

**Interrater-Reliabilität der Uroflowmetrie bei neurogener
Dysfunktion des unteren Harntraktes durch
Multiple Sklerose**

Dissertation

zur Erlangung des Doktorgrades (Dr. med.)

der Medizinischen Fakultät

der Rheinischen Friedrich-Wilhelms-Universität

Bonn

Julia Rieger

aus Bonn

2026

Angefertigt mit der Genehmigung
der Medizinischen Fakultät der Universität Bonn

1. Gutachterin: Prof. Dr. med. Ruth Kirschner-Hermanns
2. Gutachter: Prof. Dr. Ingo Schmidt-Wolf

Tag der mündlichen Prüfung: 10.02.2026

Aus der Klinik und Poliklinik für Urologie und Kinderurologie

Inhaltsverzeichnis

Abkürzungsverzeichnis	6
1. Einleitung.....	7
1.1 Multiple Sklerose.....	7
1.1.1 Grundlagen	7
1.1.2 NLUTD und deren Folgen bei MS.....	8
1.1.3 Therapiemöglichkeiten neurogener Dysfunktion des unteren Harntraktes..	9
1.2 Diagnostik neurogener Dysfunktion des unteren Harntraktes	13
1.2.1 Anamnese.....	13
1.2.2 Miktionstagebuch	14
1.2.3 Grundlagen der Uroflowmetrie	16
1.2.4 Grundlagen der Urodynamik.....	18
1.3 Interrater-Reliabilität.....	19
1.4 Ziel der Arbeit.....	20
2. Material und Methoden.....	22
2.1 Studiendesign	22
2.1.1 Charakteristik.....	22
2.1.2 Ablauf.....	22
2.1.3 Endpunkte.....	23
2.1.4 Ein- und Ausschlusskriterien.....	24
2.1.5 Untersuchung der Uroflowmetrie	24
2.2 Statistische Analyse.....	25
2.3 Teilnehmende Zentren	26
2.4 Ethische Überlegungen und Einwilligung	26

3.	Ergebnisse.....	27
3.1	Beschreibung der Stichprobe.....	27
3.2	Interrater-Reliabilität der Verdachtsdiagnose nach Anamnese und MTB..	28
3.2.1	Übereinstimmung der Einzeldiagnosen	28
3.2.2	Übereinstimmung aller Diagnosen	29
3.3	Interrater-Reliabilität der Verdachtsdiagnosen nach Anamnese, MTB und UF	30
3.3.1	Übereinstimmung der Einzeldiagnosen	30
3.3.2	Übereinstimmung der Diagnose-Kombinationen.....	31
3.4	Interrater-Reliabilität der Therapievorschlage nach UF.....	33
3.4.1	Übereinstimmung der Einzeltherapievorschlage.....	33
3.4.2	Übereinstimmung der Therapie-Kombinationen.....	34
3.5	Vergleich Diagnosen verblindete Rater mit Originalbefund.....	36
3.5.1	Übereinstimmung der Einzeldiagnosen	36
3.5.2	Übereinstimmung der Diagnose-Kombinationen.....	36
4.	Diskussion	37
4.1	Hauptergebnisse	37
4.2	Mögliche Erklärung niedriger Interrater-Reliabilität	37
4.2.1	Variabilität durch Anzahl der Diagnosen und Therapien	37
4.2.2	Mehrfachantworten	39
4.2.3	Ähnliche Kurvenformen.....	39
4.2.4	Einflussfaktor Mensch	40
4.3	Ergebnisse im wissenschaftlichen Kontext	42
4.4	Interrater-Reliabilität in der Medizin	43
4.5	Grenzen dieser Arbeit	45

4.6	Ausblick.....	47
4.6.1	Implikationen für die Praxis	47
4.6.2	Implikationen für weitere Untersuchungen	48
4.6.3	Intrarater-Reliabilität.....	48
5.	Zusammenfassung	50
6.	Abbildungsverzeichnis	52
7.	Tabellenverzeichnis.....	53
8.	Literaturverzeichnis	54
9.	Erklärung zum Eigenanteil.....	63
10.	Veröffentlichung.....	64
11.	Anmerkung	65
12.	Danksagung	66

Abkürzungsverzeichnis

BF/P	Biofeedback/Physiotherapie
BFS	Blasenfunktionsstörung
BOO	Subvesikale Obstruktion
BTX	Botulinumtoxin A
DO	Detrusorüberaktivität
DSD	Detrusor-Sphinkter-Dyssynergie
DU	Detrusorunteraktivität
EDSS	Expanded disability status scale
HWI	Harnwegsinfekt
IRR	Interrater-Reliabilität
ISK	Intermittierender Selbstkatheterismus
LUTS	Lower urinary tract symptoms
MS	Multiple Sklerose
MTB	Miktionsstagebuch
NLUTD	Neurogenic lower urinary tract dysfunction
PmMS	Personen mit Multipler Sklerose
UAW	Unerwünschte Arzneimittelwirkung
UD	Urodynamik
UF	Uroflowmetrie

1. Einleitung

1.1 Multiple Sklerose

1.1.1 Grundlagen

Die Multiple Sklerose (MS) ist eine Erkrankung des zentralen Nervensystems. Sie zeichnet sich durch Entzündung, Gliose und Demyelinisierung aus und stellt eines der häufigsten neurologischen Krankheitsbilder in Deutschland dar (Mattle, 2013).

Weltweit liegt die Prävalenz bei etwa 2,5 Millionen Fällen, wobei Frauen im Durchschnitt doppelt so häufig betroffen sind wie Männer (Weih et al., 2020).

Meist tritt MS im jungen Erwachsenenalter auf (Sakakibara, 2019).

Während die Inzidenz in den letzten Jahren stieg, konnte die Schwere der Krankheitsverläufe durch neue Therapiemöglichkeiten gemildert werden (Koch-Henriksen und Magyari, 2021).

Die Pathogenese ist noch nicht vollständig geklärt. Diskutiert werden neben immunologischen Vorgängen auch genetische, umweltbedingte und autoimmune Faktoren (Mattle, 2013).

Die Diagnose einer MS kann gestellt werden, wenn die neurologischen Symptome objektivierbar sind, mindestens 24 Stunden andauern und die weiße Hirnsubstanz betreffen. Außerdem müssen sie zeitlich und örtlich disseminiert auftreten (Mattle, 2013).

Die Krankheit kann schubförmig, primär progredient oder sekundär progredient verlaufen (Mattle, 2013).

Die Symptome hängen vom Ort der Entzündung ab. Häufig treten sensorische Beeinträchtigungen, Muskelschwäche, Hirnstammdysfunktion und Sehstörungen auf. Die Krankheit verläuft mit dem Alter progressiv und führt zu Einschränkungen der Koordination, des Gangbilds sowie der Blasen-, Mastdarm- und Sexualfunktion (Zurawski und Stankiewicz, 2018).

Der akute Schub wird mit hoch dosierten Glukokortikoiden wie Methylprednisolon behandelt. Zur Vorbeugung von Schüben sowie zur Reduktion der Krankheitsprogression können immunmodulatorische und immunsuppressive Substanzen wie zum Beispiel

Beta-Interferone verwendet werden (Hemmer B. et al., 2022). Außerdem ist eine symptomatische Therapie sinnvoll, um die fortschreitende Beeinträchtigung im Alltag zu verlangsamen und die Lebensqualität zu erhalten (Henze et al., 2018).

MS ist zurzeit nicht heilbar (Weih et al., 2020).

1.1.2 NLUTD und deren Folgen bei MS

20 % der Personen mit Multipler Sklerose (PmMS) leiden bereits während ihrer Erstdiagnose unter Blasenfunktionsstörungen (BFS). Bei bis zu 75 % treten Miktionsstörungen im Krankheitsverlauf auf (Mattle, 2013).

Ein Großteil der PmMS leidet unter Detrusorüberaktivität (DO, 65 %). Häufig sind außerdem Detrusorunteraktivität (DU, 25 %) und Detrusor-Sphinkter-Dyssynergie (DSD, 35 %) sowie subvesikale Obstruktion (BOO) (Giannantoni et al., 1999).

Die genannten Pathologien können zu imperativem Harndrang, vermehrtem und nächtlichen Wasserlassen, Inkontinenz und verschlechterter Blasenentleerung führen (Aharony et al., 2017).

Die Störungen können durch Läsionen in den auf- und absteigenden Bahnen des Rückenmarks erklärt werden (Sakakibara, 2019). Eine Schädigung unterhalb des sakralen Miktionszentrums resultiert in einer schlaffen Lähmung (hypoaktiver Sphinkter oder Detrusor), während eine Läsion oberhalb meist zu einer spastischen Lähmung führt (hyperaktiver Sphinkter oder Detrusor). Durch die verschiedene Lokalisation und Aktivität der Läsionen bei MS können Mischformen auftreten (Haensch et al., 2016).

Ein erhöhter intravesikaler Druck, welcher durch DO und DSD hervorgerufen wird, kann Veränderungen in der Blasenwand herbeiführen. Hieraus ergibt sich für die Personen ein erhöhtes Risiko für vesikoureterorenenalen Reflux, Hydronephrose, Nierenfunktionsstörungen und terminales Nierenversagen. Personen mit langsam progressiven neurologischen Erkrankungen wie der MS haben dafür jedoch ein niedrigeres Risiko als Patienten mit Rückenmarksverletzungen oder Spina bifida (Panicker et al., 2015).

Darüber hinaus besteht bei PmMS erhöhte Gefahr für Harnwegsinfektionen (HWI), welche mit Pyelonephritiden (Foxman, 2003) oder mit Urolithiasis assoziiert sein können (Rached et al., 2011).

Inkontinenz und weitere Symptome der neurogenen unteren Harntrakt-Schädigung stellen eine große Einschränkung der Lebensqualität für die Betroffenen dar (Sakakibara, 2019). Eine verminderte Blasensensitivität, eine geringere Blasenkapazität (Giannantoni et al., 1999) sowie Restharn sind beschrieben (Sakakibara, 2019).

Längere MS-Verläufe können zu fortschreitenden sekundären Schäden des Harntraktes führen. Eine Vermeidung dieser wäre durch eine frühe Intervention möglich (Jaekel et al., 2022). Es ist von Bedeutung, dass die Therapie gut auf die spezifische Symptomatik angepasst ist, da mit jedem gescheiterten Therapieversuch die Chance auf einen erfolgreichen sinkt (Jaekel et al., 2022).

Insgesamt verursachen neurogene Funktionsstörungen der unteren Harnwege (NLUTD) hohe Kosten im Gesundheitswesen (Zurawski und Stankiewicz, 2018).

1.1.3 Therapiemöglichkeiten neurogener Dysfunktion des unteren Harntraktes

Die Behandlungsmöglichkeiten der neurogenen Funktionsstörung des unteren Harntraktes sind vielfältig (Blok et al., 2023). Eine individualisierte, auf den Patienten zugeschnittene Therapie ist wichtig (Panicker et al., 2015).

Wichtigstes Ziel der Behandlung ist der Erhalt der Nierenfunktion durch Wiederherstellung physiologischer intravesikaler Druckverhältnisse (Gust et al., 2014). Darüber hinaus ist der Erhalt der Speicherfunktion der Harnblase und damit Verbesserung von Kontinenz, sozialer Teilhabe und Steigerung der Lebensqualität entscheidend. Weiterhin steht die restharnfreie und druckarme Blasenentleerung, der Schutz vor HWI sowie Schutz vor weiteren gesundheitlichen Folgeschäden und Krankenhausaufenthalten im Fokus der Therapie (Kutzenberger et al., 2022; Musco et al., 2022).

Neben konservativen, nicht medikamentösen Therapien stehen eine Reihe von pharmakologischen, interventionellen sowie chirurgischen Therapien zur Verfügung (Haensch et al., 2016).

1.1.3.1 Konservative Therapie

Durch die gezielte Analyse und Anpassung des Trink- und Miktionsverhaltens kann bereits eine Linderung der Symptomatik erreicht werden. Ergänzend dazu können verschiedene Hilfsmittel wie Vorlagen, Pessare, Tropfenfänger oder Klebesysteme zum Einsatz kommen. Eine ausführliche Beratung und Schulung der PmMS ist dabei essenziell (Kurze und Jaekel, 2024; Hemmer B. et al., 2023).

Zusätzlich kommen Beckenbodentraining, gegebenenfalls mit Elektrostimulation, sowie EMG-Biofeedback (TENS-Geräte) in Frage (Lúcio et al., 2016). So kann die Aktivität des Detrusormuskels vermindert werden, was sowohl zur Verbesserung der Kontinenz als auch zur Erhöhung der Blasenkapazität beiträgt (Kurze und Jaekel, 2024).

Blasenfunktion und Lebensqualität können auch durch eine strukturierte urologische Rehabilitation mit verschiedenen Diagnostik- und Trainingseinheiten verbessert werden (Khan et al., 2010). Verhaltenstherapeutische Ansätze, wie beispielsweise die „Miktion nach der Uhr“, helfen durch die Entwicklung eines Toilettenplans dabei, unfreiwillige Miktion zu vermeiden (Kurze und Jaekel, 2024).

1.1.3.2 Pharmakotherapie

Bei einer überaktiven Blase kommen in erster Linie Anticholinergika zum Einsatz (Blok et al., 2023). Sie sind Inhibitoren der muskarinischen Rezeptoren und reduzieren so die Detrusor-Hyperaktivität. Da diese Rezeptoren auch in anderen Organen vorkommen, verursacht das Medikament periphere und zentrale unerwünschte Arzneimittelwirkungen (UAW) (Kutzenberger et al., 2022). Besonders bei älteren Patienten kann es zu zentralnervösen Nebenwirkungen wie Konzentrationsschwäche, Schläfrigkeit und sogar Halluzinationen kommen (Madersbacher, 2006). Obwohl neuere Anticholinergika weniger UAW zeigen, stellt beispielsweise Mundtrockenheit bei fast einem Drittel der Patienten noch ein Problem dar (Nabi et al., 2006).

Bei Unverträglichkeiten oder Unwirksamkeit von Anticholinergika können an deren Stelle β 3-Sympathomimetika wie Mirabegron stattdessen oder zusätzlich zum Einsatz kommen. Sie sind jedoch nicht für die Therapie der NLUTD bei MS zugelassen (off-label-use) (Kutzenberger et al., 2022).

Darüber hinaus werden Alpha-Blocker bei Personen mit NLUTD eingesetzt, jedoch zeigen sie nicht bei allen Betroffenen Effektivität. Sie sollen durch ihre sympatholytische Wirkung zu einer Erschlaffung von Harnblasenhals und Urethra führen (Schneider et al., 2019) und senken damit den subvesikalen Auslasswiderstand (Blok et al., 2023).

1.1.3.3 Invasive Therapie

Eine weitere Möglichkeit zur Vermeidung von UAW der oralen Anticholinergika besteht in der intravesikalen Gabe des Anticholinergikums Oxybutynin. Bei dieser Art der Applikation wird das Medikament unter Umgehung der Leberpassage verstoffwechselt. Dadurch entstehen weniger Metabolite, die zu UAW führen und ein größerer Teil des Wirkstoffes kann direkt an der Blase wirken (Blok et al., 2023).

Seit 2011 ist Botulinumtoxin A (BTX) zur Applikation in den Musculus detrusor vesicae zugelassen (Cooley und Kielb, 2019). Studien zeigen, dass es sich dabei um eine sichere Therapieform handelt, welche den Blasendruck und das Miktionsvolumen verbessert, das Vorkommen von HWIs reduziert und die Lebensqualität steigert (Cooley und Kielb, 2019). Das Medikament wird im Rahmen einer Zystoskopie in verschiedene Stellen des Detrusor-Muskels injiziert (Scheiner et al., 2006). Bei neurogenen Formen der BFS werden 200-300 Einheiten empfohlen (Cruz et al., 2011).

