UHASSELT

KNOWLEDGE IN ACTION

Faculteit Revalidatiewetenschappen

master in de revalidatiewetenschappen en de kinesitherapie

Masterthesis

A validity analysis of outcome measures for pain in Multiple Sclerosis

Gunter De Groodt

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

A validity analysis of outcome measures for pain in Multiple Sclerosis

Gunter De Groodt

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

PROMOTOR :

Prof. dr. Peter FEYS

BEGELEIDER :

Mevrouw Cigdem YILMAZER

Contents

RESEARCH CONTEXT

This master's thesis is written in relation to a study by doctoral student Mrs. Çiğdem Yılmazer (internal mentor) from a team led by prof. dr. Peter Feys (internal supervisor), whose research domain centres around neurological rehabilitation. Mrs. Yılmazer's currently ongoing study is titled, "*Reliability and validity of pain outcome measures in persons with Multiple Sclerosis*.", with project code B1152022000014. It was approved by the ethical committee on 19th December 2022. A colleague of Mrs. Yılmazer is conducting an identical study in Chile. Participants from both countries are included in this master's thesis.

Study timeline: The original study by Mrs. Yılmazer was submitted for ethical approval during the autumn of 2022 and it was approved in December. Standardization of the testing procedures was finalized in February 2023. Recruitment of patients with Multiple Sclerosis (MS) and data collection began immediately afterwards and continued until late Spring of 2024. The results were processed and analyzed from January 2024 until the summer of 2024. The author of this master's thesis joined the research team at the end of 2023, when most of the data had already been collected, so the author's main contributions consist of data entry and data processing while writing the master's thesis.

This master's thesis focuses on the validity analysis of pain outcome measures in MS patients; two other master's students will conduct a reliability analysis as part of their master's thesis. To ensure a standard of quality, this master's thesis will adhere to a study design checklist provided by the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN). This is a monothesis, so its contents are written by a single student, with integration of feedback provided by the internal mentor and internal supervisor.

ABSTRACT

Background: The role of pain in people with Multiple Sclerosis (pwMS) has been studied extensively, but many commonly used outcome measures (OMs) for pain have not been validated in pwMS.

Objectives: To assess the validity of commonly used pain OMs in MS, differentiating between neuropathic pain (NP) and non-NP.

Methods: Factor analysis (FA) and hypothesis testing for concurrent validity were used in a sample of 60 Belgian and 50 Chilean pwMS with a score of ≤6 on the Expanded Disability Status Scale. The pain OMs were: Brief Pain Inventory – Short Form (BPI-SF), Douleur Neuropathique 4 (DN4), Neuropathic Pain Scale (NPS), Neuropathic Pain Symptom Inventory (NPSI) and PainDETECT (PDQ).

Results: FA confirmed the factor structure of the pain OMs. Strong correlations (r = >.5) exist between the pain OMs. Weak-to-moderate correlations (r = .3 - .499) exist between the pain OMs (except the DN4) and the clinical OMs for anxiety, cognition, depression, fatigue, sleep quality and stress. No-to-weak correlations (r = <.3) exist between the pain OMs and performance-based OMs for cognitive processing, dexterity, spasticity, strength, and walking mobility. A cluster of the DN4 and PDQ (cut-off ≥13) to detect NP demonstrated good sensitivity (92.3%) but poor specificity (37.1%).

Conclusion: FA provides evidence for the validity of all pain OMs. Concurrent validity could be partially established, as the DN4 seems clearly distinct from the other pain OMs. Including all five pain OMs in test batteries may not be necessary in pwMS.

Keywords: Multiple Sclerosis, Neuropathic Pain, validity.

INTRODUCTION

Multiple Sclerosis (MS) is a disease in which pathological immunoactivity induces irreversible degeneration of primarily white matter in the central nervous system (Dighriri et al., 2023; McIsaac et al., 2019). Damage can occur anywhere in the brain and spinal cord, which is why MS is characterized by a wide variety of motor, sensory and other symptoms (McIsaac et al., 2019). Pain is a commonly reported symptom in MS, with prevalence rates from 29% to 86% (O'Connor et al., 2008).

A Danish study involving 1114 participants found that pain in people with MS (pwMS) is more frequently described as moderate or severe and more often leads to the use of analgesics, compared with age-matched and sex-matched controls without MS (Svendsen et al., 2003). Furthermore, pain in pwMS has been associated with increased levels of anxiety and depression (Boeschoten et al., 2017; Łabuz-Roszak et al., 2019; Marck et al., 2017), lower employment (Ehde et al., 2006; Shahrbanian et al., 2013), greater levels of fatigue (Marck et al., 2017; O'Connor et al., 2008), worse physical functioning (Kratz et al., 2017; O'Connor et al., 2008), decreased quality of life (QoL) (Gil-González et al., 2020; Łabuz-Roszak et al., 2019; O'Connor et al., 2008), poorer sleep quality (Ehde et al., 2006; Merlino et al., 2009) and increased interference in activities of daily living (ADL) (Ehde et al., 2003; O'Connor et al., 2008). Additionally, a systematic review that investigated the relationship between pain and cognition in pwMS found that pain intensity was associated with poorer attention, memory, orientation and processing speed, but not with other cognitive domains (Brown et al., 2023). It is important to note that these associations with pain do not exist in a vacuum. For example, one study found that the impact of pain on depression in pwMS is influenced by anxiety, fatigue and sleep quality (Amtmann et al., 2015). This implies that the interpretation of pain associations in pwMS is probably quite complex and warrants exercising caution.

Special attention should be given to the concept of neuropathic pain (NP) in pwMS, which is any pain that is caused by a lesion to or a disease of the somatosensory system, typically presenting with altered and/or unpleasant sensations, such as burning or shooting pain (*Neuropathic pain*, 2022; Widerström-Noga et al., 2017). Two reviews that attempted to further classify NP by mechanism in pwMS both identified the same three subtypes of pain: (dysesthethic/ongoing) extremity pain, Lhermitte's sign and trigeminal neuralgia (O'Connor et al., 2008; Truini et al., 2013). The prevalence of NP in pwMS is not clear, however, with the estimated prevalence ranging from 21% to 58% (Ouyang et

al., 2024), and the literature about the impact of NP in pwMS is quite limited; a systematic review found that NP in pwMS is associated with increased levels of anxiety and depression, and that NP is more prevalent in women and in patients with the secondary progressive type of MS, compared with pwMS that suffer from other types of chronic pain (Rodrigues et al., 2023). Furthermore, another review of 52 articles found that the presence of NP is associated with greater health-related QoL impairments in six conditions, including MS (Jensen et al., 2007).

Commonly employed outcome measures (OMs) in healthcare to measure pain are the so-called patient-reported outcome measures (PROMs), in which data is gathered by asking the patient to provide it, such as in questionnaires. Even though PROMs reflect subjective experiences, they can be valuable tools when attempting to better understand the impact of symptoms and the perspective of the patient, which is a cornerstone of patient-centered healthcare (McGinley & Lapin, 2022). A 2014 systematic review provides an overview of the validation status of commonly used PROMs in neurological conditions. For MS, only two PROMs for pain have been validated: the Neuropathic Pain Scale (NPS), which assesses pain type and quantifies pain intensity, and the Brief Pain Inventory – Short Form (BPI-SF), which assesses pain severity and its interference with daily functioning and emotional wellbeing (Tyson & Brown, 2014). This master's thesis aims to determine the validity of commonly used (PR)OMs for pain in pwMS by investigating the following hypotheses that are formulated for this master's thesis:

1. Factor analysis of the investigated pain OMs reveals a factor structure in pwMS that is identical to the structure reported in the existing literature.
2. At least moderately strong correlations exist between the investigated pain OMs.
3. Strong correlations exist between pain OMs that measure the same pain construct.
4. At least moderately strong correlations exist between pain OMs and other clinical OMs.
5. Participants classified as having neuropathic pain show worse clinical and functional outcomes than patients classified as having non-neuropathic pain.
6. No significant differences exist between the Belgian and the Chilean samples.

METHODS

Study design and procedure

This master's thesis comprises an observational and comparative study that examines the validity of several pain OMs in pwMS from Belgium and Chile. The test battery, which consists of clinical examinations and self-reported questionnaires, was administered in the participants' native language and each participant was asked to complete all pain-related tests and questionnaires twice, at two testing moments 3-8 days apart. Subgroups for between-group analysis were created for country (Belgium or Chile) and pain type (neuropathic pain (NP) or non-neuropathic pain (NP)). The allocation to the pain subgroups was based on a former diagnosis by a neurologist.

Recruitment

60 participants were recruited from Belgium and 50 participants from Chile. Recruitment of the Belgian participants happened from March 2023 until April 2024 under medical supervision at the regional MS centers located in Overpelt (MS Center Noorderhart), Melsbroek (National MS Center) and in Bruges at the AZ Sint-Jan hospital. Additional participants were recruited at a private physiotherapy facility and through the regional Move-to-Sport network. The Chilean participants were recruited at the Pontificia Universidad Católica de Chile and at the Complejo Asistencial Hospital Dr. Sótero del Río, both of which are situated in Santiago.

Participants

The inclusion criteria were: at least 18 years old; a definite diagnosis of MS, based on the McDonald criteria; a score of 6.0 or lower on the Expanded Disability Status Scale (EDSS); pain in the previous month; and being able to understand and answer questions. The exclusion criteria were: presence of another diagnosed neurological disorder, presence of a major diagnosed musculoskeletal disorder, the presence of only headache, a MS relapse in the previous month, and cognitive incapacity to participate in the tests and questionnaires. These inclusion and exclusion criteria were applied to both testing moments.

Data collection

The clinical tests were conducted by trained researchers and the questionnaires were administered either by a researcher or by a master's student under the supervision of a researcher. Throughout this process, the researchers and the students were blinded to the diagnosis of pain type. The sessions took place in a comfortable and quiet environment. The test battery assesses various aspects of pain as the primary OMs, whereas additional clinical aspects not related to pain are assessed as secondary OMs. In addition to general patient data, these secondary OMs include cognitive processing, fatigue, manual dexterity, mental health, muscle strength, physical activity, sleep, spasticity, stress, and walking mobility. The questionnaires of the pain OMs were administered during both testing moments. The remaining data were collected during the first testing moment and, if necessary, completed during the second testing moment. Each of the two sessions had an expected duration of 90-120 minutes, with additional but optional intermissions as desired by the participants.

Pain outcome measures

The following OMs were used to assess pain: Brief Pain Inventory – Short Form (BPI-SF), Neuropathic Pain Scale (NPS), Neuropathic Pain Symptom Inventory (NPSI), PainDETECT (PDQ), and Doleur Neuropathique 4 (DN4).

The BPI-SF is a PROM that contains nine items and it is the shortened form of the 17-item BPI that was developed by the Pain Research Group from the University of Wisconsin Medical School (Cleeland, 2009). The BPI's original purpose was to explore pain in cancer patients, but it has since also proven useful for the assessment of chronic pain in other medical conditions (Poquet & Lin, 2016). Its first item asks if the patient has experienced any pain today, the second item consists of a full-body diagram to identify the patient's painful body parts, the next four items gauge the patient's perceived pain intensity, the next two items enquire about currently used pain treatments and medication, and the final item of the BPI-SF is a scale that consists of seven subitems that measure pain interference with ADL and with psychosocial functioning in the previous 24 hours (Cleeland, 2009). Factor analysis has identified two factors (*pain interference* and *pain severity*) in patients with cancer, as well as in patients with chronic non-malignant pain (Cleeland, 2009; Tan et al., 2004). Using a three-factor structure (*activity interference*, *affective interference* and *pain intensity* as factors) emerged as a slightly superior approach in HIV patients, but another study that compares multiple factor constructs (one,

two or three) found that the original two-factor approach has greater validity in non-cancer patients (Lapane et al., 2014). The pain-interference scale of the BPI-SF has been validated for MS, demonstrating moderate-to-good associations with pain intensity and good internal consistency (Osborne et al., 2006; Tyson & Brown, 2014).

The NPS is a 10-item PROM developed by the University of Washington School of Medicine. The first seven items quantify the intensity of pain in general and more-specific intensity of pain (general intensity, sharp, hot, dull, cold) by using a numerical scale, the eighth item is about pain duration and frequency, the ninth item enquires about the extent to which the pain is uncomfortable, and the tenth item makes a distinction between superficial and deep pain (Galer & Jensen, 1997). A study pwMS identified three factors: *alien perception*, *familiar* and *superficial* (Rog et al., 2007). The NPS has been validated for MS, demonstrating moderate associations with other measures of pain intensity and with body pain, weak associations with mood and EDSS score, a weak-to-moderate construct validity, and a high internal consistency (Rog et al., 2007; Tyson & Brown, 2014).

