UHASSELT

KNOWLEDGE IN ACTION

Faculteit Revalidatiewetenschappen

master in de revalidatiewetenschappen en de kinesitherapie

Masterthesis

Between-day reliability and diagnostic accuracy of outcome measures for pain in persons with Multiple Sclerosis

Joni Lemmens
Lotte Vanbuel

Scriptie ingediend tot het behalen van de graad van master in de revalidatiewetenschappen en de kinesitherapie, afstudeerrichting revalidatiewetenschappen en kinesitherapie bij neurologische aandoeningen

PROMOTOR :

Prof. dr. Peter FEYS

BEGELEIDER :

Mevrouw Cigdem YILMAZER

UHASSELT

KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

2023
2024

UHASSELT

KNOWLEDGE IN ACTION

Faculteit Revalidatiewetenschappen

master in de revalidatiewetenschappen en de kinesitherapie

Masterthesis

Between-day reliability and diagnostic accuracy of outcome measures for pain in persons with Multiple Sclerosis

Joni Lemmens
Lotte Vanbuel

Scriptie ingediend tot het behalen van de graad van master in de revalidatiewetenschappen en de kinesitherapie, afstudeerrichting revalidatiewetenschappen en kinesitherapie bij neurologische aandoeningen

PROMOTOR :

Prof. dr. Peter FEYS

BEGELEIDER :

Mevrouw Cigdem YILMAZER

ACKNOWLEDGEMENTS

We want to thank all those who contributed to the successful completion of our master thesis. Drs. Cigdem Yilmazer, our supervisor, deserves special mention for her guidance, support, and insightful feedback throughout this journey. We also express our appreciation to Prof. Dr. Peter Feys, our promotor, for his oversight of the study and feedback during the final stages of thesis refinement.

To our beloved friends, families and supporters, we are deeply grateful for their encouragement and constructive feedback. Additionally, we extend our thanks to the REVAL Center at the Faculty of Rehabilitation Sciences, University of Hasselt, the Rehabilitation and MS Center in Pelt, the National Multiple Sclerosis Centre (NMSC) Melsbroek, Fit Up in Kontich and AZ Sint-Jan in Brugge for generously providing facilities for our research.

Finally, we extend our sincerest appreciation to all the participants who volunteered to be part of this study. Without their willingness and cooperation, this research would not have been possible.

RESEARCH CONTEXT

This master's thesis is in the Multiple Sclerosis (MS) research domain within the Faculty of Rehabilitation Sciences and Physiotherapy at Hasselt University. Specifically, the study focuses on pain-related outcome measures' reliability and diagnostic accuracy in MS patients. Our research question was discussed and determined in consultation with our supervisor and mentor.

As will be further elaborated, pain is a common symptom in MS, yet the assessment of pain is still a significant challenge. Numerous (mostly partly subjective) tools are available for assessing pain. However, it remains unclear which measurement instruments are the most reliable. Hence, the aim of this research is to demonstrate to clinicians which pain outcome measures are the most reliable for clinical application and to investigate fluctuations in pain perception within individuals over a week. Understanding how pain perception varies depending on the day, location, and different pain types is crucial for improving pain management strategies in individuals with MS.

This master's thesis aligns with an ongoing doctoral study at Hasselt University under the supervision of Drs. Cigdem Yilmazer. The research question guiding her study is: "Reliability and validity of pain outcome measures in persons with Multiple Sclerosis." (Project number: B1152022000014). The approval date by the ethics committee is 19/12/2022. The overarching principal investigator of both studies is Professor Dr. Peter Feys.

Patient recruitment took place in Belgium and Chile and was conducted in collaboration with several partners. In Belgium, there was a collaboration for recruitment with the Rehabilitation and MS Center in Pelt, the National MS Center in Melsbroek, the private practice Fit Up in Kontich and AZ Sint-Jan Brugge. In Chile, there was a collaboration with Hospital Dr. Sotero Del Rio.

This master’s thesis was performed during the final year of the master's program at the University of Hasselt, situated at the campus in Diepenbeek. Lotte Vanbuel and Joni Lemmens, both master's students at the University of Hasselt, undertook the authorship of this master's thesis under the guidance of Drs. Cigdem Yilmazer and promotor, Prof. Dr. Peter Feys. Both students actively participated in participant recruitment, data collection, data verification, and comprehensive data analysis. Together, they collaborated on statistical analysis and writing, with periodic evaluations by Drs. Cigdem Yilmazer. The final version underwent a thorough evaluation by Prof. Dr. Peter Feys as well.

ABSTRACT

Background: Pain is a common symptom in Multiple Sclerosis (MS), affecting quality of life and functional abilities. Reliable assessment tools are essential for accurate pain management in this population. However, the between-day reliability and accuracy of pain questionnaires in MS remains underexplored.

Objectives: This study investigates the between-day reliability of patient-reported outcome measures for pain and diagnostic accuracy of screening tools in MS.

Methods: Between-day reliability was evaluated using the intra-class correlation (ICC) and standard error of measurement (SEM). Diagnostic accuracy of the DN4 and PainDETECT was assessed using sensitivity, specificity, positive and negative predictive values (PPV, NPV), positive and negative likelihood ratios (LR+, LR–), Area Under the Curve (AUC) and Cohen's Kappa.

Results: Our findings suggest promising between-day reliability for the pain questionnaires in persons with MS (PwMS). ICC values indicate good to excellent reliability for most measures, with minimal systematic bias observed in Bland-Altman plots. Regarding diagnostic accuracy, both DN4 and PainDETECT exhibited limited clinical utility. These findings underscore the importance of incorporating clinical evaluation by healthcare professionals alongside questionnaire-based assessments for accurate pain diagnosis in PwMS.

Conclusion: The study highlights the importance of comprehensive pain assessment; while all outcome measures seem to be reliable for valuable information, it concludes that tools like DN4 and PainDETECT are useful but still not accurate enough to diagnose pain independently, stressing the ongoing significance of physician evaluation in pain diagnosis.

Keywords: Multiple sclerosis, pain, outcome measures, between-day reliability, accuracy

INTRODUCTION

Multiple Sclerosis (MS) is a neurodegenerative and autoimmune disease of the central nervous system that results in impairments of the related region, such as spasticity, sensory deficits, muscle weakness, fatigue, incontinence, and pain. (Thompson J.A., 2018) The prevalence of pain in persons with MS (PwMS) can vary widely depending on the course and stage of the disease, individual differences, and the specific types of pain experienced. In literature, it can be found that pain in PwMS has an overall prevalence of 63%, which can vary from 29% to 86% (Yilmazer C., 2020).

The International Association for the Study of Pain (IASP) defines pain as: "An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage". Kratz A. et al. conducted research on the prevalence and pathophysiology of pain in PwMS and they distinguish the following chronic pain types: nociceptive pain (41%), a mix of neuropathic and nociplastic pain (27%), nociplastic pain (23%) and neuropathic pain (9%). (Kratz A., 2021). The IASP defines nociceptive pain as "Pain that arises from actual or threatened damage to non-neural tissue and is due to the activation of nociceptors." Nociceptive pain is typically described as a sharp, aching, or throbbing pain, and it is often well-localized to the area of tissue damage. The pain serves as a protective mechanism, alerting the individual to potential harm, and it is usually proportional to the extent of tissue injury. Apart from the high prevalence of nociceptive pain in the sample size of the study by Kratz et al, the primary focus is still mainly on neuropathic pain in MS patients. Neuropathic pain is described by the IASP as: 'Pain arising directly from a lesion or disease affecting the somatosensory system.' Two types of neuropathic pain can be distinguished based on the location of the symptoms: peripheral neuropathic and central neuropathic pain. Symptoms often related to neuropathic pain are burning, electrical shock-like, tingling or shooting sensations. Nociplastic pain refers to pain that arises from altered nociception without clear evidence of actual or threatened tissue damage causing the sensation. (Buldys et al., 2023)

Measuring pain presents a unique challenge due to its subjective nature, but patient-reported outcome measures for pain provide valuable information. While the Neuropathic Pain Scale (NPS) (Rog J. et al., 2007) and the Brief Pain Inventory (BPI) (Osborne L.T, 2006) have been validated for use within the MS population, the reliability of other questionnaires such as the Douleur Neuropathique 4 (DN4), PainDETECT, Neuropathic Pain Symptom Inventory (NPSI), and Nordic Musculoskeletal Questionnaire remains uncertain for PwMS.

