Received: 10 November 2022 | Accepted: 10 February 2023

DOI: 10.1002/nau.25157

CLINICAL ARTICLE

Neurourology Urodynamics | SUFU | WILEY

The test–retest reproducibility of the multiple array probe Leiden in men with lower urinary tract symptoms

Martina Beverini[1,2] | **Selma Goes**[3] | **Lambertus P. W. Witte**[4] | **Gommert A. van Koeveringe**[1] | **Nienke van der Laan**[3] | **Grietje E. Knol-de Vries**[3] | **Marco H. Blanker**[3]

[1]Department of Urology, Maastricht, University Medical Center, Maastricht, The Netherlands

[2]Istituto di Ricovero e Cura a Carattere Scientifico, Ospedale Policlinico San Martino, Genoa, Italy

[3]Department of General Practice and Elderly Care Medicine, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

[4]Department of Urology, Isala Hospital, Zwolle, The Netherlands

Correspondence
Marco H. Blanker, Department of General Practice and Elderly Care Medicine, University Medical Center Groningen, University of Groningen, P.O. Box 998, 9700 AZ, Groningen, The Netherlands.
Email: m.h.blanker@umcg.nl

Abstract

Background: We aimed to study the test–retest reliability of the Multiple Array Probe Leiden (MAPLe), a multiple electrode probe designed to acquire and discriminate electromyography signals in the pelvic floor muscles, in men with lower urinary tract symptoms (LUTS).

Methods: Adult male patients with LUTS with sufficient knowledge of Dutch language, but without complications (e.g., urinary tract infection), or previous urologic cancer and/or urologic surgery were enrolled. In the initial study, next to physical examination and uroflowmetry, all men underwent MAPLe assessment at baseline and after 6 weeks. Second, participants were reinvited for a new assessment using a stricter protocol. A time interval of 2 h (M2) and 1 week (M3) after baseline (M1) allowed the calculation of the intraday agreement (M1 vs. M2), and the interday agreement (M1 vs. M3) for all 13 MAPLe variables.

Results: The outcomes of the initial study in 21 men suggested a poor test–retest reliability. The second study in 23 men showed a good test–retest reliability with intraclass correlations ranging from 0.61 (0.12–0.86) to 0.91 (0.81–0.96). The agreement was generally higher for the intraday determinations than for the interday determinations.

Conclusions: This study revealed a good test–retest reliability of the MAPLe device in men with LUTS, when using a strict protocol. With a less strict protocol, the test–retest reliability of MAPLe was poor in this sample. To make valid interpretations of this device in a clinical or research setting, a strict protocol is needed.

KEYWORDS
electromyography, male LUTS, pelvic floor muscles

Martina Beverini and Selma Goes contributed equally to this study.

1 | INTRODUCTION

Male lower urinary tract symptoms (LUTS) have a multifactorial origin that includes prostate and bladder dysfunction, but also pelvic floor disorders.[1,2] The latter has not been studied in detail in men, in contrast to women, although this problem has similar prevalence. This could partially be explained by the lack of methods to assess male pelvic floor muscles (PFM).

The Multiple Array Probe Leiden (MAPLe) is a multiple-electrode probe designed to acquire and discriminate electromyography (EMG) signals from the different sides and layers of the PFM.[3] Before this device can be used in the clinical or scientific evaluation of patients, more information about its reliability is needed. Until now, information about the test–retest reproducibility has only been obtained in healthy individuals, suggesting moderate reliability.[3]

As patient's characteristics could influence test–retest reliability, and in clinical practice, MAPLe will not be used in healthy individuals, we have conducted a test–retest reliability study in a clinical sample of men with LUTS. The outcomes of that study were presented at a scientific meeting and suggested a poor reliability.[4] Further debate of these outcomes suggested that the results could have been flawed, due to the study protocol that included a long period between baseline and follow-up and different assessors. As such, our conclusions cannot imply that this device is not suitable. Therefore, we decided to redo the study with an adjusted protocol. Here, we present the outcomes of both the initial and the renewed study. We aim to provide information about the test–retest reproducibility of the MAPLe device in men with LUTS.

2 | MATERIALS & METHODS

In 2018, an observational cohort study was conducted at the urology outpatient department of a large non-academic teaching hospital in The Netherlands. Details of that study were published elsewhere.[5] In short, adult male patients with LUTS (International Prostate Symptom Score [IPSS][6] 8 or higher), without complications (e.g., urinary tract infection) were enrolled. LUTS medications were allowed during the study. Exclusion criteria: previous urologic cancer and/or urologic surgery or insufficient knowledge of Dutch language. In addition, in the second study, IPSS was filled out again. Six weeks after the baseline assessment, a second assessment took place with the MAPLe device. This study is further referred to as the "initial study."

For the second study, 54 out of 57 patients from the initial study (to clarify, two patients died, while one patient was not invited due to an unpleasant experience in the initial study) were re-invited for a test–retest study with a much shorter follow-up time (2 h and 1 week after the baseline). Both studies will be explained with the outcomes of the initial study presented in the Supporting Information File, and the second study in the main text of this article.