The NPSI is a PROM that measures symptoms of NP experienced in the past 24 hours on a numerical scale. It was developed by the INSERM E-322 unit of the pain treatment and evaluation center at the Ambroise-Paré hospital in France. The NPSI consists of 12 items: five items measure pain intensity for different types of pain (burning, electric, pressure, squeezing, stabbing), three items measure if the pain is caused or aggravated by different types of tactile stimuli (light touch, pressure and contact), two items detect the possible presence of unusual feelings in the painful area (making a distinction between tingles and pins and needles), and two more items measure the duration (expressed in hours) of spontaneous pain and pain flare-ups in the past 24 hours. The resulting overall score can be further divided into subscales for superficial spontaneous burning pain, deep spontaneous pressing pain, paroxysmal pain, evoked pain, and paresthesia/dysesthesia (Bouhassira et al., 2004). The authors of the NPSI describe a five-factor structure, which is the generally accepted structure (Bouhassira et al., 2004). The NPSI has not been validated in pwMS.

The PDQ is a PROM developed from multicentre research in Germany and it identifies neuropathic aspects of pain. Although the PDQ was originally developed for patients with low back pain, it is used in other patient populations as well (Freynhagen et al., 2006). The PDQ consists of four parts: In part one, three items are used to ask patients to rate their current pain, their strongest pain (past four weeks) and their average pain (past four weeks) on a numerical scale (0-10). The second part

shows four graphs of different courses of pain from which patients need to select the pain course that best matches theirs. The different pain courses are: persistent pain with slight fluctuations, persistent pain with pain attacks, pain attacks without pain between them, and pain attacks with pain between them. Part three of the questionnaire consists of a body chart on which patients can indicate any radiating pain that they may have experienced. The fourth part comprises seven items that assess the quality of NP symptoms, including different stimuli, and patients provide answers for each item by using a Likert scale. The overall score can be used to determine whether the pain is likely neuropathic (score of ≥19) or not (score of ≤12) (Freynhagen et al., 2006). The authors of the PDQ identified two factors: *pain course pattern* as part two from above and *pain quality* as part four from above (Freynhagen et al., 2006). Given that part four (with the Likert scale) contains seven items, it may be worth investigating if separate factors can be extracted from those seven items. The PDQ has not been validated in pwMS.

The DN4 is a 10-item OM developed by the INSERM E-322 unit of the pain treatment and evaluation center at the Ambroise-Paré hospital in France. The DN4 consists of two parts: questions for the patients (seven items) and a clinical examination of the patient (three items). In the first three items, the patient is asked to indicate if the pain has any of the following characteristics: burning, painful cold, electric shocks. The four remaining items of part one detect if the experienced pain is associated with tingling, pins and needles, numbness or itching. The clinical examination assesses the presence of hypoesthesia to touch and pinprick, and if pain can be caused or aggravated by brushing over the affected area. A total score of ≥4 indicates the presence of NP (Bouhassira et al., 2005). During the creation process, the authors of the DN4 performed factor analysis only on a preliminary version of the DN4 but not on the final 10-item version, so its factor structure is not clear. The DN4 has not been validated in the MS population.

Other outcome measures

Cognitive processing

The Symbol-Digit Modality Test (SDMT) is a timed cognitive test that requires patients to match nine symbols with a corresponding number from one to nine. The test continues for 90 seconds, and the number of correct matches is used as the score. The SDMT was originally developed to detect brain

damage in children (Smith, 1973), but it has since been identified as a valid and reliable instrument to measure cognitive processing speed in pwMS as well (Benedict et al., 2017).

Dexterity

Manual dexterity is measured by using the Nine-Hole Peg Test (NHPT). It is a timed test that requires patients to place nine pegs into tiny holes and to remove them as quickly as possible, using a single hand. The test is performed twice with both arms separately, after which the average completion time is used as the score for each arm. The NHPT is considered the gold standard for measuring manual dexterity and it has been validated in the MS population, showing excellent concurrent and discriminative validity (Feys et al., 2017).

Fatigue

The Modified Fatigue Impact Scale (MFIS) is a shortened version of the Fatigue Impact Scale that was developed as a PROM to measure the impact of fatigue on QoL in pwMS (Fisk et al., 1994). The MFIS consists of 21 items that assess cognitive (10 items), physical (nine items) and psychosocial (two items) functioning in people with complaints of fatigue. Each item is scored on a Likert scale from 0 (never) to 4 (always). A higher total score corresponds with a higher degree of self-perceived fatigue. The MFIS has demonstrated acceptable discriminative and known-groups validity in pwMS (Amtmann et al., 2012), but one source found that the MFIS is not very responsive to change in this population (Rietberg et al., 2010).

Mental health

The Hospital Anxiety and Depression Scale (HADS) consists of an anxiety subscale and a depression subscale, which each contain seven items with four possible answers that are scored on a scale ranging from 0 to 3. One study found that the HADS has good convergent validity and diagnostic validity in pwMS (Jerković et al., 2021), whereas another study reports that the HADS is a valid instrument for the detection of major depression and generalized anxiety disorder in pwMS, but not for other anxiety and depression disorders in this population (Honarmand & Feinstein, 2009).

Muscle strength

The Motricity Index (MI) was originally developed for the assessment of motor skills in people with vascular hemiplegia (Demeurisse et al., 1980). The patient is asked to perform six movements or tasks that require the use of different joints: pinch grip, elbow flexion, shoulder abduction, ankle dorsiflexion, knee extension and hip flexion. Scoring ranges from 0 to 33, but the scoring interpretation is different for the arms and legs. A perfect score indicates the presence of normal strength. The MI has not been validated in pwMS.

Physical activity

The International Physical Activity Questionnaire (IPAQ) is an OM that consists of five sections that measure a person's level of physical activity (PA) across different life domains in the past seven days: work-related PA, transportation-related PA, household-related PA, and leisure-related PA; and the fifth section measures a patient's self-reported average daily sitting time, with a distinction between week days and weekend days (Craig et al., 2003). The scoring system further distinguishes between time spent performing light, moderate and vigorous PA, which can subsequently be used to calculate the patient's total weekly metabolic equivalent (MET-value) as an indicator of total weekly energy expenditure. A recent study, however, determined that the IPAQ is a valid instrument only for measuring PA in people with mild MS (Wanitschek et al., 2024).

Sleep

The Pittsburgh Sleep Quality Index (PSQI) was developed by the University of Pittsburgh for use in psychiatric practice and research (Buysse et al., 1989). The PSQI measures sleep quality throughout the past month, taking into consideration total time spent in bed, total time spent sleeping, physical comfort, the presence of a bedpartner, cognitive and social functioning, and the use of sleeping medication. A high score indicates poor sleep quality. The PSQI has been identified as a valid tool for the assessment of sleep quality in pwMS (Jerković et al., 2022).

Spasticity

Spasticity is assessed by using the modified Ashworth Scale (mAS). The mAS is adapted from the original Ashworth Scale, which was designed to measure spasticity in a study with MS patients

(Ashworth, 1964). Scores on the mAS items range from 0 (no increased muscle tone) to 4 (affected body party is rigid in flexion or extension). The instrument is used to measure muscle tone of the elbow flexors, elbow extensors, wrist flexors, hip flexors, knee extensors and plantar flexors (Bohannon & Smith, 1987). The use of the mAS in patients with MS, however, has been criticized due to a lack of validation in this population and due to the existence of other instruments with better psychometric properties (Harb & Kishner, 2023; Hugos & Cameron, 2019; MohanaSundaram et al., 2023). Nevertheless, the mAS remains a popular instrument in MS research (Harb & Kishner, 2023).

Stress

The Perceived Stress Scale (PSS) is a 10-item PROM, adapted from the original 14-item PSS that was developed to measure perceived stress in students undertaking a smoking-cessation programme (Cohen et al., 1983). The PSS asks how patients might have felt in certain situations in the past month, or how frequently patients experienced a negative emotion, such as anger or irritation. Each item is scored on a Likert scale, ranging from 0 (never) to 4 (very often). The PSS has been validated in pwMS, demonstrating acceptable convergent validity (Wu & Amtmann, 2013).

Walking mobility

The Timed 25-Foot Walk (T25-FW) is a test that measures the fastest safe walking speed over a distance of 25 feet (7.62m) on a flat surface with no obstacles. An average of two consecutive trials is used as the score for this test. Lower scores indicate better walking mobility. The T25-FW has been validated in pwMS, demonstrating good validity for measuring ambulation and some predictive validity for EDSS progression (Kalinowski et al., 2022; Motl et al., 2017).

Data analysis

Exploratory Factor Analysis (EFA) was performed for pain OMs whose factor structure could not be clearly derived from the author's original publication (DN4 and PDQ). This was done by using principal component analysis with varimax rotation and Kaiser normalization, based on the following criteria utilized by the research team: the Kaiser-Meyer-Olkin (KMO) measure of sample adequacy assesses

whether the sample is appropriate for factor analysis. A value of >.6 is considered acceptable. The Bartlett's test of sphericity assesses whether the correlation matrix is factorable (i.e. significantly different from the identity matrix (p <.05)) and therefore suited for EFA. Factor loadings were used to assess if an observed variable is an acceptable (>.4) indicator of a latent factor, with factors requiring an eigenvalue of >1 to be retained in the final factor solution (Beavers et al., 2013; Samuels, 2017).

Confirmatory Factor Analysis (CFA) was performed for all other pain OMs to assess their construct validity. This was conducted by Mrs. Yılmazer because the software applications that are freely available to students are not suited for CFA (e.g. SPSS) or require an extensive learning process (e.g. R programming language). Interpretation of the CFA results provided by Mrs. Yılmazer, however, was performed by the author of this master's thesis. The overall fit of the CFA model was evaluated by using the following model fit indices and cut-off scores (Kyndt & Onghena, 2014): chi-square test of model fit (χ^2 p-value >.05 and $\chi^2/df \leq 3$), comparative fit index (CFI; >.9), Tucker-Lewis index (TLI; >.9), and the Root Mean Square Error of Approximation (RMSEA; values closer to 0 indicate good fit).

Hypothesis testing was conducted to explore the concurrent validity between the different pain OMs on the one hand, and between the pain OMs and other clinical OMs on the other hand. Normality was assessed by using the Shapiro-Wilk test. To compare independent means, an independent t-test or a Mann-Whitney U test was performed. Correlations of continuous variables were assessed by using either Pearson correlation (normal distribution) or Spearman's Rho correlation (non-normal distribution). To assess the relationship between categorical variables, the chi-square test of independence was performed. To assess the relationship between a continuous and a categorical variable, either an independent t-test or a one-way ANOVA was used for normally distributed data, or a Mann-Whitney U test or a Kruskal-Wallis test was used for non-normally distributed data. The interpretation of the correlation coefficients (r) is as follows: <.1 = no correlation, .1 to .299 = weak correlation, .3 to .499 = moderate correlation, ≥.5 = strong correlation (Pallant, 2020). The mean and the standard deviation (*SD*) are reported for normally distributed data, and the median and the interquartile range (*IQR*) are reported for non-normally distributed continuous data. Frequencies are reported for categorical data. Missing data were excluded from the analysis.

All data were analyzed in IBM SPSS Statistics, version 29.0.2.0 (IBM Corporation, Armonk, NY, USA) for Windows. The threshold for significance was set at a value of $p < .05$ (two-tailed).

Ethical considerations

The overarching study received approval from the UHasselt Committee for Medical Ethics and from the ethical committees belonging to the facilities from which MS patients were recruited. Participation in the study was based on the principles of informed consent: After granting permission, all participants were provided with a document that explained the purpose of the study, the contents of the test battery, the procedures to be followed, the possible risks, the rights of the participants (including the absolute right to withdraw from the study at any time without being required to provide a reason), and details on how to contact the people who are involved in conducting or overseeing the study. Participants were given sufficient time to read the document and to consider their potential participation, about which they were allowed to ask questions at any time. Participation was completely voluntary. No financial or other type of compensation was provided in exchange for participating in the study. All data collected and processed as part of this study are stored and handled in accordance with the General Data Protection Regulation (GDPR) and UHasselt's relevant policies.

COSMIN guidelines

To ensure a minimum standard of quality, this master's thesis adheres to a study design checklist provided by the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) (Mokkink et al., 2019). More specifically, this master's thesis follows the general recommendations for study design, as well as the guidelines for hypothesis testing. The checklist has been added to this master's thesis as appendix A.

RESULTS

Participants

The total sample consists of 60 Belgian participants and 50 Chilean participants, with a combined average age of 45.34 years (*SD* = 12.47). Of the 110 participants, 76 (69.1%) are female and 34 (30.9%) are male. The clinical diagnosis of NP was present in 65 (59.1%) participants, absent in 35 (31.8%) participants, and unspecified for the remaining 10 (9.1%) participants. The relapsing-remitting (RR) type of MS was the most prevalent MS type in the sample, with 79 (71.8%) of participants having the RR MS type. Only 13 (11.8%) participants were diagnosed with the primary progressive (PP) MS type and 18 (16.4%) participants with the secondary progressive (SP) MS type. The median EDSS score was 3.00 [*IQR* = 2.60] and the median disease duration in years was 8 [*IQR* = 11]. Fifty-seven (51.8%) participants were employed, 52 (47.3%) unemployed and one participant (.9%) did not provide their employment status. The highest education level obtained was secondary education for 45 (40.9%) participants, college for 48 (43.6%) participants, university for 15 (13.6%) participants, and two (1.8%) participants did not disclose this information.