The choice to determine specifically between-day reliability is based on the fluctuating nature of pain, the between-day variability. Another research by Kratz A. et al supported this principle and concluded that the chronic symptoms that occur in multiple sclerosis (pain, fatigue, ...) are not a static phenomenon but that these symptoms can change daily. (Kratz A., 2017). In MS, the fluctuations in reported pain levels are influenced by various factors, including the type of pain, the nature of the assessment questions, and the specific dimensions of pain being measured. Nociplastic pain is linked to higher pain intensity and more frequent chronic overlapping pain conditions. For all pain types, there was a high frequency of pain medication usage, yet relief from these medications ranged from poor to modest. While non-steroidal anti-inflammatory drug (NSAID) use was most common among individuals with nociplastic pain (80%), the highest pain relief ratings for NSAIDs were observed in those with nociceptive pain. These findings emphasize the importance of a multidimensional approach to pain assessment in MS, with a focus on identifying specific pain phenotypes. A more detailed understanding of pain as a complex condition in MS could lead to better therapeutic strategies. (Kratz A., 2021)

Although pain is one of the most disabling symptoms of MS, it is an invisible symptom. It has a major impact on the quality of life (QoL), disability level, employment and mental health. As discussed by Amtmann D. et al., pain in MS patients has been linked to heightened levels of anxiety, depression, and disruptions in sleep patterns. These symptoms, in turn, contribute to the possible development of depression. Additionally, emerging evidence suggests that depression may play a role in exacerbating MS symptoms, including pain, although the precise cause-effect relationship remains unclear. Hence, pain symptoms must be addressed as quickly and correctly as possible to avoid escalating. (Amtmann D., 2018)

METHODS

Participants

Recruitment

The sample size of this study was based on the Consensus-based standards for the selection of health measurement instruments (COSMIN) Study Design checklist for patient-reported outcome measurement instruments (Mokkink et al., 2010). This guideline reflects that at least 50 participants are needed to make an adequate conclusion about questionnaire reliability (Mokkink et al., 2019). Based on this guideline, it was aimed to include at least 50 participants per country.

For this study, patient recruitment took place across multiple locations, both in Belgium and in Chile. In Belgium, participants were recruited from various rehabilitation centers, including Rehabilitations and MS Center Noorderhart in Pelt, National MS Center Melsbroek, Fit-up Physiotherapy Center in Kontich and AZ Sint-Jan Brugge. Furthermore, patients were also recruited via social media announcements, supported by a short video providing all the necessary information. Additionally, patient recruitment extended beyond Belgium to include participants from Chile at the Hospital Dr. Sotero Del Rio.

Participant selection

PwMS were included when they (a) had an age of 18 or older, (b) they had to be diagnosed with MS (c) with an Expanded Disability Status Scale (EDSS) of 6 or below and (d) they had to have experienced pain in the past month. (e) They also had to be able to understand and answer the questionnaires.

Exclusion criteria were (a) being diagnosed with a major musculoskeletal disorder or any other neurological disorder than MS, (b) a relapse in the month prior to the first testing and (c) if a headache is the only pain they experience.

All participants received written information and signed the informed consent form before the first testing. The study protocol was approved by the Medical Ethical Committee of University of Hasselt on 19/12/2022 as well as by the local ethical committees of the recruitment centers and was also registered at the ClinicalTrials.gov website (clinical trial number: NCT05742152).

Study-design and procedure

Every participant attended two sessions and both descriptive data as pain-related data were collected. There had to be an interval of three to eight days between the two moments of measurement. The collection of descriptive data (mental health, fatigue, sleep, spasticity, stress and activities of daily living) was divided between the first and second sessions. During the two sessions, pain measures were repeated.

Table 1
Study Design

Session 1 (± 60 minutes)	Session 2 (± 60 minutes)
Descriptive outcome measures: IPAQ, MAS, MI, Pittsburg Sleep, MHI, HADS, mFIS, Perceived Stress Scale, 9HPT, SDMT, T25FW.	**Descriptive outcome measures:** Collecting incomplete descriptive results of the first session.
Pain-related outcome measures: Algometer, DN4, PainDETECT, Nordic-Musculoskeletal Questionnaire, NPSI, NPS, BPI-SF.	**Pain-related outcome measures:** Algometer, DN4, PainDETECT, Nordic-Musculoskeletal Questionnaire, NPSI, NPS, BPI-SF.

Abbreviations : IPAQ, International Physical Activity Questionnaire ; MAS, Modified Ashworth Scale ; MI, Motricity Index ; MHI, Mental Health Inventory ; HADS, Hospital Anxiety and Depression Scale ; mFIS, modified Fatigue Impact Scale ; 9HPT, Nine Hole Peg Test ; SDMT, Symbol Digit modality test ; T25FW, Timed 25 foot Walk ; DN4, Douleur Neuropathique 4 ; NPSI, Neuropathic Pain Symptom Inventory ; NPS, Neuropathic Pain Scale ; BPI-SF, Brief Pain Inventory - Short form

Descriptive tests

During the first measurement, some characteristics of the patients were questioned: age, gender, MS type, duration of the disease, education level, employment status, EDSS score, and pain medication. In addition to this, the following descriptive tests and questionnaires were conducted:

Nine Hole Peg Test - 9HPT

According to Feys P. et al. (2016) the Nine Hole Peg Test is regarded as a gold standard for assessing manual dexterity in PwMS. The patient must take 9 pegs from a container as quickly as possible and insert them into the openings on the pegboard. The patient must then remove the pegs and place them back in the container. The duration of the action is measured and is performed twice with the dominant hand and twice with the non-dominant hand. It is important that the patient does not take the pegs all at once, but one by one.

Symbol Digit Modalities Test - SDMT

The Symbol Digit Modality Test is an assessment tool used to evaluate cognitive function, specifically processing speed. The oral version of the SDMT was used. The test usually consists of a key or legend and a response sheet. The key contains a list of symbols paired with corresponding numbers. Participants were asked to couple as many symbols of the rows to a number corresponding to the 'key' in 90 seconds. The number of correct answers after 90 seconds was noted. A higher score indicates better cognitive processing speed and attention. (Benedict R. et al., 2017)

Timed 25 Foot Walk - T25FW

The T25FW is a test used to gain insight into the mobility and walking speed in individuals with neurological disorders, particularly Multiple Sclerosis. The person being tested is instructed to stand at the starting point. On the testers' signal, the individual is asked to walk 25 feet as quickly, but as safely, as possible. The timer begins as soon as the person begins to walk and is stopped as soon as any part of the person's body crosses the 25-foot mark. Two trials were performed of which the time was noted in seconds. In accordance with the guidelines of the National MS Society, we established a maximum value of 180 seconds for the test. The mean of the two trials was calculated and a shorter time indicates better mobility and walking speed, while a longer time suggests slower mobility. (Kalinowski A., 2022)

International Physical Activity Questionnaire - IPAQ

The IPAQ is a measurement tool to assess the intensity of physical activity and the amount of sitting hours that is part of the participants' daily life. The scoring is based on the total physical activity, expressed as MET-min/week and the hours spent sitting. Based on the total score the following cut-offs are used to distinguish between categories of physical activity: low activity (<600 MET-minutes/week), moderate activity (600-1500 MET-minutes/week) and high activity (>1500 MET-minutes/week). (IPAQ, 2005)

Modified Ashworth Scale - MAS

This measurement tool was developed to assess the presence of resistance against passive movement of an extremity. The following muscle groups were tested: elbow flexors and extensors, wrist-flexors, hipflexors, knee-extensors and plantarflexors. The scoring according to Bohannon and Smith (1987) was used.

Table 2
The Modified Ashworth Scale Scoring by Bohannon and Smith

Score	Definition
0	No increase in tone
1	Slight increase in muscle tone, manifested by a catch and release, or by minimal resistance at the end of the range of motion when the affected part(s) is moved into flexion and extension
1+	Slight increase in muscle tone, manifested by a catch, followed by minimal resistance throughout the remainder (less than half) of the ROM
2	More marked increase in muscle tone through most of the ROM, but affected part(s) easily moved
3	Considerable increase in muscle tone, passive movement difficult
4	Affected part(s) rigid in flexion or extension

Note. Reprinted from W. Bohannon, R. & B. Smith, M, 1987
Abbrevation: ROM, Range of Motion

Motricity Index - MI

The Motricity Index is mostly used post-stroke, but it can be applied in the MS population too. In this context, it was used to evaluate voluntary movements and isometric muscle strength. The individual being evaluated is asked to perform a series of standardized movements such as: pinch grip, elbow flexion, shoulder abduction, ankle dorsiflexion, knee extension and hip flexion, all in seated position. The score for the muscle strength during each movement could vary between 0, 9, 14, 19, 25 and 33 depending on the patient's ability to generate muscle strength during the specified movements. The scores for individual muscle groups are added up to obtain a total Motricity Index score.