2.1 | Measurements and procedure

Contrary to the first protocol, in which measurements were done at baseline and after 6 weeks; for the second study, patients were asked to participate in two measurements on Day 1 (baseline [M1] and 2 h after [M2]) then one extra measurement on Day 2 (1 week later, M3). Every participant completed the IPSS questionnaire at baseline to check study eligibility, next to the International Consultation on Incontinence Questionnaire on Male LUTS (ICIQ-MLUTS) and International Index of Erectile Function (IIEF-5).[7,8] Also, participants signed a new consent form.

In the initial protocol, different trained assessors performed the MAPLe measurements, while in the second study, a single researcher performed all MAPLe measurements, to avoid differences in the assessment. This researcher was a senior medical student (N. v. d. L.) who was trained by the MAPLe producer (Novuqare) and learned how to use the probe and to do measurements under the supervision of a dedicated pelvic floor physiotherapist. Novuqare had no other involvement in this study neither did they sponsor it.

Before each measurement, participants were asked to void and then to lay in the left lateral position with their hips and knees at approximately 70 degrees of flexion. A reference electrode was placed on the anterior superior right iliac spine. The probe was then inserted in the anus by the researcher, to obtain a correct positioning, the most caudal electrodes had to be located at the level of the external anal sphincter. The positioning of the probe was verified and was held by the researcher throughout the measurements. At the beginning, the participants were asked to briefly contract the abdominal, gluteal, and leg muscles, then, allowed to remain as relaxed as possible throughout the period of measurements to avoid co-contractions. Then, they practiced tightening of the PFM with five contractions in which the participants received feedback. After this, the actual MAPLe measurement was started. Participants were instructed not to talk, laugh, sneeze, or cough during the measurement; in case of these eventualities, the device was stopped and restarted after 60 s of rest. Furthermore, during the measurement, the participants no longer received any

feedback and the measurement screen was not visible to them.

After each measurement, participants filled out a questionnaire on how their had experienced with the procedure (possible pain, PFM fatigue, and a comparison with previous measurements).

MAPLe is a cylindrical-shaped probe (length 65 mm, diameter 15 mm) on which 24 electrodes are placed at 6 levels on 4 sides; the device measures EMG signals from the different sides and layers of the pelvic floor musculature. In the current study, four consecutive tasks were performed: 1 min rest, 10 maximal voluntary contraction (MVC) of 3 s each (contraction are explained to patients as "hold and relax"), three maximal endurance contractions of 30 s each, and two Valsalva maneuvers of 5 s each (it was practice once before taking of readings). There was a minimum of 30 s of rest in between each task.

A total of 13 variables are provided for every patient: average EMG value at rest and at MVC, peak EMG value at MVC, onset time and offset time MVC, average EMG value and peak EMG value at endurance with onset time and offset time, average EMG value and peak EMG value at Valsalva with onset time and offset time. All EMG values were measured in microvoltage and time intervals in second.

We defined the rest measurement as the mean value of 1 min rest measures, MVC as the mean value of 10 MVCs, and the endurance measurement as the mean value of 3 contractions. The Valsalva value is the average of the two Valsalva maneuvers. Onset and offset time are the time it takes to get from rest to maximum contraction and the time it takes to get from maximum contraction to rest, respectively.

2.2 | Statistical analysis

We used IBM SPSS Statistics 27.0 for all analyses. The demographic and the questionnaires results were presented with descriptive statistics. The test−retest reproducibility was investigated as follows: the intraday agreement was calculated between M1 and M2, and the interday agreement between M1 and M3 for all 13 MAPLe variables.

Bland−Altman plots were made to visualize the outcomes and identify any systematic differences.[9] These plots are used to describe the agreement between two quantitative measurements. The graph is a scatter plot, in which the *Y*-axis represents the difference between the two measurements and the X-axis represents the average. In a good agreement, the scattering of points is diminished, and points lie close to the line that represents the median. Bland−Altman plots were depicted as follows: dashed line as the zero-line, red line as the median difference, and green lines as the 10th and 90th percentiles of the differences.

Next, for normally distributed variables, the test–retest reliability was assessed using an intraclass correlation coefficient (ICC), applying the "Two-way random effects, single measures, absolute agreement" model. Values of ICC lower than 0.5 were considered an index of poor reliability, those between 0.5 and 0.75 moderate and above 0.75 good. Notably, ICC > 0.9 was rated as excellent reliability.[10]

For the not normally distributed outcomes, the test −retest reliability was determined using Spearman's correlation coefficient (SCC). The SCC ranges from −1 to +1, where −1 and +1 represent perfect consistency and 0 represents no consistency.[11]

3 | RESULTS

Of the 54 men invited for the second study, 35 agreed to participate, 18 refused participation, and 1 did not respond. Twelve participants were excluded: nine did not meet the inclusion criteria, while three could not be scheduled because the study was terminated earlier, due to the Coronavirus disease (COVID) measures. A flowchart of the inclusion is shown in Figure 1.

Baseline characteristics of the remaining 23 participants are shown in Table 1. The median age of the participants was 70 (interquartile range [IQR]: 63−75) and their median MLUTS score was 16 (IQR: 14−20), with a voiding subscore of 9 (IQR: 5−10) and an incontinence subscore of 4 (IQR: 2−6). Twelve participants (52.2%) were taking urinary tract medication (Table 1). Three previously had a transurethral resection of the prostate (TURP) and one had both TURP and laser treatment of the prostate. Eight (34.8%) had received pelvic physiotherapy or reported PMF exercises.