Zur Therapie der Blasenentleerungsstörung, welche als Folge der Erkrankung oder auch als Folge der Therapiemaßnahmen (Botulinumtoxin A, Blasenaugmentation) entstehen kann (Kurze und Jaekel, 2024), steht der Katheterismus in verschiedenen Formen zur Verfügung (Blok et al., 2023).

Eine Variante besteht in der intermittierenden Selbstkatheterisierung (ISK) mit dem Ziel der aseptischen Blasenentleerung in regelmäßigen Abständen. Eine Überdehnung der Blase (Gust et al., 2014), die Schädigung des oberen Harntraktes sowie Inkontinenz können so verhindert werden. Ein weiterer Vorteil besteht darin, dass der Patient keinen

Dauerkatheter benötigt und damit Infekte und andere Komplikationen vermindert werden können (Klingler HC et al.).

Voraussetzung sind jedoch die notwendige motorische Hand- und kognitive Globalfunktion (Jacq et al., 2021). Es entstehen hohe Kosten (Lamin und Newman, 2016) und die Suche nach einem Ort, der den ISK ermöglicht, kann für den Patienten unter Umständen aufwendig sein.

Bei Patienten, bei denen der ISK aus kognitiven, physischen oder psychischen Gründen nicht umsetzbar ist, kann eine Dauer-Katheterisierung erfolgen. Diese ist auf urethralem oder suprapubischem Weg möglich, wobei die suprapubische der transurethralen Form vorgezogen werden sollte (Hautmann, 2014). Diese Art der Blasenentleerung ist mit Komplikationen verbunden (Blok et al., 2023). Dazu zählen ein erhöhtes Risiko für HWIs, die Verletzung der Harnröhre, die Verstopfung des Katheters, welche wiederum zu einer schmerzhaften Dehnung der Blase, Urosepsis oder Nierenversagen führen kann, sowie die Irritation der Blasenschleimhaut (Jeffery und Mundy, 2020).

Darüber hinaus werden bei Dauerkatheterträgern vermehrt Plattenepithelkarzinome beobachtet (Hautmann, 2014).

Die sakrale Neuromodulation stellt eine minimalinvasive Therapiemöglichkeit für die neurogene Dysfunktion des unteren Harntraktes dar. Von einem im oberen Gesäßbereich implantierten Schrittmacher werden dabei elektrische Pulse abgegeben, welche die Nerven des Plexus sacralis, insbesondere die S3-Nervenwurzel über eine dauerhaft implantierte Elektrode, stimulieren (van Ophoven, 2018).

Der genaue Wirkmechanismus ist zurzeit noch nicht geklärt. Es wird angenommen, dass eine vornehmlich afferente Stimulation der Sakralnerven spinale Reflexe auslöst und neuronale Netzwerke im Gehirn neu moduliert werden. Detrusor und Sphinkter werden dabei vermutlich nicht direkt stimuliert (van Ophoven, 2018).

Das Implantat führt zu keinerlei Einschränkung des Patienten in seiner Aktivität und stellt besonders für therapierefraktäre neurogene BFS eine vielversprechende Behandlungsoption dar (Kessler et al., 2012).

1.1.3.4 Chirurgische Therapie

Sollten die oben genannten Therapien versagen (Haensch et al., 2016), zum Beispiel wenn weiterhin ein erhöhter intravesikulärer Druck besteht, oder nicht ausreichen im Sinne einer weiterhin bestehenden Gefahr für die Nierenfunktion (Verpoorten und Buyse, 2008), kann eine chirurgische Rekonstruktion mittels Blasenaugmentation in Betracht gezogen werden (Haensch et al., 2016).

Es können verschiedene Darmabschnitte verwendet werden. Dabei wird meist ein Teil des Ileums gewählt. Das Verfahren führt zu langanhaltend guten Ergebnissen und Patientenzufriedenheit, geht aber auch mit gewissen Risiken wie einer gehäuften Malignom Entstehung oder Elektrolytverschiebungen einher (Çetinel et al., 2016).

Das Verfahren kommt hauptsächlich bei therapierefraktärer DO, low-compliance Blase und DSD zum Einsatz (Haensch et al., 2016).

Bei Patienten mit insuffizientem Sphinkter kann Kontinenz darüber hinaus durch eine Operation zur Steigerung des Blasenaustrittswiderstandes, zum Beispiel durch einen suburethral gelegten Faszienzügel oder ein artifizielles Sphinktersystem, erreicht werden (Verpoorten und Buyse, 2008).

Bei DSD besteht neben dem Katheterismus die Möglichkeit, eine komplette Sphinkterotomie mit der Folge der Harninkontinenz durchzuführen. Bei Kontraindikationen der kontinenzerhaltenden rekonstruktiven Maßnahmen muss zum Schutz des oberen Harntraktes außerdem die Ileum-Conduit Anlage in Betracht gezogen werden, wobei der Harn über ein inkontinentes Stoma abgeleitet wird (Haensch et al., 2016).

1.2 Diagnostik neurogener Dysfunktion des unteren Harntraktes

1.2.1 Anamnese

Die Anamnese stellt den ersten diagnostischen Schritt dar und sollte sich nach den aktuellen Beschwerden des Patienten der genaueren Blasenfunktions-Anamnese zuwenden. Dabei werden die tägliche und nächtliche Miktionsfrequenz sowie etwaige

Schmerzen und andere Missempfindungen während des Wasserlassens erfragt (Palmtag et al., 2004).

Um diese Aussagen präzisieren zu können, beispielsweise durch genauere Angaben zur Trinkmenge, kann im nächsten Schritt die Führung eines Miktionstagebuchs (MTB) veranlasst werden (siehe Abbildung 1 und Kapitel 1.2.2) (Primus und Heidler, 2003).

Vegetative Symptome wie Schwitzen oder Schwindel können Hinweise auf eine neurologische Funktionsstörung geben. Des Weiteren sollte nach Restharngefühl, verzögertem Miktionsgefühl, ungewollten Miktionsunterbrechungen und einer möglichen Selbsterklärung des Patienten hierfür gefragt werden. Zusätzlich ist wichtig, dass der Patient über potenziellen ungewollten Urinabgang, dessen situativen Kontext sowie über imperativen Harndrang berichtet (Palmtag et al., 2004).

Neben einer gynäkologischen Anamnese mit Blick auf Schwangerschaften, Geburten und deren Komplikationen spielen auch urologische Vorerkrankungen eine wichtige Rolle (Palmtag et al., 2004).

Da HWI oft BFS verschlechtern und wiederum auch Folge einer BFS sein können, sollte zusätzlich immer nach bestehenden oder abgelaufenen HWI gefragt und der Urin untersucht werden (Primus und Heidler, 2003).

Eingenommene Begleitmedikation und Komorbiditäten können im Zusammenhang mit der Blasenfunktionsproblematik stehen und sollten deshalb erfasst werden. Darüber hinaus sind die Stuhlanamnese und die Frage nach früheren operativen Eingriffen in der Beckenregion wichtig (Primus G und Heidler H et al., 2003).

Der Anamnese schließen sich eine neurologische und urologische Untersuchung an (Haensch et al., 2020).

1.2.2 Miktionstagebuch

Das MTB ist ein vom Patienten oder einer betreuenden Person geführtes Protokoll, welches der semiquantitativen Analyse von neurogenen BFS dient (Kirschner-Hermanns, 2016). Darin werden über 2-3 Tage (Konstantinidis et al., 2016) die Miktionsvolumina und Miktionszeitpunkte, die Trinkmenge mit Uhrzeit, der Schweregrad des Harndrangs,

1.2.3 Grundlagen der Uroflowmetrie

Die Uroflowmetrie (UF) wird zur initialen objektiven Beurteilung bei Personen mit lower urinary tract symptoms (LUTS) empfohlen. Sie ist das am häufigsten verwendete Verfahren der Urodynamik (UD) (Gammie et al., 2018).

Dabei wird die Menge Urin, die pro Sekunde ausgeschieden wird, gemessen (=Harnflussrate) (Gust et al., 2014).

Es existieren verschiedene Verfahren:

Beim Überlaufsystem läuft der Urin durch einen Trichter in ein Rohr, welches unten eine sehr kleine Öffnung hat. Je höher der Harnfluss ist, desto mehr seitliche Öffnungen im Rohr können den Urin in ein Gefäß ableiten.

Das gravimetrische Prinzip basiert auf dem Wiegen des Urins und wird heutzutage am häufigsten verwendet. Die Gewichtszunahme pro Sekunde wird dabei ausgewertet (Palmtag et al., 2004). Dieses Verfahren wurde auch in dieser Arbeit verwendet.

Beim rotationsdynamischen Prinzip wird der Harnstrahl auf eine rotierende Scheibe geleitet, welche durch das Auftreffen unterschiedlich stark abgebremst wird. Das Gerät versucht die Bremsung durch höheren Stromfluss auszugleichen. Der Energieverbrauch kann in einer Funktion aus Urinmenge und Zeiteinheit dargestellt werden (Palmtag et al., 2004).

Durch die Harnflusskurve lassen sich Rückschlüsse auf mögliche Funktionsstörungen ziehen (Gust et al., 2014). Des Weiteren unterstützt sie die Entscheidung, wann eine weitere, genauere und gegebenenfalls invasivere Diagnostikmethode notwendig wird (van Batavia et al., 2023).

Für eine suffiziente Beurteilung der Untersuchung sind mindestens 150ml Miktionsvolumen nötig (Margreiter und Shariat, 2016).

Beim gesunden Patienten ist die Kurve annähernd glockenförmig und der höchste Punkt ist leicht nach links verschoben (Palmtag et al., 2004). Der Gipfel stellt dabei den maximalen Harnfluss (Q_{max}) dar, welcher physiologischerweise über 20ml/s liegt. Errechnen lässt sich außerdem der durchschnittliche Harnfluss (Q_{ave}) (Gust et al., 2014).

Die Flussanstiegszeit beschreibt die Steigung der Kurve bis zum Maximum. Während die Flusszeit nur die Zeit beschreibt, in der Urin fließt, bezeichnet die Miktionszeit die gesamte Dauer und kann bei unterbrochener Miktion dementsprechend länger als die addierten Flusszeiten ausfallen (Gust et al., 2014).

Ergänzend sollte der Restharn stets sonographisch bestimmt werden (Gust et al., 2014).

Eine auffällige Flusskurve stellt die undulierende Flusskurve dar. Sie ist langgestreckt, hat mehrere Gipfel und einen verringerten Spitzenflusswert. Sie entsteht bei BOO oder Detrusorhypokontraktilität.

Bei der intermittierenden Flusskurve sind Flusspausen zu erkennen; die einzelnen Kurven werden immer kleiner. Sie lässt sich durch zu kurze Detrusor-Kontraktionen erklären (Palmtag et al., 2004).

Ist die Kurve abgeflacht und plateauförmig, weist also eine reduzierte maximale Uroflowrate auf, besteht der Hinweis auf eine Obstruktion im Bereich der Harnröhre (Schmid et al., 2020).

Die Uroflowmetrie zeichnet sich im Gegensatz zur der im folgenden Kapitel erläuterten Urodynamik durch eine einfache und nicht-invasive Durchführung aus (Palmtag et al., 2004) und ist kostengünstig (Gammie et al., 2018).

Einen Nachteil hingegen stellt die limitierte Reproduzierbarkeit dar, da das Ergebnis beispielsweise von psychologischen Faktoren und der Uhrzeit beeinflusst wird. Außerdem kann die Beurteilung nur zusammen mit der Anamnese erfolgen (Karl et al., 1986). Darüber hinaus ist die technische Genauigkeit der Uroflowmeter begrenzt (Palmtag et al., 2004).

Besonders bei neurogenen Blasenstörungen und unklaren Befunden in der UF sollte deshalb im Anschluss eine genauere urodynamische Untersuchung als Druck-Fluss-Studie erfolgen (Karl et al., 1986).

1.2.4 Grundlagen der Urodynamik

Neben einer genauen klinischen Beurteilung und einer nicht invasiven Diagnostik beispielsweise mittels UF ist die anschließende invasive UD in Form einer Video-UD der Goldstandard (Blok et al., 2023).

Mit der UD lassen sich urologische Symptome objektivieren, indem sie unter Messbedingungen reproduziert werden. So können Rückschlüsse auf die zugrundeliegende Funktionsstörung des unteren Harntraktes wie Speicher- und Entleerungsstörungen sowie mögliche assoziierte Erkrankungen gezogen werden (Palmtag et al., 2004).

Es werden Blasenfüllungsvolumen, der jeweilige intravesikale und intraabdominelle Druck sowie deren zeitliche Änderung und elektromyographische Signale gemessen. Im Anschluss lassen sich daraus die pathophysiologisch relevanten Zielparameter berechnen (Palmtag et al., 2004).

Dem Patienten werden eine transurethrale und eine transrektale Drucksonde eingeführt, welche den intravesikalen sowie den abdominellen Druck (P_{ves} und P_{abd}) während der Blasenfüllung und der Miktion messen. Aus deren Differenz kann anschließend der Detrusordruck (P_{det}) berechnet werden (Margreiter und Shariat, 2016).

Wichtig zu beachten ist, dass der bei der UD ermittelte Wert als Überdruck gegenüber dem Umgebungsdruck und nicht absolut gemessen wird. Dadurch lassen sich die Werte auch zwischen verschiedenen Kliniken und Patienten vergleichen (Palmtag et al., 2004).

Durch die Ableitung eines perianalen EMGs können Rückschlüsse auf die Aktivität der Beckenbodenmuskulatur und des Sphinkters gezogen werden (Gust et al., 2014).

Zu Beginn wird die Blase langsam mit einer körperwarmen Kochsalzlösung gefüllt (Gust et al., 2014). Dabei wird der Patient dazu aufgefordert, den ersten Harndrang, starken Harndrang, Schmerzen und jegliche Form der Inkontinenz mitzuteilen (Palmtag et al., 2004).

Physiologischerweise wird der Detrusordruck durch die Compliance der Blase und eine funktionierende neuronale Verschaltung während der Füllung auf einem konstant

niedrigen Niveau gehalten. Bei der Miktion steigt der intravesikale Druck durch die Kontraktion des Detrusors. Beckenboden und externer Sphinkter relaxieren hingegen und die Blase entleert sich (Gust et al., 2014).

Die urodynamische Messung stellt keinen automatischen Untersuchungsvorgang dar und ist somit abhängig vom ausführenden Fachpersonal sowie von der Mitarbeit der untersuchten Person (Palmtag et al., 2004). Darüber hinaus ist sie invasiv, zeitaufwendig und dadurch teuer (Valentini und Nelson, 2006).

Zu den urodynamischen Messverfahren gehören all diejenigen, die mit einer geeigneten Methode die Funktion beziehungsweise Dysfunktion des unteren Harntraktes bewerten (Rosier et al., 2017). Das sind im Einzelnen UF, Zystomanometrie mit Druck-Fluss Studie sowie Beckenboden-EMG und Urethradruckprofil (Schultz-Lampel et al., 2022). Häufig wird jedoch im deutschen Sprachraum unter UD nur die Druck-Fluss-Studie verstanden.

1.3 Interrater-Reliabilität

Es hat sich gezeigt, dass der Rater (in diesem Falle der befundende Mediziner) eine wichtige Quelle für Messfehler darstellt. Um die Ausprägung dieser Fehler in bestimmten Situationen festzustellen, werden seit langer Zeit Studien zur Interrater-Reliabilität (IRR) durchgeführt (Landis und Koch, 1977).