Pittsburgh Sleep Quality Index - PSQI

PwMS commonly experience sleep disturbances, which can exacerbate pain and affect overall well-being. The PSQI is a validated tool for evaluating various components of sleep quality, such as sleep duration, latency, efficiency, disturbances, and daytime dysfunction. (Jerkovic A., 2022)

Hospital Anxiety and Depression Scale - HADS

The Hospital Anxiety and Depression Scale (HADS) is designed to evaluate symptoms of anxiety and depression among patients within medical settings. Comprising 14 questions, the questionnaire is divided evenly, with seven items dedicated to probing anxiety-related symptoms and the remaining seven aimed at assessing depression. (Pais-Ribeiro, J.L., 2018) For accurate scoring, it is imperative to summate the scores for anxiety and depression separately. The following cut-off values are used for interpreting the subscores: a score lower than 7 indicates the absence of an anxiety disorder or depression, while a score ranging from 8 to 10 suggests mild symptoms of anxiety or depression, scores of 11 to 14 indicate moderate anxiety/depressive symptoms and scores of 15 to 21 suggest severe anxiety/depressive symptoms. (Stern, F.A., 2014)

Modified Fatigue Impact Scale - mFIS

PwMS often experience fatigue symptoms, and there exists also a phenomenon called "MS-related fatigue". Unlike 'normal' fatigue, MS-related fatigue is triggered by minor exertion, exacerbated by heat and it often impacts physical functioning. Because this type of fatigue is highly fluctuating and subjective, it is a challenge to identify. Since the MFIS is also included in the Multiple Sclerosis Quality of Life Inventory, it is often used in clinical practice. The MFIS examines the impact of fatigue on physical, cognitive and psychological functioning by asking 21 questions. (Kos D., 2003)

Perceived Stress Scale - PSS

The Perceived Stress Scale was developed to measure the extent to which situations in a person's life are perceived as stressful. The ten questions in this scale ask about feelings and thoughts during the last month. This test was used to get an idea of how stress can affect the patient's pain and other symptoms. (Wu, S.M., 2013) The following interpretation applies to scores on the PSS: scores ranging from 0 to 13 indicate low levels of stress, scores between 14 and 26 suggest moderate levels of stress, and scores above 27 (up to 40) indicate high perceived stress. (Lee, 2012)

Pain outcome measures

Algometer

To ensure standardized and reliable measurements, the protocol for pressure pain threshold (PPT) assessment was established in consultation with an experienced researcher at the University of Hasselt, with the investigator receiving comprehensive training beforehand.

All participants underwent testing following a uniform protocol, with each test conducted at predetermined anatomical reference points sourced from relevant literature. (Fischer, A.A. (1987) The algometer was applied bilaterally and perpendicularly at specific sites, including the muscle belly of the trapezius, thumbnails, low back (5 cm lateral of L5), and quadriceps muscle. These sites were chosen due to their relevance to the multifocal nature of pain often observed in individuals with MS.

Before initiating the protocol, participants received detailed instructions on the measurement procedure, and a trial run was performed on the palmar hands for familiarization. Pressure was applied at a consistent rate of 10 kPa/s using a probe size of 1 cm^2. Participants were instructed to indicate when the pressure sensation transitioned into a disturbing or painful sensation by activating a switch. Each measurement was repeated three times at each anatomical location, with a 30-second interval between repetitions. The mean of the three measurements was then calculated for statistical analysis. Participants were provided with rest as needed during the procedure, and all measurements were conducted in consistent environments to minimize potential environmental influences.

Douleur Neuropathique en 4 questions- DN4

The DN4 is a widely used screening tool for assessing neuropathic pain and is designed to help clinicians differentiate neuropathic pain from other types of pain, such as nociceptive pain. The DN4 consists of four questions with ten items, each scored to determine the likelihood of neuropathic pain. These questions assess both sensory symptoms and signs. The maximum score is 10 points. Typically, if a patient scores 4 or higher on the DN4, it suggests a high likelihood of neuropathic pain. We used the same cut-off score for this study. (Bouhassira et al., 2005)

PainDETECT

The PainDETECT is a pain questionnaire to assess the likelihood of neuropathic pain in patients. The questionnaire consists of several questions about the characteristics of the patient's pain. Some items found on the questionnaire are: pain course pattern, pain quality (burning, prickling, tingling, or stabbing), radiating pain, hypersensitivity to touch or pain responding to temperature changes. Based on the patient's responses, a score is generated to estimate the likelihood of neuropathic pain. The questionnaire is typically scored out of 38 points, and a higher score indicates a higher likelihood of neuropathic pain. Additionally, two cut-off points, namely 12 and 19, are used: scores ≤ 12 suggest a neuropathic component is unlikely, scores between 13 and 18 show an unclear neuropathic component, while scores ≥ 19 suggest a probable neuropathic component and this study used the cut-off score of ≥ 19. (Freynhagen, R., 2006)

Nordic-Musculoskeletal Questionnaire

The Nordic Musculoskeletal Questionnaire self-reporting tool is designed to assess musculoskeletal symptoms and discomfort in different parts of the body. The NMQ typically consists of a series of questions that focus on various body regions, particularly those that are susceptible to musculoskeletal problems, such as the neck, shoulders, back (upper and lower), elbows, wrists/hands, hips/thighs, knees and ankles/feet. This self-reported questionnaire helps the individuals to describe their discomfort or pain in these different areas and it gathers information about the impact of these symptoms on their daily life activities of the past 12 months or the last seven days. (Kuorinka, I., 1987)

Neuropathic Pain Symptom Inventory - NPSI

The Neuropathic Pain Symptom Inventory (NPSI) is a questionnaire used to assess the symptoms and characteristics of neuropathic pain. The NPSI was developed to help understand and evaluate the nature and severity of neuropathic pain symptoms in patients. The NPSI consists of a series of questions that ask patients about the various sensory and affective components of their neuropathic pain. Patients are asked to rate the intensity and frequency of specific pain sensations, as well as the associated impact on their daily life and emotional well-being. (Bouhassira, D., 2004)

Neuropathic Pain Scale - NPS

The Neuropathic Pain Scale is a questionnaire that aims to assess the severity and characteristics of neuropathic pain. It typically includes questions about the intensity, quality, and location of the pain. The scale also includes questions related to the temporal pattern of pain, and sensory symptoms such as tingling, burning, or numbness. Scores are based on patient responses to questions about pain intensity. Zero indicates no pain, ten indicates the most pain imaginable. (Galer, B.S., 1997)

Brief Pain Inventory, Short form - BPI-SF

This questionnaire consisting of eleven items, aims to assess the severity of pain and the impact of pain on daily life. On a scale of zero to ten, the participant must rate his worst, least, current and average pain. Pain treatments are also questioned, as well as the impact of the pain on mood, walking ability, sleep, etc.

Statistical analysis

The statistical analyses of this master's thesis were performed using JMP pro 16 and 17 and Microsoft Excel. A significance level of 0.05 was used for all tests applied to the data. To check the normality of the data, the Shapiro-Wilk test was used and to compare the outcomes of the descriptive tests between the 2 patient groups (PwMS with and without neuropathic pain), a t-test was used when the data was normally distributed and if not, a Mann-Whitney U test was used.

Between-day reliability was determined by calculating the ICC (Intraclass Correlation Coefficient) and the Standard error of measurement (SEM = SD x √(1-ICC)) and smallest detectable change (SDC = 1.96 × √2 × SEM) were calculated to determine the measurement error. Bland-Altman plots and ROC (Receiver Operating Characteristics) curves were developed to provide an even better visual representation of reliability and accuracy. The interpretation of the ICC was outlined as such: below 0.5 indicates poor reliability, from 0.5 to 0.75 suggests moderate reliability, between 0.75 and 0.9 signifies good reliability, and above 0.9 reflects excellent reliability (Koo & Li, 2016). To measure the accuracy of the diagnostic tests (i.e. PainDetect and DN4), the following values were calculated: sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV) and the positive and negative likelihood ratio.

RESULTS

Descriptive outcomes

110 PwMS participated in the first measurement moment of between-day reliability and of these participants, 100 participated in both the test and retest moment. Pain diagnoses of 10 participants were missing, so these patients were excluded from analyses distinguishing between PwMS with neuropathic pain and without. Table 3 summarizes the characteristics of the participants.

Table 3
Characteristics of the participants from both Chile and Belgium

		All PwMS (N = 110)	PwMS with NP pain (N = 65)	PwMS without NP pain (N = 35)
Gender N (%)	Male	34 (31.91)	17 (26.15)	11 (31.43)
	Female	76 (69.09)	48 (73.85)	24 (68.57)
MS-type N (%)	Relapsing-Remitting	79 (71.82)	46 (70.77)	27 (77.14)
	Primary progressive	13 (11.82)	7 (10.77)	4 (11.43)
	Secondary progressive	18 (16.36)	12 (18.46)	4 (11.43)
Age groups mean ± SD		45 ± 12.74	44 ± 12	43.89 ± 12.87
EDSS score N (%)	0-3.5	70 (63.64)	40 (61.54)	24 (68.57)
	4-6	40 (36.36)	25 (38.46)	11 (31.43)
Duration of disease N (%)	<5 years	35 (31.82)	23 (35.38)	10 (28.57)
	5-15 years	53 (48.18)	29 (44.62)	21 (60.00)
	> 15 years	22 (20.00)	13 (20.00)	4 (11.43)
Employment status N (%)	Employed	57 (51.82)	35 (53.85)	19 (54.29)
	Unemployed	53 (48.18)	30 (46.15)	16 (45.71)

Abbreviations: PwMS, Persons with Multiple Sclerosis ; NP, neuropathic ; EDSS, Expanded Disability Status Scale

Overall, there were more female than male participants (69.09% vs 31.91%) and most PwMS suffered from relapsing-remitting MS (71.82%). The mean age of the entire sample size is 45 years; most participants had an EDSS score smaller than 3.5 and were diagnosed between 5 and 15 years ago. The distribution between participants still working and no longer working is approximately equal. Furthermore, no significant differences in characteristics were found between PwMS with neuropathic pain and without neuropathic pain.

