for this purpose. The answering options listed in the form are shown in Figure 1. All raters were blinded to the original clinical findings.



**Figure 1.** Visualisation of the potential answers on the evaluation form. “+/-”—a combination of both single choices is optional, but not mandatory.

The primary endpoint was the IRR regarding the following items:

- (A) Suspected diagnosis from medical history and BDs;
- (B) Suspected diagnosis from medical history, BDs, and UF;
- (C) Therapy suggestion from medical history, BDs, and UF.

For all primary endpoints, the agreement between 2 raters was determined using Cohen’s kappa ( $\kappa_C$ ) and the agreement between all raters was assessed using Fleiss’ kappa ( $\kappa_F$ ). For the primary endpoints B and C, the answers could be obtained using multiple-choice assessment (see Figure 1). The possibility of multiple-choice assessment resulted in the following analyses:

**Analysis A:** IRR for the suspected diagnoses from the medical history and BDs;

**Analysis B-1:** IRR for each single suspected diagnosis from medical history, BDs, and UF;

**Analysis B-2:** IRR for the combined suspected diagnoses from medical history, BDs, and UF;

**Analysis C-1:** IRR for each single therapy suggestion from medical history, BDs, and UF;

**Analysis C-2:** IRR for the combined therapy suggestions from medical history, BDs, and UF.

On a scale of 0.0–1.0, a kappa statistic of 0–0.2 indicates slight agreement, 0.21–0.4 indicates fair agreement, 0.41–0.6 indicates moderate agreement, 0.61–0.8 indicates substantial agreement, and 0.81–1.0 indicates nearly perfect agreement [18].

All analyses were performed with the statistical programming language R (R Core Team 2019) (R version 4.2.2 (09 March 2024) [19].

The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee for ethical approval (EK 313/13-University Hospital Bonn).

### 3. Results

#### 3.1. Patient and Disease Characteristics

The study included 92 PwMS, of which 64 (69.6%) were female and 28 (30.4%) male. Of these, 9 PwMS (9.8%) showed the primary progressive, 69 PwMS (75.0%) showed the relapsing–remitting, and 14 PwMS (15.2%) showed the secondary progressive form of MS. Bladder emptying was spontaneous in 89 cases (96.7%); in 1 case (1.1%), it required triggering; and it was conducted in 2 cases (2.2%) by residual urine catheterisation after spontaneous micturition. The descriptive analysis of the demographic variables is shown in Table 1.

**Table 1.** Descriptive analysis of the demographic variables.

	Mean (SD)	Median (IQR)	Min; Max	Missing N (%)
Age of patients in years	47.5 (9.4)	48 (40.8; 55.2)	25; 72	0 (0)
Duration of MS in months	118.7 (96.7)	102.5 (35.8; 168.2)	1; 445	0 (0)
Expanded Disability Status Scale	3.9 (1.4)	4 (2.9; 4.5)	1.5; 8	12 (13.04)

Of the 92 PwMS, 20 (21.7%) had no urological symptoms at the time of the analysis, 3 (3.3%) had symptoms lasting 1–6 months, 1 (1.1%) for 6–12 months, 15 (16.3%) for 1 to 2 years, 15 (16.3%) for 3 to 5 years, 12 (13%) for 6–10 years, and 5 (5.4%) over 10 years. Overall, 21 PwMS (22.8%) did not specify the duration of urological symptoms. An overview of the distribution and type of urological symptoms is shown in Table 2.

**Table 2.** Overview of the urinary symptoms of the included PwMS (number of PwMS (N) = 92, no missings).

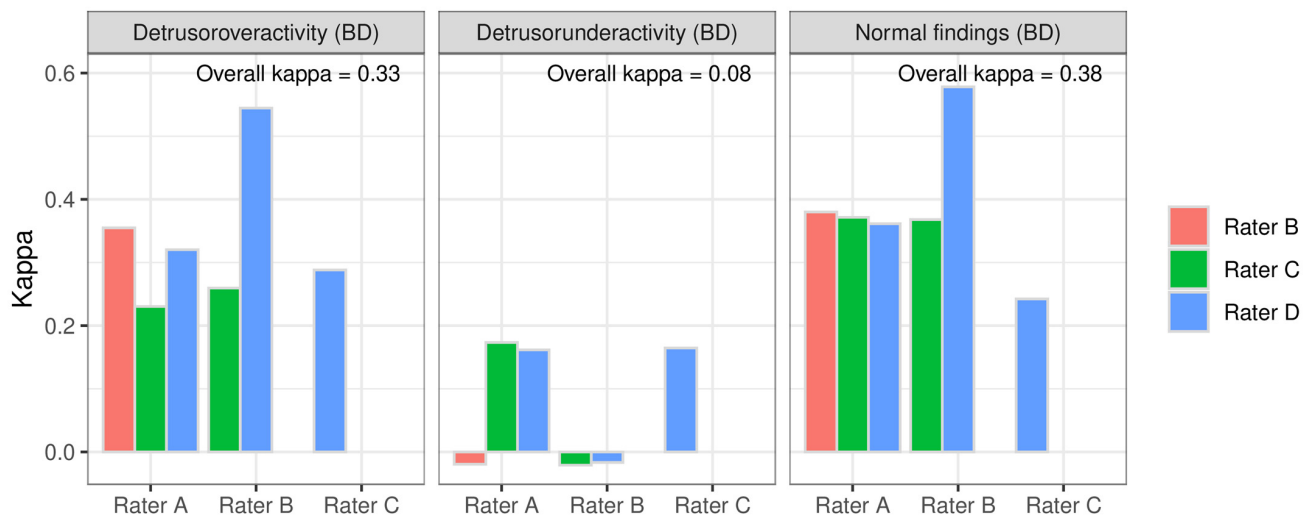
Neurogenic Lower Urinary Tract Symptoms	Yes N (%)	No N (%)
Staged micturition/interrupted urine flow	18 (19.6)	74 (80.4)
Initiation delay	11 (12)	81 (88)
Weak urine stream	11 (12)	81 (88)
Residual urine sensation	18 (19.6)	74 (80.4)
Desire to press	15 (16.3)	77 (83.7)
Drizzle	5 (5.4)	87 (94.6)
Urinary retention	1 (1.1)	91 (98.9)
Nocturia	37 (40.2)	55 (59.8)
Pollakiuria	45 (48.9)	47 (51.1)
Urgency	52 (56.5)	40 (43.5)
Incontinence	50 (54.3)	42 (45.7)