All but two of the Bland−Altman plots for intraday and interday measurements showed no systematic differences (see Figure A1A-Z); only the intraday and interday plots of the resting measurements showed that the median difference was positive (Figure A1A-B), this means that the PFM activity during the resting measurement was higher at M1 than at M2 and M3.

The ICC ranged from 0.61 (0.12−0.86) to 0.91 (0.81−0.96) (Table 2). The agreement was generally higher for the intraday determinations than for the interday determinations. An excellent intraday agreement was seen in the mean microvoltage of both the MVC measurement (ICC: 0.91 [0.81−0.96]) and the endurance measurement (ICC: 0.90 [0.79−0.96]).

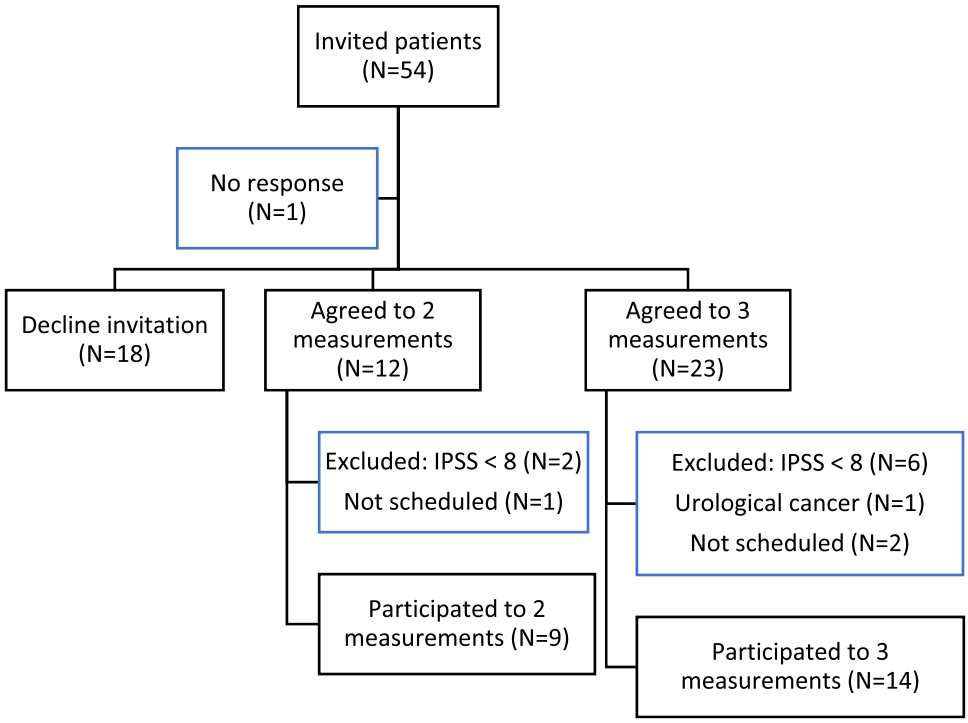

FIGURE 1 Flowchart Inclusion.

Moderate intraday agreement was seen at the peak microvoltage of both the MVC measurement (ICC: 0.67 [0.38−0.85]) and the endurance measurement (ICC: 0.64 [0.32−0.83]).

The SCC varied widely, from −0.17 to 0.84 (Table 2). Not all SCC values were statistically significant. The lowest correlation was seen at the intraday endurance offset time (SCC: −0.17). The endurance onset time (SCC intraday 0.70; interday 0.84) and the Valsalva maneuver offset time (SCC intraday 0.76; interday 0.70) had the highest correlation. Also in the Valsalva maneuver, a stronger correlation was seen among the intraday measurements compared with the interday measurements.

The outcomes of the test−retest reproducibility in the initial study (with 21 subjects) showed lower ICC and SCC for most comparisons (Supporting Information File).

4 | DISCUSSION

We found a good test−retest reliability of the MAPLe device in men with LUTS, when using a strict protocol including a single assessor and a short time interval. With a less strict protocol, the test−retest reliability of MAPLe was poor in this sample.

In the validation study of the MAPLe, conducted on healthy male volunteers, the authors have described the assessment to be moderate based on an ICC of 0.53−0.70.[3] Differences could be explained by the choice of the interpretation model for the ICC, for which several models are mentioned in literature. The impact of such models on the conclusions is illustrated by the interpretation of the ICC values found in our study ranging from 0.61 to 0.91. An ICC of 0.91 is interpreted as good according to all models, for example, as excellent, high, or substantial.[10,12,13] The lower values, however, have a very variable interpretation. An ICC of 0.61 is rated as unacceptable or questionable, but also as good.[10,12,13] The same dilemma occurs when interpreting the SCC.[11,14]

Another explanation of the differences between both studies is the lack of information in the validation study.[3] In fact, the 95% confidence interval of the ICC and the time interval between the test and retest measurements were not mentioned in that study.