Table 4 demonstrates the mean score and standard deviation for the descriptive tests that are normally distributed and the median score and interquartile range for the descriptive tests with a non-normal distribution.

Table 4
Results descriptive tests

	PwMS with NP pain (N = 65)	PwMS without NP pain (N = 35)	P-value
9HPT (median - IQR)	20.15 - 7.1	19.71 - 5.22	0.3918
T25FW (median – IQR)	4.8 - 3.31	4.81 - 1.65	0.3644
SDMT (mean ± SD)	53.03 ± 13.53	53 ± 12.21	0.9906
HADS (median – IQR)	13.95 - 8.29	12.44 - 7.57	0.4441
PSS (mean ± SD)	14.93 ± 7.3	14.53 ± 8.37	0.8214
MAS (median – IQR)	0 – 1	0 – 1	0.6571
MI-L (median – IQR)	100 – 10	100 – 6	0.5314
MI-R (median – IQR)	100 – 4.38	100 – 8	0.7747
MFIS (mean ± SD)	42.25 ± 15.90	42.82 ± 21.28	0.8995

Abbreviations: 9HPT, Nine Hole Peg Test ; T25FW, Timed 25 foot Walk ; SDMT, Symbol Digit modality test ; HADS, Hospital Anxiety and Depression Scale ; PSS, Perceived Stress Scale ; mAS, modified Ashworth Scale ; MI-L, Motricity Index - left ; MI-R, Motricity Index - Right ; mFIS, modified Fatigue Impact Scale ; PwMS, Persons with Multiple Sclerosis ; NP, Neuropathic ; IQR, Interquartile Range ; SD, Standard Deviation

The p-value was also determined to assess whether there is a significant difference in these outcome measures between PwMS with and without neuropathic pain. There were no significant differences found between the two groups but important to note is that the mean score of the PSS in both groups is greater than 14, indicating a level of moderate stress in patients with and without neuropathic pain. It is also noteworthy that the mean mFIS score in both groups is higher than the cut-off score of 38, indicating that both PwMS with and without neuropathic pain are suffering from fatigue and that this fatigue has a significant impact on their quality of life. (Larson, R. (2013)

Before assessing the reliability and accuracy of the pain questionnaires, an overview of the general results of these questionnaires can be found in Table 5. Our findings reveal that participants scored above the designated cut-off score on both the PainDetect and DN4 questionnaires, indicating the presence of neuropathic pain. With an average score of 15.1 on the PainDetect questionnaire and an average score of 6.5 on the DN4 questionnaire, participants exceeded the established cut-off points (19 for PainDETECT and 4 for DN4) for identifying possible neuropathic pain.

Table 5

Overview general results pain outcome measures

Pain questionnaire	**Mean and SD** (if normally distributed) **Median and IQR** (if not normally distributed)	**Cut-off value / score interpretation**
NPS	Mean score : 37.52 SD : ± 15.004	Total score range from 0-100 100 represents the highest degree of neuropathic like symptoms (Lape, C.E., 2020)
PainDETECT	Mean score : 15.1 SD : ± 7.651	Total score range from 0-38 <12 : nociceptive pain 13-18 : possible neuropathic pain ≥ 19 : >90% likelihood of neuropathic pain (Packham, L.T., 2017)
DN4	Median score : 6.5 IQR : 4	Total score range from 0-10 A score ≥4 is interpreted as neuropathic pain (Perez, C., 2007)
NPSI	Median score : 27 IQR : 32.75	Total score range from 0-100 High scores indicate neuropathic pain (Bouassira D., 2004)
BPI	Median score: 22 IQR : 24.5	Two subscores : interference and severity Higher scores indicate more severe pain or greater interference with daily activities. (Cleeland, S.C., 1991)

Abbreviations : SD, Standard Deviation ; IQR, Interquartile Range ; NPS, Neuropathic Pain Scale ; DN4, Douleur Neuropathique 4 ; NPSI, Neuropathic Pain Symptom Inventory ; BPI, Brief Pain Inventory

Between-day reliability

ICCs were computed to assess the between-day reliability of the pain outcome measures across Belgium and Chile and are shown in Table 6.

Our findings for the whole sample indicate strong agreement in neuropathic pain diagnosis between the two groups. Across the measures, the ICC values were robust: DN4 = 0.88, PainDETECT = 0.88, NPS = 0.86, NPSI = 0.89, BPI (interference) = 0.91, and BPI (severity) = 0.93. These results demonstrate consistently high levels of reliability, indicating strong agreement and consistency among raters.

Table 6

Test-retest Reliability of Pain-related Outcome Measures in PwMS

	ICC (95% CI)	Test (mean + SD)	Retest (mean + SD)	SEM	SDC
TOTAL (N=110)					
DN4	0.88 (0.82 – 0.91)	5.47 ± 2.72	5.36 ± 2.68	0.44	1.22
PainDetect	0.88 (0.83 – 0.92)	16.33 ± 7.83	15.53 ± 7.82	1.24	3.43
NPS	0.86 (0.80 – 0.90)	43.46 ± 18.37	38.52 ± 22.33	3.44	9.54
NPSI	0.89 (0.84 – 0.92)	35.36 ± 21.51	33.60 ± 20.36	3.05	8.45
BPI-SF (Interference)	0.91 (0.79 – 0.89)	28.08 ± 18.34	23.32 ± 17.80	0.53	1.46
BPI-SF (Severity)	0.93 (0.82 – 0.91)	5.02 ± 2.05	4.45 ± 1.78	0.31	0.85
CHILE (N=50)					
DN4	0.99 (0.98 – 0.99)	5.75 ± 2.77	6.18 ± 2.82	0.03	0.10
PainDetect	0.98 (0.98 – 0.99)	15.1 ± 7.65	15.16 ± 8.07	0.17	0.46
NPS	0.96 (0.81 – 0.98)	37.52 ± 15.00	40.38 ± 16.23	0.68	1.88
NPSI	0.98 (0.98 – 0.99)	31.02 ± 17.99	31.14 ± 17.43	0.44	1.21
BPI-SF (Interference)	0.98 (0.97 – 0.99)	25.48 ± 17.95	24.4 ± 16.87	0.05	0.15
BPI-SF (Severity)	0.86 (0.76 – 0.91)	4.15 ± 1.72	4.08 ± 1.53	0.31	0.86
BELGIUM (N=60)					
DN4	0.64 (0.45 – 0.78)	4.62 ± 2.43	4.28 ± 2.07	1.04	2.89
PainDetect	0.78 (0.64 – 0.86)	17.05 ± 7.58	15.68 ± 7.25	2.53	7.02
NPS	0.79 (0.66 – 0.87)	48.17 ± 19.14	38.20 ± 25.61	4.94	13.71
NPSI	0.82 (0.71 – 0.89)	39.20 ± 22.66	36.14 ± 21.84	5.37	14.89
BPI-SF (Interference)	0.73 (0.57 – 0.83)	4.38 ± 2.65	3.89 ± 2.46	0.98	2.72
BPI-SF (Severity)	0.83 (0.72 – 0.90)	5.06 ± 2.07	4.82 ± 1.97	0.44	1.22

Abbreviations: DN4, Douleur Neuropathique 4 ; NPS, Neuropathic Pain Scale ; NPSI, Neuropathic Pain Symptom Inventory ; BPI-SF, Brief Pain Inventory - Short form ; ICC, Intraclass Correlation Coefficient ; 95% CI, 95% Confidence Interval ; SD, Standard Deviation ; SEM, Standard Error of Measurement ; SDC, Smallest Detectable Change

Our analysis of the Standard Error of Measurement (SEM) for the entire sample provides valuable insights into the precision of our measures. The SEM for the DN4 questionnaire was 0.44, and for the PainDETECT questionnaire, it was 1.24, indicating relatively low levels of measurement error for both tools. However, when examining other measures, such as the NPS (SEM = 3.44) and NPSI (SEM = 3.05), higher SEM values were observed, suggesting greater measurement error. Notably, the BPI Interference (SEM = 0.53) and BPI severity (SEM = 0.31) demonstrated a high precision level in assessing pain interference and severity.