#### 3.2. Interrater Reliability of the Suspected Diagnoses

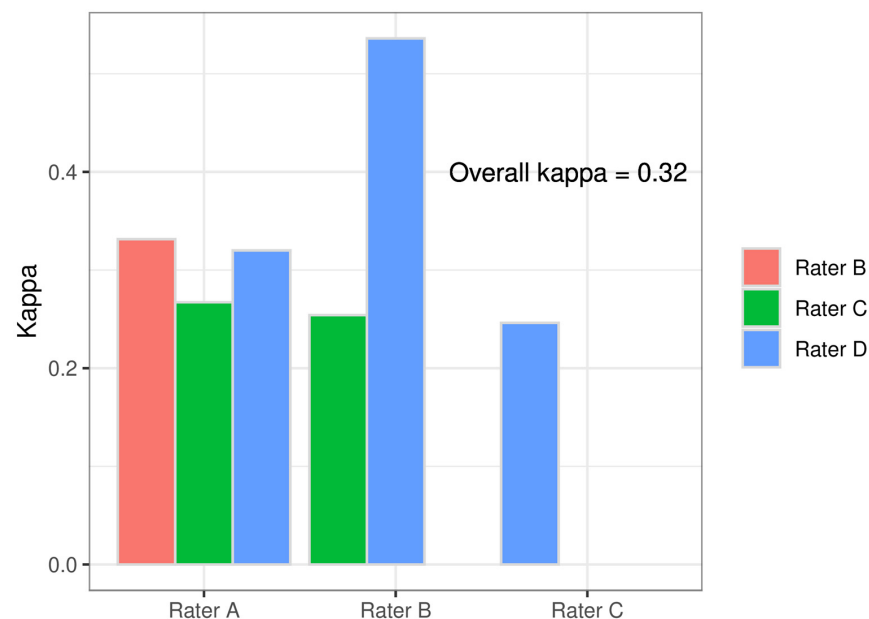
##### 3.2.1. Analysis A

For the suspected diagnosis from medical history and BDs, the raters had 3 options: normal findings (NF), detrusor overactivity (DO) and detrusor underactivity (DU) for single selection. The paired comparisons for each singular suspected diagnosis were as follows:

for NFs,  $\kappa_C = 0.24$  to  $\kappa_C = 0.58$ ; for DO,  $\kappa_C = 0.23$  to  $\kappa_C = 0.54$ ; and for DU,  $\kappa_C = -0.02$  to  $\kappa_C = 0.17$  (Figure 2). The paired comparison between the raters for all suspected diagnoses showed Cohens' kappa values of 0.25 to 0.54 and thus fair-to-moderate agreement [18] (Figure 3). The results of the comparison of all raters (Fleiss' kappa) are shown in Figures 2 and 3. DU achieved the lowest IRR.