Die IRR gibt an, in welchem Maße die Bewerter derselben Variable denselben Wert zuweisen (McHugh, 2012).

Sie wird in vielen Bereichen zur Qualitätsmessung von Beobachtungen und Beurteilungen verwendet, so auch in der Medizin- und Gesundheitsforschung, beispielsweise bei dem Vergleich von Diagnosen (Zhao et al., 2022).

Die ursprüngliche Methode zur Bewertung der IRR mittels der prozentualen Übereinstimmung wurde 1960 durch Cohen kritisiert, da sie die zufälligen Übereinstimmungen nicht mit einbezieht. Das von ihm eingeführte Cohens Kappa (κ_C) ist hingegen in der Lage, zu berücksichtigen, dass Bewerter auch bei Unsicherheit einige Variablen richtig raten (McHugh, 2012).

Cohens Kappa hat sich mittlerweile als Maß zur Beurteilung von Urteiler-Übereinstimmung etabliert (Wongpakaran et al., 2013).

1.4 Ziel der Arbeit

Personen mit Multipler Sklerose leiden in ihrem Krankheitsverlauf häufig an neurogen bedingten Beschwerden des unteren Harntraktes (Mumenthaler, 2008). Diese werden unter dem Begriff NLUTD zusammengefasst (Panicker et al., 2015).

Da die NLUTD mit belastenden Symptomen und einer verringerten Lebensqualität einhergeht, kommt einer frühzeitigen Diagnostik eine besondere Bedeutung zu. Untersuchungen haben gezeigt, dass eine inakkurate Diagnostik zu einem verspäteten Therapiebeginn und somit gegebenenfalls zu einem Fortschreiten der Krankheit sowie zu vermehrten Komplikationen führen können (Gärtner et al., 2015).

Die UF ist eine bewährte Untersuchungsmethode in der Urologie (Cornu et al., 2023). Sie findet unter anderem bei neurogenen Funktionsstörungen des unteren Harntraktes zur Primärdiagnostik und zur Verlaufskontrolle Anwendung und wird dabei vor der Durchführung weiterer Untersuchungen bei NLUTD empfohlen (Blok et al., 2023). Manche Autoren ziehen sogar die Verwendung von UF anstelle der UD in Betracht (El Helou et al., 2021).

Bei der UF wird der Harnstrahl als Zeit-Fluss-Messung aufgezeichnet. Für verschiedene Störungen der Blasenentleerung ergeben sich hierbei typische Kurvenformen (Gust et al., 2014).

Zur Erzielung eines optimalen klinischen Nutzens bedarf die UF einer präzisen klinischen Interpretation durch geschulte Mediziner (van Batavia et al., 2023). Die inhärente Subjektivität der Untersuchung hat jedoch starken Einfluss auf die Vergleichbarkeit zwischen den einzelnen Ratern (van Batavia et al., 2023).

Während Arbeiten zur standardisierten Durchführung der Harnstrahlmessung bereits existieren (Gammie und Drake, 2018) und Parameter der Untersuchung, wie beispielsweise der Einfluss der Miktionsposition und des Blasenfüllungsvolumens,

untersucht wurden (Jong et al., 2014), sind Studien zur Rater-Übereinstimmung der UF selten. Sie beschränken sich meist auf die Zuordnung der UF zu einer bestimmten Kurvenform und nicht zu einer Diagnose. Studien zur IRR bei PmMS besonders hinsichtlich einer Verdachtsdiagnose und Therapieempfehlung existieren nicht.

Die Urodynamik, welche zurzeit den Goldstandard der Diagnostik der NLUTD darstellt (Blok et al., 2023), ist nicht nur invasiv, sondern auch aufwendig und kostenintensiv und demzufolge eingeschränkt verfügbar (Valentini und Nelson, 2006).

Aus diesem Grund wurde in dieser Arbeit der Stellenwert der UF bei der Diagnose- und Therapiefindung genauer untersucht. Es wurde dazu prospektiv analysiert, ob sich verschiedene Rater auf Basis der Kenntnis von Anamnese, MTB und UF für die gleiche Diagnose, beziehungsweise für die gleiche Therapie entscheiden (siehe auch Jaekel et al., 2024).

Zusätzlich sollen die zum Zwecke dieser Arbeit neu erhobenen Befunde mit den Originalbefunden der genutzten Datenbank verglichen und mögliche Unterschiede analysiert werden.

2. Material und Methoden

2.1 Studiendesign

2.1.1 Charakteristik

Es handelt sich um eine prospektive, nicht randomisierte Untersuchung auf Basis einer bereits erhobenen Datenbank. Es liegt eine Blindung der studienbezogenen Befundung von Anamnese, MTB und UF vor.

2.1.2 Ablauf

Die Daten wurden während des stationären Aufenthaltes der Patienten am neurologischen Rehabilitationszentrum Godeshöhe erhoben und in der Datenbank dokumentiert, bevor sie zum Zwecke dieser Arbeit neu betrachtet wurden. Die Datenbank ist im Deutschen Register Klinischer Studien unter der Nummer DRKS00024744 registriert. Alle in der Datenbank enthaltenen Personen haben einer Auswertung ihrer Daten schriftlich zugestimmt (siehe auch Abschnitt 2.4, Ethische Überlegungen und Einwilligung).

Vier studienbezogene Rater mit langjähriger Erfahrung aus vier neuro-urologischen Zentren erhielten anonymisiert folgende Unterlagen von allen eingeschlossenen Patienten der Datenbank:

- Allgemeine und neurourologische Anamnese (Ausschnitt neuro-urologischer Befund Godeshöhe)
- Miktionstagebuch (Ausschnitt neuro-urologischer Befund Godeshöhe, siehe Abb. 1)
- Uroflowmetriekurve mit zugehörigen Messwerten als Papiaausdruck
- Auswertungsbogen für Verdachtsdiagnose nach Kenntnis von Anamnese und MTB sowie zusätzlicher Kenntnis der UF und für empfohlene Therapie (dargestellt in Abb. 2)

Auswertungsbogen

Patient Nr.:

Befunder: A B C D

Verdachtsdiagnose nach Anamnese und MTB:

Normalbefund	<input type="checkbox"/>				
Detrusorüberaktivität	<input type="checkbox"/>				} <input type="text" value="Einfachantwort"/>
Detrusorunteraktivität	<input type="checkbox"/>				

Verdachtsdiagnose nach Uroflow:

Normalbefund	<input type="checkbox"/>				
Detrusorüberaktivität	<input type="checkbox"/>				} <input type="text" value="Einfachantwort"/>
Detrusorunteraktivität	<input type="checkbox"/>				
Detrusor-Sphinkter-Dyssynergie	<input type="checkbox"/>				<input type="text" value="+/-"/>
Subvesikale Obstruktion	<input type="checkbox"/>				} <input type="text" value="Einfachantwort"/>

Empfohlene Therapie:

Keine	<input type="checkbox"/>				
Katheter	<input type="checkbox"/>				} <input type="text" value="Einfachantwort"/>
alpha-Blocker	<input type="checkbox"/>				
Detrusordämpfende Medikation	<input type="checkbox"/>				} <input type="text" value="Mehrfachauswahl"/>
Botulinumtoxin A	<input type="checkbox"/>				
Sakrale Neuromodulation	<input type="checkbox"/>				
Biofeedback/Physiotherapie	<input type="checkbox"/>				

Abbildung 2: Auswertungsbogen zur standardisierten Erfassung der Verdachtsdiagnosen und Therapievorschläge der vier Rater.

2.1.3 Endpunkte

Im Einzelnen wurden die folgenden primären Endpunkte betrachtet:

- Die Interrater-Reliabilität der Verdachtsdiagnosen nach der Befundung mittels Anamnese und MTB
- Die Interrater-Reliabilität der Verdachtsdiagnosen nach der Befundung mittels Anamnese, MTB und UF

- Die Interrater-Reliabilität der Therapievorschlage, die sich aus der Befundung mittels Anamnese, MTB und Uroflowmetrie ergeben

Als sekundarer Endpunkt wurde untersucht:

- Der Vergleich der Diagnosen der verblindeten Rater mit dem Originalbefund aus der Datenbank

2.1.4 Ein- und Ausschlusskriterien

Die eingeschlossenen Personen waren mindestens 18 Jahre alt und besaen die gesicherte Diagnose einer MS. Es wurden nur Patienten eingeschlossen, die eine vollstandige Diagnostik bestehend aus Anamnese, MTB und UF (Miktionsvolumen mindestens 150ml) erhalten haben. Es lag fur alle eingeschlossenen Personen eine schriftliche Zustimmung zur Auswertung der Daten vor und es bestand zum Zeitpunkt der Untersuchungen keine floride oder unbehandelte Harnwegsinfektion.

Nicht eingeschlossen wurden Patienten mit vorangegangener Harnwegschirurgie, wie zum Beispiel Harnblasenaugmentation, Zystostomie, radikaler Prostatektomie oder transurethraler Resektion der Prostata. Auerdem wurden Patienten ohne schriftliche Zustimmung zur Auswertung der Daten ausgeschlossen.

2.1.5 Untersuchung der Uroflowmetrie

Die Indikation zur UF wurde nach einer Erstuntersuchung durch einen Arzt gestellt. Die Durchfuhrung fand an einem separaten Termin statt. Die Untersuchung erfolgte durch ein Team bestehend aus drei Pflegekraften und zwei Arzten. Nach einer ausfuhrlichen Erluterung des Vorgehens wurde die Untersuchung gema der Praferenzen der PmMS entweder im Sitzen oder im Stehen und bei gefullter Blase und vorhandenem Harndrang durchgefuhrt.

Die Messung wurde mit dem MMS Nexam Pro Urodynamic System (Laborie/Medical Measurement Systems B.V., Enschede, Niederlande) vorgenommen, sobald die PmMS starken Harndrang verspurte. Im Anschluss erfolgte eine Quantifizierung des Restharns

mittels transabdomineller Sonographie unter Verwendung des GE LOGIQ S7 Pro Ultraschallsystems (GE Ultrasound Korea Inc., Seongnam-si, Südkorea).

2.2 Statistische Analyse

Zur Fragestellung lagen 92 Datensätze mit durchgeführter, gültiger UF, MTB und Anamnese vor.

Die statistische Auswertung erfolgte mit Unterstützung durch einen externen Statistiker.

Die Analysen wurden mit der statistischen Programmiersprache R (R Core Team 2019, R version 4.2.2 (2022-10-31)) auf einem aarch64-apple-darwin20 System unter macOS 14.1 durchgeführt.

Für die Übereinstimmung zwischen jeweils zwei Ratern bezüglich Diagnosen, Therapievorschlügen oder deren Kombinationen wurde die Übereinstimmungsrate mittels Cohens Kappa mit einem Konfidenzintervall von 95 % berechnet.

Ein Wert von $\kappa = 1$ entspricht vollständiger Übereinstimmung, wohingegen ein Wert von 0 bei einer zufälligen Übereinstimmung angenommen wird. Werte unter 0 können entstehen, wenn die Übereinstimmung schlechter als der Zufall ist. Es wurden folgende Grenzwerte angenommen:

- $\kappa < 0$: schlechte Übereinstimmung
- $0 < \kappa \leq 0,2$: geringe Übereinstimmung
- $0,21 < \kappa \leq 0,4$: ausreichende Übereinstimmung
- $0,41 < \kappa \leq 0,6$: mittelmäßige Übereinstimmung
- $0,61 < \kappa \leq 0,8$: beachtliche Übereinstimmung
- $0,81 < \kappa \leq 1$: (fast) vollkommene Übereinstimmung

(Landis und Koch, 1977).

Mit Fleiss' Kappa (κ_F) wurde die Übereinstimmung von mehr als 2 Ratern berechnet. Es werden die gleichen Grenzwerte wie bei Cohens Kappa verwendet.

2.3 Teilnehmende Zentren

Die in die Studie eingeschlossenen Personen wurden in der Neuro-Urologie des Johanniter-Rehabilitationszentrum Godeshöhe zur weiteren Diagnostik und Therapie vorgestellt.

Anschließend wurden sie durch Ärzte der neuro-urologischen Abteilung des REHAB Basel (Klinik für Neurorehabilitation und Paraplegiologie), des Universitätsklinikums Ulm, des Universitätsklinikums Münster und des Johanniter-Rehabilitationszentrum Godeshöhe verblindet befundet.

2.4 Ethische Überlegungen und Einwilligung

Die teilnehmenden Patienten werden keinem zusätzlichen Risiko ausgesetzt. Es liegt ein aktuelles Ethikvotum der Universitätsklinik Bonn zur Datenbank (EK 313/13-University Hospital Bonn, 07.04.2014) und Datennutzung vor.

3. Ergebnisse

3.1 Beschreibung der Stichprobe

Die Patientenkohorte bestand aus 92 PmMS.

Das Alter der PmMS betrug im Durchschnitt 47,5 Jahre ($SD = 9,4$), wobei die jüngste Person 25 Jahre und die älteste Person 72 Jahre alt war.

69,6 % ($N = 64$) der Teilnehmenden waren weiblich und 30,4 % ($N = 28$) männlich.

Die Erkrankungsdauer ergibt sich aus der Differenz des Datums der Untersuchung am neurologischen Rehabilitationszentrum Godeshöhe und dem der Erstdiagnose. Sie lag zwischen einem Monat und 445 Monaten und betrug im Durchschnitt 118,7 Monate.

Bei den meisten Patienten lag eine schubförmig remittierende MS (75 %, $N = 69$) vor. 15,2 % ($N = 14$) hatten eine sekundär progrediente MS und 9,8 % ($N = 9$) eine primär progrediente MS.

96,7 % der Teilnehmenden ($N = 89$) waren in der Lage, die Blase willentlich zu entleeren. 2,2 % ($N = 2$) nutzten die intermittierende Katheterisierung des Restharns und 1,1 % ($N = 1$) triggerten die Entleerung.

Es wurde die Expanded disability status scale (EDSS) bestimmt, welche ein Maß für die Schwere der MS darstellt und von 0 bis 10 möglichen Punkten reicht (Meyer-Moock et al., 2014). Hier lag der Punktwert durchschnittlich bei 3,9 und die Werte lagen in einem Bereich von 1,5 bis 8. Bei 13 % ($N = 12$) PmMS wurde der Score nicht angegeben.

Von den eingeschlossenen PmMS gaben 21,7 % ($N = 20$) keine Symptome an. Bei den Patienten, die Symptome bemerkten, bestanden diese seit 1 bis 2 Jahren in 16,3 % ($N = 15$) und seit 3 bis 5 Jahren in 16,3 % ($N = 15$). Bei 13 % ($N = 12$) bestanden sie seit 6 bis 10 Jahren. 5,4 % ($N = 5$) der PmMS gaben Beschwerden seit über 10 Jahren an. 22,8 % ($N = 21$) der PmMS gaben nicht an, seit wann Symptome bestanden.

54,3 % ($N = 50$) der Teilnehmenden litten zum Zeitpunkt der Untersuchung unter Inkontinenz. Bei 26,1 % ($N = 24$) der PmMS trat sie häufig auf, bei 27,2 % ($N = 25$) der PmMS selten. 1,1 % ($N = 1$) der PmMS gab die Häufigkeit als unbekannt an und 45,7 % ($N = 42$) machten hierzu keine Angabe.

Die Inkontinenz trat am häufigsten bei Harndrang (25 %, N = 23), seltener bei Husten, Niesen oder körperlicher Betätigung (13 %, N = 12) auf. 45,7 % (N = 42) der PmMS gaben nicht an, in welchem Zusammenhang die Symptome auftraten.