For the Belgian sample, the ICC values varied across the different measurement instruments. The DN4 instrument demonstrated moderate between-day reliability with an ICC of 0.64. PainDETECT showed slightly higher reliability with an ICC of 0.78, suggesting good consistency in scores across different days. The NPS, NPSI and BPI instruments demonstrated good between-day reliability in the Belgian sample, with ICC ranging from 0.73 to 0.83. In contrast, the Chilean sample exhibited consistently high ICC values across all measurement instruments. The DN4, PainDETECT, NPS, NPSI, and BPI instruments all demonstrated excellent between-day reliability. These results suggest strong consistency in scores across different days within the Chilean sample, indicating minimal variability attributable to measurement error.

A comparison of ICC values between Belgium and Chile reveals notable differences in between-day reliability across the two countries. While the Belgian sample exhibited variability in reliability across different measurement instruments, with some instruments demonstrating moderate to high reliability and others showing lower reliability, the Chilean sample consistently demonstrated high reliability across all instruments. The ICC values for the Chilean sample were uniformly high, indicating excellent between-day reliability for all measurement instruments.

In Table 7, the ICCs were also calculated to evaluate the between-day reliability of algometer measurements across various anatomical sites. The results indicate strong reliability for pressure pain threshold (PPT) measurements taken at different anatomical locations. For the Belgian sample, the ICC values for PPT measurements ranged from 0.82 to 0.91 across different anatomical sites. Specifically, the PPT measurements for Trapezius muscles (left: ICC = 0.83, right: ICC = 0.82), Thumbnail (left: ICC = 0.84, right: ICC = 0.87), Low back (left: ICC = 0.90, right: ICC = 0.87), and Quadriceps muscles (left: ICC = 0.89, right: ICC = 0.91) demonstrated good to excellent between-day reliability. These findings suggest consistent responses to pressure stimuli across multiple days for participants in the Belgian sample.

Table 7

Test-retest Reliability of Algometer in PwMS (Belgium)

	ICC 95% CI	SEM	SDC
PPT Trapezius left	0.83 (0.70 – 0.91)	36.42	100.96
PPT Trapezius right	0.82 (0.70 – 0.89)	41.81	115.92
PPT Thumbnail left	0.84 (0.72 – 0.91)	51.40	142.47
PPT Thumbnail right	0.87 (0.78 – 0.93)	45.04	124.83
PPT Low back left	0.90 (0.83 – 0.94)	39.44	109.33
PPT Low back right	0.87 (0.78 – 0.92)	56.45	156.46
PPT Quadriceps left	0.89 (0.82 – 0.94)	41.54	115.15
PPT Quadriceps right	0.91 (0.85 – 0.95)	35.34	97.96

Abbreviations: PPT, Pressure Pain Threshold ; ICC, Intraclass Correlation Coefficient ; 95% CI, 95% Confidence Interval ; SEM, Standard Error of Measurement ; SDC, Smallest Detectable Change

SEM values varied across different body regions, indicating varying levels of measurement error. Regions like the trapezius muscles and quadriceps exhibit lower SEM values, indicating higher precision in PPT measurement. This suggests a relatively consistent determination of pain thresholds in these areas. Conversely, slightly higher SEM values for regions like the low back and thumbnail imply greater variability and potential error in PPT measurements.

Specifically for the NMQ, the test-retest reliability was examined using Cohen's Kappa (see Table 8). Based on the Kappa scores of the total sample, the NMQ has a moderate to almost perfect test-retest reliability. When specifically examining the subdivision by body region, it can be concluded that certain regions exhibit higher reliability compared to others. Mainly the questions related to the low back, hips and knees show good retest reliability in contrast to the neck region, which appears to be the least reliable.

Table 8

Test-retest reliability of the Nordic Musculoskeletal Questionnaire (NMQ) using Cohen's Kappa

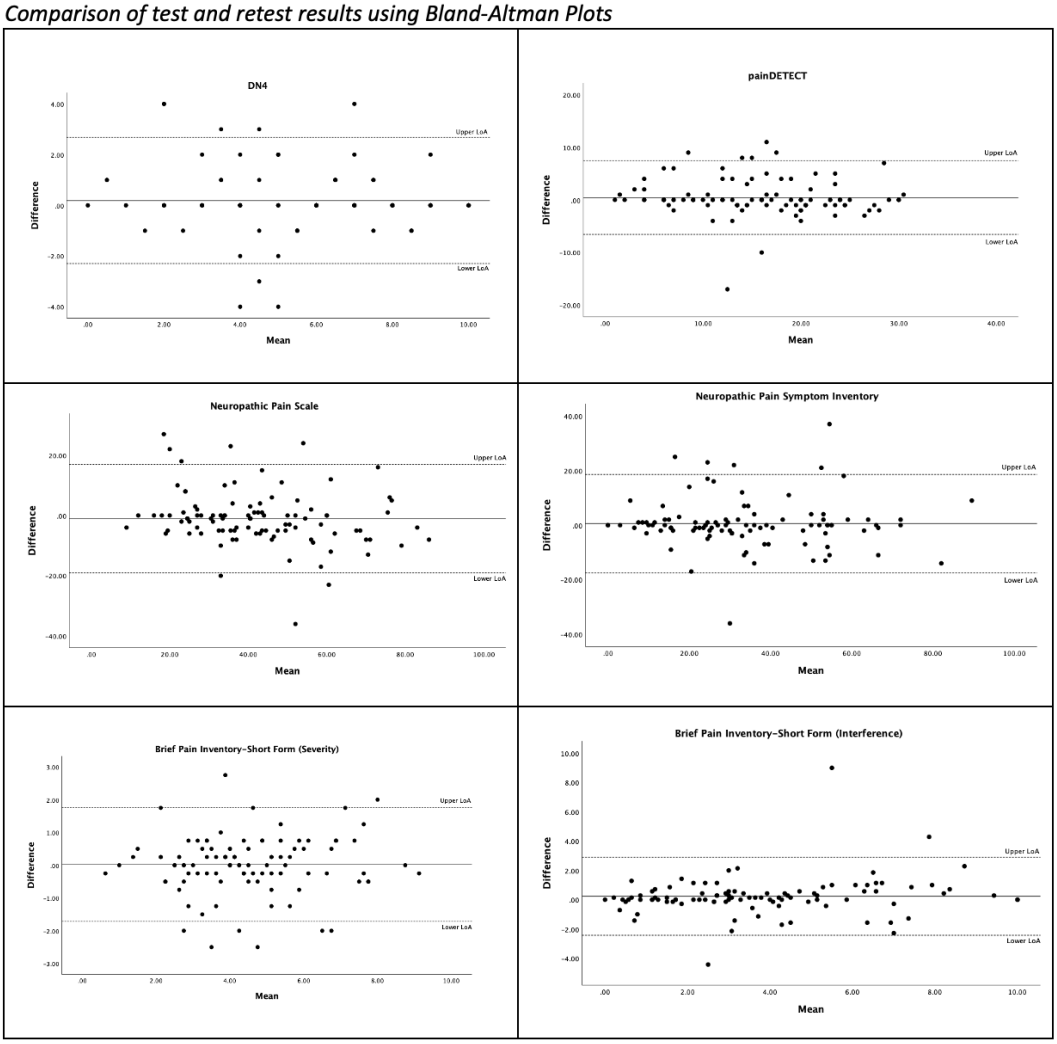
	Total n=100			**Belgium n=50**			**Chile n=50**		
	Any trouble during the last 12 months?	Any time during the last 12 months being prevented form doing normal work because of the trouble?	Any trouble during the last 7 days?	Any trouble during the last 12 months?	Any time during the last 12 months being prevented form doing normal work because of the trouble?	Any trouble during the last 7 days?	Any trouble during the last 12 months?	Any time during the last 12 months being prevented form doing normal work because of the trouble?	Any trouble during the last 7 days?
	K 95% CI	K 95% CI	K 95% CI	K 95% CI	K 95% CI	K 95% CI	K	K	K
Neck	0.70 (0.55-0.85) p<.001	0.54 (0.36-0.71), p<.001	0.60 (0.42-0.77), p<.001	0.43 (0.14-0.72), p<.001	0.27 (0.01-0.54), p=0.05	0.35 (0.09-0.61), p=.011	0.84 (0.69-0.98), p <.001	0.78 (0.55-1.01), p<.001	1 p<.001
Shoulders	0.81 (0.70-0.93) p<.001	0.70 (0.56-0.85), p<.001	0.66 (0.49-0.82), p<.001	0.56 (0.31-0.81), p<.001	0.48 (0.23-0.72), p<.001	0.51 (0.28-0.75), p<.001	0.95 (0.87-1.03), p <.001	1 (1-1), p<.001	0.77 (0.48-1.07), p<.001
Elbows	0.75 (0.60-0.89), p<.001	0.55 (0.33-0.77), p<.001	0.78 (0.60-0.96), p<.001	0.62 (0.38-0.85), p<.001	0.28 (-0.03-0.60), p=0.04	0.64 (0.36-0.92), p<.001	0.89 (0.75-1.03), p <.001	0.91 (0.74-1.08), p<.001	1 p<.001
Wrists/hands	0.80 (0.68-0.91), p<.001	0.73 (0.58-0.87), p<.001	0.70 (0.54-0.87), p<.001	0.63 (0.42-0.84), p<.001	0.53 (0.30-0.77), p<.001	0.64 (0.42-0.86), p<.001	0.95 (0.88-1.03), p <.001	1 p<.001	0.77 (0.48-1.07, p<.001
Upper back	0.73 (0.60-0.87), p<.001	0.63 (0.46-0.80), p<.001	0.67 (0.49-0.85), p<.001	0.64 (0.42-0.85), p<.001	0.41 (0.15-0.66), p=.003	0.48 (0.22-0.74), p<.001	0.83 (0.68-0.99), p< .001	0.93 (0.80-1.06), p<.001	1 p<.001
Low back	0.87 (0.76-0.98), p<.001	0.85 (0.75-0.95), p<.001	0.75 (0.62-0.89), p<.001	0.73 (0.48-0.97), p<.001	0.74 (0.55-0.93), p<.001	0.60 (0.37-0.82), p<.001	0.95 (0.87-1.03), p <.001	0.94 (0.84-1.05), p<.001	0.92 (0.79-1.06), p<.001
Hips/thighs	0.85 (0.75-0.95), p<.001	0.77 (0.64-0.90), p<.001	0.74 (0.59-0.89), p<.001	0.75 (0.57-0.93), p<.001	0.59 (0.37-0.81), p<.001	0.56 (0.33-0.80), p<.001	0.95 (0.88-1.03), p <.001	1 p<.001	1 p<.001
Knees	0.82 (0.70-0.93), p<.001	0.71 (0.56-0.87), p<.001	0.86 (0.75-0.97), p<.001	0.68 (0.47-0.88), p<.001	0.55 (0.31-0.79), p<.001	0.78 (0.60-0.96), p<.001	0.96 (0.88-1.03), p<.001	0.93 (0.80-1.06), p<.001	1 p<.001
Ankles/feet	0.81 (0.70-0.93) p<.001	0.59 (0.41-0.77), p<.001	0.73 (0.57-0.88), p<.001	0.67 (0.47-0.87), p<.001	0.38 (0.11-0.64), p= .006	0.56 (0.33-0.80), p<.001	0.96 (0.88-1.03), p<.001	0.92 (0.77-1.07), p<.001	1 p<.001