Among the Belgian and Chilean samples, the mean age was significantly lower in the Chilean sample (p = <.001), with 51.08 (*SD* = 11.61) years in the Belgian sample and 38.48 (*SD* = 10.49) years in the Chilean sample. The median Chilean EDSS score was significantly lower (p = .001) as well, with a median of 3.75 [*IQR* = 3.30] in the Belgian sample and 2.00 [IQR: 2.10] in the Chilean sample. Also the average disease duration was significantly shorter (p = <.001) among the Chilean participants, with a median disease duration of 12 [*IQR* = 11] years in the Belgian sample and 4.5 [*IQR* = 6] years in the Chilean sample. Additionally, the employment status and education level differed significantly (both p = <.001) between the Belgian and Chilean samples: For employment status, 17 (28.3%) Belgians vs 40 (80%) Chileans indicated being currently employed, and 42 (70%) Belgians vs 10 (20%) Chileans indicated being currently unemployed. For education level, 27 (45%) Belgians vs 18 (36%) Chileans reported secondary education as their highest-obtained education level, 17 (28.3%) Belgians vs 31(62%) Chileans indicated college, and 14 (23.3%) Belgians vs one (2%) Chilean indicated university. A summary of these findings is listed in the left half of table 1.

Table 1

Summary of Patient Characteristics

	All participants	Country subgroups			Pain subgroups		
	Total (n = 91-110)	Belgium (n = 41-60)	Chile (n = 50)	Difference (p-value)	Neuropathic (n = 65)	Non-neuropathic (n = 35)	Difference (p-value)
General characteristics							
Age (y)	45.34 (12.47)	51.08 (11.61)	38.48 (10.49)	<.001*	44.00 (11.86)	44.63 (12.55)	.805
EDSS score	3.00 [2.60]	3.75 [3.30]	2.00 [2.10]	.001*	3.00 [2.8]	2.50 [2.50]	.584
Disease duration (y)	8 [11]	12 [11]	4.50 [6]	<.001*	7 [11]	8 [8]	.531
Gender (n (%))							
Female	76 (69.1%)	42 (70%)	34 (68%)	.821	48 (73.8%)	24 (68.6%)	.575
Male	34 (30.9%)	18 (30%)	16 (32%)		17 (26.2%)	11 (31.4%)	
MS type (n (%))							
PP	13 (11.8%)	10 (16.7%)	3 (6%)	.210	7 (10.8%)	4 (11.4%)	.657
RR	79 (71.8%)	40 (66.7%)	39 (78%)		46 (70.8%)	27 (77.1%)	
SP	18 (16.4%)	10 (16.7%)	8 (16%)		12 (18.5%)	4 (11.4%)	
Pain diagnosis (n (%))							
Neuropathic pain	65 (59.1%)	30 (50%)	35 (70%)		n/a	n/a	
Non-neuropathic pain	35 (31.8%)	20 (33.3%)	15 (30%)		n/a	n/a	
Not specified	10 (9.1%)	10 (16.7%)	0 (0%)		n/a	n/a	
Employment status (n (%))							
Employed	57 (51.8%)	17 (28.3%)	40 (80%)	<.001*	35 (53.8%)	19 (54.3%)	.969
Unemployed	52 (47.3%)	42 (70%)	10 (20%)		29 (44.6%)	16 (45.7%)	
Not specified	1 (.9%)	1 (1.7%)	0 (0%)		1 (1.5%)	0 (0%)	
Education level (n (%))							
Secondary	45 (40.9%)	27 (45%)	18 (36%)	<.001*	24 (36.9%)	15 (42.9%)	.013*
College	48 (43.6%)	17 (28.3%)	31 (62%)		35 (5.8%)	10 (28.6%)	
University	15 (13.6%)	14 (23.3%)	1 (2%)		5 (7.7%)	9 (25.7%)	
Not specified	2 (1.8%)	2 (3.3%)	0 (0%		1 (1.5%)	1 (2.9%)	

Note. For normally distributed continuous outcomes: mean with *SD* between round brackets; for non-normally distributed continuous outcomes: median with *IQR* between square brackets. For categorical data, frequencies with corresponding percentage between round brackets. EDSS = Expanded Disability Status Scale; MS = Multiple Sclerosis; n/a = not applicable; PP = Primary Progressive; RR = Relapsing Remitting; SP = Secondary Progressive.

* $p < .05$

Between the the subgroups of NP and non-NP, only the education level differed significantly (p = .013), with 24 (36.9%) NP participants vs 15 (42.9%) non-NP participants indicating secondary education as their highest-obtained level of education, 35 (5.8%) NP participants vs 10 (28.6%) non-NP participants indicating college, and five (7.7%) NP participants vs nine (25.7%) non-NP participants indicating university. A summary of all participant characteristics with subgroup differences is listed in the right half of table 1.

Exploratory Factor Analysis

For the DN4, the KMO value of the total sample is .741 (>.5) and the result of the Bartlett's test of sphericity is significant: x^2 (n = 110) = 246.829 (p = < .001). This means that the use of EFA is appropriate in this sample. EFA has identified three factors for the DN4: factor one with an eigenvalue of 3.234, consisting of items 8, 9 and 10 (*hypoesthesia to touch*, *hypoesthesia to prick*, and *brushing*); factor two with an eigenvalue of 1.354, consisting of items 2, 3, 4 and 6 (*painful cold*, *electric shocks*, *tingling*, and *numbness*); and factor three with an eigenvalue of 1.244, consisting of items 1, 5 and 7 (*burning*, *pins and needles*, and *itching*). Combined, these three factors explain 58.31% of the total variance. The left half of table 2 shows an overview of the EFA results for the DN4.

For the PDQ, the KMO value of the total sample is .787 (>.5) and the result of the Bartlett's test of sphericity is significant: x^2 (n = 110) = 190.618 (p = < .001). This means that the use of EFA is appropriate in this sample. EFA has identified two factors for the PDQ: factor one with an eigenvalue of 3.069, consisting of items 1, 2, 4 and 6 (*burning*, *tingling*, *electric shocks*, and *numbness*); and factor two with an eigenvalue of 1.100, consisting of items 3, 5 and 7 (*painful light touch*, *painful cold*, and *slight pressure*). Combined, these three factors explain 59.55% of the total variance. The right half of table 2 shows an overview of the EFA results for the PDQ.

Table 2.

Results of Exploratory Factor Analysis

	Components				Components	
DN4	1	2	3	PDQ	1	2
1. Burning	.301	.381	**.468**	1. Burning	**.647**	.098
2. Painful cold	-.059	**.441**	.405	2. Tingling	**.798**	.146
3. Electric shocks	-.035	**.483**	.463	3. Painful light touch	.361	**.697**
4. Tingling	.292	**.792**	.128	4. Electric shocks	**.682**	.163
5. Pins and needles	.059	.036	**.777**	5. Painful cold	.307	**.710**
6. Numbness	.161	**.859**	-.125	6. Numbness	**.728**	.301
7. Itching	.274	-.038	**.684**	7. Slight pressure	-.015	**.869**
8. Hypoesthesia to touch	**.830**	.174	.079			
9. Hypoesthesia to prick	**.767**	.113	.067			
10. Brushing	**.736**	.057	.166			
Cronbach α[a]	.724	.650	.419		.721	.706
Mean inter-item correlation[a]	.466	.319	.265		.399	.445
Eigenvalue	3.234	1.354	1.244		3.069	1.100
% of Variance	32.33	13.53	12.43		43.84	15.71
% Total Variance	58.31				59.55	
KMO (sample adequacy)	.741				.787	
Bartlett's test	246.829 (p <.001)				190.618 (p <.001)	

Note. DN4 = Douleur Neuropathique 4; PDQ = PainDETECT; KMO = Kaiser-Meyer-Olkin.

[a] Of items in bold.

Confirmatory Factor Analysis

The CFA results are displayed in table 3. All indices confirm the factor structure of the DN4 and the PDQ. For each of the BPI-SF, the NPS and the NPSI, one or more of the indices does not reach the threshold of acceptability, indicating that caution should be exercised with accepting the factor structure. Nevertheless, χ^2 and TLI are influenced by sample size, so given the relatively small sample in the present study, more attention should be given to CFI, RMSEA and χ^2/df, which are not affected by sample size (Schermelleh-Engel et al., 2003). With the exception of the RMSEA for the BPI-SF and the NPSI, all three of these fit indices suggest acceptable model fit for the five pain OMs, so it seems reasonable to accept the factor structure proposed by the authors of the BPI-SF, the NPS and the NPSI as well.

Table 3.

Confirmatory Factor analysis: Indices for Goodness of Fit for the Pain Outcome Measures

	χ^2 p-value	CFI	TLI	RMSEA	χ^2/df
BPI-SF	.00	**.92**	**.90**	.11	**2.49**
DN4	**.07**	**.94**	**.91**	**.05**	**1.38**
PDQ	**.22**	**.98**	**.96**	**.04**	**1.26**
NPS	**.02**	**.93**	**.90**	**.07**	**1.61**
NPSI	.00	**.93**	.88	.09	**1.95**
Acceptable range	*p* >.05	>.90	>.90	<.09	≤3

Note. BPI-SF = Brief Pain Inventory – Short Form; DN4 = Douleur Neuropathique 4; PDQ = PainDETECT; NPS = Neuropathic Pain Scale; NPSI = Neuropathic Pain Symptom Inventory; CFI = Comparative Fit Index; df = degrees of freedom; RMSEA = Root Mean Square Error of Approximation; TLI = Tucker-Lewis Index.

Table 4 contains an overview of the factor structure of each pain OM, based on the literature provided in the Methods section and with integration of the EFA and CFA results.

Table 4

Overview of items and factors of the pain outcome measures

Item	BSI-SF	DN4	NPS	NPSI	PDQ
1	Pain today	Burning	Int.: general	Int.: burning	Int.: now
2	Body chart	Cold	Int.: sharp	Int.: squeezing	Int.: most
3	Int.: worst	Electric shocks	Int.: burning	Int.: pressure	Int.: average
4	Int.: least	Tingling	Int.: dull	Spontaneous pain	Pain pattern
5	Int.: average	Paresthesia	Int.: cold	Int.: electric	Body chart
6	Int.: now	Numbness	Int.: sensitivity	Int.: stabbing	Burning
7	Medication	Itching	Int.: itchy	Pain attacks	Tingling
8	Alleviation	Hypoesthesia	Time quality	Prov.: brushing	Prov.: touch
9	ADL	Hypoesthesia	Int.: unpleasant	Prov.: pressure	Pain attacks
10	Psychosocial	Prov. brushing	10a: Int.: deep	Prov.: cold	Prov.: cold/heat
11	ADL		10b.: int.: surface	Int.: paresthesia	Numbness
12	ADL			Int: tingling	Prov.: pressure
13	Psychosocial				
14	Sleep				
15	Psychosocial				

Note. Each color, except white, represents items that form a factor within the outcome measure. BPI-SF = Brief Pain Inventory – Short Form; DN4 = Douleur Neuropathique 4; PDQ = PainDETECT; NPS = Neuropathic Pain Scale; NPSI = Neuropathic Pain Symptom Inventory; ADL = activities of daily living; int. = intensity; prov. = provocation.

Concurrent validity among pain outcome measures

A correlation matrix was made for the BPI-SF (severity scale and interference scale), the DN4, the NPS, the NPSI and the PDQ (see table 5). All correlations between the OMs were significant (p = <.05), with the correlation coefficients ranging from .215 to .791. The weakest correlations were found for the DN4 with the BPI-I (r(108) = .215) and for the DN4 with the BPI-S (r(108) = .221), suggesting that the DN4 correlates only weakly with the BPI scales. The strongest correlations were found for the NPSI with the PDQ (r(108) = .791), for the NPSI with the NPS (r(108) = .788), for the NPSI with the BPI-S (r(108) = .737), and for the PDQ with the NPS (r(108) = .735). The NPSI thus correlates strongly with three other pain OMs.

Table 5

Correlations between Pain Questionnaires

	BPI-I	BPI-S	NPS	NPSI	PDQ
BPI-I					
BPI-S	.676***				
NPS	.587***	.665***			
NPSI	.607***	.737***	.788***		
PDQ	.470***	.570***	.735***	.791***	
DN4	.215*	.221*	.406***	.475***	.652***

Note. BPI-I = Brief Pain Inventory – Short Form: interference scale; BPI-S = Brief Pain Inventory – Short Form: severity scale; DN4 = Douleur Neuropathique 4; PDQ = PainDETECT; NPS = Neuropathic Pain Scale; NPSI = Neuropathic Pain Symptom Inventory.