56,5 % (N = 52) gaben eine Drangsymptomatik und 48,9 % (N = 45) eine Pollakisurie (erhöhte Miktionsfrequenz) an. 40,2 % (N = 37) der PmMS litten unter Nykturie (nächtliches Wasserlassen), 1,1 % (N = 1) gaben einen Harnverhalt und 5,4 % (N = 5) ein Nachträufeln an. Ein Restharngefühl bestand bei 19,6 % (N = 18) der Teilnehmenden. 16,3 % (N = 15) mussten zur erfolgreichen Miktionspressen, 12 % (N = 11) hatten Startschwierigkeiten, 19,6 % (N = 18) gaben einen unterbrochenen und 12 % (N = 11) einen schwachen Harnstrahl an.

3.2 Interrater-Reliabilität der Verdachtsdiagnose nach Anamnese und MTB

Initial wurde die Übereinstimmung der Verdachtsdiagnosen der verschiedenen Rater bei Kenntnis von Anamnese und MTB der Patienten betrachtet. Dabei wurde sowohl die Übereinstimmung der Einzeldiagnosen als auch die Übereinstimmung aller Diagnosen untersucht.

3.2.1 Übereinstimmung der Einzeldiagnosen

Es konnte zwischen den drei Diagnosen „Normalbefund“, „Detrusorüberaktivität“ und „Detrusorunteraktivität“ gewählt werden.

Bei der Diagnose Normalbefund zeigten sich Werte von $\kappa_C = 0,24$ bis $\kappa_C = 0,58$ und bei der Diagnose DO Werte von $\kappa_C = 0,23$ bis $\kappa_C = 0,54$. In diesen beiden Kategorien wurden also ausreichende bis mittelmäßige Übereinstimmungen erreicht.

Bei der DU zeigte sich die niedrigste Übereinstimmung mit einer Interrater-Reliabilität von $\kappa_C = -0,02$ bis $\kappa_C = 0,17$, also schlechter bis geringer Übereinstimmung.

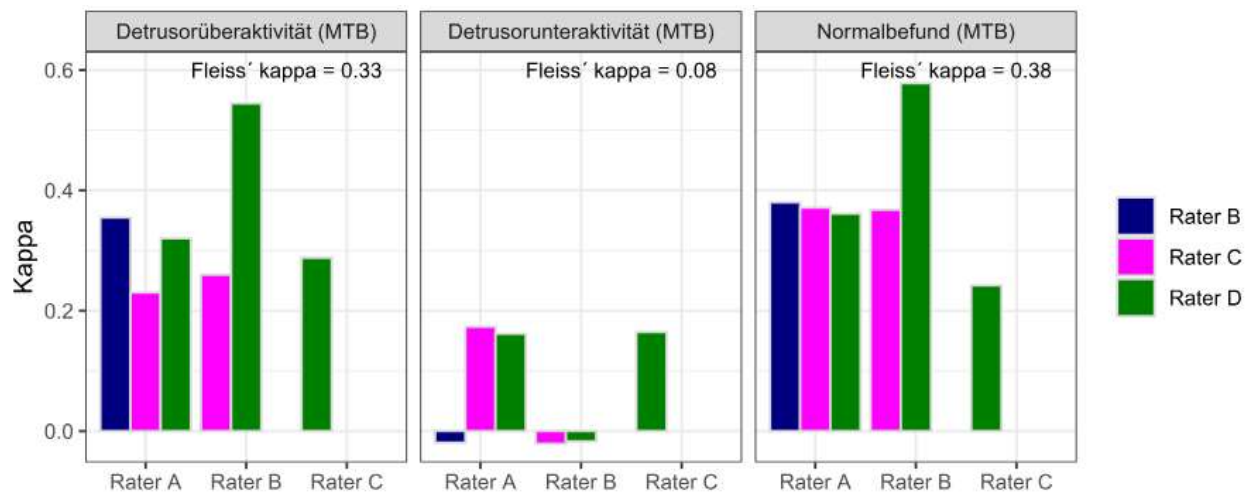


Abbildung 3: Übereinstimmung der Einzeldiagnosen der Rater nach Cohens und Fleiss' Kappa mittels Anamnese und Miktionstagebuch (modifiziert nach Jaekel et al., 2024).

Beim Vergleich der Diagnosen aller Rater (Fleiss' Kappa) zeigte sich die höchste Übereinstimmung wiederum bei Normalbefund ($\kappa_F = 0,38$) und DO ($\kappa_F = 0,33$) als ausreichend und bei der DU ($\kappa_F = 0,08$) als gering übereinstimmend.

3.2.2 Übereinstimmung aller Diagnosen

Die Übereinstimmung der Rater-Paare hinsichtlich aller drei möglichen Diagnosen wurde als nächster Schritt zusammen ausgewertet. Hierbei lag eine ausreichende Übereinstimmung ($\kappa_C = 0,24$ bis $\kappa_C = 0,33$) vor. Ein Rater-Paar zeigte eine mittelmäßige Übereinstimmung ($\kappa_C = 0,54$).

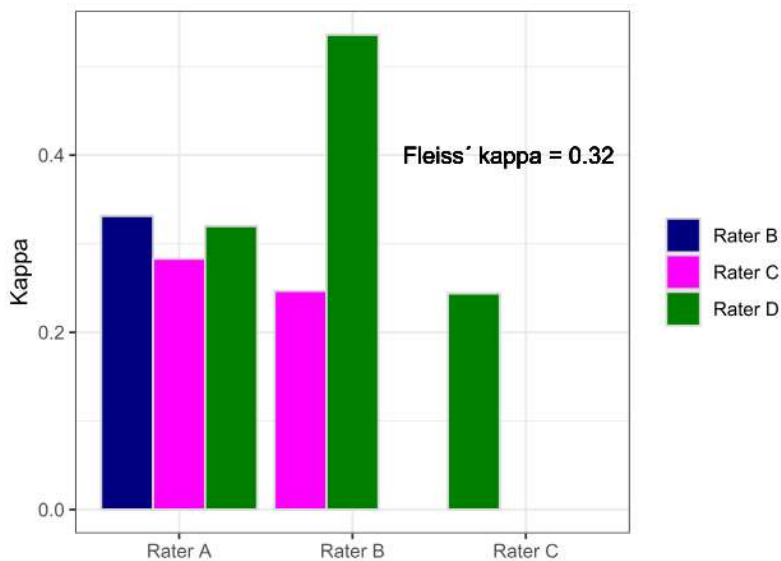


Abbildung 4: Übereinstimmung aller Diagnosen der Rater nach Cohens und Fleiss' Kappa (modifiziert nach Jaekel et al., 2024).

Diese Beobachtungen bestätigten sich mit einem Fleiss' Kappa von 0,32, also einer ausreichenden Übereinstimmung zwischen allen Ratern.

3.3 Interrater-Reliabilität der Verdachtsdiagnosen nach Anamnese, MTB und UF

Da die Analyse hier durch das Vorliegen von Mehrfachantworten erschwert wurde, erfolgte die Bewertung der Übereinstimmung der Einzeldiagnosen zunächst analog zu 3.2.1. Anschließend wurden Diagnose-Kategorien aus den Mehrfachantworten gebildet und deren Übereinstimmung betrachtet. Es wurde festgelegt, dass die Rater sich hier zwischen mehreren Antwortmöglichkeiten entscheiden konnten, da die UF gegenüber der Analyse nach lediglich Anamnese und MTB eine dezidierte Beurteilung ermöglicht.

3.3.1 Übereinstimmung der Einzeldiagnosen

Es konnte aus den fünf Diagnosen „Normalbefund“, „DO“, „DU“, „DSD“ und „BOO“ gewählt werden.

Die Übereinstimmungen reichten hierbei von $\kappa_C = 0,01$ bis $\kappa_C = 0,32$. Die höchsten Werte zeigten sich mit $\kappa_C = 0,32$ bei BOO und mit $\kappa_C = 0,29$ bei Normalbefund. Bei DO, DU und Normalbefund zeigte sich das niedrigste Cohens Kappa mit $\kappa_C = 0,01$.

Insgesamt ist die Übereinstimmung der einzelnen Rater-Paare folglich den Kategorien „gering“ bis „ausreichend“ Übereinstimmung zuzuordnen.

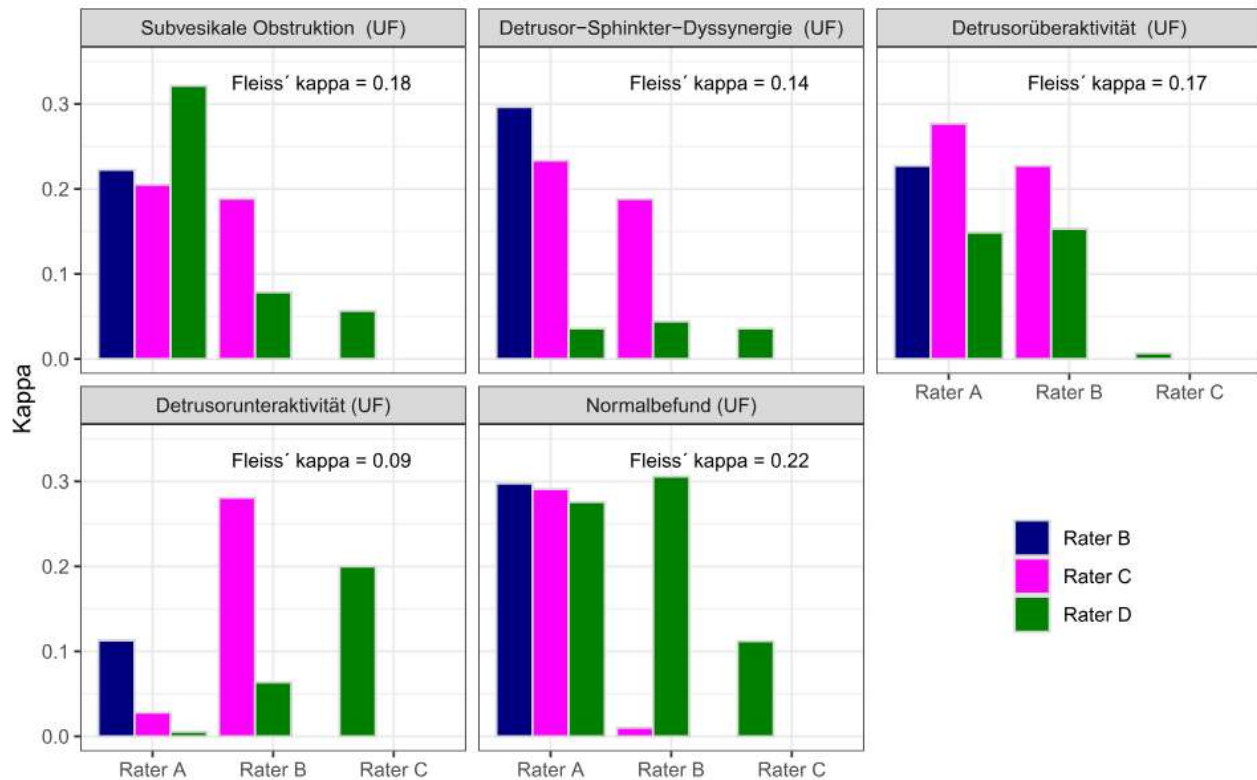


Abbildung 5: Übereinstimmung der Einzeldiagnosen der Rater nach Cohens und Fleiss' Kappa mittels Anamnese, Miktionsstagebuch und Uroflowmetrie (modifiziert nach Jaekel et al., 2024).

Für die Übereinstimmung aller Rater zeigten sich Übereinstimmungen von $\kappa_F = 0,09$ bis $\kappa_F = 0,22$. Bei der Diagnose Normalbefund ($\kappa_F = 0,22$) war die Übereinstimmung somit ausreichend. Alle anderen Diagnosen zeigten geringe Übereinstimmung.

3.3.2 Übereinstimmung der Diagnose-Kombinationen

Da Mehrfachantworten erlaubt waren, wurden die gewählten Diagnose-Kombinationen anschließend separat betrachtet, wobei hier auch das Auswählen nur einer Diagnose als Diagnose-Kombination bezeichnet wurde. Hierdurch ergaben sich 14 Kombinationsmöglichkeiten.

In absteigender Häufigkeit lauten die am häufigsten ausgewählten Kombinationen somit DO (23,6 %, N = 86), DU (22,5 %, N = 82), Normalbefund (16,2 %, N = 59), DO + DSD (11,8 %, N = 43), DSD (7,4 %, N = 27), BOO (7,4 %, N = 27), DO + BOO (5,5 %, N = 20), DU + DSD (3,6 %, N = 13), Normalbefund + BOO (0,5 %, N = 2) und DU + BOO (0,5 %, N = 2). Für die übrigen Kombinationen entschied sich jeweils 1 Rater.

Hierbei ergaben sich zwischen den Ratern insgesamt Werte von $\kappa_C = 0,09$ bis $\kappa_C = 0,19$, also geringer Übereinstimmung.

Tabelle 1: Übereinstimmung der Diagnosekombinationen nach Cohens Kappa zwischen den Ratern paarweise (modifiziert nach Jaekel et al., 2024).

Übereinstimmungsrate, Cohens Kappa, 95 % KI			
Rater	B	C	D
A	30,4 %	27,2 %	23,9 %
	0,19	0,15	0,14
	0,09, 0,28	0,06, 0,25	0,06, 0,23
B		25 %	21,7 %
		0,12	0,1
		0,02, 0,21	0,01, 0,19
C			26,1 %
			0,09
			-0,01, 0,18

Für die Übereinstimmungsbeurteilung auf Basis aller Rater mit Fleiss' Kappa ergab sich insgesamt für alle Kombinationsmöglichkeiten mit $\kappa_F = 0,12$ ebenso geringe Übereinstimmung. Bei den Diagnosegruppen zeigten sich lediglich bei der Antwort DO und Normalbefund Werte für Fleiss' Kappa mit ausreichender Übereinstimmung.

Bei den Diagnosekombinationen (maximal $\kappa_C = 0,19$) ergab sich somit eine etwas niedrigere Übereinstimmung als bei den Einzeldiagnosen (maximal $\kappa_C = 0,22$).

3.4 Interrater-Reliabilität der Therapievorschlage nach UF

Dieser Abschnitt hat die ubereinstimmung der Therapievorschlage zum Gegenstand. Dabei wird analog zu den Diagnosevorschlagen verfahren.

3.4.1 ubereinstimmung der Einzeltherapievorschlage

Zunachst wurden wieder die ubereinstimmungen zwischen den einzelnen Ratern untersucht.

Insgesamt lag bei dieser Fragestellung der Wertebereich zwischen $\kappa_C = -0,06$ und $\kappa_C = 0,71$. Die hochste ubereinstimmung zeigte sich bei den Therapievorschlagen BTX ($\kappa_C = 0,71$), Katheter ($\kappa_C = 0,45$) und keine Therapie ($\kappa_C = 0,38$).

Bei den Diagnosen Biofeedback/Physiotherapie (BF/P) und Sakrale Neuromodulation zeigten sich die niedrigsten Werte mit geringer ubereinstimmung ($\kappa_C = 0,07$ und $\kappa_C = -0,06$).

Grotenteils wurde bei den Therapievorschlagen ausreichende ubereinstimmung gemessen, nur in wenigen Fallen zeigte sich mittelmaige oder beachtliche ubereinstimmung.