Although the ICC and SCC suggested a reliable test−retest, we noted a considerable spread of the differences between the measurements in some of the Bland−Altman plots. Since not so much research has been done into the use of EMG in men with LUTS, there is no gold standard against which these differences can be tested. However, when the variation of the differences is almost as great as the variation of the mean measured value, the accuracy is questionable.

Based on the high ICC values and the relatively small spread of the differences in the Bland−Altman plots, it was concluded that the average microvoltages of the rest,

TABLE 1 Characteristic of the 23 participants.

	Median (IQR)	Number (%)
Age	70 (63–75)	
BMI		
Normal weight (BMI 18.5–24.9)		8 (34.8%)
Overweight (BMI 25–29.9)		13 (56.5%)
Obese (BMI 30–39.9)		2 (8.7%)
MLUTS score	16 (14–20)	
Voiding subscore	9 (5–10)	
Incontinence subscore	4 (2–6)	
Influence of urinary complaints on daily life (0–10)	5 (2–8)	
Medication		
Alpha-Blockers		8 (34.8%)
5-alpha-reductase		2 (8.7%)
Tolterodin		1 (4.3%)
Mirabegron and alpha-blockers		1 (4.3%)
Previous surgery		
TURP		3 (13%)
TURP and laser treatment		1 (4.3%)
Previous pelvic physiotherapy		6 (26.1%)
Perform the pelvic floor exercise		6 (26.1%)
Stool		
Normal stool frequency		15 (65.2%)
Normal stool consistence		15 (65.2%)
Difficulty defecating		6 (26.1%)
Involuntary stool loss		4 (17.4%)
Feeling that mucous membrane/tissue is coming out of the anus		3 (13%)
Sexual activity with the partner		13 (56.5%)
IIEF-5 (N = 13)		
No erectile dysfunction		1 (7.7%)
Mild erectile dysfunction		7 (53.8%)
Mild to moderate erectile dysfunction		3 (23.1%)
Moderate erectile dysfunction		1 (7.7%)
Severe erectile dysfunction		1 (7.7%)
Pain in the pelvic area		
Pain in 1 location		6 (26.1%)
Pain in 2–3 locations		2 (8.7%)
Pain in 5 locations		1 (4.3%)
Pain score (0–10)	0 (0–4.5)	

Abbreviations: BMI, body mass index; IQR, interquartile range; IIEF-5, International index of erectile function; MLUTS, male lower urinary tract symptoms; N, number; TURP, transurethral resection of the prostate.

MVC and endurance measurements in particular have good reliability. The peak microvoltages have lower ICC or SCC values and a larger spread of the differences in the Bland–Altman plots. As a result, the reliability of the peak microvoltages is worse than the reliability of the average microvoltages. The onset and offset times have low SCC values, with the exception of the endurance onset time and the press measurement offset time. In combination with a relatively large spread of the differences in a large number of the Bland–Altman plots, the onset and offset variables have a poorer reliability.

Notably, there were generally better results for the intraday measurements than for the interday measurements. The Bland–Altman plots showed no systematic difference between the measurements, whereby it is unlikely that the decrease in reliability in the interday measurements in the current study is caused by fatigue or a learning effect. Moreover, to avoid a possible learning effect, practice contractions were performed before the start of the measurement. Indeed, when a learning effect occurred, contractions were expected to be stronger at M2 and M3; however, in the Bland–Altman plots, resting measurements showed that the median PFM activity was lower at M2 and M3 compared with M1. It is concluded from this that no learning effect occurred between the measurements. Furthermore, it is important to consider that reliability decreases with time, therefore, as noticed in the initial study, measurements with a large time interval in between may not be reliable.

Compared with the initial study, the strengths of the current study were the more rigorous protocol and the shorter interval between the measurements that made the results more dependable. The limitation, however, is the small sample size, thus results should be interpreted with some caution.

The poorer outcomes when using a less strict protocol urges the users of the MAPLe device to be

TABLE 2 Test–retest reliability of the main MAPLe outcomes, according to intraclass correlation and spearman correlation coefficient.

	Intraday (M1–M2)		Interday (M1–M3)	
	ICC (CI 95%)	SCC	ICC (CI 95%)	SCC
Rest average (μV)	0.81 (0.44–0.93)		0.73 (0.37–0.91)	
MVC average (μV)	0.91 (0.81–0.96)		0.77 (0.44–0.92)	
MVC peak (μV)	0.67 (0.38–0.85)		0.78 (0.43–0.92)	
MVC onset time		0.62**		0.41
MVC offset time		0.62**		0.70**
Endurance average (μV)	0.90 (0.79–0.96)		0.87 (0.61–0.96)	
Endurance peak (μV)	0.64 (0.32–0.83)		0.61 (0.12–0.86)	
Endurance onset time		0.70**		0.84**
Endurance offset time		−0.17		0.45
Valsalva average (μV)		0.78**		0.66**
Valsalva peak (μV)		0.62**		0.51
Valsalva onset time		0.58**		0.53
Valsalva offset time		0.76**		0.70**

Abbreviations: CI, 95% confidence interval; ICC, intraclass correlation coefficient; M1, first measurement; M2, second measurement; M3, third measurement; MVC, maximal voluntary contraction; SCC, Spearman's correlation coefficient; μV, microvoltage.