* p < .05, ** p <.01, *** p <.001

The DN4 correlates strongly with the PDQ ($r(108) = .652$), which is to be expected, given that both pain OMs attempt to classify patients as having either having NP or non-NP. Nevertheless, the fact that the correlation is not excellent or near perfect indicates that both OMs are not completely identical in their classification method and that both OMs contain unique elements.

DN4 and PDQ as screening tools for neuropathic pain

Using the neurologists' diagnosis of NP as the gold standard, the sensitivity and specificity of the DN4 and PDQ can be calculated. For PDQ scores ranging from 13 to 18, it is unclear if a patient's pain is likely NP or non-NP. Sensitivity and specificity were therefore also calculated by using ≥13 as the cut-off score, in addition to the normal cut-off score of ≥19. The results are displayed in table 6. The PDQ with its normal cut-off score of ≥19 shows extremely poor sensitivity (50.8%) in the current sample, but this increases to an acceptable 80.0% when using ≥13 as the cut-off score instead. The DN4 demonstrates acceptable sensitivity (89.2%) as well. Combining the DN4 with the PDQ with a cut-off score of ≥13 increases the sensitivity to 92.3%. Combining the DN4 with the PDQ with a cut-off score of ≥19 provides no improvements, compared with using just the DN4. Closer inspection of the dataset reveals that this is because the PDQ with a cut-off score of ≥19 never succeeds in correctly detecting NP in cases where the DN4 incorrectly fails to detect NP. The specificity of the DN4, PDQ and all combinations is poor, ranging from 42.9% to 71.4%.

Table 6

Sensitivity and Specificity for Detection of Neuropathic Pain

	Sensitivity	Specificity
DN4	89.2%	42.9%
PDQ (cut-off ≥19)	50.8%	71.4%
PDQ (cut-off ≥13)	80.0%	54.3%
DN4 + PDQ (cut-off ≥19)	89.2%	42.9%
DN4 + PDQ (cut-off ≥13)	92.3%	37.1%

Note. DN4 = Douleur Neuropathique 4; PDQ = PainDETECT.

Concurrent validity between pain outcome measures and other outcome measures

Another correlation matrix was made to compare the pain OMs with the OMs of other clinical outcomes (see table 7). The DN4 correlates significantly only with the IPAQ (p = .049), but the correlation coefficient is weak (r(99) = -.196).

The interference scale of the BPI-SF correlates significantly with the HADS total score (r(103) = .394, p = <.001), the anxiety scale of the HADS (r(103) = .328; p = <.001), the depression scale of the HADS (r(103) = .399, p = <.001), the MFIS total score (r(105) = .398, p = <.001), the cognitive scale of the MFIS (r(105) = .292, p = <.001), the physical scale of the MFIS (r(105) = .474, p = <.001), the psychosocial scale of the MFIS (r(105) = .352, p = <.001), the PSQI (r(100) = 369, p = <.001), the PSS (r(89) = .427, p = <.001), and the T25FW (r(107) = .214, p = .025).

The severity scale of the BPI-SF correlates significantly with the HADS total score (r(103) = .325, p = <.001), the anxiety scale of the HADS (r(103) = .246, p = .011), the depression scale of the HADS (r(103) = .339, p = <.001), the MFIS total score (r(105) = .238, p = .013), the cognitive scale of the MFIS (r(105) = .199, p = .040), the physical scale of the MFIS (r(105) = .276, p = .004), the PSQI (r(100) = .252, p = .011), the PSS (r(89) = .284, p = .006), and the T25FW (r(107) = .213, p = .026).

The NPS correlates significantly with the HADS total score (r(103) = .218, p = .026), the depression scale of the HADS (r(103) = .271, p = .005), the MFIS total score (r(105) = .327, p = <.001), the cognitive scale of the MFIS (r(105) = .283, p = .003), the physical scale of the MFIS (r(105) = .338, p = <.001), the PSS (r(89) = .332, p = .001), and the T25FW (r(107) = .200, p = .037).

The NPSI correlates significantly with the HADS total score (r(103) = .279, p = .004), the anxiety scale of the HADS (r(103) = .231, p = .018), the depression scale of the HADS (r(103) = .285, p = .003), the MFIS total score (r(105) = .347, p = <.001), the cognitive scale of the MFIS (r(105) = .326, p = <.001), the physical scale of the MFIS (r(105) = .341 p = <.001), the psychosocial scale of the MFIS (r(105) = .249, p = .010), the PSQI (r(100) = .210, p = .034), the PSS (r(89) = .217, p = .039), and the T25FW (r(107) = .248, p = .009).

The PDQ correlates significantly with the HADS total score (r(103) = .312, p = .001), the anxiety scale of the HADS (r(103) = .243, p = .013), the depression scale of the HADS (r(103) = .305, p = .002), the MFIS total score (r(105) = .341, p = <.001), the cognitive scale of the MFIS (r(105) = .327, p = <.001),

the physical scale of the MFIS ($r(105) = .267$, $p = .047$), the psychosocial scale of the MFIS ($r(105) = .192$, $p = .013$), the PSQI ($r(100) = .221$, $p = .025$), and the PSS ($r(89) = .271$, $p = .009$).

Table 7

Correlations between Pain Outcome Measures and Other Outcome Measures

	BPI-I	BPI-S	NPS	NPSI	DN4	PDQ
HADS: total (0-42)	.394***	.325***	.218*	.279**	.144	.312**
HADS: anxiety (0-21)	.328***	.246*	.115	.231*	.109	.243*
HADS: depression (0-21)	.399***	.339***	.271**	.285**	.138	.305**
IPAQ (n)	-.065	.052	.030	.028	-.196*	.019
mAS (0-4)	.079	-.135	-.138	-.115	.170	-.088
MFIS: total (0-84)	.398***	.238*	.327***	.347***	.110	.341***
MFIS: cognitive (0-40)	.292***	.199*	.283**	.326***	.114	.327***
MFIS: physical (0-36)	.474***	.276**	.338***	.341***	.066	.267*
MFIS: psychosocial (0-8)	.352***	.186	.189	.249*	.073	.192*
MI (0-100)	-.108	.028	-.028	-.039	-.106	-.033
NHPT (s)	.046	.110	.107	.140	.063	.049
PSQI (0-21)	.369***	.252*	.146	.210*	.005	.221*
PSS (0-40)	.427***	.284**	.332**	.217*	-.042	.271**
SDMT (N)	-.094	-.159	-.041	-.118	-.186	-.157
T25FW (s)	.214*	.213*	.200*	.248**	-.020	.059

Note. Ranges and units in left column. Only significant correlations have been colour-coded. BPI-I = Brief Pain Inventory – Short Form: interference scale; BPI-S = Brief Pain Inventory – Short Form: severity scale; DN4 = Douleur Neuropathique 4; PDQ = PainDETECT; NPS = Neuropathic Pain Scale; NPSI = Neuropathic Pain Symptom Inventory; HADS = Hospital Anxiety and Depression Scale; IPAQ = International Physical Activity Questionnaire; mAS = modified Ashworth Scale; MFIS = Modified Fatigue Impact Scale; MI = Motricity Index; NHPT = Nine-Hole Peg Test; PSQI = Pittsburgh Sleep Quality Index; PSS = Perceived Stress Scale; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.

* $p < .05$, ** $p < .01$, *** $p < .001$

Subgroup analysis: Neuropathic pain vs non-neuropathic pain

For the pain OMs, participants in the NP group scored significantly higher on the severity scale of the BPI-SF (p = .038), the DN4 (p = <.001), the NPSI (p = .016) and the PDQ (p = .002). For the severity scale of the BPI-SF, the NP group scored 4.90 (*SD* = 1.86) vs 3.29 [*IQR* = 4.71] for the non-NP group. For the DN4, the NP group scored 7 [*IQR* = 4] vs 3.83 (*SD* = 2.64) for the non-NP group. For the NPSI, the NP group scored 38.06 (*SD* = 19.74) vs 23 [*IQR* = 38] for the non-NP group. For the PDQ, the NP group scored 18.08 (*SD* = 7.20) vs 13.09 (*SD* = 8.02) for the non-NP group.

For the other OMs, only the IPAQ produced significantly different results (p = .026) between the NP group and the non-NP group. In the NP group, 40 (61.5%) people scored in the *low activity* category vs 11 (31.4%) in the non-NP group, 10 (15.4%) people in the NP group scored in the *moderate activity* category vs 11 (31.4%) in the non-NP group, 11 (16.9%) people in the NP group scored in the *high activity* category vs eight (22.9%) in the non-NP group, and no IPAQ category was specified for four (6.2%) people in the NP group and for five (14.3%) people in the non-NP group.

An overview of the subgroup analysis for the NP group vs the non-NP group is listed in the right half of table 8.

Subgroup analysis: Belgium vs Chile

For the pain OMs, Belgian participants scored significantly higher on the NPS (p = .003) and the severity scale of the BPI-SF (p = .014), but significantly lower on the DN4 (p = .003), compared with Chilean participants. It is worth noting that the NPSI scored barely above the significance threshold (p = .051), with the Belgian sample obtaining higher scores. For the severity scale of the BPI-SF, Belgians scored 4.98 (*SD* = 2.63) vs 0.50 [*IQR* = 1.88] in the Chilean sample. For the DN4, Belgians scored 4.62 (*SD* = 2.43) vs 6.50 [*IQR* = 4] in the Chilean sample. For the NPS, Belgians scored 48.17 (*SD* = 19.15) vs 37.52 (*SD* = 15.00) in the Chilean sample.

Table 8

Results of Subgroup Analysis for Country and Pain Type

	All participants	Country subgroups			Pain subgroups		
	Total (n = 91-110)	Belgium (n = 41-60)	Chile (n = 50)	Difference (p-value)	Neuropathic (n = 65)	Non-neuropathic (n = 35)	Difference (p-value)
Pain outcome measures							
BPI-SF: interference (0-10)	3.43 [3.82]	4.25 (2.63)	3.14 [3.50]	.202	3.86 [3.50]	3.29 [4.71]	.198
BPI-SF: severity (0-10)	4.25 [2.50]	4.98 (2.00)	0.50 [1.88]	.014*	4.90 (1.86)	4.25 [3.00]	.038*
DN4 (0-10)	5 [3]	4.62 (2.43)	6.50 [4]	.003*	7 [4]	3.83 (2.64)	<.001*
NPS (0-100)	41.50 [24.00]	48.17 (19.15)	37.52 (15.00)	.003*	45.63 (16.70)	39.43 (20.79)	.060
NPSI (0-100)	30.50 [30.00]	39.20 (22.66)	27.00 [33]	.051	38.06 (19.74)	23 [38]	.016*
PainDETECT (0-38)	16.18 (7.66)	17.08 (7.61)	15.10 (7.65)	.177	18.08 (7.20)	13.09 (8.02)	.002*
Other outcome measures							
HADS total (0-42)	12 [11]	14 [12]	11 [12]	.295	13 [11]	12.44 (7.57)	.442
HADS: anxiety (0-21)	8 [7]	7 [7]	7.66 (4.03)	.913	7.90 (4.31)	7.66 (4.43)	.804
HADS: depression (0-21)	4 [7]	6 [8]	3 [5]	.063	5 [8]	4 [7]	.229
IPAQ (n (%))							
Low	55 (50%)	19 (31.7%)	36 (72%)	<.001*	40 (61.5%)	11 (31.4%)	.026*
Moderate	24 (21.8%)	10 (16.7%)	14 (28%)		10 (15.4%)	11 (31.4%)	
High	22 (20.0%)	22 (36.7%)	0 (0%)		11 (16.9%)	8 (22.9%)	
Not specified	9 (8.2%)	9 (15%)	0 (0%)		4 (6.2%)	5 (14.3%)	
mAS (0-4)	0 [1]	0 [0]	1 [1]	<.001*	0 [1]	0 [1]	.217
MFIS: total (0-84)	42.47 (17.38)	46.77 (12.97)	37.56 (20.38)	.007*	42.25 (15.90)	42.82 (21.28)	.883
MFIS: cognitive (0-40)	18.02 (9.80)	19.93 (8.25)	15 [18]	.034*	17.77 (8.98)	19.00 (11.82)	.601
MFIS: physical (0-36)	22 [11]	22 [7]	18.48 (9.09)	.036*	20.83 (7.32)	20.12 (9.60)	.599
MFIS: psychosocial (0-8)	4 [4]	4 [3]	3 [4]	.052	4 [4]	3.70 (2.31)	.985
MI: left (0-100)	100 [8.5]	100 [13.00]	100 [4]	.353	100 [10.0]	100 [6.0]	.662
MI: right (0-100)	100 [4]	100 [12.30]	100 [4]	.261	100 [4.4]	100 [8.0]	.882
NHPT: dominant (s)	19.30 [5.31]	20.78 [6.52]	18.25 [4.90]	.008*	19.30 [5.65]	18.99 [5.32]	.499
NHPT: non-dominant (s)	20.73 [6.70]	21.71 [6.80]	19.25 [4.50]	.011*	20.75 [7.00]	19.50 [3.65]	.179
NHPT: mean (s)	20.25 [6.10]	22.10 [8.51]	18.85 [4.41]	.002*	20.15 [7.10]	19.71 [5.22]	.390
PSQI (0-21)	8.56 (4.25)	8.54 (4.23)	8.58 (4.28)	.961	9.02 (4.14)	8.13 (4.06)	.337
PSS (0-40)	14.75 (7.62)	17.51 (7.43)	12.48 (7.06)	.001*	14.93 (7.30)	14.53 (8.37)	.814
SDMT (N)	53.03 (12.89)	53.02 (12.10)	53.04 (13.74)	.994	53.03 (13.53)	53.00 (12.21)	.991
T25FW (s)	5.00 [2.94]	5.65 [3.08]	4.10 [1.11]	<.001*	4.80 [3.31]	4.81 [1.65]	.362

Note. Ranges and units in left column. BPI-SF = Brief Pain Inventory – Short Form; DN4 = Douleur Neuropathique 4; NPS = Neuropathic Pain Scale; NPSI = Neuropathic Pain Symptom Inventory; HADS = Hospital Anxiety and Depression Scale; IPAQ = International Physical Activity Questionnaire; mAS = modified Ashworth Scale; MFIS = Modified Fatigue Impact Scale; MI = Motricity Index; NHPT = Nine-Hole Peg Test; PSQI = Pittsburgh Sleep Quality Index; PSS = Perceived Stress Scale; SDMT = Symbol Digit Modalities Test; T25FW = Timed 25-Foot Walk.