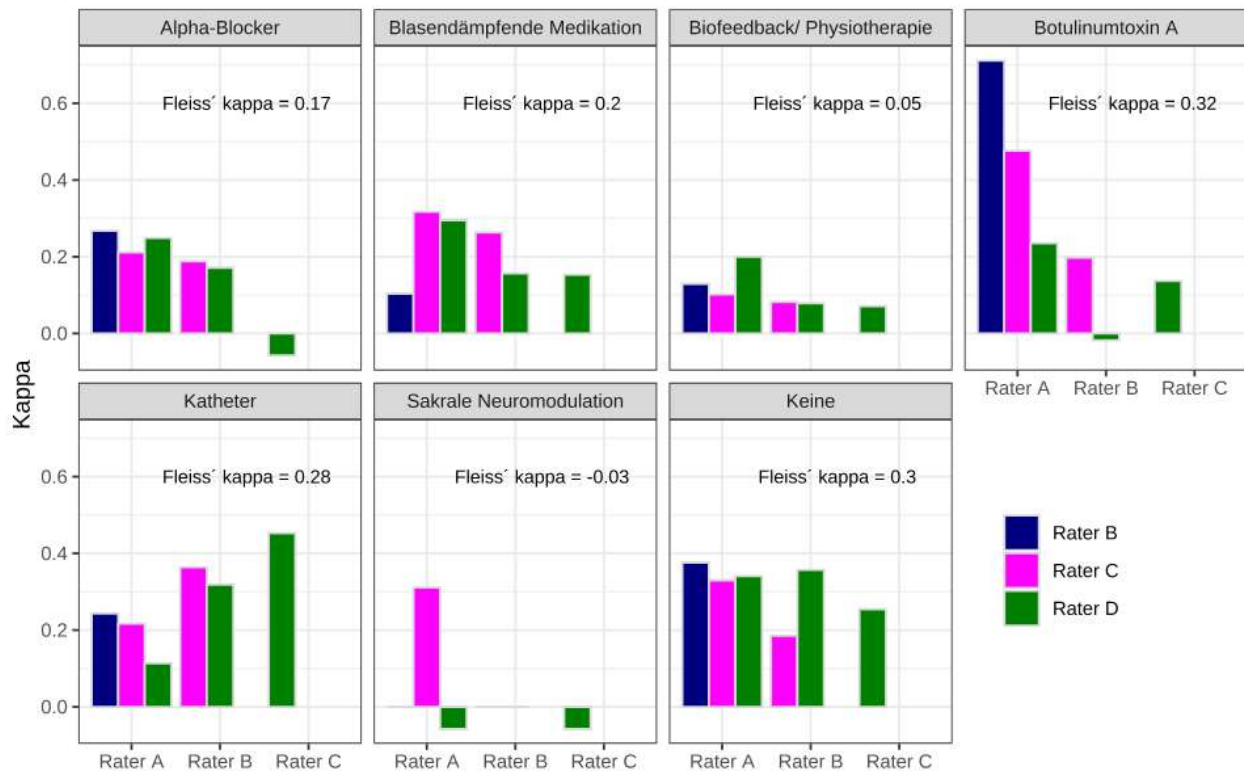


Abbildung 6: Übereinstimmung der Einzeltherapievorschläge der Rater nach Cohens und Fleiss' Kappa (modifiziert nach Jaekel et al., 2024).

Die Übereinstimmung aller Therapievorschläge wurde wieder mit Fleiss' Kappa berechnet.

Hierbei zeigte sich BTX erneut mit der höchsten Übereinstimmung ($\kappa_F = 0,32$), also ausreichender Übereinstimmung. Die Auswahlmöglichkeiten keine Therapie und Katheter erreichten ebenfalls eine ausreichende Übereinstimmung ($\kappa_F = 0,3$, $\kappa_F = 0,28$). Alpha-Blocker, detrusordämpfende Medikation und BF/P zeigten geringe Übereinstimmung ($\kappa_F = 0,17$, $\kappa_F = 0,2$, $\kappa_F = 0,05$) und sakrale Neuromodulation die geringste Übereinstimmung ($\kappa_F = -0,03$).

3.4.2 Übereinstimmung der Therapie-Kombinationen

Da für jede PmMS mehrere Therapievorschläge gemacht werden konnten, wurden anschließend wieder die Auswahl-Kombinationen betrachtet. Die am häufigsten vergebenen Kombinationen waren in absteigender Reihenfolge keine Therapie (21,4 %,

N = 78), detrusordämpfende Medikation + BF/P (11,2 %, N = 41), detrusordämpfende Medikation (9,9 %, N = 36), Alpha-Blocker + BF/P (9,0 %, N = 33), BF/P (8,8 %, N = 32), Katheter + detrusordämpfende Medikation (5,5 %, N = 20) und Alpha-Blocker (4,9 %, N = 18).

Tabelle 2: Übereinstimmung der Rater paarweise hinsichtlich der Therapie-Kombinationen (modifiziert nach Jaekel et al., 2024).

Übereinstimmungsrate, Cohens Kappa, 95 % KI

Rater	B	C	D
A	15,2 %	17,4 %	19,6 %
	0,09	0,1	0,1
	0,02, 0,16	0,03, 0,16	0,02, 0,18
B		17,4 %	10,9 %
		0,09	0,05
		0,02, 0,17	0, 0,11
C			17,4 %
			0,1
			0,03, 0,17

Insgesamt ergaben sich hierbei Werte von $\kappa_C = 0,05$ bis $\kappa_C = 0,1$, also ausschließlich Werte mit geringer Übereinstimmung.

Die Übereinstimmung aller Rater zeigte für die folgenden Diagnosekombinationen beziehungsweise Einzeldiagnosen für Fleiss' Kappa die höchsten Werte:

- Keine Therapie: $\kappa_F = 0,31$ (ausreichende Übereinstimmung)
- Katheter + BTX: $\kappa_F = 0,16$ (geringe Übereinstimmung)
- Alpha-Blocker + BF/P: $\kappa_F = 0,08$ (geringe Übereinstimmung)
- Detrusordämpfende Medikation + BF/P: $\kappa_F = 0,06$ (geringe Übereinstimmung)

Es ergab sich ein gesamtes Fleiss' Kappa von 0,08 (geringe Übereinstimmung).

3.5 Vergleich Diagnosen verblindete Rater mit Originalbefund

In diesem Abschnitt werden die Diagnosen der verblindeten Rater (Rater A - Rater D) mit dem Originalbefund aus der Datenbank (Rater E) verglichen. Die Originalbefunde wurden durch verschiedene Ärzte des neurologischen Rehabilitationszentrums Godeshöhe erstellt. Durch Rater E sind nur die Diagnosen Normalbefund, DO und DU vergeben worden, somit wurden hier nur diese drei Diagnosen verglichen.

3.5.1 Übereinstimmung der Einzeldiagnosen

Es zeigten sich Werte von $\kappa_C = 0,03$ bis $\kappa_C = 0,24$ zwischen den Ratern, also geringe bis ausreichende Übereinstimmung.

3.5.2 Übereinstimmung der Diagnose-Kombinationen

Betrachtet man die Diagnose-Kombinationen, so ergaben sich zwischen den verblindeten Ratern und Rater E Werte von $\kappa_C = 0,02$ bis $\kappa_C = 0,17$. Dies bedeutet geringe Übereinstimmung.

4. Diskussion

4.1 Hauptergebnisse

Ziel dieser Arbeit war die Untersuchung der IRR der UF bei PmMS und NLUTD hinsichtlich verschiedener Endpunkte, um den Stellenwert der UF bei dieser Personengruppe besser beurteilen zu können. Hierfür bewerteten 4 Rater die Untersuchungen von 92 Patienten prospektiv (siehe auch Jaekel et al., 2024).

Es wurde sowohl die IRR der Diagnose auf Grundlage von Anamnese und MTB bestimmt als auch die IRR der potenziellen Diagnose und die Therapiewahl bei zusätzlichem Vorliegen einer UF untersucht.

Hierbei zeigten sich je nach Fragestellung Werte für Cohens Kappa von schlechter ($\kappa_C = -0,06$) bis beachtlicher Übereinstimmung ($\kappa_C = 0,71$) und für Fleiss' Kappa für die Übereinstimmung aller Rater von schlechter ($\kappa_F = -0,03$) bis ausreichender Übereinstimmung ($\kappa_F = 0,32$).

4.2 Mögliche Erklärung niedriger Interrater-Reliabilität

4.2.1 Variabilität durch Anzahl der Diagnosen und Therapien

In dieser Arbeit wurden die Übereinstimmungen der Befunde der UF-Kurven im Hinblick auf eine Vielzahl verschiedener Diagnosen sowie eine Reihe unterschiedlicher Therapiemöglichkeiten untersucht. Die Untersuchungen der Interrater-Reliabilität sind somit sehr detailliert.

So standen beispielsweise bei der Verdachtsdiagnose nach Anamnese, MTB und UF fünf Auswahlmöglichkeiten zur Verfügung. Hier ergab sich eine IRR von maximal $\kappa_F = 0,22$ (ausreichende Übereinstimmung) bei dem Vergleich der Einzeldiagnosen.

Studien, in denen die Auswahl der Antwortmöglichkeiten zum Beispiel auf lediglich zwei beschränkt wurde, liefern bessere Ergebnisse für die IRR: So ergaben sich beispielsweise Werte für Kappa von 0,7 bei einer Entscheidung der Rater zwischen einer normalen und einer pathologischen UF-Kurve bei Kindern (van Batavia et al., 2023). Chang und Yang

zeigten bei der Bewertung von Kinder-UF-Kurven als "normal" oder "abnormal" sogar Werte für Kappa bis 0,81 (Chang und Yang, 2008).

Diese Art der Diagnosestellung lässt sich mit der klinischen Praxis allerdings schwer vergleichen: Es ist nicht ausreichend, auf Basis einer dichotomen Entscheidung eine Kurve als pathologisch zu definieren. Um eine Therapieentscheidung zu treffen, muss meist die genaue Diagnose bekannt sein und wie in dieser Studie simuliert, sind in der Praxis meist mehr als nur zwei Verdachtsdiagnosen zu betrachten.

In einem zweiten Schritt der Studie von Chang und Yang sollten die pathologischen Kurven genauer beschrieben werden. Hierbei zeigte sich lediglich eine IRR von $\kappa = 0,07$, was geringer Übereinstimmung entspricht (Chang und Yang, 2008). Auch van Batavia et al. fanden niedrigere Werte für Kappa ($\kappa = 0,34$) bei der genaueren Einteilung der Ergebnisse anhand der Kurvenform (beispielsweise glockenförmig, plateauförmig, turmförmig, staccato oder unterbrochen) (van Batavia et al., 2023).

Der Aufbau dieser Untersuchungen gleicht somit eher dieser Arbeit und auch der klinischen Praxis. Auch die Ergebnisse für die IRR sind bei dieser Form der Fragestellung mit denen dieser Arbeit vergleichbar.

Beim Vergleich der verschiedenen Einzelanalysen dieser Arbeit wird die Abhängigkeit der IRR von der Anzahl der Antwortmöglichkeiten deutlich:

Bei der Diagnosestellung mittels Anamnese und MTB konnten die Rater lediglich zwischen drei Diagnosen wählen. Es ergab sich eine höhere IRR als bei der Beurteilung der potenziellen Diagnosen mit der zusätzlichen Betrachtung von UF, bei der die Rater zwischen fünf Möglichkeiten wählen konnten (maximal $\kappa_F = 0,38$ versus maximal $\kappa_F = 0,22$).

Insgesamt zeigt sich somit eine hohe Variabilität der IRR je nach Aufbau der Fragestellung.

4.2.2 Mehrfachantworten

In dieser Arbeit war die Mehrfachauswahl in den Kategorien Diagnose nach UF und Therapie nach UF möglich, um dem klinischen Entscheidungsprozess in der Praxis zu entsprechen. Zusätzlich gab es keine Beschränkung bezüglich der Minimalbeziehungsweise Maximalanzahl der Antworten. Hieraus ergibt sich eine große Anzahl von Kombinationsmöglichkeiten, die durch die große Zahl der Antwortmöglichkeiten weiter verstärkt wird. Dies macht eine Übereinstimmung der Rater und somit eine hohe IRR unwahrscheinlicher.

Beispielsweise lag die IRR bei der Verdachtsdiagnose nach Anamnese, MTB und UF bei der Betrachtung der Diagnose-Kombinationen höchstens bei $\kappa_C = 0,19$ (geringe Übereinstimmung).

Andere Studien erlaubten den Ratern lediglich Einfachantworten und erreichten somit auch höhere Werte für die IRR (van Batavia et al., 2023; Faasse et al., 2015).

Darüber hinaus kann die Option der Mehrfachantwort durch die Rater unterschiedlich interpretiert worden sein: Möglicherweise wurde/n durch manche Rater nur die relevanteste/n Diagnose/n angekreuzt und durch andere alle möglichen Diagnosen bei diesem Patienten. Ebenso verhält es sich mit den Therapiemöglichkeiten.

Auch hier entspricht das mögliche Vorliegen mehrerer Diagnosen nebeneinander eher der klinischen Praxis: Ein Patient kann sowohl eine DO als auch eine BOO haben. Auch die gleichzeitige Anwendung mehrerer Therapien (sogar in größerem Umfang als hier zur Auswahl stehend) ist möglich und durchaus üblich.

4.2.3 Ähnliche Kurvenformen

Eine weitere Erklärung der geringen IRR besteht darin, dass die Diagnosen teilweise mit sehr ähnlichen Kurvenformen einhergehen. Ein Grund für die besonders niedrige IRR bei der Diagnose DU könnte somit darin bestehen, dass die Kurve nur schwer von der Kurve der BOO unterschieden werden kann (Lee et al., 2016; Chow et al., 2021; Arevalo-Vega et al., 2023).

4.2.4 Einflussfaktor Mensch

In der UF spielt, wie bei fast jedem medizinischen Diagnoseverfahren, der Arzt, beziehungsweise die Ärztin mit ihren individuellen Eigenschaften als Mensch eine Rolle. Aspekte wie die gesammelte Erfahrung, das Geschlecht, das Erlernen der ärztlichen Fähigkeiten an unterschiedlichen Standorten und der unterschiedliche Erfolg hierbei, aber auch sein Befinden zum Zeitpunkt der Untersuchung und die Umgebungsfaktoren sind sehr unterschiedlich und können somit die Diagnosestellung beeinflussen (Cruppé et al., 2011; McKinlay et al., 2006; McKinlay et al., 2002).

Das Ausmaß der Variabilität durch den Rater ist je nach Diagnostikverfahren unterschiedlich groß. Beinhaltet das Verfahren viele Anteile, die objektiv beurteilbar sind, wie beispielsweise Messgrößen oder eine zu erkennende Bildstruktur, kann davon ausgegangen werden, dass die Reliabilität bei diesem Verfahren höher ist als bei einer Untersuchung, welche zahlreiche subjektive Anteile enthält.

Ein Verfahren, das eine eher objektive Beurteilung ermöglicht, ist das EKG. In einer Studie zur EKG-Interpretation bei Sportlern konnte beispielsweise eine IRR von $\kappa = 0,63$ erreicht werden. Spezialisten erhielten sogar einen Wert von $\kappa = 0,69$, was einer beachtlichen Übereinstimmung entspricht (Magee et al., 2014). Dies könnte darauf zurückgeführt werden, dass das EKG Bildstrukturen enthält, die objektiv als physiologisch oder pathologisch eingeordnet werden können, wie zum Beispiel die Ausprägung einer bestimmten Zacke, die Hebung oder die Senkung bestimmter Abschnitte. Außerdem enthält das EKG viele Messgrößen, wie beispielsweise die PQ-Zeit oder die QT-Zeit, welche objektiv mit Normwerten verglichen werden können und somit unabhängig vom Rater immer gleich bewertet werden können.

Die UF lässt sich in diesem Spektrum eher im subjektiven Bereich einordnen.