Abbreviations: K, Cohen's Kappa ; 95% CI, 95% Confidence Interval

Another method to evaluate and visualize between-day reliability and, more specifically, the degree of agreement between the two testing moments is by creating Bland-Altman plots, as depicted in Figure 1. Based on the plots and limits of agreements (LoA), it can be concluded that pain-outcome measures in PwMS exhibited minimal bias between days, as evidenced by the narrow limits of agreement and consistent distribution of data points around the mean difference line without following a certain pattern. This suggests that participants provided consistent responses to these questionnaires over time, indicating good between-day reliability in assessing neuropathic pain symptoms.

Figure 1

Comparison of test and retest results using Bland-Altman Plots



Abbreviations: LoA , Limits of Agreement ; DN4 , Douleur Neuropathique 4

Accuracy

Accuracy analyses were performed on the DN4 and PainDETECT to determine the diagnostic validity and whether their diagnosis of neuropathic pain was consistent with that of the neurologists involved. The results are summarized in Table 9 for the total sample size (N=100). The sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated to evaluate the ability of the two questionnaires to identify cases of neuropathic pain correctly.

In the context of these specific questionnaires, sensitivity tells us how well the questionnaire identifies PwMS who truly have neuropathic pain and specificity tells us how well the questionnaire correctly identifies PwMS without neuropathic pain. The DN4, with a sensitivity score of 0.89 seems to detect almost 90% of the PwMS with neuropathic pain correctly (true positives) but nevertheless, approximately 10% of PwMS received a negative score on the DN4 while receiving a positive diagnosis from the neurologist (false negatives). A sensitivity score of 0.50 for the PainDETECT questionnaire indicates that the questionnaire correctly identifies only half of the PwMS who actually have neuropathic pain, meaning that also 50% of the PwMS are incorrectly diagnosed with neuropathic pain using the PainDETECT. The specificity scores for the DN4 and PainDETECT, respectively 0.429 and 0.714, indicate that the PainDETECT is more accurate at correctly ruling out PwMS without neuropathic pain.

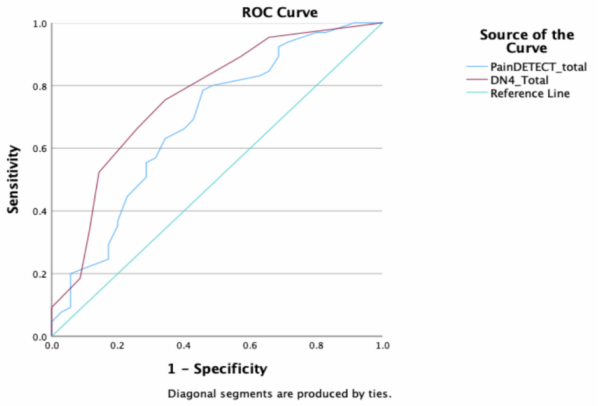
According to the study by Parikh R. et al. a PPV close to 100 means that this diagnostic tool is doing good and that it might even become a gold standard. (Parikh, R., 2008) The PPV's of the DN4 and PainDETECT, 0.74 and 0.76, tell us that approximately 75% of the PwMS with a positive result on one of the two tests are effectively suffering neuropathic pain. On the other hand, the NPV of 0.68 (DN4) and 0.43 (PainDETECT) indicate that of all the PwMS with a negative result on the questionnaires 68% and 43% are truly not experiencing neuropathic pain. This implies that primarily based on the PainDETECT, a considerable number of PwMS receive an incorrect negative result.

Additionally, the positive and negative likelihood ratios (LR+ and LR-) give an indication of the diagnostic performance of the DN4 and PainDETECT. A LR+ of 1.75 means that PwMS with neuropathic pain are 1.75 times more likely to receive a score of 4 or higher on the DN4 than PwMS without neuropathic pain and a LR- of 0.25 suggests that PwMS without neuropathic pain are 0.25 times more likely to have a negative result on the DN4. When compared with the LR+ and LR- of the PainDETECT, respectively 1.77 and 0.68, it is particularly noticeable that PwMS without neuropathic pain are more likely to receive a negative score on the PainDETECT than on the DN4.

The AUC values for both the DN4 and PainDETECT were calculated based on receiver operating characteristic (ROC) curves (the summarizing ROC curve can be found in Figure 2), with the pain diagnosis by the neurologist used as the reference standard. The AUC values, based on the whole sample size, are 0.76 (DN4) and 0.68 (PainDETECT). These results indicate that both the DN4 and PainDETECT are considered to be of limited clinical utility because their AUC scores are smaller than 0.80 (Mandrekar N.J, 2010).

Finally, also Cohen's Kappa was calculated for both DN4 and PainDETECT and these results (0.19 and 0.35) indicate that there was slight to moderate agreement between the diagnostic result of the DN4 and the pain diagnosis of the neurologist and a fair agreement between the total score of the PainDETECT and the diagnosis of the neurologist.

Figure 2
ROC-curves of DN4 and PainDETECT



Abbrevations: ROC, Receiver Operating Characteristics

Table 9
Overview of the diagnostic accuracy of the DN4 and painDETECT in PwMS

		Total sample (Belgium + Chile) N = 100
PainDETECT (Point estimate 95% CI)	Sensitivity	0.50 (0.38-0.63)
	Specificity	0.71 (0.53-0.85)
	PPV	0.76 (0.61-0.88)
	NPV	0.43 (0.30-0.57)
	LR+	1.77 (0.99-3.16)
	LR-	0.68 (0.49-0.95)
	AUC (95% CI)	0.68 (0.57 – 0.80)
	Kappa	0.35 (0.16-0.54), p < .001
DN4 (Point estimate 95% CI)	Sensitivity	0.89 (0.79-0.95)
	Specificity	0.42 (0.26-0.60)
	PPV	0.74 (0.63-0.83)
	NPV	0.68 (0.45-0.86)
	LR+	1.56 (1.15-2.10)
	LR-	0.25 (0.11-0.55)
	AUC (95% CI)	0.76 (0.66 – 0.87)
	Kappa	0.19 (0.02-0.36), p = .032

Abbreviations: DN4 , Douleur Neuropathique 4 ; PwMS : Persons with Multiple Sclerosis ; PPV , Positive Predictive Value ; NPV , Negative Predictive Value ; LR+ , Positive Likelihood Ratio ; LR- , Negative Likelihood Ratio ; AUC , Area Under The Curve.