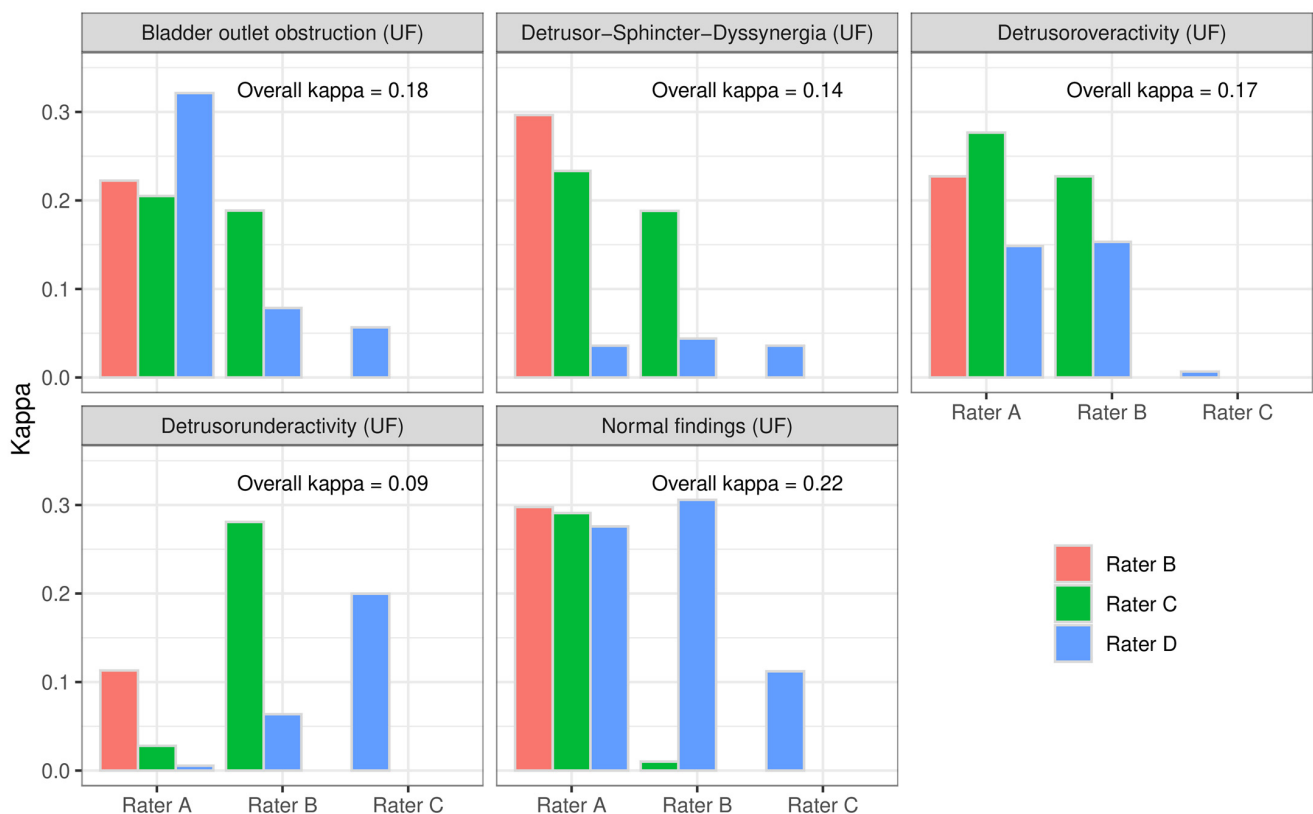
**Figure 2.** Interrater reliability of single suspected diagnoses from anamnesis and bladder diary between the raters paired and overall.



**Figure 3.** Interrater reliability of all suspected diagnoses from anamnesis and bladder diary between the raters paired and overall.

### 3.2.2. Analysis B

The raters had 5 options for the suspected diagnosis from medical history, BDs, and UF. In addition to NFs, DO, DU, bladder outlet obstruction (BOO), and detrusor-sphincter-dyssynergia (DSD) could be selected. This resulted in multiple-choice answers and thus 14 combinations for suspected diagnoses. The IRR was notably lower for combinations compared to the individual diagnoses with a maximum  $\kappa_C = 0.19$ . Figure 4 shows the paired interrater reliability of the single suspected diagnoses (analysis B-1). As in analysis A, the diagnosis DU achieved the lowest IRR. Table 3 shows the paired interrater reliability of the combinations for suspected diagnoses (analysis B-2).



**Figure 4.** Interrater reliability of the single suspected diagnoses between the paired raters and overall (analysis B-1).

**Table 3.** Overview of interrater reliability of combined suspected diagnoses between the raters paired and overall (analysis B-2).