Zwar fließen auch hier bestimmte Messgrößen, welche direkt durch das Gerät gemessen werden, mit in die Beurteilung ein, wie beispielsweise der maximale Harnfluss und das Miktionsvolumen. Darüber hinaus ist die Durchführung der Messung an sich

standardisiert: Die Messung erfolgt durch das UF-Gerät und nicht durch den behandelnden Arzt selbst.

Allerdings liegt der Fokus der Befundung auf der Bewertung der Form der Kurve, welche subjektiv durch den Rater erfolgt. Laut früheren Untersuchungen stellt die Form der UF-Kurve den wichtigsten Faktor bei der Interpretation und Beurteilung der UF dar (Griffiths und Scholtmeijer, 1984; Mattsson und Spångberg, 1994). Dieser Teil der Bewertung konnte bisher nicht standardisiert werden.

Zwar existieren für bestimmte Kurvenformen konkrete Vorschläge, welche Diagnosen ihnen zugeordnet werden könnte. So wird beispielsweise postuliert, dass eine abgeflachte, plateauförmige Kurve mit verringertem maximalen Harnfluss auf eine Obstruktion der Harnröhre hinweist (Schmid et al., 2020). Allerdings zeigt die Praxis, dass die Kurven nur in seltenen Fällen genau einer Vorlage zuzuordnen sind, denn die Kurven sind vielfältig und können Mischbilder enthalten. Während ein Rater die Kurve als glockenförmig interpretiert, könnte der nächste sie als turmförmig einordnen (Franco et al., 2016).

Welches Flussmuster der Betrachter also in der UF erkennt, beziehungsweise was er diesem Flussmuster auf den ersten Blick entnimmt, ist somit trotz der Mitbetrachtung gewisser Grenzwerte sehr subjektiv (van Batavia et al., 2023; Vijverberg et al., 2011).

Um den diagnostischen Wert der UF bezüglich dieser Unsicherheit zu verbessern, wurden bereits Untersuchungen zu sogenannter Patterning-Software durchgeführt. Hierbei werden UF-Kurven mittels eines Programmes, welches künstliche Intelligenz verwendet, mit physiologischen UF-Befunden verglichen und nach bestimmten Algorithmen ausgewertet (Kanematsu et al., 2013). Das Programm verspricht also eine Objektivierung des Diagnostik-Vorgangs.

Choo et al. entwickelten beispielsweise einen Algorithmus zur automatischen Analyse von UF-Kurven mittels künstlicher Intelligenz. Hier kann allerdings bisher nur zwischen pathologischem und physiologischem Befund unterschieden werden (Choo et al., 2022).

In Zukunft könnte diese Methode jedoch nach deren Weiterentwicklung eine Möglichkeit zur Verbesserung der IRR und somit zur Güte der UF darstellen.

Darüber hinaus ist beschrieben worden, dass Bezeichnungen wie "staccato-förmig" oder "unterbrochen", die für die Formen der Kurven verwendet werden, missverständlich sind und häufig keine konsistente Terminologie verwendet wird (Gammie und Drake, 2018).

Es sollten also Wege gefunden werden, die Form-Analyse der UF-Kurven begrifflich zu standardisieren, um sie sicherer mit der zugrundeliegenden Blasenentleerungsstörung assoziieren zu können (Gammie und Drake, 2018).

Es lässt sich schlussfolgern, dass der einzelne Rater einen signifikanten Einfluss auf die Ergebnisse der Untersuchungen ausübt. Folglich kann nicht sichergestellt werden, dass mittels UF zuverlässige, reproduzierbare und vergleichbare Daten als einzige Grundlage für die Behandlung von PmMS und NLUTD erhoben werden können, obwohl dies in der Medizindiagnostik von grundlegender Bedeutung ist.

4.3 Ergebnisse im wissenschaftlichen Kontext

Mit dieser Arbeit ist die Interrater-Reliabilität der UF von neurogenen BFS bei PmMS erstmalig systematisch untersucht worden. Es existieren hingegen viele Studien, welche die UF in anderen Kontexten, beispielsweise bei anderen Patientengruppen untersucht.

Viele Autoren kommen zu dem Schluss, dass die UF als Diagnostikmethode aufgrund niedriger IRR nicht ausreichend oder nur eingeschränkt nutzbar ist (Netto et al., 2020; Faasse et al., 2015; Liu et al., 2014; van de Beek et al., 1997).

Andere Autoren (Chang und Yang, 2008) kommen zu dem Ergebnis, dass die UF lediglich für eine Unterscheidung zwischen abnormaler und normaler Kurve, nicht aber für eine spezifische Klassifizierung der Kurven gut geeignet ist. Dies rechtfertigt den Einsatz der UF zumindest als Screening-Methode.

Es besteht ein Konsensus, dass die UF besonders bezüglich der Befundung besser standardisiert werden muss (Gammie et al., 2018; Chang und Yang, 2008; Faasse et al., 2015)

Es lässt sich zusammenfassen, dass die UF bei der Diagnostik von Blasenfunktionsstörungen bei NLUTD ergänzend genutzt werden kann, beispielsweise

zum Screening oder zum Vergleich von Vorbefunden mit Befunden nach Therapie. Zur alleinigen Diagnosestellung und Therapiegrundlage sollte sie in dieser Form keine Verwendung finden.

4.4 Interrater-Reliabilität in der Medizin

Bei der Untersuchung der IRR in dieser Arbeit fand die Maßzahl Cohens Kappa Anwendung.

Es handelt sich hierbei um das am häufigsten eingesetzte Maß zur Bewertung der Übereinstimmung zwischen zwei Ratern. Allerdings werden die potenziellen Schwierigkeiten und Limitationen dieses Indikators bei der Analyse häufig nicht ausreichend berücksichtigt (Wirtz und Kutschmann, 2007).

In dieser Arbeit wurden subjektive Einschätzungen von Ratern erhoben, die darüber Aufschluss geben sollten, welche von mehreren Merkmalsalternativen zutreffen. Dies impliziert die Verwendung eines polytomen Antwortformates.

Ein Kritikpunkt besteht darin, dass bei diesem Vorgehen alle nicht übereinstimmenden Beurteilungen als gleich gravierend bewertet werden. Es empfiehlt sich daher eher, das gewichtete Cohens Kappa (κ_w) zu verwenden, wenn davon ausgegangen wird, dass eine bestimmte Fehlbeurteilung mit schwerwiegenderen Konsequenzen einhergeht (Wirtz und Kutschmann, 2007).

Bezogen auf diese Arbeit müsste also zusätzlich eine Beurteilung stattfinden, welche der fünf Diagnosen, beziehungsweise welche der sieben Therapieoptionen unter keinen Umständen als falsch klassifiziert werden darf. Sei beispielsweise das Nicht-Erkennen einer BOO besonders kritisch, könnte diese Diagnose durch das gewichtete Cohens Kappa als doppelt so gravierend eingeordnet werden, wie nicht übereinstimmende Befunde, welche die BOO nicht betreffen (Wirtz und Kutschmann, 2007).

κ_w kann also flexibel an spezifische Fragestellungen angepasst werden und ermöglicht somit die Berücksichtigung von unterschiedlich bedeutsamen Inkongruenzen zwischen Ratern (Wirtz und Kutschmann, 2007).

Ein weiterer Kritikpunkt besteht darin, dass κ (und ebenso κ_{ω}) von der Grundhäufigkeit, mit der die Merkmale auftreten, abhängt. Variiert deren Häufigkeit stark, nehmen κ und κ_{ω} beispielweise niedrigere Werte an, auch wenn ansonsten gleiche Bedingungen herrschen (Wirtz und Kutschmann, 2007). Die Auswirkungen der Prävalenz auf die IRR konnten in dieser Studie nicht bewertet werden, da das Studiendesign nicht dazu angelegt war und kein Vergleich der Ergebnisse mit denen aus UD-Studien durchgeführt wurde.

Neben der Betrachtung des Kappa-Wertes selbst sollten auch dessen Grenzwerte, die zur Interpretation unumgänglich sind, genauer betrachtet werden.

Es ist unklar, welche Kappa-Werte in der Gesundheitsforschung als annehmbar gelten. Nach Cohens Interpretation könnten die Grenzen für gesundheitsbezogene Studien zu nachsichtig sein (McHugh, 2012).

In der Literatur werden unterschiedliche Grenzwerte für Cohens Kappa verwendet. Da in dieser Arbeit nicht wie in anderen auf klinische Erfahrungswerte zurückgegriffen werden konnte, wurden diejenigen von Landis und Koch verwendet (Landis und Koch, 1977).

Die beiden Autoren erklären die Notwendigkeit dieser Grenzwerte, beziehungsweise Kategorien, zum Ermöglichen einer einheitlichen Nomenklatur (Landis und Koch, 1977). Ab welchem Grenzwert, beziehungsweise ab welcher Kategorie eine ausreichende Sicherheit für eine medizinische Diagnostikmethode anzunehmen ist, bleibt unklar.

Die Verwendung der Autoren eines Beispiels aus dem medizinischen Bereich in ihrem Artikel weist zumindest darauf hin, dass sie bei ihren Überlegungen durchaus die Verwendung von κ im medizinischen Kontext angedacht haben (Landis und Koch, 1977).

Im Folgenden soll die Bedeutung der Grenzwerte inhaltlich betrachtet werden. Der einfache Umkehrschluss bietet eine hilfreich Perspektive (Rädiker und Kuckartz, 2019): Mit den in dieser Arbeit verwendeten Grenzwerten läge eine fast vollkommene Übereinstimmung dann vor, wenn die Rater in über 80 % der Fälle in ihren Ergebnissen übereinstimmen. In diesem Fall hätten also wiederum 20 % der Patienten eine abweichende, gegebenenfalls falsche Diagnose, beziehungsweise Therapie erhalten. Es muss diskutiert werden, ob dies ein annehmbarer Teil ist.

Für den Vergleich der Diagnosen nach UF in dieser Arbeit ergab sich der höchste Wert für die Diagnose Normalbefund mit einem Fleiss' Kappa von 0,22, was laut den hier verwendeten Grenzwerten als ausreichend klassifiziert wurde. Damit ergibt sich ein Anteil von ungefähr 78 % der Patienten, die eine von der Allgemeinheit abweichende, potenziell falsche Diagnose erhalten haben. Es muss in Frage gestellt werden, ob ein solch großer Anteil, besonders im Bereich der Medizin, noch als „ausreichend“ anzusehen ist. Dies wäre nur möglich, wenn die Entscheidung für eine abweichende Diagnose keine große Bedeutung im Hinblick auf Therapieentscheidung und Outcome hätte.

4.5 Grenzen dieser Arbeit

In dieser Arbeit wurde für jeden Patienten nur eine UF im Rahmen der Erstvorstellung am Johanniter-Rehabilitationszentrum Godeshöhe betrachtet.

Die UF kann abhängig von der Miktionsmenge und den äußeren Umständen des Tests zu unterschiedlichen Ergebnissen führen (Gammie und Drake, 2018).

Laut Primus et al. wird nur bei pathologischen Werten, wie beispielsweise einem verringerten Maximalfluss, eine Wiederholung der Untersuchung empfohlen (Primus und Heidler, 2003). Andere Autoren postulieren, dass die standardmäßige Durchführung mehrerer vergleichbarer UFs, beispielsweise drei Stück, den größten informativen Nutzen bringt (Gammie und Drake, 2018). Somit wäre die Beurteilung nur einer UF, wie in dieser Arbeit, unzureichend.

In der klinischen Praxis ist es jedoch schwer umsetzbar, für jede Person mehrere UF durchzuführen, da für jede neue Blasenfüllung teilweise erhebliche Wartezeiten anfallen.

Darüber hinaus erfasst die Studie keine Begründungen der Rater für ihre Entscheidungen. Mit unterschiedlichen Therapien können unterschiedliche Follow-Up-Strategien verbunden sein, die auf ihre Weise sinnvoll sein können. Zusätzlich zur IRR müsste also der Therapieerfolg gemessen werden.

Eine weitere Grenze besteht darin, dass die Rater in dieser Arbeit lediglich die verblindete Anamnese, das MTB und die UF-Kurve der Patienten erhielten. Sie trafen die Patienten nicht persönlich und hatten somit keine Möglichkeit, sich ein Bild über den Allgemeinzustand des Patienten zu machen, sie also über die Fakten der Untersuchungsmethode hinaus einzuschätzen oder nach eigenem Schema zu befragen. Dies ist in der klinischen Praxis eher ungewöhnlich.

Aus diesem Grund wurde mittels Rater E ein Rater zur Untersuchung hinzugezogen, welcher während der Erstuntersuchung direkten Kontakt zu den Patienten in der Klinik hatte.

Möglicherweise konnte er hierbei, wenn auch unbewusst, mehr über sie erfahren, einen besseren Eindruck von ihnen gewinnen und diesen, auf Basis ihm bereits bekannter Fälle verarbeiten und einordnen. Er konnte so in der Anamnese seine Fragen individueller anpassen und die Patienten sogar selbst untersuchen.

Eine Einschränkung besteht darin, dass für Rater E nur die Diagnosen Normalbefund, DO und DU erfasst wurden, da die Studie diesen Rater retrospektiv eingeschlossen hat.

Es ist nicht bekannt, ob die kontaktlose Bewertung zu bedeutend anderen Ergebnissen führt als die mit Patientenkontakt. Diese Überlegung wurde mit kleinen Fallzahlen für die Video-Urodynamik untersucht. Hierbei zeigte sich kein signifikanter Vorteil einer live-Befundung gegenüber einer post-hoc Befundung (Smith et al., 2009).

Möglicherweise ergibt sich durch die ärztliche Intuition bei Patientenkontakt die Genauigkeit der Diagnosestellung sowie eine Verbesserung des Erkennens von falschen Diagnosen. Hinweise zeigen, dass insbesondere das Vorliegen von Erfahrungswerten bei diesem nicht-analytischen Denken eine Rolle spielt (Vanstone et al., 2019).

Auch wenn ärztliches Personal der Rolle und dem Vorteil der Intuition unterschiedlich viel Bedeutung zuweisen (van den Brink et al., 2019), ist das Vertrauen auf ärztliche Intuition weit verbreitet und sollte deshalb als Aspekt in der Diagnosefindung nicht vernachlässigt werden (Adler, 2022).

Einen Hinweis gibt der Vergleich des Raters E mit den verblindeten Ratern A-D: Hier zeigt sich eine etwas größere Diskrepanz ($\kappa = \text{maximal } 0,24$) als zwischen den verblindeten

Ratern allein ($\kappa = \text{maximal } 0,32$). Ein kausaler Zusammenhang lässt sich an dieser Stelle jedoch nicht belegen.

Zusammenfassend muss jedoch betont werden, dass die genannten Limitationen für alle Rater A-D gleich waren.

4.6 Ausblick

4.6.1 Implikationen für die Praxis

Während für die Durchführung der UF bereits standardisierte Empfehlungen bestehen (Rosier et al., 2017), existieren diese für die Beurteilung der UF noch nicht.

Vorschläge dazu wurden bereits für die Interpretation von Kinder-UF-Kurven untersucht. Hierbei werden Symptome und UF-EMG-Befunde für fünf verschiedene, nicht neurologische Diagnosen genannt, die auf die richtige Diagnose hinweisen (van Batavia et al., 2023).

Die Anwendung dieser Kriterien wurde hinsichtlich der IRR überprüft (Faasse et al., 2015): Insgesamt konnte ein Kappa von 0,46 ($KI = 0,38 - 0,54$), also ausreichender Übereinstimmung erreicht werden (Faasse et al., 2015). Somit wurde eine etwas bessere IRR als in den Untersuchungen in dieser Arbeit bei Erwachsenen mit NLUTD gefunden.