DISCUSSION

The objective of this study was to investigate the between-day reliability of pain outcome measures in PwMS and to compare the reliability and accuracy of multiple patient-reported outcomes in order to determine which one(s) have the most value within clinical practice. Exploring the between-day reliability allows us to capture the dynamic nature of pain in MS, recognizing that pain levels may fluctuate from day to day. This insight is essential for developing interventions that address not only the overall pain burden but also the day-to-day fluctuations that individuals with MS may experience. Because this is the first study to investigate the reliability of all these pain outcome measures in MS, we aim to contribute valuable insights to the field of MS pain management, ultimately improving the quality of care and quality of life for individuals living with this condition. Moreover, this study forms a critical part of a larger research project focused on the validity and responsiveness of measurement instruments in MS.

Results show promising findings regarding the between-day reliability of the pain outcome measures utilized, as well as the reliability of PPT measurements obtained with the algometer. Notably, all pain outcome measures demonstrated good between-day reliability, indicating consistent assessment of pain-related constructs over consecutive days. All of this could be deduced from the relatively high ICC values along with the corresponding low SEM values and narrow limits of agreement observed in the Bland-Altman plots. Furthermore, the PPT measurements exhibited reliable outcomes, suggesting stability in pressure pain threshold assessments across different testing sessions. Moreover, our findings expand upon previous research by demonstrating the reliability of additional pain outcome measures beyond those previously identified in the MS population. While previous studies have highlighted the utility of measures such as the NPS and BPI in the MS population, our study suggests the presence of other reliable outcome measures that can complement and enhance pain assessment in this population. This highlights the importance of considering a comprehensive set of outcome measures to capture the multidimensional nature of pain experiences in individuals with MS.

When comparing between-day reliability of pain-related outcome measures in persons with MS with other pathologies we found similar results. For instance, Ferreira A.C.L. (2023) investigated the validity and reliability of the BPI-SF in older adults with various pain types and came to the conclusion that it had an excellent between-day reliability for both intensity and interference score, namely ICCs of 0.90 and 0.96 (in comparison to 0.93 and 0.91 in our results). Furthermore, the reliability and validity of the PainDETECT and DN4 were also determined in patients with suspected cervical or lumbar radiculopathy by Epping R. et al (2017). The ICC's they found for both the PainDETECT and DN4 of 0.91 and 0.86 were similar to our results (ICC of 0.88 for both questionnaires), meaning good between-day reliability.

Additionally, when comparing results from Belgium and Chile, some interesting results were found regarding the between-day reliability of pain outcome measures in PwMS. Notably, we observed substantial variation in the ICC scores across different pain outcome measures in Belgium, whereas all measures in Chile demonstrated excellent ICC scores. The discrepancy in ICC scores between the two countries can be attributed to several factors.

A limitation of this study is the variance in intervals between the two testing sessions across countries, as well as differences in the administration methods utilized, which may have introduced some variability into our results. With a four-day interval in Chile and a variable interval in Belgium of three to seven days, there is a discrepancy in the interval between assessments and this may have influenced the reliability of the pain outcome measures. A shorter interval, as observed in Chile, could lead to better retention of responses and decreased variability compared to a longer interval, as in Belgium. Furthermore, the administration method of the outcome measures differed. In Chile, investigators read all questions aloud to the patients, ensuring standardized administration and minimizing potential interpretation errors. Conversely, in Belgium, patients were asked to read the questions independently, introducing variability in comprehension and response.

Furthermore, our results regarding the diagnostic accuracy of the DN4 and PainDETECT in PwMS, based on the whole sample size, demonstrated that the DN4 has a much better sensitivity than the PainDETECT, meaning that the DN4 is more accurate in correctly identifying PwMS with neuropathic pain (89% true positives versus only 50%). It is, however, noteworthy that the PainDETECT exhibits a higher specificity score (71% compared to 42% of the DN4), indicating that the PainDETECT demonstrates better accuracy in identifying PwMS who do not have neuropathic pain (more true negatives). This implies that while the PainDETECT excels in correctly identifying PwMS without neuropathic pain, the DN4's strength lies in accurately pinpointing those who do have neuropathic pain. In addition to sensitivity and specificity, the PPV, NPV, LR+, and LR- also offer important insights regarding the diagnostic accuracy of these questionnaires. From these metrics, we can infer that a positive outcome on both the DN4 and PainDETECT is relatively indicative of neuropathic pain. However, it is evident that the DN4 exhibits greater reliability than the PainDETECT in distinguishing individuals with Multiple Sclerosis (PwMS) who do not have neuropathic pain.

Our results regarding the diagnostic accuracy of the DN4 and PainDETECT in PwMS are quite consistent with findings from other studies regarding other pathologies. In individuals with suspected cervical or lumbar radiculopathy, the accuracy of these questionnaires for diagnosing neuropathic pain was examined by Epping R. et al and they concluded that both questionnaires showed limited overall diagnostic accuracy. (Epping, R., 2017) The results from this study align with our conclusion based on the AUC values of 0.76 (DN4) and 0.68 (PainDETECT), which are both smaller than 0.80, indicating that both questionnaires have limited clinical utility.

A limitation concerning our diagnostic accuracy analysis relates to the diagnosis of neuropathic pain established by one of the involved neurologists. To date, this diagnosis relies on a medical history and a basic clinical examination, but there is still no gold standard available, potentially leading to variations in approach among different neurologists. This, in turn, will possibly have had an impact on our accuracy analysis since we have used the diagnosis provided by the physicians as the reference value. The fact that multiple, different neurologists were involved in diagnosing pain within our sample size also plays a role in this context.

CONCLUSION

In conclusion, this study aimed to investigate the between-day reliability of pain outcome measures in PwMS and to compare the reliability and accuracy of multiple patient-reported outcomes to determine their clinical value. By exploring between-day reliability, we aimed to capture the dynamic nature of pain in MS, recognizing its fluctuations over time, crucial for developing interventions addressing both overall pain burden and daily variations. Being the first of its kind in MS, our study contributes significant insights into pain management, ultimately aiming to enhance care and quality of life for those with this condition. Our findings present promising results regarding the between-day reliability of pain outcome measures and Pressure Pain Threshold (PPT) measurements obtained with an algometer. All pain outcome measures demonstrated good reliability, suggesting consistent assessment of pain-related constructs over consecutive days. Additionally, our study expands upon previous research by identifying reliable outcome measures beyond those previously recognized in the MS population, emphasizing the importance of a comprehensive approach to pain assessment. Regarding the diagnostic accuracy of the DN4 and PainDETECT, we must align with previous literature and also come to the conclusion that they are still insufficiently accurate to base a pain diagnosis on solely. Anamnesis and clinical examination by a physician remain essential in this regard.

REFERENCES

1. Amtmann, D., Bamer, A. M., Kim, J., Chung, H., & Salem, R. (2018). People with multiple sclerosis report significantly worse symptoms and health related quality of life than the US general population as measured by PROMIS and NeuroQoL outcome measures. *Disabil Health J*, *11*(1), 99-107. https://doi.org/10.1016/j.dhjo.2017.04.008
2. Benedict, R. H., DeLuca, J., Phillips, G., LaRocca, N., Hudson, L. D., Rudick, R., & Multiple Sclerosis Outcome Assessments, C. (2017). Validity of the Symbol Digit Modalities Test as a cognition performance outcome measure for multiple sclerosis. *Mult Scler*, *23*(5), 721-733. https://doi.org/10.1177/1352458517690821
3. Bohannon, R. W., & Smith, M. B. (1987). Interrater reliability of a modified Ashworth scale of muscle spasticity. *Phys Ther*, *67*(2), 206-207. https://doi.org/10.1093/ptj/67.2.206
4. Bouhassira, D., Attal, N., Alchaar, H., Boureau, F., Brochet, B., Bruxelle, J., Cunin, G., Fermanian, J., Ginies, P., Grun-Overdyking, A., Jafari-Schluep, H., Lanteri-Minet, M., Laurent, B., Mick, G., Serrie, A., Valade, D., & Vicaut, E. (2005). Comparison of pain syndromes associated with nervous or somatic lesions and development of a new neuropathic pain diagnostic questionnaire (DN4). *Pain*, *114*(1-2), 29-36. https://doi.org/10.1016/j.pain.2004.12.010
5. Bouhassira, D., Attal, N., Fermanian, J., Alchaar, H., Gautron, M., Masquelier, E., Rostaing, S., Lanteri-Minet, M., Collin, E., Grisart, J., & Boureau, F. (2004). Development and validation of the Neuropathic Pain Symptom Inventory. *Pain*, *108*(3), 248-257. https://doi.org/10.1016/j.pain.2003.12.024
6. Buldys, K., Gornicki, T., Kalka, D., Szuster, E., Biernikiewicz, M., Markuszewski, L., & Sobieszczanska, M. (2023). What Do We Know about Nociplastic Pain? *Healthcare (Basel)*, *11*(12).https://doi.org/10.3390/healthcare11121794
7. Epping, R., Verhagen, A. P., Hoebink, E. A., Rooker, S., & Scholten-Peeters, G. G. M. (2017). The diagnostic accuracy and test-retest reliability of the Dutch PainDETECT and the DN4 screening tools for neuropathic pain in patients with suspected cervical or lumbar radiculopathy. *Musculoskelet Sci Pract*,*30*, 72-79. https://doi.org/10.1016/j.msksp.2017.05.010