*Significance level $p < 0.05$.

**Significance level $p < 0.01$.

well trained for this assessment and apply comparable conditions with each measure. Otherwise, changes seen in individual patients could reflect measurement errors instead of actual changes in PFM function over time.

5 | CONCLUSION

The test–retest reproducibility of the MAPLe in men with LUTS was good for most, but not all, outcomes. In particular, the average microvoltages of the rest measurement, MVC measurement and endurance measurement are reliable. The reliability of the other variables is questionable. Moreover, this study emphasized the need for a strict protocol in daily practice and research settings; in fact, the results showed that the reliability decreases as the time interval between measurements increases, as well as when a less strict protocol was applied.

ACKNOWLEDGMENTS

All costs were covered by the department of the corresponding author.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author (Marco H. Blanker), upon reasonable request.

ETHICS STATEMENT

The Medical Ethical Committee of the Isala Hospital Zwolle approved the study under number NL64332.075.17. Eligible participants provided written informed consent.

ORCID

Martina Beverini http://orcid.org/0000-0001-5812-0900
Lambertus P. W. Witte https://orcid.org/0000-0002-1567-4847
Gommert A. van Koeveringe http://orcid.org/0000-0003-2328-8265
Grietje E. Knol-de Vries http://orcid.org/0000-0002-7833-5260
Marco H. Blanker http://orcid.org/0000-0002-1086-8730

REFERENCES

1. Chapple CR, Wein AJ, Abrams P, et al. Lower urinary tract symptoms revisited: a broader clinical perspective. *Eur Urol*. 2008;54(3):563-569. doi:10.1016/j.eururo.2008.03.109
2. Gravas S, Cornu JN, Gacci M, et al. EAU guidelines on Management of non-neurogenic male lower urinary tract

symptoms (LUTS), incl. benign prostatic obstruction (BPO). 2022. Accessed on February 1, 2023. https://uroweb.org/guidelines/management-of-non-neurogenic-male-luts

3. Voorham-van der Zalm PJ, Voorham JC, van den Bos TWL, et al. Reliability and differentiation of pelvic floor muscle electromyography measurements in healthy volunteers using a new device: the multiple array probe Leiden (MAPLe): reliability of the multiple array probe. *Neurourol Urodyn*. 2013;32(4):341-348. doi:10.1002/nau.22311
4. Goes S, Notenboom-Nas F, Knol-de Vries G, Witte L, Blanker M. ICS 2017 abstracts. *Neurourol Urodyn*. 2017;36(S3):1. doi:10.1002/nau.23386
5. Vrolijks RO, Notenboom-Nas FJM, Boer D, et al. Exploring pelvic floor muscle activity in men with lower urinary tract symptoms. *Neurourol Urodyn*. 2020;39(2):732-737. doi:10.1002/nau.24267
6. Barry MJ, Fowler FJ, Jr., O'Leary MP, et al. The American urological association symptom index for benign prostatic hyperplasia. *J Urol*. 1992;148(5):1549-1557. discussion 1564 doi:10.1016/s0022-5347(17)36966-5
7. Donovan JL, Peters TJ, Abrams P, Brookes ST, De La Rosette JJMCH, Schäfer W. Scoring the short form ICSmaleSF Questionnaire. *J Urol*. 2000;164(6):1948-1955. doi:10.1016/S0022-5347(05)66926-1
8. Rosen RC, Riley A, Wagner G, Osterloh IH, Kirkpatrick J, Mishra A. The international index of erectile function (IIEF): a multidimensional scale for assessment of erectile dysfunction. *Urology*. 1997;49(6):822-830. doi:10.1016/S0090-4295(97)00238-0
9. Martin Bland J, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*. 1986;327(8476):307-310.
10. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163. doi:10.1016/j.jcm.2016.02.012
11. Schober P, Boer C, Schwarte LA. Correlation coefficients: appropriate use and interpretation. *Anesth Analg*. 2018;126(5):1763-1768. doi:10.1213/ANE.0000000000002864
12. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428. doi:10.1037//0033-2909.86.2.420
13. De Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Measurement in Medicine: A Practical Guide*. Cambridge University Press; 2011doi:10.1017/CBO9780511996214
14. Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med*. 2018;18(3):91-93. doi:10.1016/j.tjem.2018.08.001

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Beverini M, Goes S, Witte LPW, et al. The test–retest reproducibility of the multiple array probe Leiden in men with lower urinary tract symptoms. *Neurourol Urodyn*. 2023; 1-11. doi:10.1002/nau.25157

APPENDIX A

Bland–Altman plots

(see Figure A1)