Zur Standardisierung des Prozesses der Befundung Erwachsener mit NLUTD könnte also eine ähnliche Leitlinie entworfen werden. Dies würde vermutlich dazu führen, dass die Bewertung weniger abhängig von den individuellen Ratern ist und sich somit die IRR sowohl für die Diagnostik- als auch für die Therapieentscheidung erhöhen würde.

Es sollte darauf geachtet werden, dass die Kriterien einfach zu verstehen und durchzuführen sowie möglichst objektivierbar sind (Faasse et al., 2015). So ließe sich auch der Faktor der unterschiedlichen Erfahrung, die die Rater mitbringen, verringern.

Ebenso könnte ein standardisiertes Curriculum zum Erlernen der Nutzung dieses Leitfadens und der Interpretation der UF-Kurven im Allgemeinen die IRR erhöhen.

Zusätzlich sollte die Einhaltung der Standards zur Durchführung der UF stringent überprüft werden, um eine einheitliche Grundlage zur Bewertung dieser zu ermöglichen:

Neben der Schaffung einer geeigneten Umgebung für die Untersuchung müssen die Patienten vor dem Test bestmöglich über dessen Durchführung informiert sein (Gammie und Drake, 2018).

4.6.2 Implikationen für weitere Untersuchungen

Gammie und Drake schlagen eine Kombination der UF mit anderen nicht-invasiven Untersuchungen vor (Gammie und Drake, 2018). Bis zur Entwicklung und Validierung einer solchen diagnostischen Kombination ist auf ein invasiveres Verfahren zurückzugreifen.

Bisher hat sich die Urodynamik in grenzwertigen oder unklaren Fällen der UF angeschlossen (Blok et al., 2023). Einige Leitlinien empfehlen die Durchführung einer UD erst nach Versagen der Therapieempfehlung, die auf der Basisdiagnostik des symptomatischen Patienten basiert (Fowler et al., 2009; Çetinel et al., 2013). In zahlreichen Studien wird der klinische Nutzen der UD in Diagnostik und konsekutiven Therapieentscheidung der NLUTD bei PmMS sogar bestritten (El Helou et al., 2021; Dillon und Lemack, 2014; Stoffel, 2017).

Die Implementierung einer regelhaften Durchführung einer UD nach jeder UF könnte jedoch die Genauigkeit von Diagnose- und Therapieentscheidung, also auch der Reliabilität signifikant erhöhen. So wäre möglicherweise der diagnostische Nutzen für die Patienten dem Risiko der Invasivität der Untersuchung überlegen.

Es sollte analog zu dieser Arbeit untersucht werden, wie sich die IRR der UD hinsichtlich Diagnose- und Therapieentscheidung verhält. Zusätzlich können die Diagnosen und Therapien der einzelnen Patienten zwischen den beiden Verfahren miteinander verglichen werden.

4.6.3 Intrarater-Reliabilität

Eine weitere Möglichkeit zur Untersuchung der Güte der UF bei PmMS besteht in der Betrachtung der Intrarater-Reliabilität. Bei dieser steht der einzelne Rater im Vordergrund.

Es wird untersucht, ob er sich wiederholt bei dem gleichen Patienten für die gleiche Diagnose und Therapie entscheiden würde.

Bisherige Studien zur UF bei Erwachsenen ergaben insgesamt eher niedrige Werte für die Intrarater-Reliabilität, wobei Rater mit langjähriger Erfahrung etwas besser abschnitten (Netto et al., 2020).

Eine Untersuchung der Intrarater-Reliabilität speziell bei PmMS könnte die bisherigen Ergebnisse in der Literatur für andere Patientengruppen bestätigen und einen weiteren Hinweis auf die Relevanz und die Testgüte der UF bei der Behandlung von NLUTD geben.

5. Zusammenfassung

Personen mit Multipler Sklerose (PmMS) sind häufig von neurogener Dysfunktion des unteren Harntraktes betroffen. Da dies große Auswirkungen auf die Lebensqualität der Betroffenen hat, ist eine frühzeitige, akkurate Diagnosestellung von Bedeutung.

Die Uroflowmetrie (UF) stellt ein etabliertes Verfahren in der Urologie dar, welches vor weiteren Untersuchungen bei neurogenic lower urinary tract dysfunction (NLUTD) empfohlen wird. Einige Autoren ziehen sogar in Erwägung, UF anstelle der Urodynamik zu verwenden. Der Stellenwert der Uroflowmetrie in der Diagnostik und Therapiesteuerung der NLUTD bei PmMS ist jedoch nicht systematisch untersucht.

Daher war es Ziel dieser Arbeit, die Interrater-Reliabilität (IRR) der UF im Hinblick auf potenzielle Diagnose- und Therapieempfehlungen hin zu überprüfen.

In der prospektiven Untersuchung erhielten vier in der Neuro-Urologie erfahrene Rater verblindete Anamnesen, Miktionstagebücher und Uroflowmetrien von 92 PmMS und bewerteten diese hinsichtlich der vermuteten Diagnose (Normalbefund, Detrusorüberfunktion, Detrusorunterfunktion, Detrusor-Sphinkter-Dyssynergie, subvesikale Obstruktion) und der entsprechenden Therapieempfehlung (keine Therapie, Katheter, Alpha-Blocker, detrusordämpfende Medikation, Botulinumtoxin A, sakrale Neuromodulation, Biofeedback/Physiotherapie). Anschließend wurde die IRR von Raterpaaren mit Cohens und von allen Ratern mit Fleiss' Kappa bewertet.

Bei den Einzeldiagnosen ergaben sich Werte für Cohens Kappa von 0,01 bis 0,32 sowie für Fleiss' Kappa von 0,09 bis maximal 0,22. Hinsichtlich der Therapien zeigten sich für Cohens Kappa Werte von -0,06 bis 0,71 und für Fleiss' Kappa Werte von -0,03 bis 0,32. Somit wurde größtenteils eine geringe bis ausreichende Übereinstimmung erzielt. In der Bewertung von Kombinationen war die IRR noch geringer.

Insgesamt zeigte sich also ein signifikanter Einfluss der Rater auf die Untersuchungsergebnisse. Je mehr Auswahlmöglichkeiten einer Diagnose oder Therapie zur Verfügung standen, desto schlechter war die IRR.

Zusammenfassend bedarf die Befundung der UF einer größeren Standardisierung, um die IRR und damit den Stellenwert der Untersuchung zu steigern. Diese Standardisierung

muss auf einzelnen dichotomen Entscheidungsprozessen beruhen. Die Implementierung einer Patterning-Software könnte den Prozess der Befundung weiter qualitativ steigern.

6. Abbildungsverzeichnis

Abbildung 1: Miktionstagebuch zur Dokumentation des Miktions- und Trinkverhaltens an 2 bis 3 Tagen durch den Patienten.	15
Abbildung 2: Auswertungsbogen zur standardisierten Erfassung der Verdachtsdiagnosen und Therapievorschlage der vier Rater.	23
Abbildung 3: bereinstimmung der Einzeldiagnosen der Rater nach Cohens und Fleiss' Kappa mittels Anamnese und Miktionstagebuch (modifiziert nach Jaekel et al., 2024).....	29
Abbildung 4: bereinstimmung aller Diagnosen der Rater nach Cohens und Fleiss' Kappa (modifiziert nach Jaekel et al., 2024).	30
Abbildung 5: bereinstimmung der Einzeldiagnosen der Rater nach Cohens und Fleiss' Kappa mittels Anamnese, Miktionstagebuch und Uroflowmetrie (modifiziert nach Jaekel et al., 2024).....	31
Abbildung 6: bereinstimmung der Einzeltherapievorschlage der Rater nach Cohens und Fleiss' Kappa (modifiziert nach Jaekel et al., 2024).	34

7. Tabellenverzeichnis

Tabelle 1: Übereinstimmung der Diagnosekombinationen nach Cohens Kappa zwischen den Ratern paarweise (modifiziert nach Jaekel et al., 2024).....	32
Tabelle 2: Übereinstimmung der Rater paarweise hinsichtlich der Therapie-Kombinationen (modifiziert nach Jaekel et al., 2024).....	35

8. Literaturverzeichnis

Adler I. The medical gap: intuition in medicine. *Med Health Care and Philos* 2022; 25: 361–369

Aharony SM, Lam O, Corcos J. Evaluation of lower urinary tract symptoms in multiple sclerosis patients: Review of the literature and current guidelines. *Canadian Urological Association journal = Journal de l'Association des urologues du Canada* 2017; 11: 61–64

Arevalo-Vega D, Ponce L, Valdevenito JP, Gallegos H, Dell'Oro A, Santis-Moya F, Calvo CI. 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

Blok B, Pannek J, Castro Díaz D, Del Popolo G. 2023. EAU Guidelines on Neuro-Urology. Verfügbar unter <https://uroweb.org/guidelines/neuro-urology/chapter/the-guideline> (Zugriffsdatum: 07.09.2023)

Çetinel B, Kocjancic E, Demirdağ Ç. Augmentation cystoplasty in neurogenic bladder. *Investigative and clinical urology* 2016; 57: 316–323

Çetinel B, Tarcan T, Demirkesen O, Özyurt C, Şen İ, Erdoğan S, Siva A. Management of lower urinary tract dysfunction in multiple sclerosis: a systematic review and Turkish consensus report. *Neurourology and urodynamics* 2013; 32: 1047–1057

Chang S-J, Yang SSD. Inter-observer and intra-observer agreement on interpretation of uroflowmetry curves of kindergarten children. *Journal of pediatric urology* 2008; 4: 422–427

Choo MS, Ryu HY, Lee S. Development of an Automatic Interpretation Algorithm for Uroflowmetry Results: Application of Artificial Intelligence. *International neurourology journal* 2022; 26: 69–77

Chow P-M, Hsiao S-M, Kuo H-C. Identifying occult bladder outlet obstruction in women with detrusor-underactivity-like urodynamic profiles. *Scientific reports* 2021; 11: 23242

Cooley LF, Kielb S. A Review of Botulinum Toxin A for the Treatment of Neurogenic Bladder. *PM & R : the journal of injury, function, and rehabilitation* 2019; 11: 192–200

Cornu JN, Gacci M, Hashim H. 2023. EAU Guidelines on Management of Non-neurogenic Male LUTS. Verfügbar unter uroweb.org/guidelines/management-of-non-neurogenic-male-luts (Zugriffsdatum: 06.09.2023)

Cruppé W de, Knesebeck O von dem, Gerstenberger E, Link C, Marceau L, Siegrist J, Geraedts M, McKinlay J. Hausärztliche entscheidungen bei symptomen für diabetes mellitus typ 2. *Deutsche medizinische Wochenschrift (1946)* 2011; 136: 359–364

Cruz F, Herschorn S, Aliotta P, Brin M, Thompson C, Lam W, Daniell G, Heesakkers J, Haag-Molkenteller C. Efficacy and safety of onabotulinumtoxinA in patients with urinary incontinence due to neurogenic detrusor overactivity: a randomised, double-blind, placebo-controlled trial. *European urology* 2011; 60: 742–750

Dillon BE, Lemack GE. Urodynamics in the evaluation of the patient with multiple sclerosis: when are they helpful and how do we use them? *The Urologic clinics of North America* 2014; 41: 439-44, ix

El Helou E, Sarkis J, Mjaess G, Zalaket J, Mouawad C, Sayegh N, Ghattas S, Azar C, El Helou J, Abboud H, Koussa S, Nemr E. Urodynamics in patients with multiple sclerosis: is it necessary? A randomized-controlled trial. *Scandinavian journal of urology* 2021; 55: 161–168

Faasse MA, Nosnik IP, Diaz-Saldano D, Hodgkins KS, Liu DB, Schreiber J, Yerkes EB. Uroflowmetry with pelvic floor electromyography: inter-rater agreement on diagnosis of pediatric non-neurogenic voiding disorders. *Journal of pediatric urology* 2015; 11: 198.e1-6

Fowler CJ, Panicker JN, Drake M, Harris C, Harrison SCW, Kirby M, Lucas M, Macleod N, Mangnall J, North A, Porter B, Reid S, Russell N, Watkiss K, Wells M. A UK consensus on the management of the bladder in multiple sclerosis. *Postgraduate medical journal* 2009; 85: 552–559

Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. *Disease-a-month : DM* 2003; 49: 53–70

Franco I, Franco J, Lee YS, Choi EK, Han SW. Can a quantitative means be used to predict flow patterns: Agreement between visual inspection vs. flow index derived flow patterns. *Journal of pediatric urology* 2016; 12: 218.e1-8

Gammie A, Drake MJ. The fundamentals of uroflowmetry practice, based on International Continence Society good urodynamic practices recommendations. *Neurourology and urodynamics* 2018; 37: S44-S49

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

Gärtner M, Krhut J, Zvarová K, Juráková M, Navrátilová M, Zvara P. Význam uroflowmetrie v diagnostice poruch dolních močových cest. *Ceska gynekologie* 2015; 80: 204–209

Giannantoni A, Scivoletto G, Di Stasi SM, Grasso MG, Agrò EF, Collura G, Vespasiani G. Lower urinary tract dysfunction and disability status in patients with multiple sclerosis. *Archives of Physical Medicine and Rehabilitation* 1999; 80: 437–441

Griffiths DJ, Scholtmeijer RJ. Place of the free flow curve in the urodynamic investigation of children. *British journal of urology* 1984; 56: 474–477

Gust K, Bartsch G, Haferkamp A. Funktionsstörungen des unteren Harntraktes. In: Hautmann R, Gschwend JE, Hrsg. *Urologie*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014: 111–130

Haensch et al... Diagnostik und Therapie von neurogenen Blasenstörungen. In: Michel MS, Thüroff JW, Janetschek G, Wirth MP, Hrsg. *Die Urologie*. Berlin: Springer, 2016

Haensch et al. 2020. Diagnostik und Therapie von neurogenen Blasenstörungen, S1-Leitlinie. Verfügbar unter <https://register.awmf.org/de/leitlinien/detail/030-121> (Zugriffsdatum: 05.04.2024)

Hautmann R. Operative urologische Therapie. In: Hautmann R, Gschwend JE, Hrsg. *Urologie*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014: 95–110

Hemmer B. et al. 2022. LL 030 050 Diagnose und Therapie der Multiplen Sklerose, Neuromyelitis-optica-Spektrum-Erkrankungen und MOG-IgG-assoziierten Erkrankungen - Deutsche Gesellschaft für Neurologie & e.V. Verfügbar unter <https://dgn.org/leitlinien/II-030-050-diagnose-und-therapie-der-multiplen-sklerose-neuromyelitis-optica-spektrum-erkrankungen-und-mog-igg-assoziierten-erkrankungen/> (Zugriffsdatum: 05.08.2022)

Hemmer B. et al. 2023. Diagnose und Therapie der Multiplen Sklerose, Neuromyelitis-optica-Spektrum-Erkrankungen und MOG-IgG-assoziierten Erkrankungen, S2k-Leitlinie (Zugriffsdatum: 07.09.23)

Henze T, Feneberg W, Flachenecker P, Seidel D, Albrecht H, Starck M, Meuth SG. Neues zur symptomatischen MS-Therapie: Teil 3 – Blasenfunktionsstörungen. *Der Nervenarzt* 2018; 89: 184–192

Jacq C, Hubeaux K, Ramanantsitonta J. Sclérose en plaques et autosondage intermittent. *Progres en urologie : journal de l'Association française d'urologie et de la Société française d'urologie* 2021; 31: 195–203