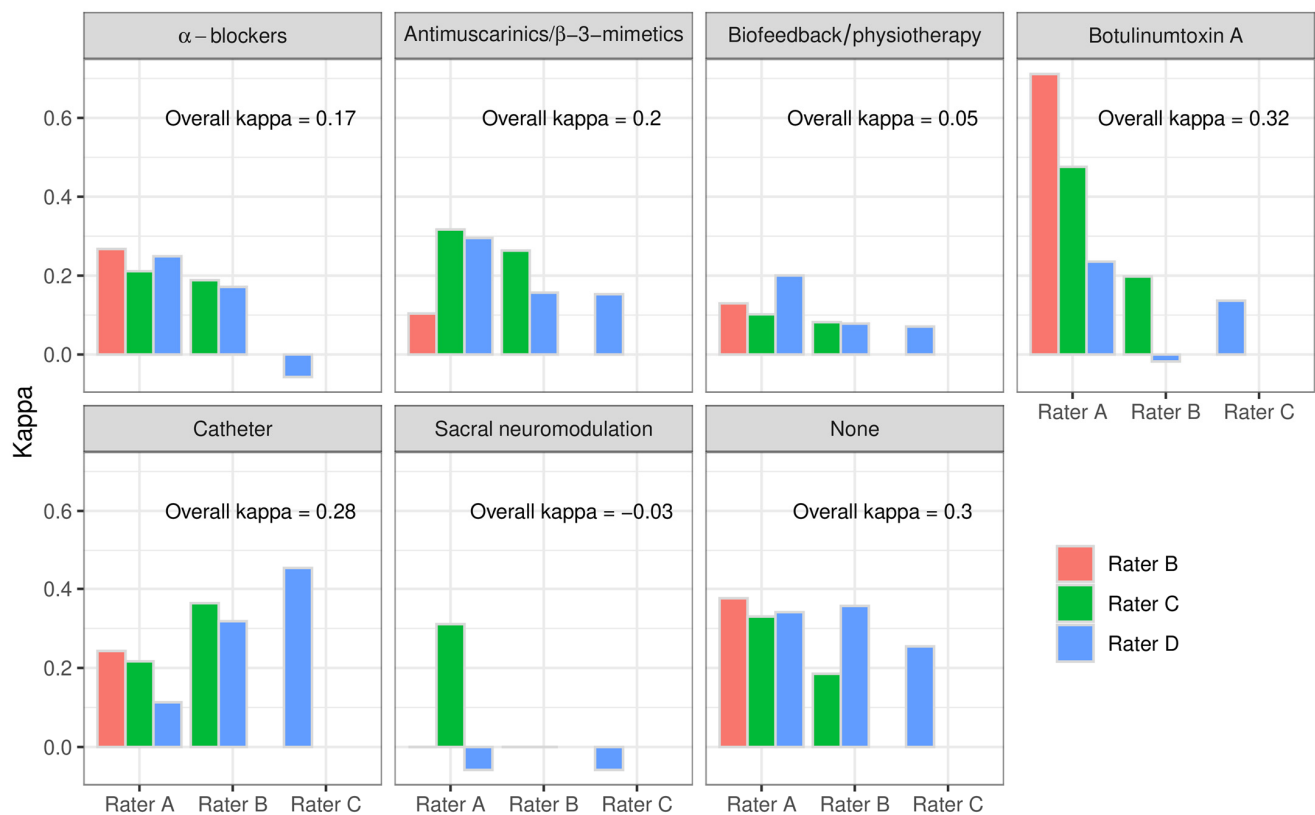
Paired Interrater Reliability of the Combined Suspected Diagnoses			
Rater	B	Match rate in % Cohen's kappa 95% CI	
		C	D
A	30.4%	22.8%	23.9%
	0.19 0.09; 0.28	0.13 0.04; 0.22	0.14 0.06; 0.23
B		22.8%	21.7%
		0.12 0.03; 0.21	0.1 0.01; 0.19
C			25%
			0.09 0; 0.18
Fleiss' kappa 0.12			

A uniform consensus with 100% agreement between all raters was found for the diagnoses NO, DU, BOO, and DSD in one data set each and for DO in 8 data sets.

### 3.3. Interrater Reliability of the Therapy Suggestions (Analysis C)

The IRR of the therapy suggestions from UF was determined to be analogous to the suspected diagnoses for each of the seven predefined therapy suggestions and for the therapy combinations that occurred. For the several treatment suggestions considered, NO

$\kappa_C = 0.38$ , KAT  $\kappa_C = 0.45$ , and BTX  $\kappa_C = 0.71$  showed the highest IRR between the rater pairs. We revealed the lowest IRR for NM  $\kappa_C = -0.06$  and P/BF  $\kappa_C = 0.07$  (Figure 5).



**Figure 5.** Interrater reliability of the single therapy suggestions between the raters paired and overall.

The possibility of therapy combinations resulted in a variety of selected combinations. The combined treatment suggestions showed a remarkably low agreement with a Fleiss' kappa of 0.08. The best agreement in this subanalysis was achieved by the therapy suggestion "no therapy" with a  $\kappa_F = 0.31$ . Table 4 provides an overview of the IRR between the rater pairs for combinations of treatment suggestions.

**Table 4.** Overview of the interrater reliability between the rater pairs for combinations of treatment suggestions (analysis C-2).

Paired Interrater Reliability of the Combined Therapy Suggestions			
Rater	Match rate in % Cohen's kappa 95% CI		
	B	C	D
A	15.2% 0.09 0.02; 0.16	17.4% 0.1 0.03; 0.16	19.6% 0.1 0.02; 0.18
B		17.4% 0.09 0.02; 0.17	10.9% 0.05 0.0; 0.11
C			17.4% 0.1 0.03; 0.17
Fleiss' kappa 0.08			



#### 4. Discussion

In everyday urological practice, it is often impossible to examine PwMS regarding NLUTD using video urodynamics. Although the examination represents the gold standard in the diagnostics of NLUTD [7], it is invasive, associated with side effects [20], costly, and limited in availability [21]. Compared to persons with traumatic or congenital damage of the spinal cord, NLUTD in PwMS results in less damage to the upper urinary tract [22] and not every PwMS is initially affected by NLUTD [23]. Therefore, the rationalised work-up of the NLUTD in PwMS is continually the subject of various expert panels and guideline recommendations [23]. However, a standardised recommendation and predictor for the presence of NLUTD—reviewed in prospective longitudinal studies—have not yet been clearly identified. At the same time, multiple sclerosis results in a high financial and personnel burden on the healthcare system [24] and leads to a severe loss in the quality of life of those affected [23]. To address all those needs, it is essential to develop a sufficient and less costly screening/preselection procedure [9]. In light of these facts, there is a growing interest in UF as a non-invasive, straightforward urodynamic procedure [1,6]. UF is widely available in urological practices and together with medical history and BDs it is often the basis of treatment decisions in daily practice. Therefore, the aim of our work was to analyse the IRR of UF not only regarding the suspected diagnosis, but also the derived clinical consequences. To this end, we had 4 neuro-urologically experienced raters assess 92 data sets based on medical history, BDs, and UF from PwMS regarding predefined suspected diagnoses and therapy suggestions.