* $p < .05$

For the other clinical outcomes, the Belgian sample scored significantly higher (i.e. worse) on the total MFIS (p = .007), the cognitive scale of the MFIS (p = .034), the physical scale of the MFIS (p = .036), the NHPT score of the dominant arm (p = .008), the NHPT score of the non-dominant arm (p = .011), the mean NHPT score (p = .002), the PSS (p = .001) and the T25FW (p = <.001). The Chilean sample scored significantly higher (i.e. worse) on the mAS (p = <.001) and demonstrated significantly lower levels of physical activity on the IPAQ (p = <.001).

For the total score of the MFIS, the mean score was 46.77 (*SD* = 19.97) in the Belgian sample and 37.56 (*SD* = 20.38) in the Chilean sample. For the cognitive scale of the MFIS, the mean score was 19.93 (*SD* = 8.25) in the Belgian sample with a median score of 15 [*IQR* = 18] in the Chilean sample. For the physical scale of the MFIS, the median score was 22 [*IQR* = 7] in the Belgian sample with a mean score of 18.48 (*SD* = 9.09) in the Chilean sample.

For the NHPT score of the dominant hand, the median score was 20.78 [*IQR* = 6.52] seconds in the Belgian sample and 18.25 [*IQR* = 4.90] seconds in the Chilean sample. For the NHPT score of the non-dominant hand, the median score was 21.71 [*IQR* = 6.80] seconds in the Belgian sample and 19.25 [*IQR* = 4.50] seconds in the Chilean sample. For the mean NHPT score (using the scores of both the dominant and the non-dominant arm), the median score was 22.10 [*IQR* = 8.51] seconds in the Belgian sample and 18.85 [*IQR* = 4.41] seconds in the Chilean sample.

For the PSS, the mean score was 17.51 (*SD* = 7.43) in the Belgian sample and 12.48 (*SD* = 7.06) in the Chilean sample.

For the T25FW, the median score was 5.65 [*IQR* = 3.08] seconds in the Belgian sample and 4.10 [IQR = *1.11*] seconds in the Chilean sample.

For the mAS, the median score of the Belgian sample was 0 [*IQR* = 0] and 1 [*IQR* = 1] in the Chilean sample.

For the IPAQ, 19 (31.7%) Belgians scored in the *low activity* category vs 36 (72%) in the Chilean sample, 10 (16.7%) Belgians scored in the *moderate activity* category vs 14 (28%) in the Chilean sample, 22 (36.7%) Belgians scored in the *high activity* category vs no one in the Chilean sample, and no IPAQ category was determined for nine (8.2%) Belgians.

An overview of the subgroup analysis for the Belgian sample vs the Chilean sample is listed in the left half of table 8.

DISCUSSION

The objective of this master's thesis was to assess the validity of the BPI-SF, DN4, NPS, NPSI and PDQ in pwMS through a combination of factor analysis and hypothesis testing, and that while differentiating between pwMS with NP and pwMS without NP, so that recommendations about the use of the pain OMs can be made in this population. In the Introduction, the following hypotheses were formulated, which will be discussed in this section of the master's thesis:

1. Factor analysis of the investigated pain OMs reveals a factor structure in pwMS that is identical to the structure reported in the existing literature (accepted).
2. At least moderately strong correlations exist between the investigated pain OMs (accepted).
3. Strong correlations exist between pain OMs that measure the same pain construct (accepted).
4. At least moderately strong correlations exist between pain OMs and other clinical OMs (partially rejected).
5. Participants classified as having neuropathic pain show worse clinical and functional outcomes than patients classified as having non-neuropathic pain (partially rejected).
6. No significant differences exist between the Belgian and the Chilean samples (inconclusive).

Factor analysis confirmed the structure of the BPI-SF, NPS and NPSI in the current sample, which is a finding that aids in supporting the validity of these instruments in pwMS. The first hypothesis is thus accepted. Furthermore, EFA identified three factors for the DN4, with Cronbach α values ranging from .42 to .72. Normally, values of <.7 indicate unacceptable internal consistency, but low Cronbach α values can also be caused by a low number (<10) of test items. In that case, it is recommended to calculate the mean inter-item correlation coefficient instead, with values above .2 indicating acceptable internal consistency (Pallant, 2020). In the current study, the mean inter-item correlation coefficients for the three DN4 factors range from .265 to .466, thus demonstrating acceptable internal consistency.

These DN4 findings, however, do not fully agree with the existing literature. For example, a Dutch multicenter study involving 228 patients with chronic pain identified four factors. Although the authors did not report the mean inter-item correlation coefficients, the worst Cronbach α values they reported (α = .37 and α = .51) are lower than the ones found in the current study, with one factor consisting of only one item (Timmerman et al., 2017). Based on these results, a three-factor model seems better suited. In addition, the factor structure of the Portuguese version of the DN4 was assessed in a sample of 101 patients with either nociceptive or neuropathic pain. The authors identified three factors, but the internal consistency of each factor was not assessed, and the internal factor structures differ from the ones found in the current study (Santos et al., 2010). It should be noted that factor analysis is better suited for larger samples (MacCallum et al., 1999) and that the sample sizes of the current study and the studies cited above may not be sufficiently large, which could explain the discrepancies between the studies. The factor structure of the DN4 therefore remains somewhat unclear. EFA identified two factors for the seven-item part of the PDQ with acceptable internal consistency (α = .72 and α = .71), which confirms the findings from a validity analysis performed for the Japanese version of the PDQ (Matsubayashi et al., 2013).

Correlation analysis between the pain OMs reveals that the correlation coefficients range from weak to strong. Of the 15 possible combinations in the correlation matrix, 10 (66.67%) combinations show strong correlations, and every pain OM correlates strongly with at least one of the five other pain OMs. This means that the second hypothesis about the existence of at least moderate correlations between the pain OMs is accepted. Excluding the DN4, all correlations between the pain OMs are of at least moderate strength. When considering all 15 possible combinations of pain OMs, however, four of the five weakest correlation coefficients include the DN4, which correlates only weakly with both subscales of the BPI-SF, and moderately with the NPS and the NPSI. The reason why this is the case might be due to the design of the DN4: First, the DN4 consists of simple yes-no statements that do not allow interpreting the intensity or severity of the symptoms, whereas the other pain OMs contain Likert scales that do allow non-binary and nuanced responses, including the interpretation of intensity and severity. Second, the last three items of the DN4 involve a clinical assessment by administering physical stimuli, whereas the other pain OMs lack such a clinical examination. These differences could explain

the weaker correlations of the DN4. Given that the DN4 seems to be clearly distinct from the other pain OMs, including it in test batteries to assess pain in pwMS is recommended.

The third hypothesis concerns strong correlations between pain OMs that measure the same pain construct. This means that DN4 and the PDQ should correlate strongly with each other, and that the NPS and the NPSI should correlate strongly with each other. The results indicate that this is the case for both the DN4-PDQ combination and the NPS-NPSI combination, which means that the third hypothesis about correlations between pain OMs that measure the same pain construct is accepted. Given that both the DN4 and the PDQ are used to screen for NP, this strong correlation is expected. Nevertheless, the absence of an even stronger correlation indicates that differences exist between the DN4 and the PDQ, which strengthens the recommendation of including the DN4 in addition to the PDQ when screening for NP in pwMS. Other notably strong combinations from the correlation matrix are NPSI - PDQ (r = .791), NPSI - BPI-S (r = .737), and NPS - PDQ (r = .735). In situations where the available time is limited, it may therefore be worth evaluating if one or more of these pain OMs can be excluded from the test battery when assessing pain in pwMS. Given that three of the four strongest pairings include the NPSI, however, it would be advisable to retain at least the NPSI in such test batteries.

Sensitivity and specificity analysis for the classification of NP indicates that the PDQ with its normal cut-off score of ≥19 should not be used in isolation to detect NP due to its extremely poor sensitivity (50.8%). Using a cut-off score of ≥13 instead ≥19 improves the sensitivity to 80.0%, but this is still relatively low. The DN4 shows acceptable sensitivity (89.2%), and even better sensitivity (92.3%) when combined with the PDQ with a cut-off score of >13%. The specificity of both pain OMs (and combinations) is poor, with the values indicating that DN4, PDQ and their combined use are generally better at detecting NP than at ruling out the condition. Compared with the existing literature, a systematic review about the psychometric properties of the DN4 and PDQ in various populations with chronic pain found sensitivity values for the DN4 with a cut-off score of ≥4 that range from 80% to 96% and specificity values ranging from 75.4% to 97.2%, and sensitivity values for the PDQ with a cut-off score of ≥19 that range from 71% to 79% and specificity values ranging from 83% to 93% (Fagbohun, 2021). These sensitivity values are similar to the ones found in the present study, but the specificity values are considerably higher.

Based on these findings, some conclusions can be drawn for the detection of NP in pwMS: First, when using the PDQ, it seems better to use the cut-off score of ≥13 instead of the normal cut-off score of ≥19. Second, neither the DN4 nor the PDQ should be used in isolation. Instead, it seems better to combine both the DN4 and the PDQ (with a cut-off of ≥13), and to administer them as a test cluster. Third, given that neither of these instruments (or combinations) is able to detect every single case of NP, consulting with a clinician remains strongly recommended. It should be noted, however, that the diagnosis of NP was not performed as part of the current study. Instead, information about an existing diagnosis in the past was provided by the participants or their caregivers, or the participant's historical records were consulted to determine whether a participant suffers from NP or non-NP. Older diagnoses may no longer be accurate, which means that the true sensitivity and specificity values may differ from what the analysis currently indicates.

Correlation analysis between the pain OMs and the other clinical outcomes found no-to-weak correlations with the IPAQ, mAS, MI, NHPT, SDMT and T25FW. It is notable that this includes all performance-based OMs from the test battery. A potential reason for this finding is that the clinical severity of the spasticity (mAS), muscle weakness (MI), the dexterity impairments (NHPT), the cognitive impairments (SDMT) and the ambulation impairments (T25FW) may be too low. Indeed, the scores of the mAS and the MI indicate that spasticity and muscle weakness is almost completely absent in the sample. This absence is also reflected in the NHPT scores; the average mean score for healthy people is 19.3 seconds (Mathiowetz et al., 1985), which is only a second faster than the median score in the current study, so also for the NHPT does the degree of impairment seem too low to show any clinical importance. Likewise, a study about the use of the T25FW in pwMS identified two possible clinically meaningful cut-off scores (Goldman et al., 2013), but the current sample demonstrates a T25FW completion time that is faster than both cut-off scores (≥6 seconds and ≥8 seconds). The same is true for the SDMT; a study involving 359 pwMS identified a cut-off score of 40 for the SDMT (Van Schependom et al., 2014), and the sample in the current study scores well above that. Nevertheless, it is worth reporting about a systematic analysis in which nine of the eleven studies that identified a weak-to-moderate link between cognition and pain in pwMS included three studies in which the SDMT was used to assess cognition (Brown et al., 2023). Based on these studies, a relationship between pain and SDMT performance should exist. It should be noted that none of the correlations with the mAS, MI,

NHPT and SDMT were significant. Therefore, it could also be that these correlations exist only by chance or that the sample size is not large enough to confidently interpret the correlation coefficients.