Jaekel AK, Rieger J, Butscher A-L, Möhr S, Schindler O, Queissert F, Hofmann A, Schmidt P, Kirschner-Hermanns R, Knüpfer SC. Diagnoses and Treatment Recommendations- Interrater Reliability of Uroflowmetry in People with Multiple Sclerosis. *Biomedicines* 2024; 12

Jaekel AK, Winterhagen FI, Zeller FL, Butscher A-L, Knappe FK, Schmitz F, Hauk C, Stein J, Kirschner-Hermanns RKM, Knüpfer SC. Neurogenic Lower Urinary Tract Dysfunction in Asymptomatic Patients with Multiple Sclerosis. *Biomedicines* 2022; 10

Jeffery N, Mundy A. Innovations in indwelling urethral catheterisation. *BJU international* 2020; 125: 664–668

Jong Y de, Pinckaers JHFM, Brinck RM ten, Lycklama à Nijeholt AAB, Dekkers OM. Urinating standing versus sitting: position is of influence in men with prostate enlargement. A systematic review and meta-analysis. *PloS one* 2014; 9: e101320

Kanematsu A, Tanaka S, Johnin K, Kawai S, Nakamura S, Imamura M, Yoshimura K, Higuchi Y, Yamamoto S, Okada Y, Nakai H, Ogawa O. A multi-center study of pediatric uroflowmetry data using patterning software. *Journal of pediatric urology* 2013; 9: 57–61

Karl C, Gerlach R, Hannappel J, Lehnen H. Uroflow measurements: their information yield in a long-term investigation of pre- and postoperative measurements. *Urologia internationalis* 1986; 41: 270–275

Kessler TM, Wöllner J, Kozomara M, Mordasini L, Mehnert U. Sakrale Neuromodulation bei neurogenen Blasenfunktionsstörungen. *Der Urologe. Ausg. A* 2012; 51: 179–183

- Khan F, Pallant JF, Pallant JI, Brand C, Kilpatrick TJ. A randomised controlled trial: outcomes of bladder rehabilitation in persons with multiple sclerosis. *Journal of Neurology, Neurosurgery & Psychiatry* 2010; 81: 1033–1038
- Kirschner-Hermanns R. Allgemeine Diagnostik bei neurologischen Blasenfunktionsstörungen. In: Michel MS, Thüroff JW, Janetschek G, Wirth MP, Hrsg. *Die Urologie*. Berlin: Springer, 2016: 1019–1023
- Klingler HC, Dietersdorfer F, Fink KG, Fischer M, Heidler H, Huber J, Lüftenegger W, Madersbacher H, Petrovic Z, Primus G, Schrey A. Leitlinie: Der intermittierende Katheterismus
- Koch-Henriksen N, Magyari M. Apparent changes in the epidemiology and severity of multiple sclerosis. *Nature reviews. Neurology* 2021; 17: 676–688
- Konstantinidis C, Kratiras Z, Samarinas M, Skriapas K. Optimal bladder diary duration for patients with suprapontine neurogenic lower urinary tract dysfunction. *International braz j urol : official journal of the Brazilian Society of Urology* 2016; 42: 766–772
- Kurze I, Jaekel AK. Neuro-urologische Diagnostik und Therapie nicht traumatischer/degenerativ bedingter neurogener Dysfunktion des unteren Harntraktes am Beispiel der Multiplen Sklerose. *Aktuelle Urologie* 2024; 55: 315–325
- Kutzenberger, Domurath, KirschnerHermanns, Möhr, Pretzer, Soljanik. 043-053m_S2k_Medikamentoese-Therapie-neurogene-Dysfunktion-des-unteren-Harntraktes-NLUTD_2022-02 2022
- Lamin E, Newman DK. Clean intermittent catheterization revisited. *International urology and nephrology* 2016; 48: 931–939
- Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; 33: 159–174
- Lee KS, Song PH, Ko YH. Does uroflowmetry parameter facilitate discrimination between detrusor underactivity and bladder outlet obstruction? *Investigative and clinical urology* 2016; 57: 437–441

Liu Y-B, Yang SS, 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. *Lower urinary tract symptoms* 2014; 6: 76–80

Lúcio A, D'ancona CAL, Perissinotto MC, McLean L, Damasceno BP, Moraes Lopes MHB de. Pelvic Floor Muscle Training With and Without Electrical Stimulation in the Treatment of Lower Urinary Tract Symptoms in Women With Multiple Sclerosis. *Journal of wound, ostomy, and continence nursing : official publication of The Wound, Ostomy and Continence Nurses Society* 2016; 43: 414–419

Madersbacher H. Orale Anticholinergika bei überaktiver Blase. *Der Urologe. Ausg. A* 2006; 45: 830–834

Magee C, Kazman J, Haigney M, Oriscello R, DeZee KJ, Deuster P, Depenbrock P, O'Connor FG. Reliability and validity of clinician ECG interpretation for athletes. *Annals of Noninvasive Electrocardiology : The Official Journal of the International Society for Holter and Noninvasive Electrocardiology, Inc* 2014; 19: 319–329

Margreiter M, Shariat S. Benigne Prostatahyperplasie und benignes Prostatasyndrom. In: Michel MS, Thüroff JW, Janetschek G, Wirth MP, Hrsg. *Die Urologie*. Berlin: Springer, 2016: 1279–1288

Mattle Hea. 5.1 Multiple Sklerose. In: Mattle H, Mumenthaler M, Hrsg. *Neurologie*. Stuttgart: Georg Thieme Verlag, 2013

Mattsson S, Spångberg A. Urinary flow in healthy schoolchildren. *Neurourol. Urodyn.* 1994; 13: 281–296

McHugh ML. Interrater reliability: the kappa statistic. *Biochemia medica* 2012; 22: 276–282

McKinlay J, Link C, Marceau L, O'Donnell A, Arber S, Adams A, Lutfey K. How do doctors in different countries manage the same patient? Results of a factorial experiment. *Health services research* 2006; 41: 2182–2200

McKinlay JB, Lin T, Freund K, Moskowitz M. The unexpected influence of physician attributes on clinical decisions: results of an experiment. *Journal of health and social behavior* 2002; 43: 92–106

Meyer-Moock S, Feng Y-S, Maeurer M, Dippel F-W, Kohlmann T. Systematic literature review and validity evaluation of the Expanded Disability Status Scale (EDSS) and the Multiple Sclerosis Functional Composite (MSFC) in patients with multiple sclerosis. *BMC neurology* 2014; 14: 58

Mumenthaler M. *Neurologie*. Stuttgart: Thieme 2008

Musco S, Ecclestone H, Hoen L 't, Blok BFM, Padilla-Fernández B, Del Popolo G, Groen J, Pannek J, Kessler TM, Karsenty G, Phé V, Sartori AM, Castro-Diaz D, Rizwan H. Efficacy and Safety of Surgical Treatments for Neurogenic Stress Urinary Incontinence in Adults: A Systematic Review. *European urology focus* 2022; 8: 1090–1102

Nabi G, Cody JD, Ellis G, Herbison P, Hay-Smith J. Anticholinergic drugs versus placebo for overactive bladder syndrome in adults. *The Cochrane database of systematic reviews* 2006; 2006: CD003781

Netto JMB, Hittelman A, Lambert S, Murphy K, Collette-Gardere T, Franco I. Interpretation of uroflow curves: A global survey measuring inter and intra rater reliability. *Neurourology and urodynamics* 2020; 39: 826–832

Palmtag, Goepel, Heidler. *Urodynamik*. Berlin: Springer 2004

Panicker JN, Fowler CJ, Kessler TM. Lower urinary tract dysfunction in the neurological patient: clinical assessment and management. *The Lancet Neurology* 2015; 14: 720–732

Primus G, Heidler H. Leitlinien BLasenfunktionsstörungen. *Journal für Urologie und Urogynäkologie* 2003

Rached L, Lacour B, Daudon M. Lithiase urinaire et sclérose en plaques. *Progres en urologie : journal de l'Association française d'urologie et de la Société française d'urologie* 2011; 21: 102–108

Rädiker S, Kuckartz U. Intercoder-Übereinstimmung analysieren. In: Rädiker, Hrsg. *Analyse qualitativer Daten mit MAXQDA*. Wiesbaden: Springer Fachmedien Wiesbaden, 2019: 287–303

Rosier PFWM, Schaefer W, Lose G, Goldman HB, 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

Sakakibara R. Neurogenic lower urinary tract dysfunction in multiple sclerosis, neuromyelitis optica, and related disorders. *Clinical autonomic research : official journal of the Clinical Autonomic Research Society* 2019; 29: 313–320

Scheiner D, Perucchini D, Fink D. Überaktive Blase: Möglichkeiten und Grenzen der Botoxtherapie. *Gynäkologisch-geburtshilfliche Rundschau* 2006; 46: 88–95

Schmid F, Mack M, Sulser T. Die Harnröhrenstriktur des Mannes. *Urol. Prax.* 2020; 22: 93–99

Schneider MP, Tornic J, Sýkora R, Abo Youssef N, Mordasini L, Krhut J, Chartier-Kastler E, Davies M, Gajewski J, Schurch B, Bachmann LM, Kessler TM. Alpha-blockers for treating neurogenic lower urinary tract dysfunction in patients with multiple sclerosis: A systematic review and meta-analysis. A report from the Neuro-Urology Promotion Committee of the International Continence Society (ICS). *Neurourology and urodynamics* 2019; 38: 1482–1491

Schultz-Lampel D, Goepel M, Hampel C. *Urodynamik*: Springer 2022

Smith PP, Hurtado EA, Appell RA. Post hoc interpretation of urodynamic evaluation is qualitatively different than interpretation at the time of urodynamic study. *Neurourol. Urodyn.* 2009; 28: 998–1002

Stoffel JT. Chronic Urinary Retention in Multiple Sclerosis Patients: Physiology, Systematic Review of Urodynamic Data, and Recommendations for Care. *The Urologic clinics of North America* 2017; 44: 429–439

Valentini F-A, Nelson P-P. Méthodes d'urodynamique non invasive chez l'homme. *Annales de readaptation et de medecine physique : revue scientifique de la Societe francaise de reeducation fonctionnelle de readaptation et de medecine physique* 2006; 49: 187–196

van Batavia JP, Pohl HG, Farhat WA, Chiang G, BaniHani A, Collett-Gardere T, Franco I. Is it time to reconsider how we document pediatric uroflow studies?: A study from the SPU Voiding Dysfunction task force. *Journal of pediatric urology* 2023; 19: 546–554

- van de Beek C, Stoevelaar HJ, McDonnell J, Nijs H, Casparie AF, Janknegt RA. Interpretation of Uroflowmetry Curves by Urologists. *The Journal of urology* 1997; 157: 164–168
- van den Brink N, Holbrechts B, Brand PLP, Stolper ECF, van Royen P. Role of intuitive knowledge in the diagnostic reasoning of hospital specialists: a focus group study. *BMJ open* 2019; 9: e022724
- van Ophoven A. Sakrale Neuromodulation bei refraktärer überaktiver Blase. *Der Urologe. Ausg. A* 2018; 57: 1375–1388
- Vanstone M, Monteiro S, Colvin E, Norman G, Sherbino J, Sibbald M, Dore K, Peters A. Experienced physician descriptions of intuition in clinical reasoning: a typology. *Diagnosis (Berlin, Germany)* 2019; 6: 259–268
- Verpoorten C, Buyse GM. The neurogenic bladder: medical treatment. *Pediatric nephrology (Berlin, Germany)* 2008; 23: 717–725
- Vijverberg MAW, Klijn AJ, Rabenort A, Bransen J, Kok ET, Wingens JPM, Jong TPVM de. A comparative analysis of pediatric uroflowmetry curves. *Neurourology and urodynamics* 2011; 30: 1576–1579
- Weih M, Roßnagel F, Dikow H, Wehrle K, Braune S, Bergmann A. Daten zur Multiplen Sklerose in Deutschland und ihre Abbildung im Register des ambulanten Netzwerkes NeuroTransData (NTD). *Fortschritte der Neurologie-Psychiatrie* 2020; 88: 379–385
- Wirtz M, Kutschmann M. Analyse der Beurteilerübereinstimmung für kategoriale Daten mittels Cohens Kappa und alternativer Masse. *Die Rehabilitation* 2007; 46: 370–377
- Wongpakaran N, Wongpakaran T, Wedding D, Gwet KL. A comparison of Cohen's Kappa and Gwet's AC1 when calculating inter-rater reliability coefficients: a study conducted with personality disorder samples. *BMC medical research methodology* 2013; 13: 61
- Zhao X, Feng GC, Ao SH, Liu PL. Interrater reliability estimators tested against true interrater reliabilities. *BMC medical research methodology* 2022; 22: 232
- Zurawski J, Stankiewicz J. Multiple Sclerosis Re-Examined: Essential and Emerging Clinical Concepts. *The American journal of medicine* 2018; 131: 464–472

9. Erklärung zum Eigenanteil

Die Arbeit wurde in der Klinik und Poliklinik für Urologie und Kinderurologie, Universitätsklinikum Bonn (Herr Prof. Dr. Manuel Ritter) unter Betreuung von Frau Prof. Dr. Ruth Kirschner-Hermanns am Johanniter Neurologischen Rehabilitationszentrum Godeshöhe realisiert.

Die Konzeption und Umsetzung der Arbeit erfolgte unter Anleitung von Frau Dr. med. Anke Jaekel, Oberärztin an beiden oben genannten Einrichtungen.

Die studienbezogene, prospektive Auswertung der Datensätze zum Zwecke dieser Arbeit wurde durch folgende Fachärzte durchgeführt: Dr. med. Sandra Möhr, Dr. med. Oliver Schindler, Dr. med. Fabian Queissert, Dr. med. Anke Jaekel.

Das zur Auswertung verwendete Datenmaterial der Datenbank wurde mit Unterstützung von Frau Anna-Lena Butscher zusammengestellt.

Die statistische Auswertung wurde durch Dr. Paul Schmidt, Diplom Statistiker, unterstützt.

Ich versichere, die Dissertationsschrift selbständig verfasst zu haben und keine weiteren als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

10. Veröffentlichung

Teile der Inhalte der vorliegenden Dissertationsschrift wurden bereits in der folgenden Publikation veröffentlicht: „Diagnoses and Treatment Recommendations - Interrater Reliability of Uroflowmetry in People with Multiple Sclerosis“

Jaekel AK, Rieger J, Butscher A-L, Möhr S, Schindler O, Queissert F, Hofmann A, Schmidt P, Kirschner-Hermanns R, Knüpfer SC. Diagnoses and Treatment Recommendations- Interrater Reliability of Uroflowmetry in People with Multiple Sclerosis. *Biomedicines* 2024; 12

<https://doi.org/10.3390/biomedicines12071598>

11. Anmerkung

Die Arbeit wird durch den Förderverein zur Kontinenzforschung und Kontinenzaufklärung e. V.; Karmeliterhöfe, Karmeliterstr. 10, 52064 Aachen unterstützt.

12. Danksagung

An dieser Stelle möchte ich meine aufrichtige Dankbarkeit gegenüber all jenen zum Ausdruck bringen, die mich bei der Anfertigung dieser Dissertation unterstützt haben.

Mein besonderer Dank gilt Frau Dr. med. Anke Jaekel für die hervorragende Betreuung, fachliche Expertise und das stets offene Ohr bei Fragen oder Schwierigkeiten. Ebenso Frau Prof. Dr. med. Ruth Kirschner-Hermanns, die mich als Doktormutter vertrauensvoll und zuverlässig begleitet hat.

Frau Anna-Lena Butscher danke ich für die Unterstützung bezüglich der Zusammenstellung der Patienteninformationen und der Datenbank.

Abschließend möchte ich mich bei meiner Familie für den starken Rückhalt und die unermüdliche Begleitung während der gesamten Zeit bedanken.