For the diagnoses from medical history, BDs, and UF, we obtained a fair agreement with kappa values of 0.32 ( $\kappa_C$ ) and 0.2 ( $\kappa_F$ ), respectively, for the singular consideration of the diagnoses, both between the rater pairs and in the overall rater comparison. However, many kappa values were lower, especially when combinations of diagnoses were considered. For the singular analyses of the treatment suggestions, the kappa values tended to be slightly higher. Particularly in the treatment suggestions “catheter”, “botulinum toxin” and “none”, the kappa between the rater pairs was up to 0.71 and was 0.32 in the overall comparison of all raters. When looking at treatment combinations, the IRR fell significantly to values below 0.2 for slight agreement. The fewer options were available, the higher was the raters’ agreement.

This is a methodological problem: the more choices available to raters, the lower is the probability of deciding for the same result. This is particularly the case when the decisions are not dichotomous, e.g., pathological or non-pathological, but are based on an individual interpretation by the rater. Gacci et al. [16], involving over 100 urologists specialised in functional urology, were able to show that there is a high IRR when curves and numerical uroflowmetry results are categorised as normal or abnormal. When a diagnosis was included, the IRR in the same group of investigators fell to a  $\kappa$  value of up to <0.1, except for the diagnosis “no abnormalities”, where substantial agreement (0.7) was still achieved. This could be because the decision in favour of or against a therapy is still based on a dichotomous decision-making process. The study by van de Beek et al. [25] showed analogous results as early as 1997, whereby the diagnosis “normal findings” also showed the best IRR.

In paediatric urological studies, the IRR of UF also showed similar results. Chang and Yang, 2008, showed a regularly substantial  $\kappa$  (0.68–0.81) in the differentiation of normal vs. pathological findings; in the assessment of specific, pathological waveforms, the IRR was  $\kappa = 0.07$  for the same investigators [26]. As soon as an interpretative aspect was included in the diagnosis, the IRR dropped considerably.

The study by Faasse et al. [15] also focussed on the assessment of the UF curve shape in children. The IRR of uroflow EMG was analysed and a specific selection of diagnoses was specified in the study design. Working according to fixed diagnostic criteria, this was a single selection. An agreement of  $\kappa = 0.33$  up to a maximum of 0.74 was demonstrated between the rater pairs. The slightly lower kappa values observed in our work in comparison to those reported in the studies mentioned before can be attributed



to the fact that, in addition to the querying specific curve shapes, several diagnostic suggestions may also be applicable. In our study, the storage function of the bladder could be assessed using normal findings, detrusor overactivity, and detrusor underactivity, as well as the micturition phase with DSD or BOO. This resulted in many combinations, despite the predefined answers in the form. However, these answers reflect the complexity of the diagnostic and decision-making process that results from UF in daily practice: a clear idea of the underlying diagnosis is essential to determining treatment recommendations. A simple yes/no decision as to whether a pathological condition exists is not sufficient.

The low IRR for DU was noticeable in all our results. On the one hand, we believe that this is due to the poor differentiability between DU and BOO in UF for both genders [27–30], as the shape of the curve and the measured uroflow parameters can be identical for both diagnoses. On the other hand, the clinical definition and diagnosis of DU and detrusor hypocontractility [30–34] are still not finally clarified. Additional indices and parameters are constantly being tested [28,29,31] in order to reliably differentiate the diagnosis of DU with non-invasive measures. These aspects have a negative influence on the consistency of the raters, which influences  $\kappa_C$  [14]. Another aspect of information that has an impact on  $\kappa_C$  is the prevalence of a trait. The less frequently a characteristic is actually present or the less frequently it is estimated to be present, the lower  $\kappa_C$  is [14]. We are unable to assess the influence of the prevalence of the diagnoses in our current study at this time, as we have not correlated the diagnoses with the urodynamic results. This will be the subject of a further study in which the IRR of UD and UF will be compared.

In summary, various authors consider the diagnostic value of UF to be insufficient [13] or as limited [15–17] due to the IRR achieved. Some authors consider UF to be a good screening tool between normal and abnormal micturition [26]. However, all authors see the need for standardisation in the reporting of the curves [15–17,26].

In our opinion, in its current form UF can be an adjunct to the management of NLUTD in PwMS, e.g., act as a follow-up during therapy compared to a pre-treatment assessment. However, the design of all previous studies on IRR does not allow the conclusion that having a limited IRR is also detrimental to the affected patient. The reasons for decisions are not included and could provide reasonable explanations for different decisions. Each rater has an individual horizon of experience and follow-up strategies. The IRR would need to be combined with an outcome measure of treatment success to be able to make statements about the question of unfavourable treatment success with a low IRR.