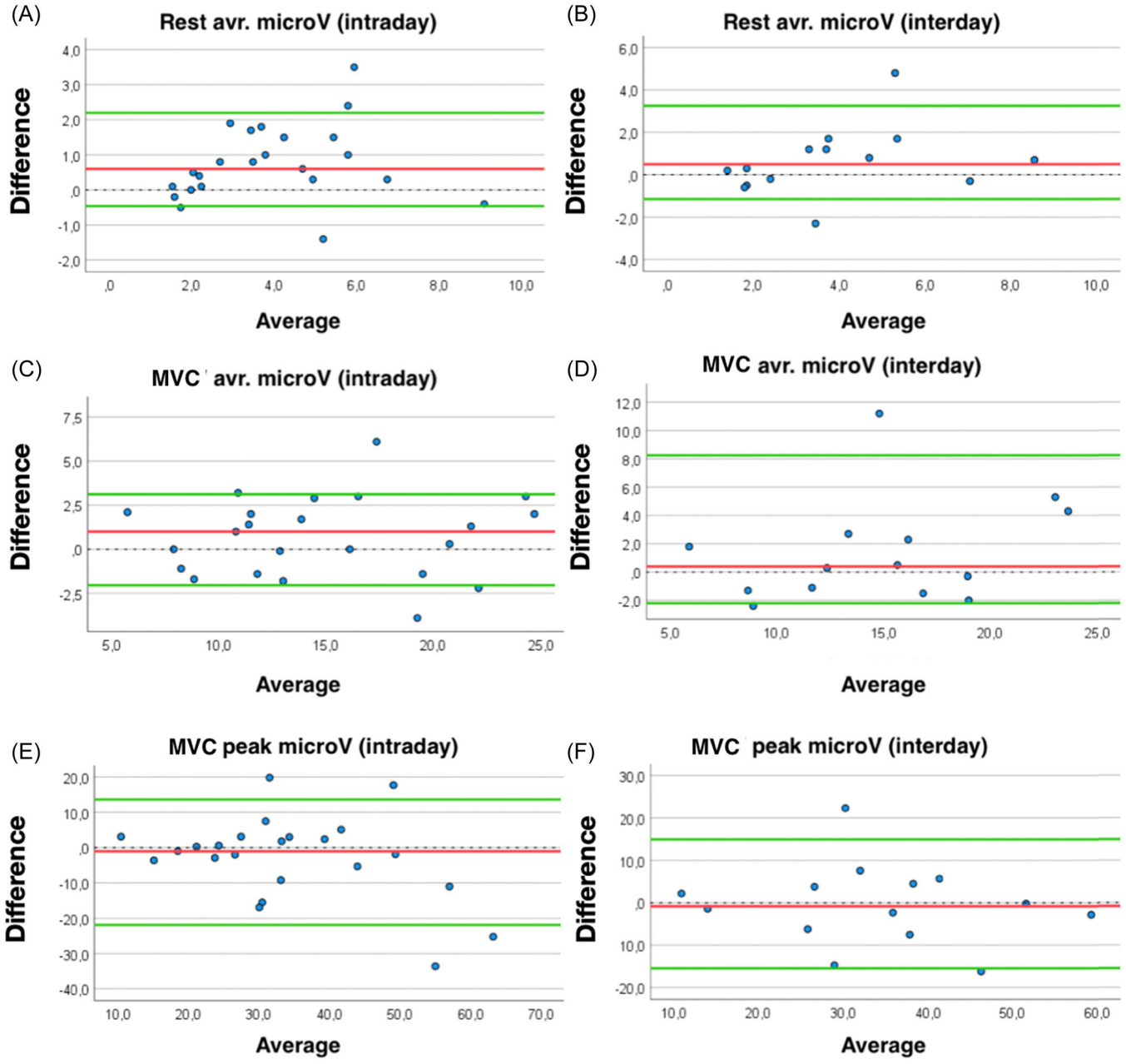

FIGURE A1 A–Z: Bland–Altman plots depicted as follow: dashed line as the zero line, red line as the median difference, and green lines as the 10th and 90th percentiles of the differences. Avr., average; μV, microvoltage; MVC, maximal voluntary contraction. Onset and offset plots are displayed in seconds, plots of the average and peak microvoltage in μV.