Furthermore, all pain OMS, except the DN4, show weak-to-moderate correlations with the HADS (total score, anxiety scale and depression scale) the MFIS (total score, cognitive scale, physical scale and psychosocial scale), the PSQI and the PSS. For anxiety and depression, a study about chronic pain in pwMS found weak-to-moderate correlations with pain severity by using the HADS (Kalia & O'Connor, 2005), so this aligns with the findings from the present study. Another study in pwMS identified pain interference as a main mediator of issues with anxiety, fatigue, and sleep quality, and that these problems are, in turn, associated with an increased risk of depressive symptoms (Amtmann et al., 2015). This means that pain, fatigue, mental health and sleep quality seem to be interrelated to some extent in pwMS, which makes the interpretation of these symptoms quite complex.

Some specific findings require closer examination. For example, The NPS and the NPSI not only correlate strongly with each other, but they also show great overlap in their correlations with the other clinical outcomes, so in situations where the available time is limited, it may be worth evaluating if either the NPS or the NPSI can be excluded from the test battery. Given that the scores of the NP group were significantly worse than the scores of the non-NP group only when using the NPSI and not when using the NPS, it may be preferable to opt for the NPSI. Furthermore, the use of the NPS in pwMS has been validated in previous research (Rog et al., 2007; Tyson & Brown, 2014), so the similarities in the results of the NPSI and NPS suggest concurrent validity of the NPSI in pwMS.

Furthermore, both the SDMT and the MFIS measure cognition, but correlations between the pain OMs and the SDMT are only weak and non-significant, whereas correlations with the MFIS are mostly of moderate strength. Although it appears evident that the weak SDMT correlations are due to a low level of cognitive impairment in the current sample, there may be another contributing reason too: Given that the SDMT is a timed performance-based task, it may be that participants are able to focus their attention on completion of the task in a way that distracts them from their experienced pain, thus preventing pain from exerting any notable influence on SDMT performance. This could explain why stronger correlations exist between the cognitive scale of the MFIS and the pain OMs (except the DN4), given that the cognitive scale of the MFIS is not a performance-based task but a self-reported questionnaire about experiences from the past four weeks, so the participants' current pain status would likely not determine the responses to the questionnaires. Previous research with functional

imaging techniques has revealed altered activation of brain regions involved in pain processing during distraction tasks (Johnson, 2005), and a person's executive capabilities do not influence their pain experience (Verhoeven et al., 2011). This means that a decreased pain perception and thus a decreased potential for pain to interfere while performing tasks that require executive functioning, such as the SDMT, would be due to a distraction effect and not due to the cognitive performance itself. This distraction effect may also explain why all other performance-based OMs correlate only weakly or not at all with the pain OMs in the current study as well. These explanations, however, assume that performance is directly related to pain experienced during the task and not to pain experienced when not performing the tasks. Another potential explanation is that the use of analgesics at the moment of testing may have influenced the relationship between pain and performance, but participants were not asked if they had taken any analgesics on the day that the test battery was administered.

Overall, the current results of hypothesis testing indicate the existence of weak-to-moderate correlations between pain on the one hand, and anxiety, depression, fatigue, pain interference, perceived stress, and sleep quality on the other hand. As discussed in the Introduction, the existence of these significant correlations is in agreement with the literature about pain in pwMS. Nevertheless, the hypothesis that at least moderately strong correlations exist between pain OMs and other clinical outcomes is partially rejected by the results, on the basis that some correlations found in the current study are weak and that many correlation coefficients failed to attain statistical significance, which is the case mainly for the correlations that involve performance-based OMs.

Subgroup analysis between the pain types reveals that the NP group obtained significantly higher scores than the non-NP group for the DN4, NPSI, PDQ and the severity scale of the BPI-SF, and significantly worse scores for the IPAQ. The DN4 and the PDQ both intend to differentiate NP from non-NP, so the significantly different results between these subgroups constitute an important finding to support the validity of both instruments in the current sample.

The NPSI demonstrates significant differences between the subgroups as well, so together with the DN4 and PDQ, it may be recommended to also include the NPSI in test batteries for NP. No significant difference was found for the NPS but only by a small margin of .01 (p = .06), so a slightly bigger sample size might decrease the p-value to attain significance. The difference of six points on the

NPS between both groups may be big enough to be clinically meaningful, but no minimal clinically significant difference has been reported for the NPS (Galer & Jensen, 1997; Patel et al., 2023).
The BPI-SF is designed to assess pain regardless of pain type, so the significantly higher score on the severity scale indicates that NP is experienced as more intense than non-NP pain. This finding is in agreement with a study that found that pwMS with NP show greater levels of pain severity than pwMS with non-NP (Kalia & O'Connor, 2005). A recent study, however, found no differences in pain intensity when comparing NP pain with musculoskeletal pain in pwMS by using the McGill Pain Questionnaire, which measures pain severity, but the authors acknowledge that the results do show a trend towards higher pain intensity in the NP group (Rivel et al., 2022). These findings regarding pain severity therefore need further investigation.

The significant differences between the pain subgroups found for the IPAQ results are due to a relatively higher percentage of participants with NP being classified into the *low activity* category, and a relatively lower percentage of participants with NP being classified into the *moderate activity* and the *high activity* categories. Although no studies could be found that investigated whether pwMS with NP are less physically active or not, there is evidence that low levels of physical activity are a risk factor for NP in diabetic neuropathy (Chiang et al., 2016; Ziegler et al., 2009).

The current study found no significant differences for EDSS level between the pain subgroups. This contrasts with a cross-sectional study involving 1249 pwMS, which found that the median EDSS score was higher for pwMS with NP (Solaro et al., 2018), but the study failed to report the median EDSS scores. Confirming these findings, another study with 374 pwMS also found a higher mean EDSS score: 3.12 (*SD* = 2.15) in the NP group vs 2.48 (*SD* = 2.35) in the non-NP group (Ferraro et al., 2018). The reason why no subgroup effect was found in the current study may be due to study design because pwMS with an EDSS score of >6 were excluded from this study. A systematic review with meta-analysis that examined 24 studies with a total of 6671 pwMS created two EDSS subgroups (EDSS >3 and EDSS <3), but the prevalence of NP did not differ significantly between these subgroups (26.65% (*n* = 2152) in the EDSS >3 group and 28.00% (*n* = 4385) in the other group) (Rodrigues et al., 2023). This systematic review, however, did not report the number of participants with an EDSS score of >6, so there is not enough information to judge if including participants with an EDSS score of >6 in the current study would have produced a statistically significant difference between the pain subgroups.

Neither of the HADS scores (total, anxiety and depression) differed significantly between the pain subgroups. This is in stark contrast to the existing literature, with the same systematic review finding higher levels of anxiety and depression in pwMS with NP (Rodrigues et al., 2023). A study that examined the prevalence of anxiety and depression in 70 pwMS with subgroups of EDSS <5 and EDSS >5 found that severe anxiety and depression are more prevalent in the subgroup with an EDSS score of <5 (Hassan et al., 2023). This means that the different results between the systematic review and the current study is likely not caused by the exclusion of pwMS with an EDSS score of >6.

Also neither of the MFIS scores (total, cognitive, physical and psychosocial) differed significantly between the pain subgroups, but the literature is comparatively scarce about this, as studies involving pwMS could not be found, so further research is warranted about fatigue's relation to NP in pwMS.

Overall, the hypothesis that the NP group shows worse clinical and functional outcomes than the non-NP group is partially rejected by the results, as almost none of the clinical outcomes demonstrate statistically significant differences between both groups. Instead, it is mainly the pain OMs on which the NP group logically scores significantly worse.

Subgroup analysis between the countries found that the Belgian sample obtained significantly worse scores than the Chilean sample for the severity scale of the BPI-SF, the NPS, the MFIS total score, the MFIS cognitive and physical scales, the NHPT, the PSS and the T25FW. These differences might be explained by the fact that Belgian participants score on average significantly higher for age (51.08 years vs 38.48 years), EDSS score (3.75 vs 2.00) and disease duration (12 years vs 4.5 years). Indeed, there is some evidence that age and disease duration are associated with pain interference (Hirsh et al., 2009; O'Connor et al., 2008), and it would make sense that pwMS who have progressed further in their disease should score worse on performance-based tasks, such as the NHPT and the T25FW.

Despite the Belgian sample scoring consistently worse where differences exist, the Belgian sample did score significantly better (i.e. lower) than the Chilean sample on the DN4. This may be because the Belgian sample contains a lower proportion of NP participants (50%) than the Chilean sample (70%) does, so a lower mean DN4 score would be expected for the Belgian sample. Nevertheless, it is rather surprising that this is not reflected in the PDQ scores as well, as no significant difference was found between the mean PDQ scores of both countries. It is not immediately clear why the DN4 and the PDQ are inconsistent in this regard. Perhaps the use of Likert scales (PDQ) corrects for

the binary yes-no answers from the DN4 by bringing the extremes (yes or no) closer together in a way that leads to an image of NP that is less polarized and that better encompasses the complexity and the nuances of the participants' pain experiences. In any case, this finding does provide additional evidence that substantial differences between the DN4 and the PDQ must exist.

Overall, given the significant differences in age, EDSS score and disease duration between the Belgian and Chilean samples, no obvious conclusions can be drawn from the current sample regarding the results of the OMs. Therefore, the hypothesis about significant differences between Belgium and Chile remains inconclusive. It is advisable that future cross-cultural comparisons include age-matched, EDSS-matched and disease duration-matched participants.

The literature about the validation of pain OMs in pwMS is scarce. The current study has attempted to address this hiatus with mainly a combination of factor analysis and hypothesis testing to investigate the validity of commonly used pain OMs in pwMS. In addition, many studies have investigated the relationship between MS and various clinical outcomes, but few have differentiated between NP and non-NP, so the current study also contributes to a better understanding of pain type and its impact on a wide range of clinical outcomes in pwMS. The use of the COSMIN checklist has helped to ensure an acceptable standard of quality.

Several limitations of the current study should be mentioned as well, however. First, the current sample size is not optimal, so some of the results should be interpreted with caution. Second, the EDSS scores and the diagnosis of pain type were retrieved from historical records, so it could not be independently verified by the authors of this study if EDSS score and pain diagnosis were still correct by the time that the participants were recruited for the current study. This means that the accuracy of that information is assumed but not verified. Third, pwMS with an EDSS score of >6 were excluded from the current study, so no conclusions can be drawn for pwMS who have progressed to the later stages of the disease. Fourth, impairment levels of the non-pain clinical outcomes were generally low in the current sample, which may have prevented the emergence of stronger correlations. Fifth, hypothesis testing was limited to main pairings without further investigating the influence of confounders (e.g. the influence of fatigue on depression in relation to pain). Sixth, it is important to remember that causation cannot be inferred from the correlations that have emerged from hypothesis testing.

Future research in pwMS could clarify the composition of each DN4 factor and further explore the relationship between the DN4 and the PDQ. Also the clinical outcomes and the differences found between the NP group and the non-NP group should be investigated in more depth to address the current lack of evidence. Furthermore, the inclusion of pwMS with an EDSS score of >6 could provide a completer picture of the results obtained in the current study, at least insofar as pwMS with an EDSS score of >6 are physically able to complete the test battery. In addition, it is recommended that cross-cultural comparisons control for age, EDSS score and disease duration.

REFERENCE LIST

Amtmann, D., Askew, R. L., Kim, J., Chung, H., Ehde, D. M., Bombardier, C. H., Kraft, G. H., Jones, S. M., & Johnson, K. L. (2015). Pain affects depression through anxiety, fatigue, and sleep in multiple sclerosis. *Rehabil Psychol*, *60*(1), 81-90. https://doi.org/10.1037/rep0000027

Amtmann, D., Bamer, A. M., Noonan, V., Lang, N., Kim, J., & Cook, K. F. (2012). Comparison of the psychometric properties of two fatigue scales in multiple sclerosis. *Rehabil Psychol*, *57*(2), 159-166. https://doi.org/10.1037/a0027890

Ashworth, B. (1964). Preliminary trial of carisoprodol in multiple sclerosis. *Practitioner*, *192*, 540-542.