In the future, artificial intelligence-supported, learning algorithms based on UD studies could be used to refine the diagnostic value of UF for PwMS. These would need to be designed to allow dichotomous decisions to be made as a small detail of the whole, leading to a more complex diagnostic or therapeutic decision. Parameters with a high degree of selectivity and impact on the therapy decision must be identified for this purpose. The first artificial intelligence-assisted approaches to the diagnosis of detrusor underactivity in men are already available [32]. Choo et al. have developed algorithms for the reporting of uroflowmetry in men and women using artificial intelligence [35]. However, the curve so far can only be assessed as normal or pathological, and the results are compared with the judgements of urological specialists and not with underlying urodynamics. Incorporating uroflowmetry with other clinical predictors, working according to the methodology demonstrated by Ito et al. [36], could offer another comprehensive approach to estimating the necessity for additional diagnostics for NLUTD in PwMS. Therefore, further investigations are needed to correlate the results of UD with non-invasive diagnostics and clinical treatment outcomes in PwMS.

## 5. Conclusions

The suspected suggested diagnosis and treatment suggestions from UF are subject to a high level of individual influence by the raters. If there are several differential diagnoses and therapies to choose from, as it occurs in the daily use of UF, interrater reliability decreases considerably. This results in a variety of possible treatment decisions. To evaluate

the possible long-term relevance of UF in addition to UD, the current gold standard in diagnostics of NLUTD in PwMS, UD should be subjected to the same analysis. The results of this analysis then should be compared with those of the UF. According to our study, the standardisation of UF findings based on a dichotomous decision-making algorithm could help to improve the consistency of the assessment.

## 6. Limitations

Our study had some limitations. The main limitation was that the findings were based on medical records. The raters did not have the real patients in front of them. The influence of the patient's constitution, such as physical or mental limitations, and the effect on the decision of the rater was therefore completely absent. Furthermore, the data originate from a neurological inpatient rehabilitation clinic, which has an influence on the type of PwMS participating, as more severe or advanced cases requiring rehabilitation are usually treated in these institutions (selection bias). In summary, all raters had the same conditions, and the focus of this study was on interrater reliability and not on the most correct result for the patient.

**Author Contributions:** Conceptualization, A.K.J. and R.K.-H.; methodology, A.K.J., S.C.K. and A.H.; software, J.R.; validation, J.R. and A.K.J.; formal analysis, P.S.; investigation, S.M., F.Q., O.S. and A.K.J.; resources, A.-L.B.; data curation, A.-L.B. and J.R.; writing—original draft preparation, A.K.J.; writing—review and editing, A.H.; visualization, P.S. and J.R.; supervision, S.C.K.; project administration, S.C.K.; funding acquisition, R.K.-H. All authors have read and agreed to the published version of the manuscript.

**Funding:** The primary funding was provided by the German charity fund 'Förderverein zur Kontinenzforschung und Kontinenzaufklärung e. V.', Karmeliterhöfe, Karmeliterstr. 10, 52064 Aachen, Germany, grant number 2023/05. This work was supported by the Open Access Publication Fund of the University of Bonn.

**Institutional Review Board Statement:** This study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee for ethical approval (EK 313/13-University Hospital Bonn).

**Informed Consent Statement:** Informed consent was obtained from all the subjects involved in this study.

**Data Availability Statement:** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Gammie, A.; Rosier, P.; Li, R.; Harding, C. How can we maximize the diagnostic utility of uroflow?: ICI-RS 2017. *Neurourol. Urodyn.* **2018**, *37*, S20–S24. [CrossRef] [PubMed]
2. Radmayr, C.; Bogaert, G.; Bujons, A.; Burgu, B.; Castagnetti, M.; 't Hoen, L.A.; O'Kelly, F.; Pakkasjärvi, N.A.; Quadackers, J.; Rawashdeh, Y.F.H.; et al. *EAU Guidelines on Paediatric Urology*; EAU Guidelines Office: Arnhem, The Netherlands, 2023. Available online: <https://uroweb.org/guidelines/paediatric-urology> (accessed on 6 March 2024).
3. Cornu, J.N.; Gacci, M.; Hashim, H.; Herrmann, T.R.M.; Malde, S.; Netsch, C.; De Nunzio, C.; Rieken, M.; Sakalis, V.; Tutolo, M. *EAU Guidelines on Management of Non-Neurogenic Male LUTS*; EAU Guidelines Office: Arnhem, The Netherlands, 2023. Available online: <https://uroweb.org/guidelines/management-of-non-neurogenic-male-luts> (accessed on 6 March 2024).
4. Schäfer, W.; Abrams, P.; Liao, L.; Mattiasson, A.; Pesce, F.; Spangberg, A.; Sterling, A.M.; Zinner, N.R.; Kerrebroeck, P.V. Good urodynamic practices: Uroflowmetry, filling cystometry, and pressure-flow studies. *Neurourol. Urodyn.* **2002**, *21*, 261–274. [CrossRef] [PubMed]
5. Harding, C.K.; Lapitan, M.C.; Arlandis, S.; Bø, K.; Cobussen-Boekhorst, H.; Costantini, E.; Groen, J.; Nambiar, A.K.; Omar, M.I.; Peyronnet, B.; et al. *EAU Guidelines on Management of Non-Neurogenic Female LUTS*; EAU Guidelines Office: Arnhem, The Netherlands, 2023. Available online: <https://uroweb.org/guidelines/non-neurogenic-female-luts> (accessed on 6 March 2024).
6. El Helou, E.; Sarkis, J.; Mjaess, G.; Zalaket, J.; Mouawad, C.; Sayegh, N.; Ghattas, S.; Azar, C.; El Helou, J.; Abboud, H.; et al. Urodynamics in patients with multiple sclerosis: Is it necessary? A randomized-controlled trial. *Scand. J. Urol.* **2021**, *55*, 161–168. [CrossRef] [PubMed]