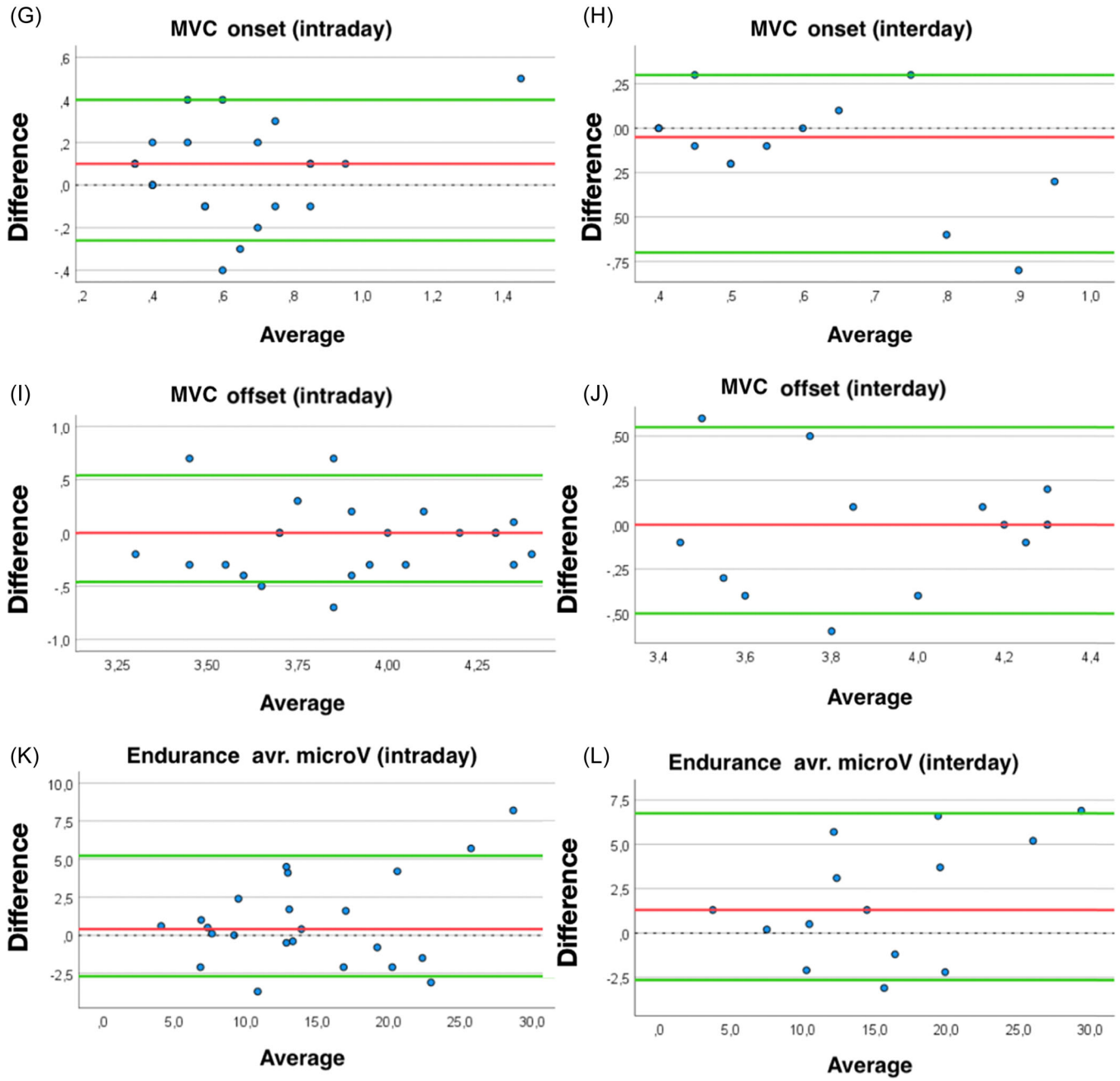

FIGURE A1 Continued

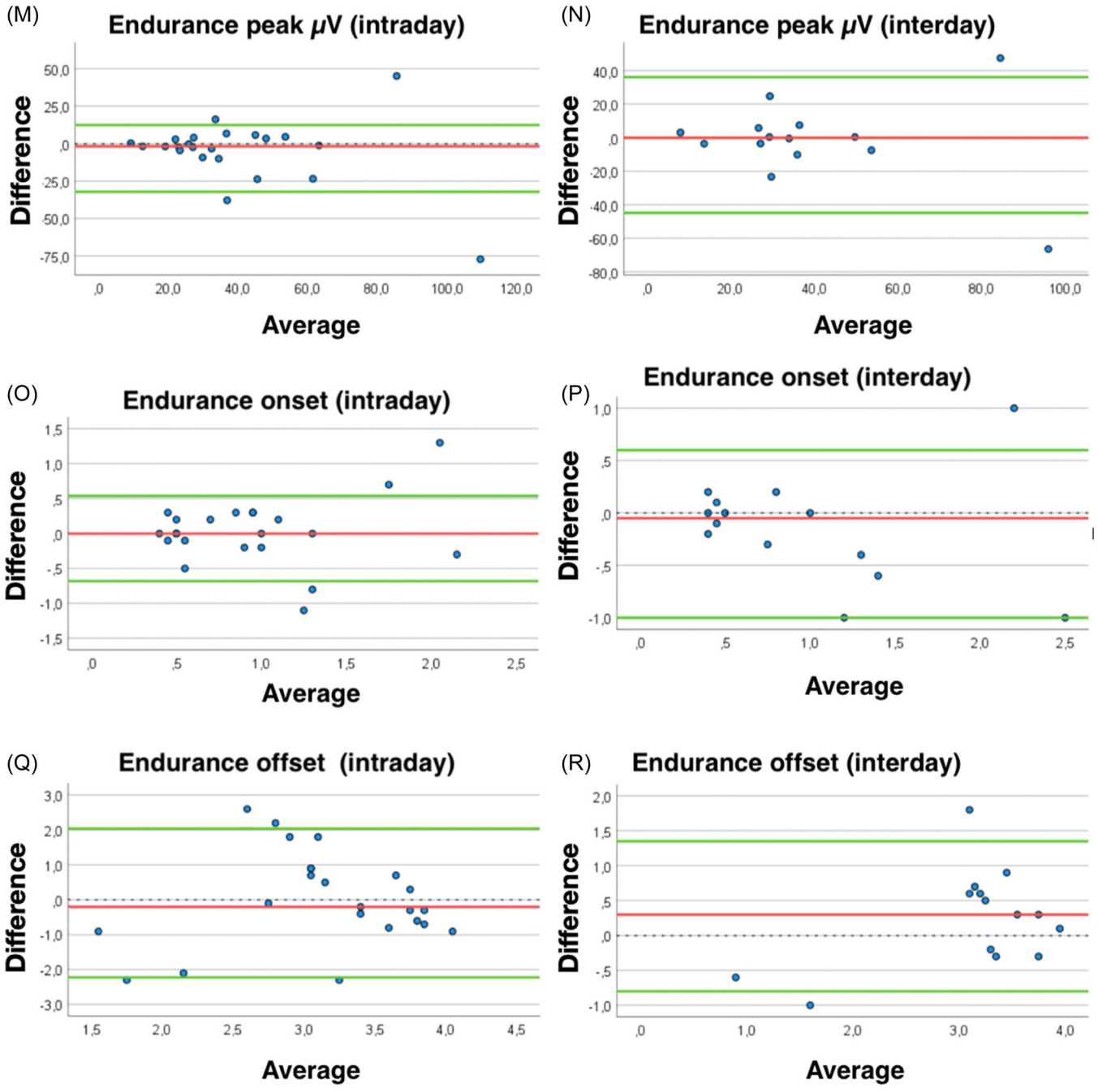