Beavers, A. S., Lounsbury, J. W., Richards, J., Huck, S. W., Skolits, G. J., & Esquivel, S. L. (2013). Practical considerations for using exploratory factor analysis in educational research. *Practical Assessment, Research and Evaluation*, *18*, 1-13. https://doi.org/10.7275/qv2q-rk76

Benedict, R. H., DeLuca, J., Phillips, G., LaRocca, N., Hudson, L. D., & Rudick, R. (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

Boeschoten, R. E., Braamse, A. M. J., Beekman, A. T. F., Cuijpers, P., van Oppen, P., Dekker, J., & Uitdehaag, B. M. J. (2017). Prevalence of depression and anxiety in multiple sclerosis: A systematic review and meta-analysis. *J Neurol Sci*, *372*, 331-341. https://doi.org/10.1016/j.jns.2016.11.067

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

Bouhassira, D., Attal, N., Alchaar, H., Boureau, F., Brochet, B., Bruxelle, J., Cunin, G., Fermanian, J., Ginies, P., Grun-Overdyking, A., Jafari-Schluep, H., Lantéri-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

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

Brown, F. S., Glasmacher, S. A., Taylor, D., Jenkins, R., Chandran, S., Gillespie, D., & Foley, P. (2023). Pain and cognitive performance in adults with multiple sclerosis: A systematic review. *Mult Scler Relat Disord*, *71*, 104584. https://doi.org/10.1016/j.msard.2023.104584

Buysse, D. J., Reynolds, C. F., 3rd, Monk, T. H., Berman, S. R., & Kupfer, D. J. (1989). The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. *Psychiatry Res*, *28*(2), 193-213. https://doi.org/10.1016/0165-1781(89)90047-4

Chiang, S. S., Lee, C. L., Liu, H. C., Wang, J. S., Lee, I. T., Song, Y. M., Fu, C. P., Li, Y. F., Sheu, W. H., & Lin, S. Y. (2016). Physical activity and albuminuria were associated with painful diabetic polyneuropathy in type 2 diabetes in an ethnic Chinese population. *Clin Chim Acta*, *462*, 55-59. https://doi.org/10.1016/j.cca.2016.08.022

Cleeland, C. S. (2009). *The Brief Pain Inventory user guide*. The University of Texas MD Anderson Cancer Center. Retrieved March 2, 2024 from https://www.mdanderson.org/research/departments-

labs-institutes/departments-divisions/symptom-research/symptom-assessment-tools/brief-pain-inventory.html

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *J Health Soc Behav*, *24*(4), 385-396. https://doi.org/10.2307/2136404

Craig, C. L., Marshall, A. L., Sjöström, M., Bauman, A. E., Booth, M. L., Ainsworth, B. E., Pratt, M., Ekelund, U., Yngve, A., Sallis, J. F., & Oja, P. (2003). International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*, *35*(8), 1381-1395. https://doi.org/10.1249/01.Mss.0000078924.61453.Fb

Demeurisse, G., Demol, O., & Robaye, E. (1980). Motor evaluation in vascular hemiplegia. *Eur Neurol*, *19*(6), 382-389. https://doi.org/10.1159/000115178

Dighriri, I. M., Aldalbahi, A. A., Albeladi, F., Tahiri, A. A., Kinani, E. M., Almohsen, R. A., Alamoudi, N. H., Alanazi, A. A., Alkhamshi, S. J., Althomali, N. A., Alrubaiei, S. N., & Altowairqi, F. K. (2023). An overview of the history, pathophysiology, and pharmacological interventions of multiple sclerosis. *Cureus*, *15*(1), e33242. https://doi.org/10.7759/cureus.33242

Ehde, D. M., Gibbons, L. E., Chwastiak, L., Bombardier, C. H., Sullivan, M. D., & Kraft, G. H. (2003). Chronic pain in a large community sample of persons with multiple sclerosis. *Mult Scler*, *9*(6), 605-611. https://doi.org/10.1191/1352458503ms939oa

Ehde, D. M., Osborne, T. L., Hanley, M. A., Jensen, M. P., & Kraft, G. H. (2006). The scope and nature of pain in persons with multiple sclerosis. *Mult Scler*, *12*(5), 629-638. https://doi.org/10.1177/1352458506071346

Fagbohun, T. R. (2021). Systematic review on the psychometric, reliability and validity properties of translated neuropathic pain screening tools (DN4, LANSS AND PDQ) 1 January 2005 – 19 July 2019. *International Journal of Medicine and Medical Research*. https://doi.org/10.11603/ijmmr.2413-6077.2021.1.12120

Ferraro, D., Plantone, D., Morselli, F., Dallari, G., Simone, A. M., Vitetta, F., Sola, P., Primiano, G., Nociti, V., Pardini, M., Mirabella, M., & Vollono, C. (2018). Systematic assessment and characterization of chronic pain in multiple sclerosis patients. *Neurol Sci*, *39*(3), 445-453. https://doi.org/10.1007/s10072-017-3217-x

Feys, P., Lamers, I., Francis, G., Benedict, R., Phillips, G., LaRocca, N., Hudson, L. D., & Rudick, R. (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

Fisk, J. D., Ritvo, P. G., Ross, L., Haase, D. A., Marrie, T. J., & Schlech, W. F. (1994). Measuring the functional impact of fatigue: Initial validation of the fatigue impact scale. *Clin Infect Dis*, *18 Suppl 1*, S79-83. https://doi.org/10.1093/clinids/18.supplement_1.s79

Freynhagen, R., Baron, R., Gockel, U., & Tölle, 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

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

Gil-González, I., Martín-Rodríguez, A., Conrad, R., & Pérez-San-Gregorio, M. (2020). Quality of life in adults with multiple sclerosis: A systematic review. *BMJ Open*, *10*(11), e041249. https://doi.org/10.1136/bmjopen-2020-041249

Goldman, M. D., Motl, R. W., Scagnelli, J., Pula, J. H., Sosnoff, J. J., & Cadavid, D. (2013). Clinically meaningful performance benchmarks in MS: Timed 25-foot walk and the real world. *Neurology*, *81*(21), 1856-1863. https://doi.org/10.1212/01.wnl.0000436065.97642.d2

Harb, A., & Kishner, K. (2023). *Modified Ashworth Scale*. StatPearls Publishing. Retrieved April 20, 2024 from https://www.ncbi.nlm.nih.gov/books/NBK554572/

Hassan, S. S., Darwish, E. S., Ahmed, G. K., Azmy, S. R., & Haridy, N. A. (2023). Relationship between disability and psychiatric outcome in multiple sclerosis patients and its determinants. *The Egyptian Journal of Neurology, Psychiatry and Neurosurgery*, *59*(1), 105. https://doi.org/10.1186/s41983-023-00702-x

Hirsh, A. T., Turner, A. P., Ehde, D. M., & Haselkorn, J. K. (2009). Prevalence and impact of pain in multiple sclerosis: Physical and psychologic contributors. *Arch Phys Med Rehabil*, *90*(4), 646-651. https://doi.org/10.1016/j.apmr.2008.10.019

Honarmand, K., & Feinstein, A. (2009). Validation of the Hospital Anxiety and Depression Scale for use with multiple sclerosis patients. *Mult Scler*, *15*(12), 1518-1524. https://doi.org/10.1177/1352458509347150

Hugos, C. L., & Cameron, M. H. (2019). Assessment and measurement of spasticity in MS: State of the evidence. *Curr Neurol Neurosci Rep*, *19*(10), 79. https://doi.org/10.1007/s11910-019-0991-2

Jensen, M. P., Chodroff, M. J., & Dworkin, R. H. (2007). The impact of neuropathic pain on health-related quality of life: Review and implications. *Neurology*, *68*(15), 1178-1182. https://doi.org/10.1212/01.wnl.0000259085.61898.9e

Jerković, A., Mikac, U., Matijaca, M., Košta, V., Ćurković Katić, A., Dolić, K., Vujović, I., Šoda, J., Đogaš, Z., Pavelin, S., & Rogić Vidaković, 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

Jerković, A., Proroković, A., Matijaca, M., Vuko, J., Poljičanin, A., Mastelić, A., Ćurković Katić, A., Košta, V., Kustura, L., Dolić, K., Đogaš, Z., & Rogić Vidaković, M. (2021). Psychometric properties of the HADS measure of anxiety and depression among multiple sclerosis patients in Croatia. *Front Psychol*, *12*, 794353. https://doi.org/10.3389/fpsyg.2021.794353

Johnson, M. H. (2005). How does distraction work in the management of pain? *Curr Pain Headache Rep*, *9*(2), 90-95. https://doi.org/10.1007/s11916-005-0044-1

Kalia, L. V., & O'Connor, P. W. (2005). Severity of chronic pain and its relationship to quality of life in multiple sclerosis. *Mult Scler*, *11*(3), 322-327. https://doi.org/10.1191/1352458505ms1168oa

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

Kratz, A. L., Braley, T. J., Foxen-Craft, E., Scott, E., Murphy, J. F., 3rd, & Murphy, S. L. (2017). How do pain, fatigue, depressive, and cognitive symptoms relate to well-being and social and physical functioning in the daily lives of individuals with multiple sclerosis? *Arch Phys Med Rehabil*, *98*(11), 2160-2166. https://doi.org/10.1016/j.apmr.2017.07.004

Kyndt, E., & Onghena, P. (2014). The integration of work and learning: Tackling the complexity with structural equation modelling. In *Discourses on professional learning: On the boundary between learning and working.* (pp. 255-291). Springer Science + Business Media. https://doi.org/10.1007/978-94-007-7012-6_14

Łabuz-Roszak, B., Niewiadomska, E., Kubicka-Bączyk, K., Skrzypek, M., Tyrpień-Golder, K., Majewska, A., Matejczyk, A., Dobrakowski, P., & Pierzchała, K. (2019). Prevalence of pain in patients with multiple sclerosis and its association with anxiety, depressive symptoms and quality of life. *Psychiatr Pol*, *53*(2), 475-486. https://doi.org/10.12740/pp/94469

Lapane, K. L., Quilliam, B. J., Benson, C., Chow, W., & Kim, M. (2014). One, two, or three? Constructs of the brief pain inventory among patients with non-cancer pain in the outpatient setting. *J Pain Symptom Manage*, *47*(2), 325-333. https://doi.org/10.1016/j.jpainsymman.2013.03.023

MacCallum, R., Widaman, K., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*, 84-99. https://doi.org/10.1037/1082-989X.4.1.84

Marck, C. H., De Livera, A. M., Weiland, T. J., Jelinek, P. L., Neate, S. L., Brown, C. R., Taylor, K. L., Khan, F., & Jelinek, G. A. (2017). Pain in people with multiple sclerosis: Associations with modifiable lifestyle factors, fatigue, depression, anxiety, and mental health quality of life. *Front Neurol*, *8*, 461. https://doi.org/10.3389/fneur.2017.00461

Mathiowetz, V., Weber, K., Kashman, N., & Volland, G. (1985). Adult norms for the Nine Hole Peg Test of finger dexterity. *Occupational Therapy Journal of Research*, *5*(1), 24-38. https://doi.org/10.1177/153944928500500102

Matsubayashi, Y., Takeshita, K., Sumitani, M., Oshima, Y., Tonosu, J., Kato, S., Ohya, J., Oichi, T., Okamoto, N., & Tanaka, S. (2013). Validity and reliability of the Japanese version of the painDETECT questionnaire: A multicenter observational study. *PLoS One*, *8*(9), e68013. https://doi.org/10.1371/journal.pone.0068013

McGinley, M. P., & Lapin, B. (2022). The value of patient-reported outcome measures for multiple sclerosis. *Mult Scler*, *28*(10), 1489-1490. https://doi.org/10.1177/13524585221111675

McIsaac, T. L., Fritz, N. E., & O'Sullivan, S. B. (2019). Multiple sclerosis. In *Physical Rehabilitation* (7th ed., pp. 662-669). F.A. Davis Company.

Merlino, G., Fratticci, L., Lenchig, C., Valente, M., Cargnelutti, D., Picello, M., Serafini, A., Dolso, P., & Gigli, G. L. (2009). Prevalence of 'poor sleep' among patients with multiple sclerosis: An independent predictor of mental and physical status. *Sleep Med*, *10*(1), 26-34. https://doi.org/10.1016/j.sleep.2007.11.004

MohanaSundaram, A., Sathanantham, S. T., Gudesblatt, M., Ganvir, S., Chinchole, V., Patil, B., & Velayutham, R. (2023). Using modified Ashworth scale for assessing multiple sclerosis-associated spasticity: A high time for a paradigm shift. *Ther Adv Neurol Disord*, *16*, 17562864231154653. https://doi.org/10.1177/17562864231154653

Mokkink, L. B., C., P. C. A., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Terwee, C. B. (2019). *COSMIN study design checklist for patient-reported outcome measurement instruments*. Retrieved March 2, 2024 from https://www.cosmin.nl/tools/checklists-assessing-methodological-study-qualities/

Motl, R. W., Cohen, J. A., Benedict, R., Phillips, G., LaRocca, N., Hudson, L. D., & Rudick, R. (2017). Validity of the timed 25-foot walk as an ambulatory performance outcome measure for multiple sclerosis. *Mult Scler*, *23*(5), 704-710. https://doi.org/10.1177/1352458517690823

Neuropathic pain. (2022). International Association for the Study of Pain. Retrieved March 23, 2024 from https://www.iasp-pain.org/advocacy/global-year/neuropathic-pain/

O'Connor, A. B., Schwid, S. R., Herrmann, D. N., Markman, J. D., & Dworkin, R. H. (2008). Pain associated with multiple sclerosis: Systematic review and proposed classification. *Pain*, *137*(1), 96-111. https://doi.org/10.1016/j.pain.2007.08.024

Osborne, T. L., Raichle, K. A., Jensen, M. P., Ehde, D. M., & Kraft, G. (2006). The reliability and validity of pain interference measures in persons with multiple sclerosis. *J Pain Symptom Manage*, *32*(3), 217-229. https://doi.org/10.1016/j.jpainsymman.2006.03.008

Ouyang, H., Li, X., Xu, H., Zhan, Y., Zheng, Z., Chen, G., Lou, Z., Chen, H., Zhang, J., Mao, H., Zhang, C., Qin, L., Zhao, Y., & Zhao, M. (2024). Risk factors of neuropathic pain in multiple sclerosis: A retrospective case-cohort study. *Front Immunol*, *15*, 1309583. https://doi.org/10.3389/fimmu.2024.1309583

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). McGraw-Hill, Open University Press.