7. Blok, B.; Castro-Diaz, D.; Del Popolo, G.; Groen, J.; Hamid, R.; Karsenty, G.; Karsenty, T.M.; Musco, S.; Padilla-Fernández, B.; Pannek, J. *EAU Guidelines on Neuro-Urology*; EAU Guidelines Office: Arnheim, The Netherlands, 2023. Available online: <https://uroweb.org/guidelines/neuro-urology> (accessed on 6 March 2024).
8. Fowler, C.J.; Panicker, J.N.; Drake, M.; Harris, C.; Harrison, S.C.W.; Kirby, M.; Lucas, M.; Macleod, N.; Mangnall, J.; North, A.; et al. A UK consensus on the management of the bladder in multiple sclerosis. *Postgrad. Med. J.* **2009**, *85*, 552–559. [[CrossRef](#)]
9. Gammie, A.; Speich, J.E.; Damaser, M.S.; Gajewski, J.B.; Abrams, P.; Rosier, P.F.W.M.; Arlandis, S.; Tarcan, T.; Finazzi Agrò, E. What developments are needed to achieve less-invasive urodynamics? ICI-RS 2019. *Neurourol. Urodyn.* **2020**, *39* (Suppl. S3), S36–S42. [[CrossRef](#)]
10. Rosier, P.F.W.M.; Schaefer, W.; Lose, G.; Goldman, H.B.; Guralnick, M.; Eustice, S.; Dickinson, T.; Hashim, H. International Continence Society Good Urodynamic Practices and Terms 2016: Urodynamics, uroflowmetry, cystometry, and pressure-flow study. *Neurourol. Urodyn.* **2017**, *36*, 1243–1260. [[CrossRef](#)]
11. Dönmez, M.İ.; Özkent, M.S.; Hamarat, M.B.; Kocalar, M. Through the zipper or pants down: Does it change uroflowmetry parameters in healthy males? *Low. Urin. Tract Symptoms* **2022**, *14*, 341–345. [[CrossRef](#)]
12. Gammie, A.; Drake, M.J. The fundamentals of uroflowmetry practice, based on International Continence Society good urodynamic practices recommendations. *Neurourol. Urodyn.* **2018**, *37*, S44–S49. [[CrossRef](#)] [[PubMed](#)]
13. Netto, J.M.B.; Hittelman, A.; Lambert, S.; Murphy, K.; Collette-Gardere, T.; Franco, I. Interpretation of uroflow curves: A global survey measuring inter and intra rater reliability. *Neurourol. Urodyn.* **2020**, *39*, 826–832. [[CrossRef](#)]
14. Wirtz, M.; Kutschmann, M. Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Masse. *Rehabilitation* **2007**, *46*, 370–377. [[CrossRef](#)]
15. Faasse, M.A.; Nosnik, I.P.; Diaz-Saldano, D.; Hodgkins, K.S.; Liu, D.B.; Schreiber, J.; Yerkes, E.B. Uroflowmetry with pelvic floor electromyography: Inter-rater agreement on diagnosis of pediatric non-neurogenic voiding disorders. *J. Pediatr. Urol.* **2015**, *11*, 198.e1–198.e6. [[CrossRef](#)] [[PubMed](#)]
16. Gacci, M.; Del Popolo, G.; Artibani, W.; Tubaro, A.; Palli, D.; Vittori, G.; Lapini, A.; Serni, S.; Carini, M. Visual assessment of uroflowmetry curves: Description and interpretation by urodynamists. *World J. Urol.* **2007**, *25*, 333–337. [[CrossRef](#)] [[PubMed](#)]
17. Liu, Y.-B.; Yang, S.S.; Hsieh, C.-H.; Lin, C.-D.; Chang, S.-J. Inter-Observer, Intra-Observer and Intra-Individual Reliability of Uroflowmetry Tests in Aged Men: A Generalizability Theory Approach. *Low. Urin. Tract Symptoms* **2014**, *6*, 76–80. [[CrossRef](#)] [[PubMed](#)]
18. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, *33*, 159. [[CrossRef](#)]
19. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2020. Available online: <https://www.R-project.org/> (accessed on 9 March 2024).
20. Klingler, H.C.; Madersbacher, S.; Djavan, B.; Schatzl, G.; Marberger, M.; Schmidbauer, C.P. Morbidity of the evaluation of the lower urinary tract with transurethral multichannel pressure-flow studies. *J. Urol.* **1998**, *159*, 191–194. [[CrossRef](#)] [[PubMed](#)]
21. Baunacke, M.; Leuchtweis, I.; Kaufmann, A.; Schmidt, M.; Groeben, C.; Borkowetz, A.; Eisenmenger, N.; Thomas, C.; Huber, J. Decreasing Number of Urodynamics in Urological and Gynaecological Clinics Reflects Decreased Importance for Surgical Indications: German Population-Based Data from 2013 to 2019. *Urol. Int.* **2022**, *106*, 1068–1074. [[CrossRef](#)] [[PubMed](#)]
22. Musco, S.; Padilla-Fernández, B.; Del Popolo, G.; Bonifazi, M.; Blok, B.F.M.; Groen, J.; 't Hoen, L.; Pannek, J.; Bonzon, J.; Kessler, T.M.; et al. Value of urodynamic findings in predicting upper urinary tract damage in neuro-urological patients: A systematic review. *Neurourol. Urodyn.* **2018**, *37*, 1522–1540. [[CrossRef](#)] [[PubMed](#)]
23. Aharony, S.M.; Lam, O.; Corcos, J. Evaluation of lower urinary tract symptoms in multiple sclerosis patients: Review of the literature and current guidelines. *Can. Urol. Assoc. J.* **2017**, *11*, 61–64. [[CrossRef](#)] [[PubMed](#)]
24. Paz-Zulueta, M.; Parás-Bravo, P.; Cantarero-Prieto, D.; Blázquez-Fernández, C.; Oterino-Durán, A. A literature review of cost-of-illness studies on the economic burden of multiple sclerosis. *Mult. Scler. Relat. Disord.* **2020**, *43*, 102162. [[CrossRef](#)]
25. Van de Beek, C.; Stoevelaar, H.J.; McDonnell, J.; Nijs, H.; Casparie, A.F.; Janknegt, R.A. Interpretation of Uroflowmetry Curves by Urologists. *J. Urol.* **1997**, *157*, 164–168. [[CrossRef](#)]
26. Chang, S.-J.; Yang, S.S.D. Inter-observer and intra-observer agreement on interpretation of uroflowmetry curves of kindergarten children. *J. Pediatr. Urol.* **2008**, *4*, 422–427. [[CrossRef](#)] [[PubMed](#)]
27. Lee, K.S.; Song, P.H.; Ko, Y.H. Does uroflowmetry parameter facilitate discrimination between detrusor underactivity and bladder outlet obstruction? *Investig. Clin. Urol.* **2016**, *57*, 437–441. [[CrossRef](#)] [[PubMed](#)]
28. Chow, P.-M.; Hsiao, S.-M.; Kuo, H.-C. Identifying occult bladder outlet obstruction in women with detrusor-underactivity-like urodynamic profiles. *Sci. Rep.* **2021**, *11*, 23242. [[CrossRef](#)] [[PubMed](#)]
29. Arevalo-Vega, D.; Ponce, L.; Valdevenito, J.P.; Gallegos, H.; Dell'Oro, A.; Santis-Moya, F.; Calvo, C.I. Defining bladder outlet obstruction and detrusor underactivity in females with overactive bladder: Are we forgetting about the free uroflowmetry? *Neurourol. Urodyn.* **2023**, *42*, 1255–1260. [[CrossRef](#)] [[PubMed](#)]
30. Ahmed, A.; Farhan, B.; Vernez, S.; Ghoniem, G.M. The challenges in the diagnosis of detrusor underactivity in clinical practice: A mini-review. *Arab J. Urol.* **2016**, *14*, 223–227. [[CrossRef](#)] [[PubMed](#)]
31. Oelke, M.; Rademakers, K.L.J.; Van Koeveeringe, G.A. Unravelling detrusor underactivity: Development of a bladder outlet resistance-Bladder contractility nomogram for adult male patients with lower urinary tract symptoms. *Neurourol. Urodyn.* **2016**, *35*, 980–986. [[CrossRef](#)] [[PubMed](#)]