FIGURE A1 Continued

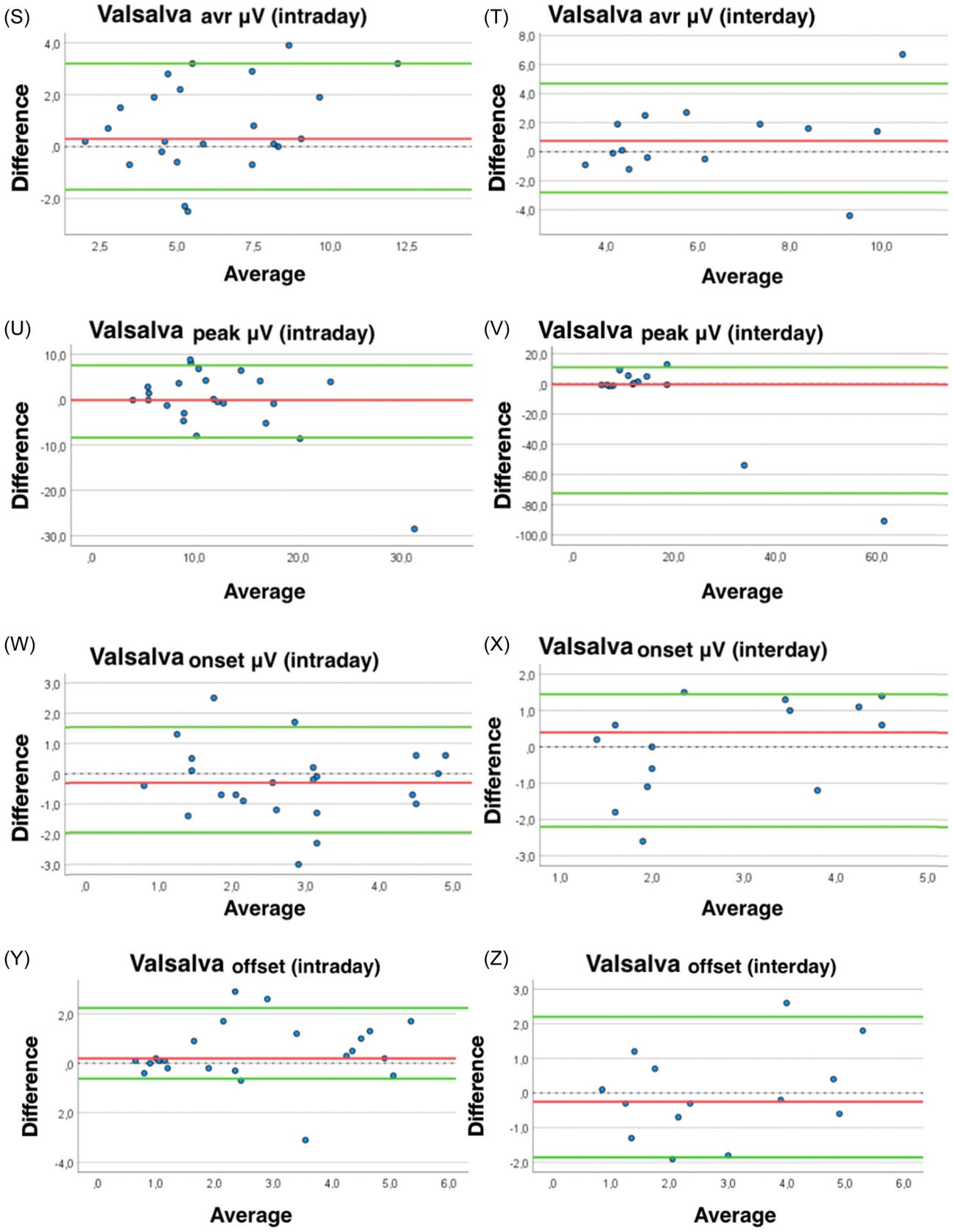

FIGURE A1 Continued