Patel, A. A., Tapper, E. B., Kanwal, F., Woodrell, C. D., Hansen, L., Lai, J. C., Rogal, S., McDermott, C., Rakoski, M., & Ufere, N. N. (2023). Targets and study design for symptom-focused trials aimed at patients with cirrhosis: An expert consensus. *Hepatology Communications*, *7*(6), e0135. https://doi.org/10.1097/hc9.0000000000000135

Poquet, N., & Lin, C. (2016). The Brief Pain Inventory (BPI). *J Physiother*, *62*(1), 52. https://doi.org/10.1016/j.jphys.2015.07.001

Rietberg, M. B., Van Wegen, E. E., & Kwakkel, G. (2010). Measuring fatigue in patients with multiple sclerosis: Reproducibility, responsiveness and concurrent validity of three Dutch self-report questionnaires. *Disabil Rehabil*, *32*(22), 1870-1876. https://doi.org/10.3109/09638281003734458

Rivel, M., Achiron, A., Dolev, M., Stern, Y., Zeilig, G., & Defrin, R. (2022). Unique features of central neuropathic pain in multiple sclerosis: Results of a cluster analysis. *Eur J Pain*, *26*(5), 1107-1122. https://doi.org/10.1002/ejp.1934

Rodrigues, P., da Silva, B., & Trevisan, G. (2023). A systematic review and meta-analysis of neuropathic pain in multiple sclerosis: Prevalence, clinical types, sex dimorphism, and increased depression and anxiety symptoms. *Neurosci Biobehav Rev*, *154*, 105401. https://doi.org/10.1016/j.neubiorev.2023.105401

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

Samuels, P. C. (2017). *Advice on Exploratory Factor Analysis*. Retrieved June 8, 2024 from https://www.researchgate.net/publication/319165677_Advice_on_Exploratory_Factor_Analysis

Santos, J. G., Brito, J. O., de Andrade, D. C., Kaziyama, V. M., Ferreira, K. A., Souza, I., Teixeira, M. J., Bouhassira, D., & Baptista, A. F. (2010). Translation to Portuguese and validation of the Douleur Neuropathique 4 questionnaire. *J Pain*, *11*(5), 484-490. https://doi.org/10.1016/j.jpain.2009.09.014

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research*, *8*(2), 23-74. https://doi.org/10.23668/psycharchives.12784

Shahrbanian, S., Auais, M., Duquette, P., Andersen, K., & Mayo, N. E. (2013). Does pain in individuals with multiple sclerosis affect employment? A systematic review and meta-analysis. *Pain Res Manag*, *18*(5), e94-e100. https://doi.org/10.1155/2013/829464

Smith, A. (1973). Symbol Digit Modalities Test. *The Clinical Neuropsychologist*.

Solaro, C., Cella, M., Signori, A., Martinelli, V., Radaelli, M., Centonze, D., Sica, F., Grasso, M. G., Clemenzi, A., Bonavita, S., Esposito, S., Patti, F., D'Amico, E., Cruccu, G., & Truini, A. (2018).

Identifying neuropathic pain in patients with multiple sclerosis: A cross-sectional multicenter study using highly specific criteria. *J Neurol*, *265*(4), 828-835. https://doi.org/10.1007/s00415-018-8758-2

Svendsen, K. B., Jensen, T. S., Overvad, K., Hansen, H. J., Koch-Henriksen, N., & Bach, F. W. (2003). Pain in patients with multiple sclerosis: A population-based study. *Archives of Neurology*, *60*(8), 1089-1094. https://doi.org/10.1001/archneur.60.8.1089

Tan, G., Jensen, M. P., Thornby, J. I., & Shanti, B. F. (2004). Validation of the Brief Pain Inventory for chronic nonmalignant pain. *J Pain*, *5*(2), 133-137. https://doi.org/10.1016/j.jpain.2003.12.005

Timmerman, H., Steegers, M. A. H., Huygen, F., Goeman, J. J., van Dasselaar, N. T., Schenkels, M. J., Wilder-Smith, O. H. G., Wolff, A. P., & Vissers, K. C. P. (2017). Investigating the validity of the DN4 in a consecutive population of patients with chronic pain. *PLoS One*, *12*(11), e0187961. https://doi.org/10.1371/journal.pone.0187961

Truini, A., Barbanti, P., Pozzilli, C., & Cruccu, G. (2013). A mechanism-based classification of pain in multiple sclerosis. *J Neurol*, *260*(2), 351-367. https://doi.org/10.1007/s00415-012-6579-2

Tyson, S. F., & Brown, P. (2014). How to measure pain in neurological conditions? A systematic review of psychometric properties and clinical utility of measurement tools. *Clin Rehabil*, *28*(7), 669-686. https://doi.org/10.1177/0269215513514231

Van Schependom, J., D'Hooghe M, B., Cleynhens, K., D'Hooge, M., Haelewyck, M. C., De Keyser, J., & Nagels, G. (2014). The Symbol Digit Modalities Test as sentinel test for cognitive impairment in multiple sclerosis. *Eur J Neurol*, *21*(9), 1219-1225, e1271-1212. https://doi.org/10.1111/ene.12463

Verhoeven, K., Van Damme, S., Eccleston, C., Van Ryckeghem, D. M., Legrain, V., & Crombez, G. (2011). Distraction from pain and executive functioning: An experimental investigation of the role of inhibition, task switching and working memory. *Eur J Pain*, *15*(8), 866-873. https://doi.org/10.1016/j.ejpain.2011.01.009

Wanitschek, A., Seebacher, B., Muehlbacher, A., Brenneis, C., & Ehling, R. (2024). Comparison of patient-reported outcomes of physical activity and accelerometry in people with multiple sclerosis and ambulatory impairment: A cross-sectional study. *Mult Scler Relat Disord*, *85*, 105532. https://doi.org/10.1016/j.msard.2024.105532

Widerström-Noga, E., Loeser, J. D., Jensen, T. S., & Finnerup, N. B. (2017). AAPT diagnostic criteria for central neuropathic pain. *J Pain*, *18*(12), 1417-1426. https://doi.org/10.1016/j.jpain.2017.06.003

Wu, S. M., & Amtmann, D. (2013). Psychometric evaluation of the Perceived Stress Scale in multiple sclerosis. *ISRN Rehabilitation*, *2013*, 608356. https://doi.org/10.1155/2013/608356

Ziegler, D., Rathmann, W., Meisinger, C., Dickhaus, T., & Mielck, A. (2009). Prevalence and risk factors of neuropathic pain in survivors of myocardial infarction with pre-diabetes and diabetes. The KORA Myocardial Infarction Registry. *Eur J Pain*, *13*(6), 582-587. https://doi.org/10.1016/j.ejpain.2008.07.007

APPENDIX A: COSMIN checklist

General recommendations for the design of a study on measurement properties	very good	adequate	doubtful	inadequate	justification
Research aim					
1 Provide a clear research aim, including (1) the name and version of the PROM, (2) the target population, and (3) the measurement properties of interest	Research aim clearly described			Research aim not clearly described	New
PROM					
2 Provide a clear description of the construct to be measured	Construct clearly described			Construct not clearly described	RoB Box 1
3 Provide a clear description of the development process of the PROM, including a description of the target population for which the PROM was developed	Development process clearly described		Development process clearly described		RoB Box 1
4 The origin of the construct should be clear: provide a theory, conceptual framework (i.e. reflective or formative model) or disease model used or clear rationale to define the construct to be measured	Origin of the construct clear		Origin of the construct not clear		RoB Box 1

5	Provide a clear description of the structure of the PROM (i.e. the number of items and subscales included in the PROM, instructions given and response options) and its scoring algorithm	Structure and scoring algorithm clearly described			Structure and scoring algorithm not clearly described	RoB Box 1
6	Provide a clear description of existing evidence on the quality of the PROM	Existing evidence on the quality of the PROM clearly described		Existing evidence on the quality of the PROM not clearly described		New
7	Provide a clear description of the context of use*	Context of use clearly described		Context of use not clearly described		RoB Box 1
Target population						
8	Provide a clear description of in- and exclusion criteria to select patients, e.g. in terms of disease condition and characteristics like age, gender, language or country, and setting (e.g. general population, primary care or hospital/rehabilitation care)	In- and exclusion criteria for patients clearly described			In- and exclusion criteria for patients not clearly described	Characteristic of study population [6]
9	Provide a clear description of the method used to select the patients for the study (e.g. convenience, consecutive, or random)	Method for patient selection clearly described		Method of patient selection not clearly described		New
10	Describe whether the selected sample is representing the target population in which the PROM will be used in terms of age, gender, important disease characteristics (e.g. severity, status, duration)	Study sample representing the target population clearly described	Assumable that the study sample is representing the target population, but not clearly described	Unclear whether the study sample is representing the target population	Study will not be performed in a sample representing the target population	RoB Box 1

Hypotheses testing for construct validity **A. Comparison with other outcome measurement instruments (convergent validity)**	very good	adequate	doubtful	inadequate	NA	Justification
Design requirements						
1 Formulate hypotheses about expected relationships between the PROM under study and other outcome measurement instrument(s)	Hypotheses formulated including the expected direction and magnitude of the correlations stated		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what is expected		Original CC
2 Provide a clear description of the construct(s) measured by the comparator instrument(s)	Construct(s) measured by the comparator instrument(s) is/are clearly described			Construct(s) measured by the comparator instrument(s) is/are not clearly described		RoB Box 9a (1)
3 Use comparator instrument(s) with sufficient measurement properties	Sufficient measurement properties of the comparator instrument(s) in a population similar to the study population	Sufficient measurement properties of the comparator instrument(s) but not sure if these apply to the study population	Some information on measurement properties (or a reference to a study on measurement properties) of the comparator instrument(s) in any study population	No information on the measurement properties of the comparator instrument(s), or evidence of insufficient measurement properties of the comparator instrument(s)		RoB Box 9a (2)

4 Perform the analysis in a sample with an appropriate number of patients (taking into account expected number of missing values)	≥100 patients	50-99 patients	30-49 patients	<30 patients	Sample size
5 Use an appropriate time schedule for assessments of the PROM of interest and comparison instruments	PROM and comparison instrument(s) will be administered at the same time	PROM and comparison instrument(s) not administered at the same time, but assumable that patient will not change in the interim period	PROM and comparison instrument(s) will not be administered at the same time, but unclear if patients will changed	PROM and comparison instrument(s) will not be administered at the same time, and patients are expected to change	New
Statistical methods					
6 Use statistical methods that are appropriate for the hypotheses to be tested	Statistical methods will be appropriate	Assumable that statistical methods will be appropriate	Statistical methods will not be optimal	Statistical methods will NOT be appropriate	RoB Box 9a (3)
7 Provide a clear description of how missing items will be handled	The way missing items will be handled is clearly described		The way missing items will be handled is not clearly described		Original CC

B. Comparison between subgroups (discriminative or known-groups validity)	very good	adequate	doubtful	inadequate	NA	Justification
Design requirements						
1 Formulate hypotheses regarding mean differences between subgroups	Hypotheses formulated including the expected directions and magnitude of the mean differences stated		Hypotheses vague or not formulated but possible to deduce what was expected	Unclear what was expected		Original CC
2 Provide an adequate description of important characteristics of the subgroups, such as disease or demographic characteristics	Adequate description of the important characteristics of the subgroups	Adequate description of most of the important characteristics of the subgroups	Poor of no description of the important characteristics of the subgroups			RoB Box 9b (5)
3 Perform the analysis in a sample with an appropriate number of patients (taking into account expected number of missing values)	≥100 patients per group	50-99 patients per group	30-49 patients per group	<30 patients per group		Sample size
Statistical methods						
4 Use statistical methods that are appropriate for the hypotheses to be tested	Statistical methods will be appropriate	Assumable that statistical methods will be appropriate	Statistical methods will not be optimal	Statistical methods will NOT be appropriate		RoB Box 9b (6)
5 Provide a clear description of how missing items will be handled	The way missing items will be handled is clearly described		The way missing items will be handled is not clearly described			Original CC