32. Matsukawa, Y.; Kameya, Y.; Takahashi, T.; Shimazu, A.; Ishida, S.; Yamada, M.; Sassa, N.; Yamamoto, T. Characteristics of uroflowmetry patterns in men with detrusor underactivity revealed by artificial intelligence. *Int. J. Urol.* **2023**, *30*, 907–912. [\[CrossRef\]](#)
33. Lee, J.; Yoo, S.; Cho, M.C.; Jeong, H.; Choo, M.S.; Son, H. Significance of a decrease in the proportion of detrusor muscle to bladder wall for non-invasive diagnosis of detrusor underactivity in men with lower urinary tract symptoms. *Sci. Rep.* **2022**, *12*, 5237. [\[CrossRef\]](#)
34. Osman, N.I.; Chapple, C.R.; Abrams, P.; Dmochowski, R.; Haab, F.; Nitti, V.; Koelbl, H.; Van Kerrebroeck, P.; Wein, A.J. Detrusor underactivity and the underactive bladder: A new clinical entity? A review of current terminology, definitions, epidemiology, aetiology, and diagnosis. *Eur. Urol.* **2014**, *65*, 389–398. [\[CrossRef\]](#)
35. Choo, M.S.; Ryu, H.Y.; Lee, S. Development of an Automatic Interpretation Algorithm for Uroflowmetry Results: Application of Artificial Intelligence. *Int. Neurourol. J.* **2022**, *26*, 69–77. [\[CrossRef\]](#)
36. Ito, H.; Sakamaki, K.; Young, G.J.; Blair, P.S.; Hashim, H.; Lane, J.A.; Kobayashi, K.; Clout, M.; Abrams, P.; Chapple, C.; et al. Predicting prostate surgery outcomes from standard clinical assessments of lower urinary tract symptoms to derive prognostic symptom and flowmetry criteria. *Eur. Urol. Focus* **2024**, *10*, 197–204. [\[CrossRef\]](#) [\[PubMed\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.