8. Fayazi, M., Dehkordi, S. N., Dadgoo, M., & Salehi, M. (2012). Test-retest reliability of Motricity Index strength assessments for lower extremity in post stroke hemiparesis. *Med J Islam Repub Iran*, *26*(1), 27-30. https://www.ncbi.nlm.nih.gov/pubmed/23483112
9. Ferreira, A. C. L., Pereira, D. S., da Silva, S. L. A., Carvalho, G. A., & Pereira, L. S. M. (2023). Validity and reliability of the short form brief pain inventory in older adults with nociceptive, neuropathic and nociplastic pain. *Geriatr Nurs*, *52*, 16-23. https://doi.org/10.1016/j.gerinurse.2023.04.011
10. Feys, P., Lamers, I., Francis, G., Benedict, R., Phillips, G., LaRocca, N., Hudson, L. D., Rudick, R., & Multiple Sclerosis Outcome Assessments, C. (2017). The Nine-Hole Peg Test as a manual dexterity performance measure for multiple sclerosis. *Mult Scler*, *23*(5), 711-720. https://doi.org/10.1177/1352458517690824
11. Fischer, A. A. (1987). Pressure algometry over normal muscles. Standard values, validity and reproducibility of pressure threshold. *Pain*, *30*(1), 115-126. https://doi.org/10.1016/0304-3959(87)90089-3
12. Freynhagen, R., Baron, R., Gockel, U., & Tolle, T. R. (2006). painDETECT: a new screening questionnaire to identify neuropathic components in patients with back pain. *Curr Med Res Opin*, *22*(10), 1911-1920. https://doi.org/10.1185/030079906X132488
13. Galer, B. S., & Jensen, M. P. (1997). Development and preliminary validation of a pain measure specific to neuropathic pain: the Neuropathic Pain Scale. *Neurology*, *48*(2), 332-338. https://doi.org/10.1212/wnl.48.2.332
14. Hardt, S., Bergau, S. A., Jacques, A., & Tampin, B. (2023). Short- and long-term test-retest reliability of the English version of the 7-item DN4 questionnaire - a screening tool for neuropathic pain. *Scand J Pain*,*23*(3), 494-500. https://doi.org/10.1515/sjpain-2022-0149
15. Jerkovic, A., Mikac, U., Matijaca, M., Kosta, V., Curkovic Katic, A., Dolic, K., Vujovic, I., Soda, J., Dogas, Z., Pavelin, S., & Rogic Vidakovic, M. (2022). Psychometric Properties of the Pittsburgh Sleep Quality Index (PSQI) in Patients with Multiple Sclerosis: Factor Structure, Reliability, Correlates, and Discrimination. *J Clin Med*, *11*(7). https://doi.org/10.3390/jcm11072037

16. Kalinowski, A., Cutter, G., Bozinov, N., Hinman, J. A., Hittle, M., Motl, R., Odden, M., & Nelson, L. M. (2022). The timed 25-foot walk in a large cohort of multiple sclerosis patients. *Mult Scler*, *28*(2), 289-299. https://doi.org/10.1177/13524585211017013
17. Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med*, *15*(2), 155-163. https://doi.org/10.1016/j.jcm.2016.02.012
18. Kos, D., Kerckhofs, E., Nagels, G., D'Hooghe, B. D., Duquet, W., Duportail, M., & Ketelaer, P. (2003). Assessing fatigue in multiple sclerosis: Dutch modified fatigue impact scale. *Acta Neurol Belg*, *103*(4), 185-191. https://www.ncbi.nlm.nih.gov/pubmed/15008502
19. Kratz, A. L., Murphy, S. L., & Braley, T. J. (2017). Ecological Momentary Assessment of Pain, Fatigue, Depressive, and Cognitive Symptoms Reveals Significant Daily Variability in Multiple Sclerosis. *Arch Phys Med Rehabil*, *98*(11), 2142-2150. https://doi.org/10.1016/j.apmr.2017.07.002
20. Kratz, A. L., Whibley, D., Alschuler, K. N., Ehde, D. M., Williams, D. A., Clauw, D. J., & Braley, T. J. (2021). Characterizing chronic pain phenotypes in multiple sclerosis: a nationwide survey study. *Pain*, *162*(5), 1426-1433. https://doi.org/10.1097/j.pain.0000000000002136
21. Kuorinka, I., Jonsson, B., Kilbom, A., Vinterberg, H., Biering-Sorensen, F., Andersson, G., & Jorgensen, K. (1987). Standardised Nordic questionnaires for the analysis of musculoskeletal symptoms. *Appl Ergon*,*18*(3), 233-237. https://doi.org/10.1016/0003-6870(87)90010-x
22. Lape, E. C., Selzer, F., Davis, A. M., Collins, J. E., Losina, E., & Katz, J. N. (2020). Psychometric properties of the Neuropathic Pain Scale (NPS) in a knee osteoarthritis population. *Osteoarthr Cartil Open*, *2*(1), 100027. https://doi.org/10.1016/j.ocarto.2020.100027
23. Larson, R. D. (2013). Psychometric properties of the modified fatigue impact scale. *Int J MS Care*, *15*(1), 15-20. https://doi.org/10.7224/1537-2073.2012-019
24. Lee, E. H. (2012). Review of the psychometric evidence of the perceived stress scale. *Asian Nurs Res (Korean Soc Nurs Sci)*, *6*(4), 121-127. https://doi.org/10.1016/j.anr.2012.08.004

25. Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *J Thorac Oncol*, *5*(9), 1315-1316. https://doi.org/10.1097/JTO.0b013e3181ec173d
26. Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res*,*19*(4), 539-549. https://doi.org/10.1007/s11136-010-9606-8
27. Mokkink, L.B., et al., COSMIN Study Design checklist for Patient-reported outcome measurement instruments. Amsterdam, The Netherlands, 2019. 2019: p. 1-32
28. Packham, T. L., Cappelleri, J. C., Sadosky, A., MacDermid, J. C., & Brunner, F. (2017). Measurement properties of painDETECT: Rasch analysis of responses from community-dwelling adults with neuropathic pain. *BMC Neurol*, *17*(1), 48. https://doi.org/10.1186/s12883-017-0825-2
29. Pais-Ribeiro, J. L., Martins da Silva, A., Vilhena, E., Moreira, I., Santos, E., & Mendonca, D. (2018). The hospital anxiety and depression scale, in patients with multiple sclerosis. *Neuropsychiatr Dis Treat*, *14*, 3193-3197. https://doi.org/10.2147/NDT.S184260
30. Parikh, R., Mathai, A., Parikh, S., Chandra Sekhar, G., & Thomas, R. (2008). Understanding and using sensitivity, specificity and predictive values. *Indian J Ophthalmol*, *56*(1), 45-50. https://doi.org/10.4103/0301-4738.37595
31. Perez, C., Galvez, R., Huelbes, S., Insausti, J., Bouhassira, D., Diaz, S., & Rejas, J. (2007). Validity and reliability of the Spanish version of the DN4 (Douleur Neuropathique 4 questions) questionnaire for differential diagnosis of pain syndromes associated to a neuropathic or somatic component. *Health Qual Life Outcomes*, *5*, 66. https://doi.org/10.1186/1477-7525-5-66
32. Rog, D. J., Nurmikko, T. J., Friede, T., & Young, C. A. (2007). Validation and reliability of the Neuropathic Pain Scale (NPS) in multiple sclerosis. *Clin J Pain*, *23*(6), 473-481. https://doi.org/10.1097/AJP.0b013e31805d0c5d
33. Santinelli, F. B., Ramari, C., Poncelet, M., Severijns, D., Kos, D., Pau, M., Kalron, A., Meyns, P., & Feys, P. (2024). Between-Day Reliability of the Gait Characteristics and Their Changes During the 6-Minute Walking Test in People With Multiple Sclerosis.

Neurorehabil Neural Repair, *38*(2), 75-86. https://doi.org/10.1177/15459683231222412

34. Stern, A. F. (2014). The hospital anxiety and depression scale. *Occup Med (Lond)*, *64*(5), 393-394. https://doi.org/10.1093/occmed/kqu024
35. Thompson, A. J., Baranzini, S. E., Geurts, J., Hemmer, B., & Ciccarelli, O. (2018). Multiple sclerosis. *Lancet*,*391*(10130), 1622-1636. https://doi.org/10.1016/S0140-6736(18)30481-1
36. Wu, S. M., & Amtmann, D. (2013). Psychometric Evaluation of the Perceived Stress Scale in Multiple Sclerosis. ISRN Rehabilitation, 2013, 1–9. https://doi.org/10.1155/2013/608356
37. Yilmazer, C., Lamers, I., Solaro, C., & Feys, P. (2022). Clinical perspective on pain in multiple sclerosis. *Mult Scler*, *28*(4), 502-511. https://doi.org/10.1177/1352458520952015